A Meta-Analysis on the Debt-Growth Relationship

D'Andrea, Sara

Sapienza University of Rome Department of Economics and Law, Sogei SpA

1 September 2022

A Meta-Analysis on the Debt-Growth Relationship

Sara D'Andrea[*†]

August 30, 2022

Abstract

We perform a meta-analysis on the relationship between public (government and external) debt and economic growth, coding 422 observations from 32 studies estimating cross-sectional or panel regressions. The average estimated effect size turns out to be negative (around -0.2). Heterogeneity is substantial and influenced mainly by within-studies variability. The moderators that allow to slightly mitigate it and that influence the estimate of the effect size concern the publication status, the journal ranking, the variables used as proxies, the level of wealth and development of the countries considered, the sample size, the region, and the estimation method. Publication bias arises, both as regards the direction of the estimated effect size, and the statistical significance of the results presented.

Keywords: Meta-Analysis, Economic Growth, Public Debt, External Debt, Systematic Review

JEL Codes: H63, O40, C1

[*]Department of Economics and Law
Sapienza University of Rome
Email address: sara.dandrea@uniroma1.it
[†]Sogei S.p.A.
Email address: sdandrea@sogei.it

Acknowledgments

The author thanks Carmelo P. Parello for his guidance in conducting the research. Special thanks go to Tomáš Havránek, Tom Stanley, Jacek Buczny, and Ayaz Zeynalov, for their valuable comments on the method.

The author would further like to thank José Alves, Javier J. Pérez, Gianni Carvelli, Raouf Boucekkine, Cecilia Garcia-Peñalosa, Lorenzo Rotunno, and Rebeca Jiménez-Rodríguez for their sugggestions in refining the paper.

The views expressed in this paper are those of the author and do not necessarily reflect those of Sapienza University of Rome or Sogei S.p.A.

1 Introduction

From the beginning of 2020, an unprecedented shock upset the growth trends of the global economy, attributing great importance to fiscal policy interventions and the presence of the public sector in the economy. Economists and scholars have long argued for the complementarity of state and market, which is fundamental in times of crisis. The Covid-19 pandemic has led the governments of the most developed nations to raise the amount of public debts as never before, to finance fiscal policies sorely needed to support and relaunch the economy. For the Eurozone, it was even allowed to suspend the rules enshrined in the Maastricht Treaty, thus to exceed the debt-to-GDP threshold of 60% and the deficit-to-GDP ratio (3%)[1], which was not even granted during the 2007-2011 crisis. During those years, some economists and policymakers considered debt as an obstacle to growth[2], as excessive debt would have raised the riskiness of a country (especially in the outlook of rating agencies) and consequently would have increased the cost of interest expenditure.

Starting with these fiscal policy reactions, we aim to study the relationship between public debt and economic growth, both in direction and magnitude. This paper explores the study of the literature on the linear relationship between public debt and economic growth by performing a meta-analysis following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, henceforth) standards (Moher et al. (2009)). Maybe, in the future, this paper could initiate a meta-data collection process, perhaps through the use of an internet-based platform, where authors of papers that meet PRISMA standards independently enter the required data, and meta-analysis estimates could update automatically.

The research questions to be answered by this meta-analysis are as follows: what is the average effect in the literature of an increase in debt on a country's economic growth? Is the direction of this link positive, negative, or zero? Is there, and to what extent, a certain degree of heterogeneity in the results of the studies analyzed? If heterogeneity exists, what influences it? Is there publication bias in this area of research? If so, in which direction?

The methodology we follow in developing a meta-analysis of the literature regarding the debt-growth relationship is based on Stanley and Jarrell (2005)'s seminal paper. In this research, we try to satisfy as much as possible the reporting guidelines established by the Meta-Analysis of Economics Research Network (MAER-Net) (Stanley et al. (2013)), recently updated by Havránek et al. (2020). Therefore, we will first discuss how the research phase of the studies was conducted, following the PRISMA statement by Moher et al. (2009), then we will define the effect size and describe the coding phase of the studies. Subsequently, the theoretical model adopted will be introduced and the results concerning the analysis with fixed effects, random effects, and the moderator analysis will be shown, together with the forest plot. Several meta-regressions will be estimated and, finally, the publication bias will be analyzed.

The first meta-analysis on the debt-growth relationship in the literature has been developed by

[1] A complete description of the European Union response to the pandemic and economic crisis is available at https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/jobs-and-economy-duringcoronaviruspandemic/state-aid-cases_en.

[2] For a good review of the debate, see Corsetti et al. (2012).

Moore and Thomas (2010), who only consider 17 studies, but mainly refer to the elasticity of growth to debt, and analyze fewer moderators than we include. Furthermore, their meta-analysis does not consider post-2010 studies, a period in which the literature on the debt-growth nexus proliferated. A second meta-analysis on the relationship between public debt and growth was proposed by Heimberger et al. (2021). This paper considers 48 primary studies and analyses the relationship between public debt and economic growth, using a standardized coefficient as effect size, and the estimated mean value is -0.14. The authors find evidence for a negative direction of publication bias. Moreover, when endogeneity is considered in the relationship, it leads to a positive link, as well as in the case of developed countries and as the journal impact factor in which the study is published increases. Finally, the estimation of the 90% threshold value (Reinhart and Rogoff (2010b)), beyond which the debt to GDP ratio becomes detrimental to growth, depends on the methodology and the data used. On average, this threshold is also 18% lower for developing countries than for developed ones.

These reference papers can be complementary to our work. We use a different effect size (the Partial Correlation Coefficient) and include both continuous and categorical moderators (which are transformed into dummy variables only for the Bayesian Model Averaging analysis). Moreover, we analyze additional moderators not considered in previous meta-analyses, such as region, external debt, income level, proxies for dependent and independent variables, and focus on non-linearities. At the methodological level, we deeply analyze heterogeneity in coded studies and apply the BMA method for metaregression.

We code 422 observations from 32 records, using the Partial Correlation Coefficient (PCC, henceforth) as effect size and including 11 categorical moderators and 7 continuous moderators.

The results of the paper are the following. First, the mean effect size is -0.2 in the multi-level analysis, while is slightly larger in absolute terms in the analysis of the whole sample and zero in the reduced sample (where the average PCC for each study is considered).

Heterogeneity is strongly present and is due to differences both within- and between-studies; within-studies variability is more important than between-studies. Heterogeneity continues to be present when analyzing moderators. Among these, the moderators that lower the level of heterogeneity and thus explain the different estimates across studies are region, income, and development level, the variables used as proxies, and the methodology used.

In particular, the average nexus is -0.3 when the dependent variable is the GDP growth rate, while it becomes 0.3 when the GDP per capita level is used. Thus, according to our meta-analysis, an increase in debt on average increases the level of GDP (with a consequent increase in individual welfare), but slows down its growth rate. Then, this correlation is on average negative for several proxies of the independent variable and becomes less negative when using the debt to GDP ratio, or the logarithm of debt to GDP.

The average PCC is -0.3 for developing countries, but it is slightly positive for developed countries, implying greater efficiency in the allocation of debt-financed expenditure in the latter. The different regression estimation methods, ceteris paribus, lead to different PCC estimates.

Concerning the estimation method, when endogeneity is accounted for, the average PCC is positive

(0.2).
Moreover, in countries with high-income levels, the relationship is positive, while in countries with medium and low-income levels it is negative. This indicates that high-income countries spend debt to finance high-return investments, while in middle- and low-income countries it is more complicated to turn public debt into productive investment for economic growth. We also find evidence for a "middle-income trap". The average nexus decreases with sample size (in space and time) and increases when considering samples from more recent years or results reported in more high-ranked journals.
To select the most important moderators we implement multimodel inference, BMA, and stepwise regression. These methods agree on the following moderators with the highest predictive power: publication type, estimation method, presence of inflation as a control variable, and journal ranking. Finally, concerning publication bias, our analysis indicates positive asymmetry of the funnel plot. This leads to an over-representation of studies with positive effect sizes. The estimate of the correct average effect size, according to this analysis, lies between -0.5 and -0.9.
Publication bias can also be seen in the massive presence of p-values below 0.05, confirmed by the three-parameter selection model and the p-curve analysis. Statistically significant results at the 95% level tend to be published more easily than results with p-values greater than the 0.05 threshold.
The rest of the paper proceeds as follows. In Section 2, we provide a literature review on the debt-economic growth empirical relationship. In Section 3, we detail the steps of the search phase, while the coding phase description is in Section 4. In Section 5, we show the results obtained. We develop a multi-level meta-analysis, using the R software (following Harrer et al. (2021)), in which all the estimates obtained in each study are encoded in the meta-database, to obtain accurate results on the debt-growth relationship. In this Section, we show the fixed effects and the random effects estimation, the analysis of heterogeneity and variance components, the moderator analysis, the different methodological techniques of meta-regression, and the study of publication bias. Finally, Section 6 concludes.

2 Literature Review

The direction of the link between public debt and economic growth has been a debated issue for more than 200 years. With the birth of economic sciences, the phenomenon was initially approached from a philosophical point of view (Smith (1776)). Subsequently, various theories have been developed on this topic, and it is only since the 2007 crisis (to the best of our knowledge) that scholars and institutions have focused more on this issue than before, and empirical contributions have increased dramatically. In this literature review, we only focus on empirical contributions. We first discuss the literature about the linear relationship between debt and growth (Subsection 2.1) and then analyze in more detail the non-linear estimations (Subsection 2.2).

2.1 Linear Relationship

The empirical literature has supported three different theories about the linear relationship between public debt and growth. We will first analyze the contributions supporting the neutrality of the relationship (corresponding to the "Ricardian Equivalence Hypothesis" by Ricardo (1888), Barro (1989), and Buchanan (1976)), then those outlining a negative relationship (the "debt overhang" theory by Buchanan and Buchanan (1958), Modigliani (1961), Diamond (1965), and Myers (1977)), then those highlighting a positive relationship (the Keynesian theory in Elmendorf and Mankiw (1999)).

Using a panel of 24 developed nations between 1970 and 2002, Schclarek et al. (2005) finds that public debt doesn't impact GDP per capita growth; the results do not change when the same methodology is applied to a sample of 59 developing countries. Moreover, Dao et al. (2011) analyze 3 different samples and find that the external debt/GDP ratio doesn't affect growth in developing countries, but it harms Highly Indebted Poor Countries (HIPC). According to this study, the public debt/GDP ratio has a positive relationship with growth in developing countries. Panizza and Presbitero (2012) use a sample of OECD countries and their results show neutrality of the ratio of public debt to growth, perhaps because in these economies central banks act as lenders of last resort. Kourtellos et al. (2013), applying a structural threshold methodology and pooled OLS on a sample of 82 countries from 1980 to 2009, assert that debt neutrality is found in countries with good quality institutions.

Several authors estimate a negative impact of debt on economic growth. Gani (1999) uses a sample of 6 Pacific island countries from 1985 to 1992, and affirms that high external debt negatively affects growth because of excessive burden; in these countries, the need for fiscal discipline arises. The same result is achieved by Próchniak (2011), analyzing a sample of 10 CEE countries from 1993 to 2009. Afonso and Jalles (2011) uses both pooled OLS and cross-sectional time-series regressions on a sample of 155 countries from 1970 to 2008 and finds a statistically significant negative relationship between public debt and growth (via productivity). Égert (2015) highlights the same results by analyzing 20 developed countries from 1946 to 2009 (estimating a linear regression with thresholds at 30%, 60%, and 90%). Szabó (2013) argues that the impact in the short run is negative and, in the long run, the effect is negligible, after estimating the regression on 27 European Union countries between 2008 and 2014 (using the forecasted values for economic growth). Using various econometric methodologies on a sample of 111 countries from 1970 to 2010, Ahlborn and Schweickert (2018) find that the negative impact of public debt on growth varies across countries and is larger in continental countries than in liberal ones[3]. Finally, estimating an ARDL on a sample of 11 European countries (both central and peripheral) from 1961 to 2013, Gómez-Puig and Sosvilla-Rivero (2018) support the negativity of the relationship even in the long run. Furthermore, Koroglu (2019) develops a semiparametric smooth coefficient approach on a sample of 82 countries from 1980 to 2009, estimating an average negative effect of public debt on growth. The magnitude of this effect varies depending on the institutional quality.

[3]This distinction concerns the political system.

Opposed results emerge instead from the work by Abbas and Christensen (2010), who use a sample of 93 low-income countries and emerging markets from 1975 to 2004. Moderate domestic debt positively contributes to economic growth when it is scaled to GDP, but a nonlinear relationship is observed when debt is scaled to deposits. The optimal size of domestic debt depends on its composition. The main channel by which debt affects growth is the investment efficiency or factor productivity, rather than the volume of capital accumulation. Greiner (2011) observes that, in economies where the public sector is concerned with progressively decreasing the debt/GDP ratio, GDP growth rates are higher than in other countries. Finally, analyzing a sample of 19 countries from 1991 to 2009, Uzun et al. (2012) find empirical evidence of the positive effect of foreign public debt on the growth rate of GDP per capita.

2.2 Non-linear Relationship

The most recent strand of empirical literature establishes that the relationship between public debt and long-run economic growth is non-linear. We find it useful to cite the leading papers in this area of the literature, although this meta-analysis focuses on the linear relationship. A milestone is the work by Reinhart and Rogoff (2010b) (and also Reinhart and Rogoff (2010a)), who use a sample of 20 developed and 24 emerging countries from 1946 to 2009 and come to assert that the threshold above which the debt-to-GDP ratio hinders growth is 90%. Another finding of this paper is the weakness of the positive effect for debt-to-GDP levels below the threshold. In addition, for emerging countries, the threshold of foreign debt to GDP is lowered to 60%. However, these results have been criticized due to non-standardized weights of summary statistics, some coding errors, and arbitrary exclusion of key countries in the sample used (Herndon et al. (2014), Dafermos (2015)). Herndon et al. (2014) replicate part of the experiment of Reinhart and Rogoff (2010b) (using only the sample of advanced economies) and don't find evidence for non-linearities in the estimates.

Conclusions similar to those of Reinhart and Rogoff (2010b) are instead reached by Woo and Kumar (2015), who establish that on average countries with debt levels less than 30% of GDP grow more than countries in which this ratio exceeds 90%, after analyzing a sample of 38 advanced economies from 1970 to 2008. Cordella et al. (2010) use a sample of 79 developing countries between 1970 and 2002 and establish that in countries with good institutions, the relationship becomes negative when debt exceeds about 23% of GDP. This relationship becomes irrelevant for debt-to-GDP ratios greater than 80%. For countries with poor institutions, the debt threshold is lower (10%) and becomes irrelevant very quickly (15%). The empirical evidence of non-linearity in the relationship between debt and growth and the influence of the quality of institutions on this link is also confirmed by Megersa and Cassimon (2015) with a sample of 57 developing countries from 1990 to 2011 and by Bouchrara et al. (2020), using a sample of 36 countries from 1990 to 2013.

In addition, Clements et al. (2003), using a sample of 55 low-income countries (HIPC) from 1970 to 1999, affirm that when external debt is higher than 30-37% of GDP or 115-120% of exports, it starts to detriment growth. The channel through which debt affects economic growth is via resource use efficiency. Caner et al. (2010) estimate the government debt turning point at 77% of GDP for developed countries, but it lowers at 64% for emerging economies, studying a sample of 75 developing and 26 developed countries from 1980 to 2008. Checherita-Westphal and Rother (2011) find evidence of a non-linear, inverted U-shaped relationship by analyzing a sample of 12 European countries from 1970 to 2011. The relationship is positive for countries with debt-to-GDP ratios below 90-100%. Confidence intervals call for more prudent debt policies, as the turning point can start at 70-80%. The results of Cecchetti et al. (2011) establish that debt no longer positively impacts growth when it exceeds 85% of GDP. The sample used consists of 18 OECD countries from 1980 to 2008. Moreover, Padoan et al. (2012) estimate several panel regressions on a sample of 28 OECD countries from 1960 to 2011, and state that the relationship between government debt and GDP growth is weak when debt is lower than 90% of GDP, while it turns negative over this threshold. Baum et al. (2013) move the value of this threshold to 95% and, using a panel of 12 European countries from 1990 to 2010, also identify the threshold of 67%, around which the relationship between debt and growth begins to change sign (from positive to null). To conclude, Minea and Parent (2012) find that the debt-growth relationship has a U-shape: the relationship is negative for debt-to-GDP levels between 90% and 115% and positive above 115%. The authors use the same sample as Reinhart and Rogoff (2010b), whose variables were estimated from different sources. Evidence for non-linearities also arises in the work by Afonso and Alves (2014), who use a sample of 14 European countries from 1970 to 2012. They find that debt has on average a negative impact on growth, and the debt threshold is estimated at 75% of GDP. Furthermore, Mencinger et al. (2015) use a sample of 31 OECD countries from 1980 to 2010 and 5 non-OECD from 1995 to 2010, and estimate the threshold at 90-94% for developed countries, while it lowers to 44-45% for emerging economies.

Three studies mainly focus on African countries. Megersa (2014) estimates a negative relationship between debt and growth, using a sample of 22 low-income SubSaharian African countries from 1990 to 2011. When non-linear methods are applied, the debt turning point is estimated at 45% of GDP. Similar results concerning non-linearities are achieved by Lartey et al. (2018), adopting a panel OLS and GMM methodology on a sample of 50 African countries from 1980 to 2015. Then, Khanfir et al. (2019) applies a Panel Threshold Regression (PTR) on a sample of 4 North African countries from 2003 to 2012, and fixes the public debt threshold at about 43% of GDP.

More recently, Bökemeier and Clemens (2016) employ a panel of 12 Euro-zone countries and 6 non-Euro but European countries from 1970 to 2014, and find that growth is higher in the first group and if the debt to GDP ratio is below 60%. Wamboye and Tochkov (2016) estimate the external debt turning point around 64-78% of GDP, using a System-GMM approach on a sample of 33 least developed countries from 1970 to 2010. Moreover, estimating a standard linear median regression model on a sample of 20 advanced economies from 1946 to 2009, Lee et al. (2017) fix the debt overhang threshold at 30% of GDP. The debt turning point is indeed estimated at 62% of GDP by Shkolnyk and Koilo (2018), who study a sample of 11 emerging economies from 2006 to 2016. Furthermore, Karadam (2018) employs a Panel Smooth Transition Regression (PSTR) framework on a sample of 24 developed and 111 developing countries from 1970 to 2012, and finds that debt negatively affects growth in developed countries when it exceeds 106% of GDP, while the threshold is lower (88%) for developing countries. Gaies and Nabi (2019) use a sample of 67 low and middle-income countries from 1972 to 2011, affirming that external debt can boost growth through credit and investment, but over a certain threshold of indebtedness, economies become vulnerable to financial crises. The relationship has an inverted U-shape. Finally, An et al. (2020) employ a Panel Transition Regression (PTR) method on a sample of 13 ASEAN and 3 other countries from 2004 to 2015. According to their empirical evidence, debt has a positive impact on growth when it lies between 27% and 72% of GDP for the whole sample, while for only high-income countries, debt is detrimental to growth when it exceeds 66% of GDP.

3 Search Phase

Figure 1, shows the PRISMA flow diagram, following Page et al. (2021). The search for studies to be included in the meta-analysis was carried out in January 2021. We report below, for each search engine, the keywords entered and the respective number of records produced.

- For RePEc search: (“public debt”) AND (“economic growth”) AND (“regression”): 39 records;
- For JStor search: ("economic growth") AND ("public debt") AND ("regression") Filters: Only Journal articles with economic subject: 178 records;
- For Scopus: (“public debt”) AND (“economic growth”) AND (“regression”): 36 records;
- For SSRN: (“public debt”) AND (“economic growth”) AND (“regression”): 16 records.

Thus 269 records were collected, of which 16 were removed as duplicates. The total after the removal of the duplicates is 253.

For the screening phase, the following inclusion criteria were adopted: papers studying regression of economic growth on public debt or external debt, in English, French or Spanish. The exclusion criteria adopted are: papers using other methods (ARDL, VEC, cointegration tests, VAR), or surveying literature or experiments concerning public choice or voting behavior, or theoretical growth models, or concerning the sustainability of public debt, the impact of public debt on spreads and on the interest rate, the fiscal consolidation process, or if public debt is the dependent variable or investigating regional/local debt.

In total, 145 records were removed. The total after the screening phase is 108.

Then, we adopted the following eligibility criteria: studies not including squared debt as a regressor, reporting the partial correlation coefficient, or the t-stat (or standard error and coefficient, or p-value, so the t-stat can be calculated), the degrees of freedom, and the number of observations. Studies excluded meet the following reasons: studies using public expenditure as a covariate, using time series modeling (not specified in the abstract), thus using data for only 1 country, or with the same exclusion criteria applied at the screening phase (but not specified in the abstract), or

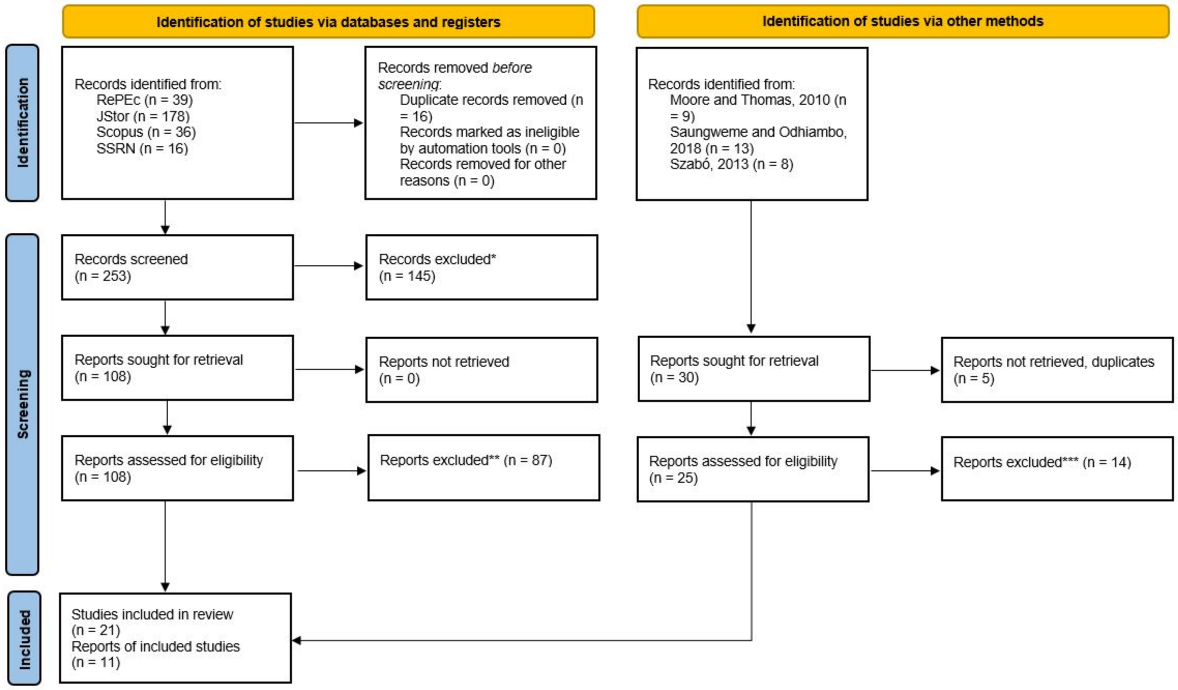

Figure 1: PRISMA Flow Chart

*Inclusion criteria: papers studying regression of economic growth on public debt or external debt, in English, French or Spanish.

Exclusion criteria: papers using other methods (ARDL, VEC, cointegration tests, VAR), or surveying literature or experiments concerning public choice or voting behavior, or theoretical growth models, or concerning the sustainability of public debt, the impact of public debt on spreads and on the interest rate, the fiscal consolidation process, or if public debt is the dependent variable or investigating regional/local debt.

**Eligibility criteria: Studies not including squared debt as a regressor, reporting the partial correlation coefficient, or the t-stat (or standard error and coefficient, or p-value, so the t-stat can be calculated), the degrees of freedom, and the number of observations.

Reasons for exclusion: studies using public expenditure as a covariate, using time series modeling (not specified in the abstract), thus using data for only 1 country, or with the same exclusion criteria applied at the screening phase (but not specified in the abstract), or estimating the Ricardian equivalence, or estimating the impact of debt growth (Δ d) on output growth, or using deficit.

***Eligibility criteria: Studies not including squared debt as a regressor, reporting the partial correlation coefficient, or the t-stat (or standard error and coefficient, or p-value, so the t-stat can be calculated), the degrees of freedom, and the number of observations.

Reasons for exclusion: studies using public expenditure as a covariate, using time series modeling (not specified in the abstract), thus using data for only 1 country, or with the same exclusion criteria applied at the screening phase (but not specified in the abstract), or estimating the Ricardian equivalence, or estimating the impact of debt growth (Δ d) on output growth, or using deficit, or estimating the impact of debt on the interest rate.

The flow diagram is drawn following Page et al. (2021).

estimating the Ricardian equivalence, or estimating the impact of debt growth (Δ d) on output growth, or using deficit.

87 records were then removed. The total after the eligibility phase is 21.

Simultaneously, 9 papers from Moore and Thomas (2010), 13 from Saungweme and Odhiambo (2018) and 8 from Szabó (2013) were included, summing up to 30 additional studies. The first source is a meta-analysis, the second is a literature review and in the third, we found a very good literature review including papers that had not been collected before[4].

The same screening and eligibility criteria were applied and finally, 1 paper were selected from Moore and Thomas (2010), 6 from Saungweme and Odhiambo (2018) and 4 from Szabó (2013), for a total of 11 additional papers (listed in Table 15 of Appendix A).

The total number of papers to be coded is 32. The list of coded studies with the associated number of effect size observations per study is available in Table 14 in Appendix A.

4 Coding Phase

We suppose to study regressions of the type:

$$EG_{i,j} = \delta_j + \beta_i PD_{i,j} + \sum_{k=1}^{K} \gamma_{i,k} X_{i,k,j} + \epsilon_i, \tag{1}$$

in which public debt (PD) is used as the explanatory variable and economic growth (EG) as the dependent variable, $i = 1, ..., N$ defines the generical i^{th} study, $j = 1, ..., J$ is the country index and $k = 1, ..., K$ is the index related to other explanatory variables.

The effect size we measure is the PCC, calculated as follows:

$$PCC_i = \frac{t_i}{\sqrt{(t_i^2 + df_i)}}, \tag{2}$$

in which t is the t-stat associated with the i^{th} β coefficient estimated from (1), df are the corresponding degrees of freedom[5], and $i = 1, ..., N$, in which N is the total number of PCC collected[6].

We use this proxy as an effect size, as already used for example by Xue et al. (2021), although the coefficient is neither mathematically interpretable as the coefficients of standard regressions, nor as an elasticity.

However, we prefer it because it also makes the results of regressions with different degrees of freedom comparable and allows us to expand the range of studies that can be included in our meta-database.

[4]Studies from Heimberger et al. (2021) could not be included because the cited work was published in November 2021, while our coding phase was conducted in January 2021. In the future, they can be integrated to extend this meta-dataset.

[5]If the estimation method is fixed effects, the country dummies are not included in the degrees of freedom.

[6]Note that the i index in (2) is different from the index in (1). From each study, in fact, a different amount of PCCs can be encoded.

As the range of values that the PCC can assume varies between -1 (maximum negative relationship) and 1 (maximum positive relationship), the Standard Error is calculated as follows:

$$SE(PCC_i) = \sqrt{\frac{(1 - PCC_i^2)}{df_i}} = \frac{PCC_i}{t_i}. \tag{3}$$

If t-stat is not reported, we use the p-value, applying *t.inv(p,value,dof)=t* (corrected for the β coefficient sign) function for Excel. If the p-value is not reported, we can use the standard error and the β coefficient, applying the following:

$$t_i = \frac{\beta_i - 0}{SE(\beta_i)}. \tag{4}$$

If t is not reported, but PCC is, we use the inverse formula:

$$t_i = \frac{\sqrt{PCC_i^2 \cdot dof_i}}{\sqrt{(1 - PCC_i)^2}}. \tag{5}$$

If df are not reported, we use the number of regressors or the first parameter of the F test (if published). Finally, if the PCC is reported, we use it[7].
Moreover, if the p-value is not reported, we use the *distrib.t.2t(value,df)=p-value* function for Excel.
If n is not reported, we impute it as:

$$n_i = nc_i \cdot \frac{ny_i}{lt_i}, \tag{6}$$

where nc is the number of countries, ny is the number of years, and lt is the length of the time span.
Table 16 in Appendix B displays the definitions of all the coded variables, along with the distinction of categorical and continuous moderators.
The analysis of the moderators, their distributions, and descriptive statistics will be shown later. Therefore, we code 11 categorical moderators and 7 continuous moderators.
Our dataset is composed of 422 observations collected from 32 studies. Figure 2 shows the histogram of the PCC and the approximate distribution. The graph shows that the distribution is double-picked (with a big peak around -0.7 and a second little peak around 0.5) and that most observations have a negative value. We, therefore, expect the average effect size to be negative.

[7]The only study reporting this statistic is Próchniak (2011).

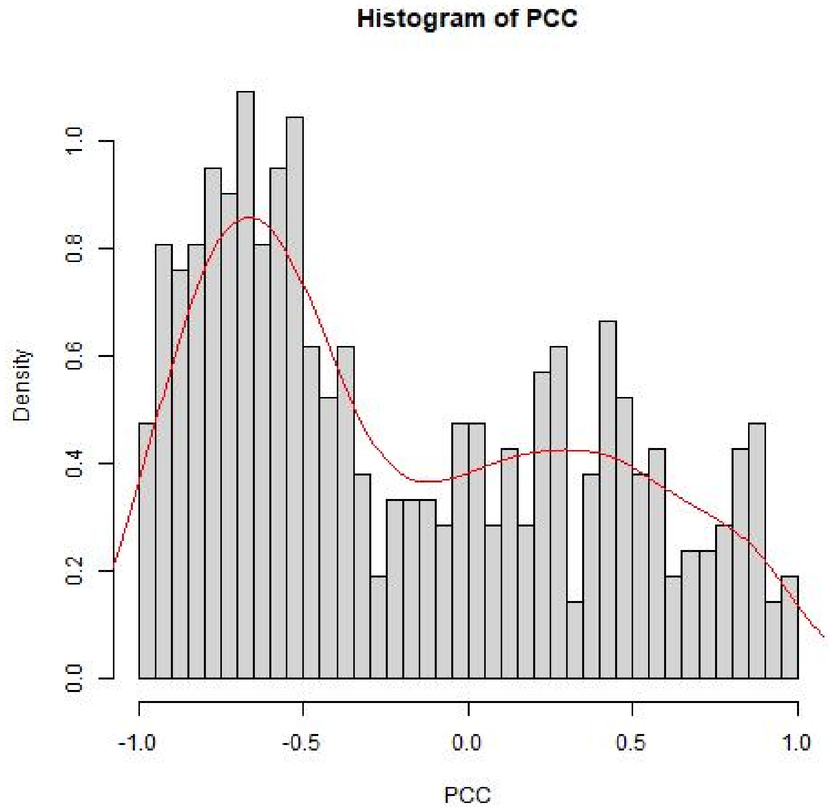

Figure 2: Distribution of PCC

This Figure shows the histogram of the PCC, represented by the gray bars with the approximate distribution of the variable (the red line).

5 Meta-Analysis Results

5.1 Fixed and Random Effects Analysis

To analyze the expected value of the PCC, we use both fixed effects and random effects model analysis (Harrer et al. (2021)). For a complete description of the model, see Appendix C.

Table 1 shows the effect size estimates (in our case, the PCC) for the fixed effects (columns 2 and 4) and random effects (columns 3 and 5) models. The Table also shows the extremes of the confidence interval for this estimate, the value of the z-statistic, and the p-value. In this research, we develop a multi-level analysis, first using the database containing all PCC observations for each study ($K = 422$), then the mean PCC for each study (one observation per study, so $K = 32$), and finally analyzing the entire sample but taking into account the study to which the PCCs belong (column 6).

Looking at the entire sample, both the estimates obtained with the fixed effects model and the random effects model present a negative value of the PCC, around -0.3 (consistent with our expectations concerning the histogram in Figure 2); both estimates are statistically significant. However, when we only consider the mean PCC for each study, we obtain different results, both statistically significant: the fixed effects model returns a positive (around 0.1) estimate, while the random effects model estimates a negative effect (-0.15). Finally, the multi-level analysis estimates a negative PCC (around -0.2). Since the multi-level analysis weighs each PCC by the number of observations belonging to the same study, we rely mainly on this method and can conclude that the relationship is negative on average. However, it is important to specify that in this meta-analysis we have focused only on studies performing linear regressions: we cannot at this stage determine whether there is nonlinearity in the relationship (and what the debt turning point is). With these estimates, we simply state that, in the sample of linear relationships analyzed, on average the PCC takes on a value around -0.2. This result is also in line with the mean effect size sign obtained by Heimberger et al. (2021), even if a different proxy is used.

Table 1: Estimated Effect Size

	$K = 422$		$K = 32$		
Model	Fixed Effects	Random Effects	Fixed Effects	Random Effects	Multi-level
PCC Estimate	-0.366***	-0.221***	0.094***	-0.152*	-0.193***
Lower Bound 95% C.I.	-0.389	-0.278	0.079	-0.315	-0.322
Upper Bound 95% C.I.	-0.344	-0.165	0.109	0.011	-0.063
z	-31.95	-7.63	11.99	-1.82	-
p-value	0.000	0.000	0.000	0.068	0.004

This Table shows the average PCC estimate, the 95% confidence intervals, the z statistics, and the p-value for the PCC. Estimates are shown for the whole sample ($K = 422$) with fixed effects and random effects in columns 2 and 3. Columns 4 and 5 reports the estimates with fixed effects and random effects for the reduced sample ($K = 32$), in which the average PCC per study is coded. In column 6, the estimates for the multi-level analysis are shown.
*** p<0.01, ** p<0.05, * p<0.1

5.2 Heterogeneity and Variance Components Analysis

One of the main purposes of this research is to identify whether heterogeneity is present in the analyzed studies and, if so, which moderators can mitigate it. This concept can have two facets, according to Rücker et al. (2008). Baseline or design-related heterogeneity occurs when the population or research design of studies differs between studies. Statistical heterogeneity, on the other hand, is quantifiable and influenced by the spread and precision of the effect size estimates coded in a meta-analysis. In the literature, several statistics are available that can quantify heterogeneity in meta-analyses. For a brief description of these statistics, see Appendix D.

Table 2 shows the results of the heterogeneity analysis for the 2 levels of sample aggregation. For the τ^2, τ, and H statistics, the extremes of the confidence interval are also reported, and for the Q, the p-value is reported.

We first note that both Q statistics are highly significant and that they have values indicating the presence of heterogeneity across studies. Moderators' analysis is needed to investigate the causes of this heterogeneity. Furthermore, by comparing the two values of the I^2 statistic with the rule of thumb by Higgins and Thompson (2002), we can state that heterogeneity is substantial. Finally, comparing the values of the statistics in the two samples, the sample in which each PCC of each study represents an observation has more heterogeneity than the sample in which only the mean PCC for each study is used. The more detailed degree of information provided by the larger sample thus increases the heterogeneity of the effect size, which was already significantly present in the sample with $K = 32$.

Table 2: Heterogeneity Estimates

	Statistics	Value	Lower bound 95% C.I.	Upper bound 95% C.I.	p-value
$K = 422$	τ^2	0.716	0.627	0.823	-
	τ	0.846	0.792	0.907	-
	I^2	0.998	-	-	-
	H	21.050	20.840	21.270	-
	Q	186,583.430	-	-	0.000
$K = 32$	τ^2	0.143	0.059	0.249	-
	τ	0.378	0.243	0.499	-
	I^2	0.822	0.656	0.889	-
	H	5.606	2.905	9.039	-
	Q	172.160	-	-	0.000

This Table shows the values of the main statistics for heterogeneity analysis (τ^2, τ I^2, H, and Q), the extreme bounds of the 95% confidence interval and the p-values when calculated. Rows 2-6 show the estimates for the whole sample ($K = 422$), while rows 7-11 show the estimates for the reduced sample ($K = 32$).

We analyze the random effects variance components, performing a multivariate meta-analysis model.

We assume that every study represents a cluster. The between-cluster variance, which is equivalent to the between-study heterogeneity variance in a conventional meta-analysis, is 0.0951. The within-cluster variance is 0.1709. We can state that the within-study heterogeneity is more important than the between-study, even if both types of variability strictly affect heterogeneity.

As Figure 15 in Appendix E also shows, Level 1 variance, i.e., sampling error variance, counts for about 17% of the total. Most of the variability in our sample is not attributable to sampling error: the within-studies (Level 2) variance counts for 53%, while the between studies (Level 3) variance counts for about 29% of the total.

5.3 Forest Plot

Figure 3 shows the forest plot representing each study's average PCC and confidence interval, ordered in descending order according to the average effect size.

The effect sizes present in the studies analyzed are very heterogeneous, confirming the previous analysis since the average values reported in the Figure vary from a maximum of about 1 to a minimum of about -1. On the whole, if only the average PCCs are observed, on 7 studies these are equal to 0, on 6 they are positive, and on the remaining 19, they are negative.

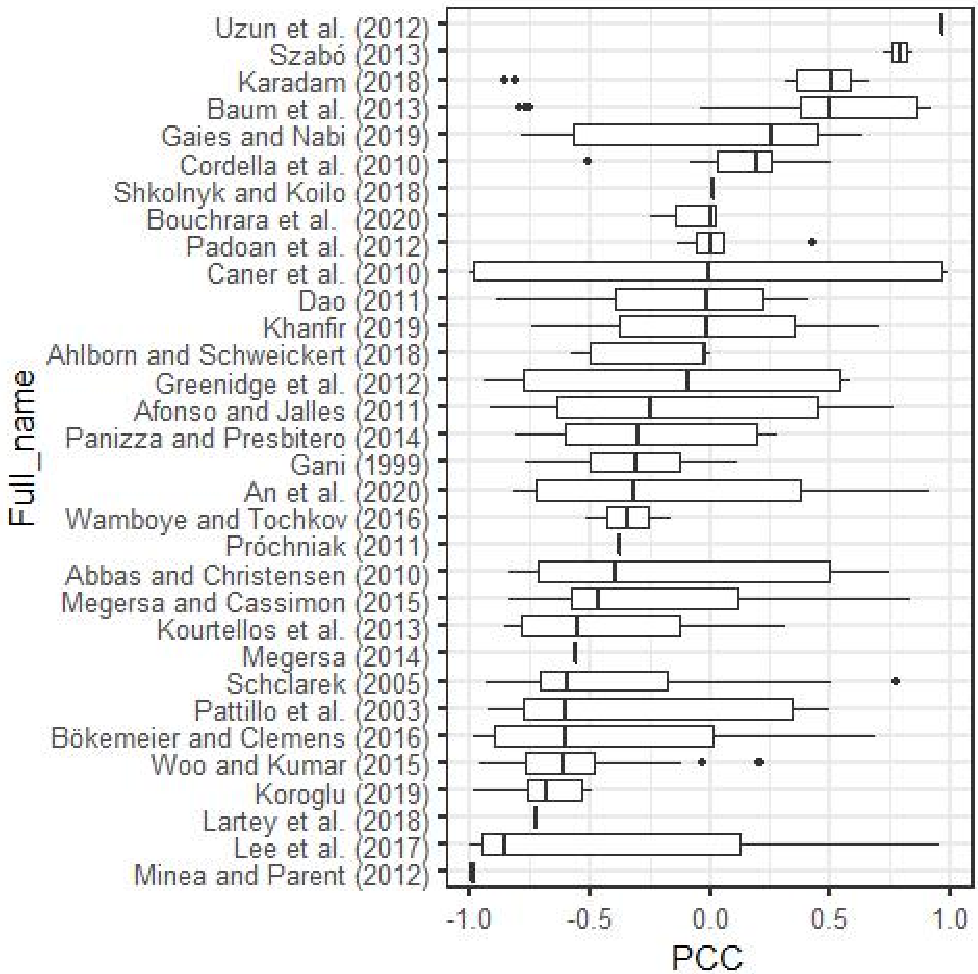

Figure 3: Forest Plot

This Figure shows the PCC forest plot. For each study, the PCC boxplot is represented. Studies are sorted in descending order based on the average PCC.

The studies also present different degrees of variability, since there are boxplots that are concentrated around the mean (such as Szabó (2013), Padoan et al. (2012), or An et al. (2020)) and some that instead vary from highly positive to highly negative values of the PCC, thus suggesting that the effect size is close to zero (for example, Caner et al. (2010), Greenidge et al. (2012), Gaies and Nabi (2019), or Abbas and Christensen (2010)). This means that some studies produce results more precise than others. For some study, we code only one observation, as in Uzun et al. (2012), Shkolnyk and Koilo (2018), Próchniak (2011), Lartey et al. (2018), and Minea and Parent (2012). In addition, the following studies report outliers: Karadam (2018), Baum et al. (2013), Cordella et al. (2010), Padoan et al. (2012), Schclarek et al. (2005), and Woo and Kumar (2015).

5.4 Moderators Analysis

In this Subsection, we analyze the coded moderators. In Subsubsection 5.4.1 we study the categorical moderators, while the continuous ones are analyzed in Subsubsection 5.4.2.

5.4.1 Categorical Moderators

We now turn to analyzing the categorical moderators on the extended dataset ($K = 422$). The list of moderators can be found in Table 16 of Appendix B. In Appendix F, Figures 16, 17, and 18 show the barplots for each categorical moderator and the relative comments.

Here, we analyze the distributions of PCCs under each moderator category. The densities are plotted in Figures 4, 5, and 6.

Almost all distributions regarding the dependent variable are double-peaked: in particular, the main peak is negative when the proxy used is GDP per capita growth rate or other measures, while it is positive when GDP per capita level is used.

For the independent variable, the PCC tends to take on negative values when the debt to exports ratio is considered, while it is double-peaked when the log of debt over GDP is used.

The distribution of PCC is double-peaked also concerning the focus on non-linearities: when these are not present the modal values are closer to 0. When these are considered, the biggest peak is negative.

Instead, the peak is only one when considering the publication status of the study: in both cases, the PCC tends to have a negative value.

In addition, PCC values are predominantly negative if the observations are from working papers or peer-reviewed journals, while they tend to be positive if they are from institutional staff papers.

The distribution is bimodal if the countries considered are developed or developing, with the first category tending toward positive values and the second toward negative values. Samples that consider both types of countries lead mainly to negative estimates.

Moreover, almost all estimation methods tend to produce negative PCCs, except for IV, which leads to a PCC estimate centered around the 0.2 value.

Samples that consider European countries estimate more positive effect size values, as opposed to samples that include Africa, Asia, or more than one region of the world.

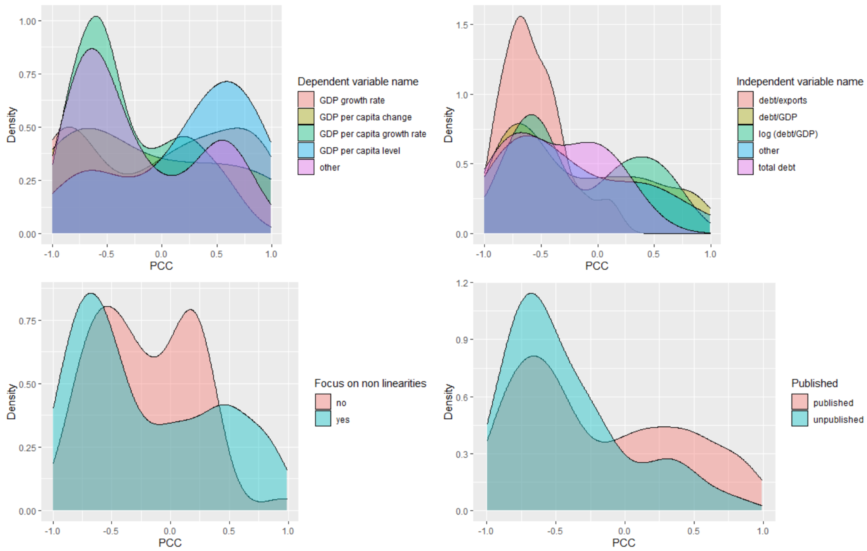

Figure 4: PCC Distribution by Categorical Moderator (1/3)

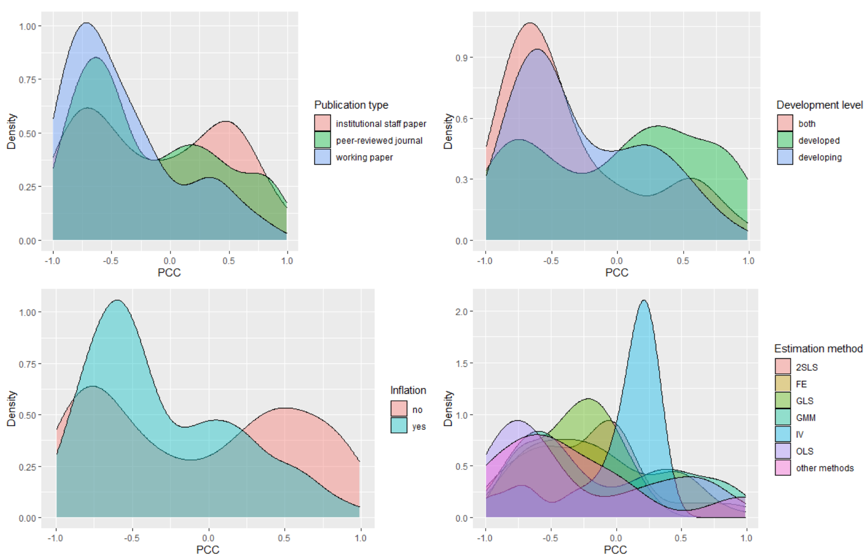

Figure 5: PCC Distribution by Categorical Moderator (2/3)

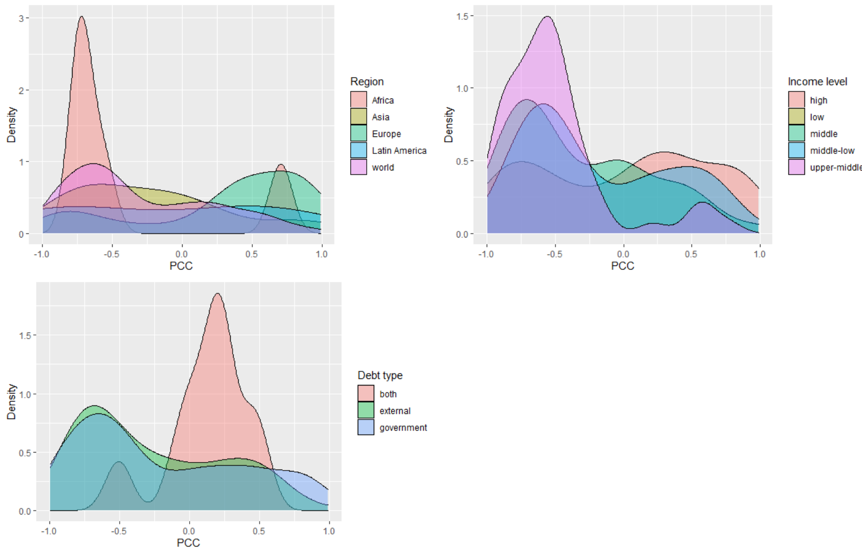

Figure 6: PCC Distribution by Categorical Moderator (3/3)

These Figures show the approximate PCC density for each category of each categorical moderator. Every category within the same moderator has a color. In particular, Figure 4 shows the distributions for the moderators "dependent variable name", "independent variable name", "focus on non linearities", and "published". Figure 5 shows the distributions for the moderators "publication type", "development level", "inflation", and "estimation method". Figure 6 shows the distributions for the moderators "region", "income level", and "debt type".

In addition, the PCC estimate is mostly positive when considering high-income countries, while it is mostly negative for the rest of the sample.

Finally, when the government or external debt are considered separately, the distribution is bimodal, with the main peak around -0.6, while when these are considered jointly, the estimated PCC tends to be positive (around 0.2).

To conclude the analysis of categorical moderators, we estimate a meta-regression for each of these, according to the following formula:

$$PCC_i = \beta_0 + \sum_{j_\ell=2}^{J_\ell} \beta_{j_\ell} \cdot X_{j_\ell,i} + \epsilon_i \tag{7}$$

, where X_{j_ℓ} is the j-th generic moderator, whose categories vary for $j_\ell = 1, ..., J_\ell$, and β_{j_ℓ} is the regression coefficient associated to each category. In this case, we identify $J = 11$ categorical moderators, but each of these has a different number of categories J_ℓ. The observation index is $i = 1, .., K$; we consider the whole sample, thus $K = 422$.

Table 3: Results of Single Regressions for Categorical Moderators (1/2)

Moderator	σ_1^2	σ_2^2	Q	F
Dependent variable	0.0977	0.1701	3125.2840	1.0937
			(0.0000)	(0.3592)
Independent variable	0.1122	0.1677	3174.2350	1.4776
			(0.0000)	(0.2081)
Focus on non linearities	0.1007	0.1705	3209.7069	0.1730
			(0.0000)	(0.6776)
Published	0.0956	0.1708	3206.7689	1.0461
			(0.0000)	(0.3070)
Publication type	0.0942	0.1705	3177.5665	1.2563
			(0.0000)	(0.2858)
Countries type	0.0989	0.1655	2966.5966	5.7771***
			(0.0000)	(0.0033)
Inflation	0.0939	0.1711	3058.2002	1.1683
			(0.0000)	(0.2804)
Estimation method	0.1160	0.1621	2878.5304	2.7107**
			(0.0000)	(0.0136)
Region	0.0855	0.1667	2595.7806	3.7957***
			(0.0000)	(0.0048)
Income level	0.1163	0.1640	2830.8386	3.0979**
			(0.0000)	(0.0157)
Debt type	0.1073	0.1672	3179.8918	2.5348*
			(0.0000)	(0.0805)

This Table shows the values of the σ_1^2, σ_2^2, Q, and F statistics in the individual regressions for each categorical moderator. P-values are in brackets.
*** p<0.01, ** p<0.05, * p<0.1

In this type of regressions, the intercept β_0 represents the estimated effect size for the subset of observations whose category is $j_\ell = 1$ of the j-th moderator[8]. In this study, the categories are sorted alphabetically, so the first coefficient represents the PCC estimate for the subgroup with the first category for each moderator. Instead, the estimate of the generic coefficient β_{j_ℓ}, represents the effect size difference between the subgroup of the first category and that of the j_ℓ-th category. Therefore, to compute the estimated PCC for the j_ℓ-th subgroup, one must sum the generic coefficient β_{j_ℓ} and the intercept β_0.

In Tables 3 and 4, we report the estimates obtained by performing a multi-level meta-regression for each categorical moderator. In particular, Table 3 shows the values of the variance between-studies σ_1^2 (column 2), the variance within-studies σ_2^2 (column 3), Cochran's Q (column 4), and Fisher's F (column 5); for the latter two statistics, we also report the relative p-values. In contrast, Table 4 shows the regression coefficient estimates with associated p-values for each categorical moderator. Analyzing the results in Table 3, we observe that Q has very high values for all meta-regressions.

[8]This is why the summation in (7) is calculated for $j_\ell = 2, ..., J_\ell$.

Heterogeneity is strongly present. For all moderators, the variance between-studies is smaller than the variance within-studies, as in Appendix E. The moderators that explain the most heterogeneity in PCC estimates (i.e., those with the lowest Q values) are also those for which the meta-regression is statistically significant at the $\alpha = 0.01$ level. To this selection of categorical moderators belong the region analyzed in the sample, the income level of the countries considered, the estimation method used, and the development level of the countries in the sample. Another statistically significant meta-regression at the $\alpha = 0.1$ level relates to the debt type.

We move on to analyze the coefficients of the meta-regressions, shown in Table 4. Regarding the proxy used as the dependent variable, the estimated average PCC when this is the GDP growth rate is -0.3; this value remains negative when GDP per capita change, GDP per capita growth rate, or other proxies are considered. On the other hand, when the GDP per capita level is used, the estimated PCC is positive at about 0.3. Therefore, the choice of the proxy for the dependent variable influences the link between debt and economic growth.

Concerning the proxy for the independent variable, when debt is scaled to exports, the average PCC is -0.4. The same is true when considering total debt or other measures, while, when debt is scaled to GDP or the log of this is used, the PCC approaches -0.2.

For published studies, the average correlation coefficient is -0.2, while it is more negative for unpublished studies.

Moreover, for samples that include both developed and developing countries, the effect size is negative (-0.3), while it becomes slightly positive when only developed countries are considered.

Among the estimation methods, the one that leads to higher PCCs is IV (0.2); all other methods lead to negative results, ranging from -0.1 in the case of FE to -0.35 in the case of GLS. This result is in line with Heimberger et al. (2021), as the IV estimation method tackles endogeneity. When endogeneity is taken into account in the estimates, the effect size is larger than the average effect size estimated in Table 1.

The region considered in the sample also leads to different estimation results for the link between public debt and growth. In the samples that include European countries, the estimated PCC is about 0.3, while for all other samples it is negative.

Income level appears to be another significant moderator, as the effect size is slightly positive in samples from high-income countries, while it decreases when the income level lowers.

Finally, the focus on non-linearities, the publication type of the study, the use of inflation as covariates, and the type of debt used do not lead to statistically significant results. Thus, these moderators do not allow us to lower the level of heterogeneity among the studies considered.

Table 4: Results of Single Regressions for Categorical Moderators (2/2)

Moderator	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Dependent variable	GDP growth rate	GDP per capita change	GDP per capita growth rate	GDP per capita level	other		
	-0.2681**	0.0570	0.0512	0.5922**	0.1095		
	(0.0248)	(0.8797)	(0.7223)	(0.0442)	(0.5226)		
Independent variable	debt/exp	debt/GDP	log(debt/GDP)	other	tot debt		
	-0.4430***	0.2906*	0.2643	0.0497	-0.0507		
	(0.0085)	(0.0577)	(0.2395)	(0.8136)	(0.8966)		
Focus on non linearities	no	yes					
	-0.1187	-0.0845					
	(0.5324)	(0.6776)					
Published	published	unpublished					
	-0.1764***	-0.2955					
	(0.0098)	(0.3070)					
Publication type	institutional staff paper	peer-reviewed journal	working paper				
	-0.1664	0.0079	-0.3616				
	(0.2191)	(0.9599)	(0.1642)				
Countries type	both	developed	developing				
	-0.3182***	0.3446***	0.0304				
	(0.0007)	(0.0018)	(0.7669)				
Inflation	no	yes					
	-0.1163	-0.1354					
	(0.2284)	(0.2804)					
Estimation method	2SLS	FE	GLS	GMM	IV	OLS	other
	-0.1868	0.0768	-0.1762	0.0580	0.4078**	-0.0953	-0.0351
	(0.2402)	(0.6425)	(0.6774)	(0.7202)	(0.0450)	(0.5548)	(0.8576)
Region	Africa	Asia	Europe	Latin America	world		
	-0.4193	0.1348	0.7434**	0.1901	0.1653		
	(0.1320)	(0.7159)	(0.0187)	(0.6861)	(0.5650)		
Income level	high	low	middle	middle-low	upper-middle		
	0.0638	-0.6223	-0.3855***	-0.3192**	-0.2075		
	(0.5461)	(0.3055)	(0.0007)	(0.0369)	(0.1263)		
Debt type	both	external	government				
	0.1357	-0.4797	-0.2915				
	(0.7087)	(0.2025)	(0.4319)				

This Table shows the coefficient estimates for the regressions for each categorical moderator. P-values are in brackets. *** p<0.01, ** p<0.05, * p<0.1

5.4.2 Continuous Moderators

In this Subsection, we analyze continuous moderators. Seven were identified, whose descriptive statistics for the entire sample ($K = 422$) are given in Table 17 in Appendix G. Figure 7 shows the distribution of each moderator on the main diagonal, the correlation coefficients for pairs of moderators in the upper right triangle, and the scatterplots with the lines that best approximate the relationships for pairs of moderators in the lower left triangle.

It is worth noting that the variables publication year and last year are highly correlated since generally the last data available at the end of the sample is used in each study. In addition, there is a very strong negative correlation between the first year of the sample and the number of years considered: this is because the variable number of years is constructed as the difference between the last and the first year plus one[9].

Looking at the main diagonal, we note that the oldest paper was published in 1999 and the most recent in 2020. Most of these are concentrated in the period 2010-2015 (following seminal papers by Reinhart and Rogoff (2010b) and Reinhart and Rogoff (2010a)), just as the problem of bursting public debts begins to cause concern among institutions and academics in more developed economies. The earliest observation in time dates back to 1880 and the last to 2008. A high percentage of studies investigate samples beginning in the 1960s and ending just before 2010. The most recent final data point was in 2016. We analyzed studies with samples ranging from 2 to 130 years (mostly between 20 and 40) and including 3 to 135 countries (many include about 20). In almost all regressions, we do not consider the lagged debt, in some cases, this can be up to 3 years, in very few even up to 33. Finally, the simple impact factors ranged from 0 to 25[10].

[9] The results for the coefficients of these two variables in the meta-regression in Section 5.5 could therefore be biased, so we supplement with a multi-model inference analysis to capture the best predictors among the moderators.

[10] We impute 0 to non-ranked journals or unpublished papers.

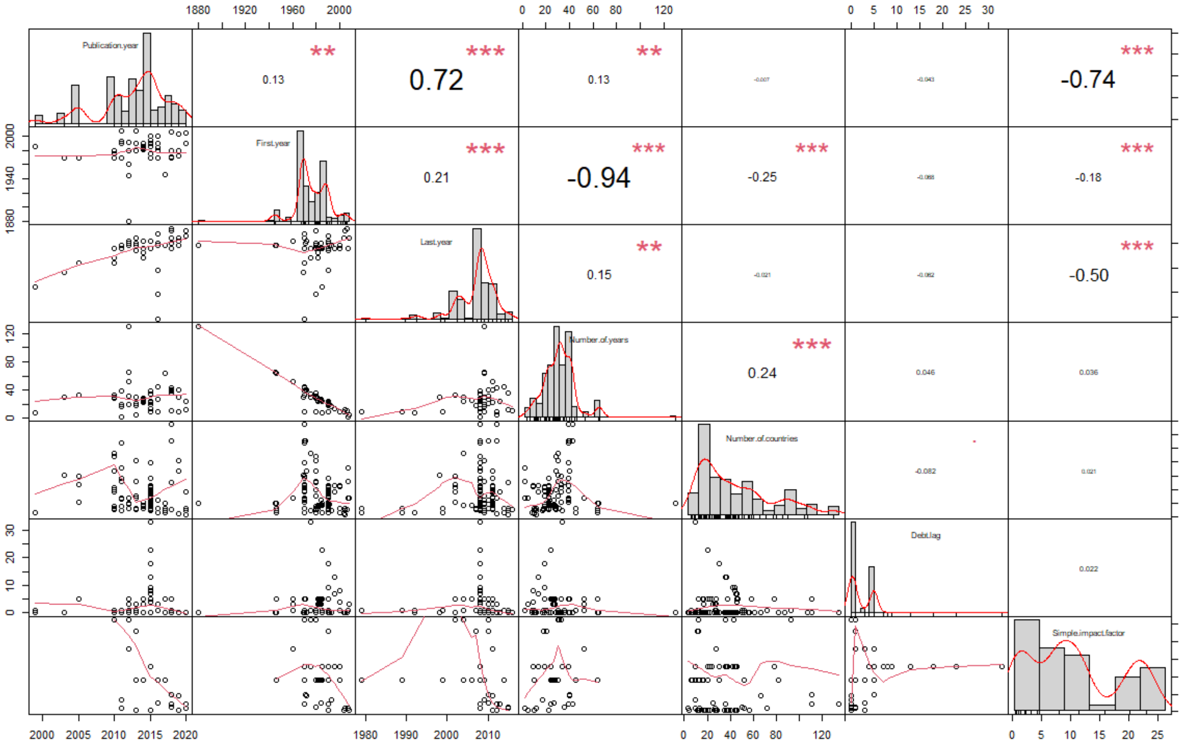

Figure 7: Correlations among Continuous Moderators

This Figure shows on the main diagonal the histograms and approximate distributions of the continuous moderators. Starting in the upper right corner, these are "publication year", "first year", "last year", "number of years", "number of countries", "debt lag", and "simple impact factor". In the upper left triangle of the main diagonal, the correlation coefficients for each pair of moderators are shown. In the bottom right triangle of the main diagonal, the scatterplots and lines that best approximate the relationship between pairs of moderators are represented.

Table 5 shows the values of the coefficients of the multi-level meta-regressions for each continuous moderator (columns 2-3), with the relative value of the variance between-studies σ_1^2 (column 4) and within-studies σ_2^2 (column 5), the Q (column 6) and the F (column 7). For the latter two statistics and coefficients, p-values are also reported[11]. When running regressions with continuous moderators, the estimated equation is as follows:

$$PCC_i = \beta_0 + \beta_j X_{j,i} + \epsilon_i, \tag{8}$$

where β_0 represents the value of the effect size when $X_j = 0$, and the generic coefficient for the j-th moderator β_j represents the slope of the regression.

In this case, the variance between-studies is smaller than the variance within-studies. Heterogeneity is again present, and the moderators that best identify it are the journal ranking, the number of years considered, and the first year. These also present significant F statistics at the $\alpha = 0.1$ level, as does the number of countries considered. The study of the other moderators (publication year,

[11] Since each regression has only one coefficient (beyond the intercept), the p-value for β_1 is the same as Fisher's F.

last year, and debt lag) does not show statistically significant results.
In particular, as the first year of the sample is more recent, the PCC between debt and growth tends to increase, whereas it tends to decrease as the sample size increases both temporally and spatially. The average PCC tends to decrease as the ranking of the journal in which it is published increases. The average PCC is 0.2 for unpublished studies or papers in unranked journals. Moreover, when the debt lag is 0 (thus, for contemporaneous relationships) the average PCC is about -0.2.
Finally, Figures 19 and 20 in Appendix G show the scatterplots between PCC and each continuous moderator with the corresponding regression lines in red.

Table 5: Results of Single Regressions for Continuous Moderators

Moderator	β_0	β_1	σ_1^2	σ_2^2	Q	F
Publication year	-2.1697	0.0010	0.1000	0.1708	3,214.4739	0.0053
	(0.9363)	(0.9419)			(0.0000)	(0.9419)
First year	-14.1181**	0.0070**	0.0800	0.1703	2,983.8933	5.8786**
	(0.0144)	(0.0157)			(0.0000)	(0.0157)
Last year	9.9689	-0.0051	0.1003	0.1704	3,214.5083	0.3280
	(0.5745)	(0.5672)			(0.0000)	(0.5672)
Number of years	0.0229	-0.0072	0.0828	0.1696	2,978.1696	6.3309**
	(0.8298)	(0.0122)			(0.0000)	(0.0122)
Number of countries	-0.0861	-0.0025*	0.1060	0.1693	3,191.1900	3.1148*
	(0.3470)	(0.0783)			(0.0000)	(0.0783)
Debt lag	-0.1809***	-0.0086	0.0930	0.1716	3,130.8807	0.5769
	(0.0074)	(0.4479)			(0.0000)	(0.4479)
Simple impact factor	0.2449**	-0.1045***	0.1322	0.1400	2,789.3876	49.8375
	(0.0117)	(0.0000)			(0.0000)	(0.0000)

This Table shows the coefficient estimates, and the values of the σ_1^2, σ_2^2, Q, and F statistics for the individual regressions of each continuous moderator. P-values are in brackets.
*** p<0.01, ** p<0.05, * p<0.1

5.5 Meta-Regression

Table 6: Heterogeneity Estimates of the Meta-Regression

Statistics	τ^2	τ	I^2	H	R^2	Q	F
Value	0.1435	0.3789	0.70	3.37	47.07	1,525.8198***	7.0231***
p-value	-	-	-	-	-	(0.0000)	(0.0000)

This Table shows the main statistics for the comprehensive model meta-regression. P-values are in brackets.
*** p<0.01, ** p<0.05, * p<0.1

In this Subsection, we perform several meta-regressions for the entire sample ($K = 422$) including all moderators, both categorical and continuous. In Subsubsection 5.5.1, we perform multimodel inference. In Subsubsection 5.5.2, we show BMA estimates of the meta-regression, while in Subsub-

section 5.5.3, we analyze the results of the OLS stepwise procedure.
Table 6 reports the estimates for heterogeneity (τ^2, τ, I^2, H, Q) and significance of the meta-regression (R^2, F).
It can be seen that, compared to the relative estimates in Table 2 in Subsection 5.2, heterogeneity has decreased. Thus, the study of these moderators, taken altogether, allows us to clarify what affects the variability in the PCC estimates between public debt and economic growth. Moreover, F shows that the regression is significant at the $\alpha = 0.01$ level, an expected result since 18 moderators are included.
The results of the meta-regression are similar but not perfectly aligned with those of the previous Subsections. We only comment on those that are statistically significant at least at the $\alpha = 0.1$. level.
The estimate of PCC to baseline (intercept) decreases when GDP per capita growth rate is taken as the dependent variable and it increases when the debt-to-GDP ratio or the log of this ratio is used as the independent; it also increases as the value of the first year of the sample increases and as the number of countries included in the sample increases. The coefficient is positive for unpublished studies but negative for working papers. The study of European samples increases the PCC to baseline. To conclude, PCC is negatively correlated with studies that consider either external or government debt individually and it decreases as the simple impact factor increases.

Table 7: Meta-Regression Results (Comprehensive Model)

Variable	Estimate	SE	t	p-value
Intercept	-14.2541	25.6483	-0.5558	0.5787
Publication year	0.0031	0.0155	0.1982	0.8430
GDP per capita change	0.0568	0.3496	0.1626	0.8709
GDP per capita growth rate	-0.3250***	0.1176	-2.7644	0.0060
GDP per capita level	0.2613	0.2109	1.2392	0.2160
Other dep. variables	-0.2313	0.1541	-1.5015	0.1340
Debt/GDP	0.2627*	0.1490	1.7628	0.0787
Log(debt/GDP)	0.3166*	0.1817	1.7424	0.0822
Other indep. variables	-0.0430	0.2219	-0.1938	0.8464
Total debt	-0.1592	0.4390	-0.3626	0.7171
First year	0.0114***	0.0031	3.6493	0.0003
Last year	-0.0071	0.0121	-0.5845	0.5592
Number of countries	0.0040**	0.0018	2.2356	0.0259
Debt lag	-0.0041	0.0112	-0.3705	0.7112
Focus on non linearities	-0.1662	0.1663	-0.9993	0.3183
Unpublished	0.9093***	0.3126	2.9093	0.0038
Peer-reviewed journal	-0.0944	0.1385	-0.6813	0.4961
Working paper	-1.0785***	0.2730	-3.9506	0.0000
Developed	0.2955	0.2615	1.1304	0.2590
Developing	0.1022	0.1493	0.6843	0.4942
Inflation	-0.1699	0.1133	-1.4995	0.1346
FE	0.0483	0.1516	0.3188	0.7501
GLS	-0.0088	0.5099	-0.0173	0.9862
GMM	0.0184	0.1492	0.1230	0.9022
IV	0.2075	0.1873	1.1080	0.2685
OLS	-0.1022	0.1433	-0.7135	0.4760
Other methods	0.1782	0.1748	1.0196	0.3086
Europe	0.6079**	0.2919	2.0824	0.0380
Latin America	0.0309	0.4759	0.0649	0.9483
World	0.2492	0.2754	0.9048	0.3661
Low income	0.2876	0.6546	0.4394	0.6606
Middle income	0.1176	0.2354	0.4995	0.6177
Middle-low income	0.1157	0.2766	0.4184	0.6759
Upper-middle income	0.1109	0.2364	0.4690	0.6394
External debt	-0.5920**	0.2396	-2.4709	0.0139
Government debt	-0.5296**	0.2289	-2.3137	0.0212
Simple impact factor	-0.0632***	0.0152	-4.1513	0.0000

This Table shows the coefficient estimate, the standard error, the t statistics and the p-value for the comprehensive model meta-regression.
*** p<0.01, ** p<0.05, * p<0.1

5.5.1 Multimodel Inference

Table 8: Multimodel Inference Coefficients

	Coefficient Estimate	SE	z	p-value
Intercept	-84.3321	82.0796	1.0274	0.3042
External debt	-0.5882**	0.2840	2.0710	0.0384
Government debt	-0.1473	0.2709	0.5438	0.5865
Developed	-1.6385***	0.4309	3.8027	0.0001
Developing	-0.1857	0.1808	1.0276	0.3042
FE	0.1415	0.1673	0.8458	0.3977
GLS	-0.0000	0.0000	0.0000	1.0000
GMM	0.2215	0.1662	1.3325	0.1827
IV	0.7207***	0.2223	3.2422	0.0012
OLS	0.0071	0.1678	0.0425	0.9661
Other methods	0.1399	0.1954	0.7163	0.4738
Low income	0.0000	0.0000	0.0000	1.0000
Middle income	-1.2981***	0.3798	3.4182	0.0006
Middle-low income	-1.4924***	0.4205	3.5492	0.0004
Upper-middle income	-1.5896***	0.4137	3.8425	0.0001
Inflation	-0.3299	0.2045	1.6135	0.1066
Number of years	-0.0053	0.0084	0.6305	0.5284
Peer-reviewed journal	-0.9860***	0.2484	3.9696	0.0001
Working paper	-2.2278***	0.4178	5.3319	0.0000
Unpublished	0.0000	0.0000	0.0000	1.0000
Publication year	0.0502	0.0393	1.2777	0.2013
Asia	-0.5412	0.4320	1.2529	0.2103
Europe	1.1518***	0.3900	2.9535	0.0031
Latin America	1.7065***	0.6195	2.7548	0.0059
World	0.1988	0.3317	0.5991	0.5491
Simple impact factor	-0.0152	0.0099	1.5379	0.1241
Non linearities	-0.1300	0.1937	0.6715	0.5019
Last year	-0.0078	0.0126	0.6232	0.5331
First year	0.0012	0.0083	0.1500	0.8807
Number of countries	0.0002	0.0012	0.1634	0.8702
Debt lag	-0.0004	0.0055	0.0666	0.9469
Debt/GDP	-0.8441*	0.4713	1.7908	0.0733
Log(debt/GDP)	-0.6709	0.4193	1.6001	0.1096
Other indep. variables	-0.2718	0.3780	0.7190	0.4722
Total debt	-0.0000	0.0000	0.0000	1.0000
GDP per capita change	-0.1445	0.3170	0.4558	0.6485
GDP per capita growth rate	0.0348	0.1140	0.3049	0.7605
GDP per capita level	0.0958	0.2099	0.4563	0.6482
Other dep. Variables	0.0583	0.1448	0.4025	0.6873

This Table shows the coefficient estimate, the standard error, the z statistics and the p-value for the multimodel inference meta-regression.
*** p<0.01, ** p<0.05, * p<0.1

We report below the results of the multimodel inference. This methodological procedure estimates all possible meta-regressions by combining the various moderators. Thus, $2^{18} = 262.144$ meta-regressions were estimated. We did not consider interactions between moderators[12]. The models were then sorted in an ascending fashion following the Akaike Information Criterion (AIC).
The results proposed may be of interest for exploratory purposes only, to compare the coefficients obtained with those estimated in the previous Subsection, and to identify which moderators have more predictive importance.
Table 8 shows the coefficient estimates, the standard errors, the z-statistic, and the relative p-value for each moderator, according to the multimodel inference algorithm. Here we will comment only on the statistically significant coefficients at least at the $\alpha = 0.1$ level. By comparing Table 8 with Table 7, we find that external debt negatively affects the PCC between debt and economic growth. The same is true for developed countries and estimates published in working papers. On the contrary, in samples from European countries, the growth-debt relationship is on average relatively better than in other countries. Moreover, the signs of the coefficients associated with the IV estimation method (-), results published in peer-reviewed journals (-), and samples of Latin American countries (+) are concordant in both the multimodel inference estimate (Table 8) and the metaregression in Table 7. It is worth noting that the meta-regression reported in Table 7 is only one of the 262.144 models fitted and analyzed to generate the results just described.
Finally, different results arise concerning the "middle-income trap", as samples from middle, middle-low, and upper-middle income countries lead to more negative PCCs. The same is true when debt/GDP is used as a proxy for the independent variable.
Figure 8 shows the importance of the predictors by ordering them from the highest to the lowest value. The reference line is set at 0.8. According to the results obtained, 10 of 18 moderators have predictor importance greater than 75%. The main predictors are: the simple impact factor (100%), the publication type, the estimation method, the region (both with 99.9% of predictor importance), the debt type (99.2%), the development level (98.9%), the income level (98.8%), the proxy for the independent variable (87.7%), and the inclusion of inflation as a control variable (85.3%). It is worth highlighting that the publication year has a predictor importance of 75.8%.

[12]They were not included, as it took days to get just the estimates without interactions. In any case, given the large number of moderators, it would be very complicated to interpret the meaning of the coefficients of the interactions.

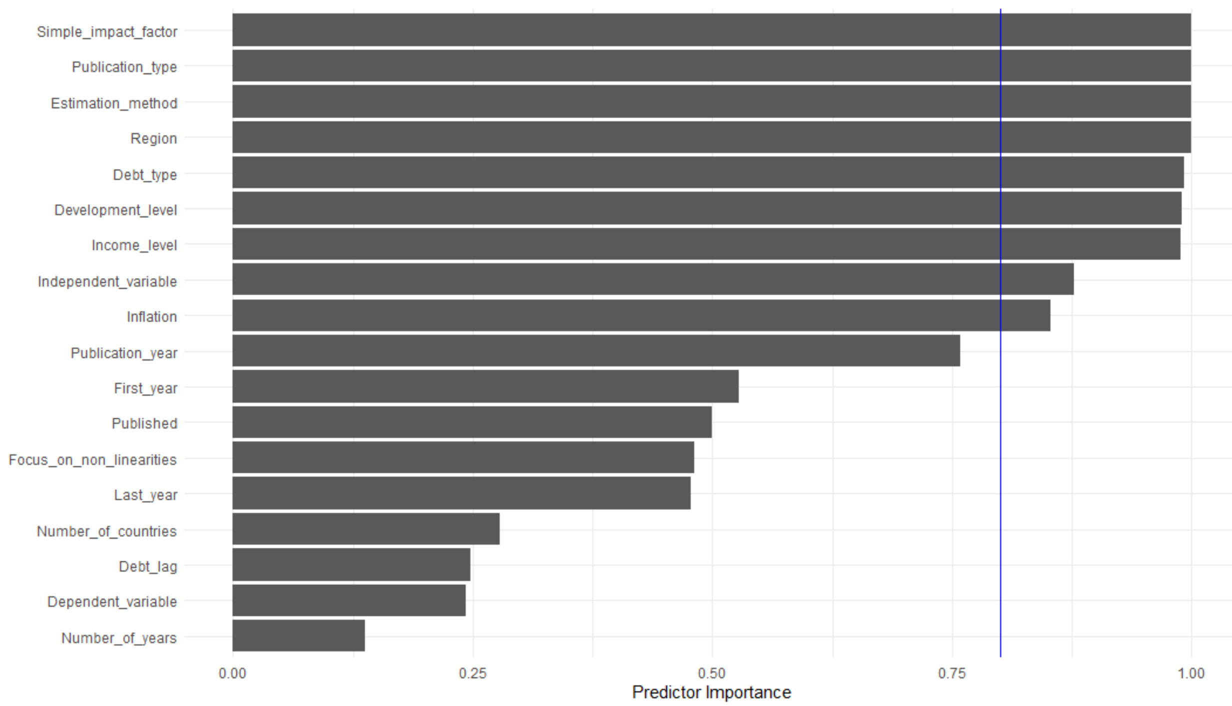

Figure 8: Predictors Importance in Multimodel Inference

This Figure shows the horizontal barplot of predictive importance for each moderator. Moderators are sorted in descending order. The vertical blue line represents the 0.8 threshold.

5.5.2 BMA

In this Subsection, we present the results obtained with the Bayesian Model Averaging (BMA, henceforth) methodology, referring to the seminal papers of Raftery (1995) and Raftery et al. (1997). To apply it, we modified the dataset in such a way as to convert all the categorical moderators into dummy variables, to create, for each categorical moderator, as many variables as categories belonging to it. To avoid problems of perfect multicollinearity, a category has been eliminated for each of these moderators. Then, we eliminate moderators with less than 10 observations, to avoid convergence problems of the algorithm. The final dataset is therefore composed of 30 possible explanatory variables.

Given the impossibility to explore the whole model space, i.e. the set of possible models $M = \{M_1, ..., M_J\}$, with $J = 2^Z = 2^{30}$, we rely on Bayesian methods. Thus, the posterior distribution of the β_z coefficient is a weighted average of the posterior distribution under each model, with weights equal to the posterior model probabilities. A description of the theoretical model is provided in Appendix H.

We use the *bms* package for the RStudio software developed by Zeugner (2011) and set the most uninformative model prior and Zellner (1986) g-prior, following Gechert et al. (2021) and Havránek et al. (2021). Thus, we use a birth-death MCMC algorithm to draw from the posterior distribution, considering 3,000,000 iterations, with a burn-in of 1,000,000. Our results refer to the best 1,000 models. We adopt the uniform model prior[13], and the Unit Information Prior (UIP)[14] for the coefficients.

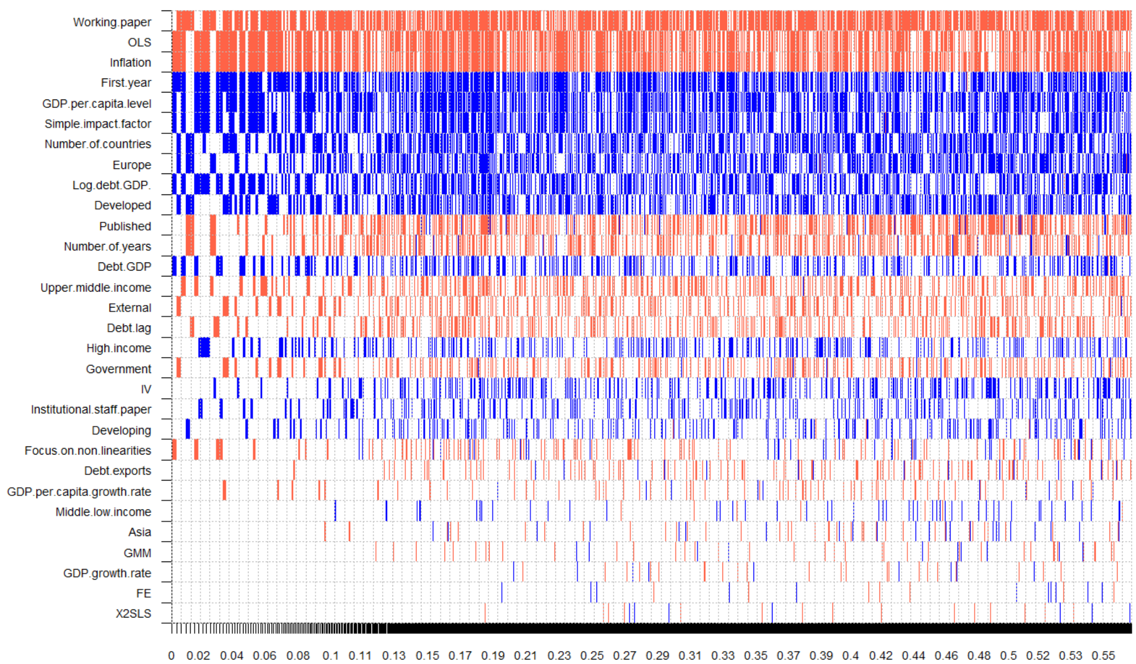

Figure 9: Model Inclusion in BMA (UIP and Uniform Model Prior)

This Figure shows the PIP of each moderator. Each row represents a variable and each column one of the J models. The columns are sorted in ascending order according to the posterior probability of the model. Variables are sorted in descending order according to PIP. Colored cells indicate the presence of the variable in the j-th model: blue cells represent a positive sign of the posterior mean, while red cells represent a negative sign of the posterior mean.

Figure 9 shows the results of the BMA. Each row represents a variable and each column one of the J models. The columns are sorted in ascending order according to the posterior model probability. Variables are sorted in descending order according to the Posterior Inclusion Probability (PIP), i.e. the probability of being included in the true final model. Colored cells indicate the presence of the variable in the j-th model; the blue color indicates a positive sign of the posterior mean, while for the negative sign the color is red. Model diagnostics and relative figures are shown in the Appendix (Table 18 and Figure 21).

[13] We assume that all models have the same weight on the posterior distribution and the prior model size is equal to $\bar{m} = \frac{Z}{2} = 15$

[14] Thus, $g = n$ and the Bayes factors approximate the BIC criterion by Schwarz (1978). See Kass and Wasserman (1995) for a complete description.

Table 9: Regression Results of BMA and OLS

	BMA			OLS		
Variable	PIP	Post. Mean	Post. SD	Estimate	SD	p-value
Intercept	1.000	-12.385		-12.291	15.631	0.432
Working paper	0.772	-0.420	0.351	-0.576***	0.204	0.005
OLS	0.659	-0.141	0.114	-0.176	0.124	0.155
Inflation	0.651	-0.209	0.172	-0.214*	0.114	0.061
First year	0.638	0.006	0.005	0.006	0.008	0.413
GDP per capita level	0.602	0.351	0.341	0.482*	0.254	0.058
Simple impact factor	0.511	0.007	0.008	0.004	0.006	0.511
Number of countries	0.510	0.002	0.002	0.005***	0.002	0.006
Europe	0.471	0.247	0.294	0.269	0.164	0.102
Log(debt/GDP)	0.465	0.180	0.231	0.359*	0.195	0.066
Developed	0.461	0.157	0.196	0.317	0.217	0.145
Published	0.415	-0.275	0.376	-0.454*	0.250	0.070
Number of years	0.336	-0.003	0.004	-0.004	0.008	0.576
Debt/GDP	0.286	0.083	0.150	0.331*	0.172	0.055
Upper-middle income	0.276	-0.060	0.110	0.048	0.138	0.727
External debt	0.250	-0.153	0.303	-0.551**	0.257	0.033
Debt lag	0.241	-0.006	0.011	-0.009	0.011	0.420
High income	0.241	0.063	0.127	0.102	0.205	0.620
Government debt	0.241	-0.151	0.304	-0.505*	0.259	0.052
IV	0.224	0.062	0.133	0.191	0.175	0.274
Institutional staff paper	0.188	0.028	0.067	0.034	0.102	0.742
Developing	0.178	0.036	0.091	0.209	0.139	0.133
Focus on non linearities	0.141	-0.020	0.095	-0.018	0.160	0.911
Debt/exports	0.091	-0.019	0.083	0.078	0.214	0.716
GDP per capita growth rate	0.076	-0.008	0.037	-0.078	0.139	0.573
Middle-low income	0.051	0.005	0.030	0.002	0.119	0.984
Asia	0.051	-0.005	0.063	0.128	0.236	0.587
GMM	0.039	-0.002	0.021	-0.036	0.126	0.777
GDP growth rate	0.025	-0.000	0.017	0.013	0.143	0.926
FE	0.024	0.001	0.015	0.015	0.134	0.913
2SLS	0.022	-0.000	0.021	-0.036	0.165	0.827
Adjusted R^2						0.282
F p-value						0.000

This Table shows the PIPs, the posterior means, and the posterior standard deviations from the BMA model in columns 2-4. In columns 5-6 the coefficient estimates, the standard deviations, and the p-values from the OLS model are reported. The Table also shows the adjusted R^2 and the F p-value for the OLS model in the last 2 rows.
*** p<0.01, ** p<0.05, * p<0.1

The corresponding numerical results are reported in Table 9, which shows for each variable the PIP, the posterior mean, and the posterior standard deviation. We rely on the 0.5 PIP threshold (Raftery (1995)) to draw the dashed line: variables with a PIP greater than this value are supposed to belong to the true model. In the same table, we compare the results of the BMA with the results

of the OLS regression (applied to the same dataset): we then report the estimated coefficients, the standard deviation, and the p-value.
Looking only at variables with PIP above the threshold, the sign of the coefficients is concordant for the two estimation methods. Moreover, all variables selected by the BMA model have predictor importance greater than 0.5 also in the multimodel inference (in Figure 8), except for the number of countries and the proxy for the dependent variable.
We can therefore confirm that the main moderators able to explain the heterogeneity present among the coded studies are: publication type, estimation method, sample size, journal ranking, and inflation. In particular, when GDP per capita level is used as a proxy for the dependent variable, the debt-GDP nexus tends to be more positive than average. On the other hand, estimates are more skewed towards a more negative linkage when using the OLS method, when published in working papers, and when considering inflation in the regression. We don't comment on the other coefficients, since they are negligible (not too far from 0).

5.5.3 Stepwise Regression

Table 10: Stepwise Regression Results

Variable	Estimate	SD	p-value
Intercept	0.247	0.201	0.219
Europe	0.363***	0.121	0.003
Working paper	-0.593***	0.169	0.000
Upper-middle income	-0.189**	0.080	0.018
Published	-0.454**	0.204	0.026
Number of years	-0.007***	0.002	0.000
Institutional staff paper	0.123*	0.071	0.083
IV	0.283**	0.115	0.014
GDP per capita level	0.642***	0.156	0.000
Log(debt/GDP)	0.410***	0.107	0.000
Debt/GDP	0.272***	0.099	0.006
Inflation	-0.248***	0.075	0.001
Simple impact factor	0.009**	0.004	0.021
OLS	-0.119**	0.057	0.038
Adjusted R^2			0.286
F p-value			0.000

This Table shows the coefficient estimates, the standard deviations and the p-values for the selected moderators from the stepwise regression model.The Table also shows the adjusted R^2 and the F p-value in the last 2 rows.
*** p<0.01, ** p<0.05, * p<0.1

For the sake of completeness, we perform a stepwise regression. Table 10 shows the results obtained: selected variables, estimated coefficients, standard deviations, and p-values are shown. The sample

size coefficient in time is negligible. Excluding it, the results agree with the previously discussed analyses, as the remaining variables all have predictor importance greater than 0.5 and a PIP greater than 0.5, except from the proxy for the dependent variable.
According to what is shown in the table, the PCC tends to assume higher values than the average when considering countries located in Europe, if the estimation is carried out with IV. The same happens if the study is published as an institutional staff paper, if the dependent variable is GDP per capita level or if the independent is the debt-to-GDP ratio or its logarithm. On the contrary, values below the average of the PCC come from estimates published, with the OLS method, as working papers, if they consider inflation and for upper-middle-income countries.

5.6 Publication Bias

Publication bias arises when the choice of whether or not to publish a study is influenced by the statistical significance of the results presented or by the direction and magnitude of the results. To assess the presence of publication bias in the analyzed dataset, we proceed first by showing the funnel plot, then the Trim-n-fill test, and then we analyze the p-curve. Several statistical tests accompany the reported Figures.

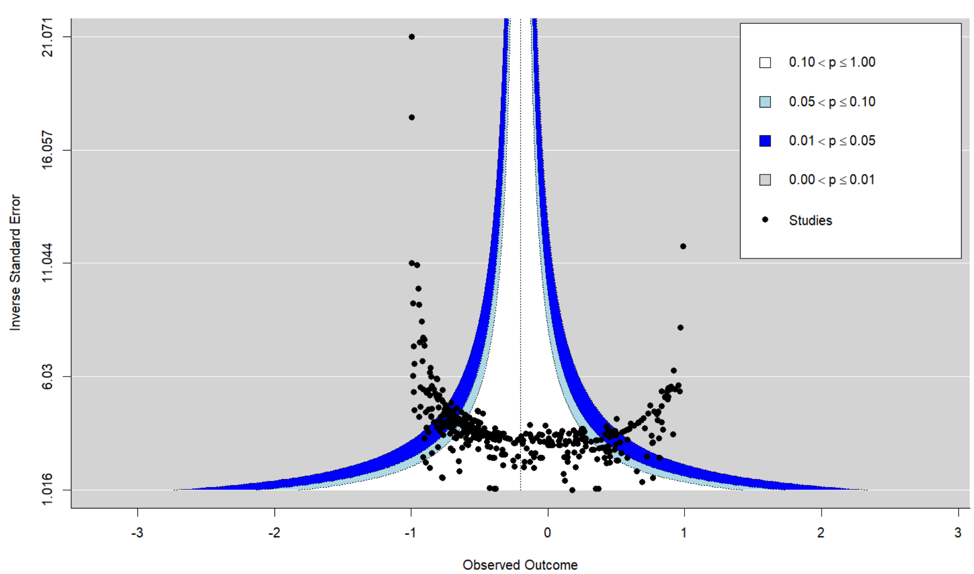

Figure 10: Funnel Plot

This Figure shows the funnel plot. On the x-axis is the PCC, while the y-axis represents the precision of the estimate, i.e., the inverse of the standard error. The white triangle with blue contours represents the ideal distribution of points if there was no publication bias. The plot is centered around the average estimated effect size (-0.2).

Figure 10 shows the contour-enhanced funnel plot, where the x-axis shows the PCC and the y-axis the inverse of the standard error (thus, the precision), from which it can be determined that there are several observations outside the reference area.

In the absence of publication bias, the points would be distributed within the triangle with blue contours, but in this case, there is a lot of scattering in the results. Therefore, looking at the entire sample ($K = 422$) it can be asserted that publication bias is present. The Figure also has both positive and negative biases, so we resort to statistical tests of symmetry to determine if the bias is symmetric or not.

To test the presence of funnel plot asymmetry, Egger's test (Egger et al. (1997)) is used. It is based on the following linear regression:

$$\frac{PCC_i}{SE_{PCC_i}} = \beta_0 + \beta_1 \cdot \frac{1}{SE_{PCC_i}} + \epsilon_i, \tag{9}$$

where $i = 1, ..., K$. Thus, the ratio of PCC to the respective standard error (the z-score) is regressed on its accuracy (the inverse of the standard error). We are interested in evaluating the sign of the intercept β_0: if this is zero, it would indicate the absence of asymmetry. This is true because the intercept shows the expected value of the z-score if the precision of the study is zero.

Table 11 shows the results of Egger's test: the estimated intercept is positive but not statistically significant. Thus, we are not yet able to assess the funnel plot asymmetry. The regression is represented graphically in Figure 11.

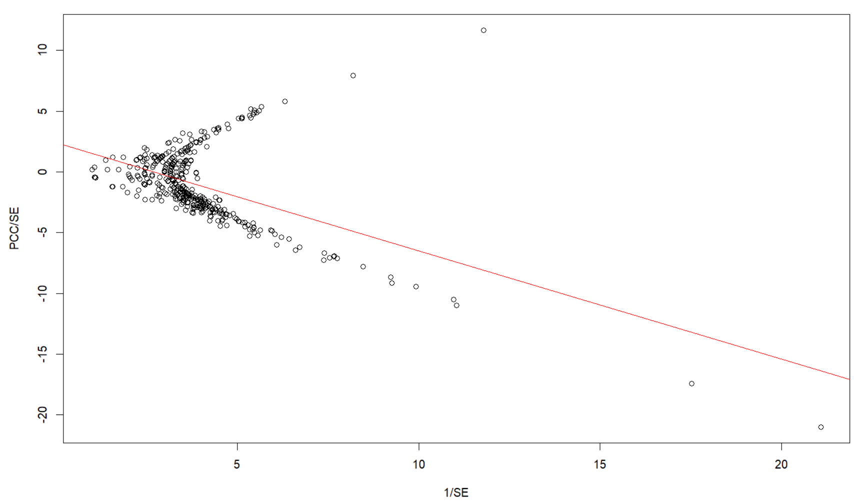

Figure 11: Scatterplot of Egger's Regression

This Figure shows the scatterplot of Egger's test, whose linear regression is represented by the red line. On the x-axis is the precision of the PCC, and on the y-axis is the z-score, i.e., the ratio of the PCC to the respective standard error.

Table 11: Egger's Test for Funnel Plot Asymmetry

Statistics	β_0	Lower Bound 95% C.I.	Upper Bound 95% C.I.	t	p-value
Value	2.855	1.320	7.030	1.340	0.181

This Table shows the intercept estimate, the extreme values of the 95% confidence interval, the t stat, and the p-value from the Egger's test regression.

Since Egger's test performs well for small-study effects, we rely on an additional method to identify publication bias. We follow Stanley and Doucouliagos (2014) and Stanley (2008), and perform the PET-PEESE estimation.
The PET (Precision-Effect Test) weighted regression is estimated according to the following equation:

$$PCC_i = \beta_0 + \beta_1 \cdot \hat{SE_{PCC_i}} + \epsilon_i, \tag{10}$$

where $\hat{SE_{PCC_i}}$ is the corrected standard error associated with the i-th PCC, and the weights are calculated as the inverse of the corrected variance of the i-th PCC, as follows:

$$w_i = \frac{1}{\hat{SE_{PCC_i}}^2}. \tag{11}$$

Since small studies tend to report highly over-estimated effects, we also calculate the PEESE (Precision-Effect Estimate with Standard Error) weighted regression:

$$PCC_i = \beta_0 + \beta_1 \cdot \hat{SE_{PCC_i}}^2 + \epsilon_i, \tag{12}$$

with the same weights as above.

Table 12: PET-PEESE Results

Parameter	Estimate	SE	t	p-value
β_0^{PET}	-0.887***	0.073	-12.204	0.000
β_1^{PET}	2.410***	0.307	7.850	0.000
β_0^{PEESE}	-0.572***	0.046	-12.505	0.000
β_1^{PEESE}	3.667***	0.608	6.030	0.000

This Table shows the coefficient estimates, the statndard errors, the t statistics and the p-values for the PET (rows 2 and 3) and PEESE (rows 4 and 5) regressions.
*** p<0.01, ** p<0.05, * p<0.1

We are interested in estimating the intercept of both (10) and (12) since these would give the limit effect, i.e. the expected effect size of a study with a null standard error/variance. In particular, the PET estimation performs well when β_0 is not statistically larger than zero, while the PEESE estimation is preferable when the intercept is statistically positive.

Table 12 shows the regressions results. All the coefficients are statistically significant. Given that the intercept estimated with the PET method is negative, we rely on this method. Therefore, the bias-corrected effect is around -0.9. In any case, given that the PEESE estimate gives a value of -0.5, the average effect size estimated in Table 1 is upward bias. This proves evidence of positive publication bias in the coded studies.

In the literature, the most widely used method for correcting funnel plot asymmetry is "Trim and Fill" by Duval and Tweedie (2000). This method consists in removing ("trimming") the extreme values and recalculating the pooled effect size using the fixed effects model. In the second step, once the pooled effect size, assumed to be the center of the funnel plot, has been calculated, the removed studies are reinserted and values are added to compensate for the extreme ones ("filling") so that the funnel plot is symmetrical to the new pooled effect size.

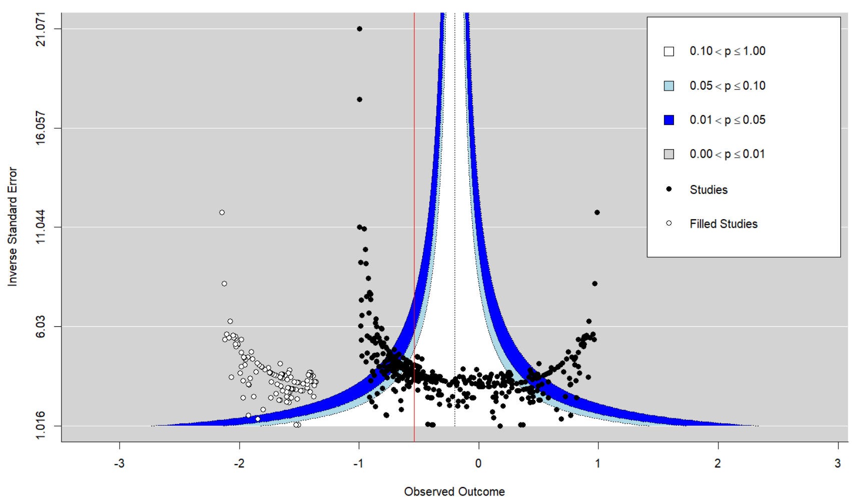

Figure 12: Contour Enhanced Funnel Plot (Trim-n-Fill)

This Figure shows the contour-enhanced funnel plot of the Trim-n-Fill test. To the funnel plot shown in Figure 10, the studies represented by the white dots were added. This funnel plot is therefore symmetrical to the estimated average correct effect size (-0.54), represented by the red vertical line.

The Trim-n-fill analysis adds 119 studies to the 422 examined, for a total of 541 observations. The new pooled effect size is -0.54, a much lower value than those presented in Table 1 of Subsection 5.1. This result is significant at level $\alpha = 0.01$. Figure 12 allows us to visualise this process. The black dashed line represents the average effect size, as in 10, while the solid red vertical line represents the corrected average effect size, around which the new funnel plot is symmetrical. The added studies are represented by empty circles. Since the previously analyzed funnel plot was asymmetrical with a predominance of studies with positive effect sizes, observations with negative values were added and the estimated average PCC have lower values than the baseline.

We perform a robustness test, using an alternative measure of effect size, Fisher's Z transformed correlation, in Appendix I.

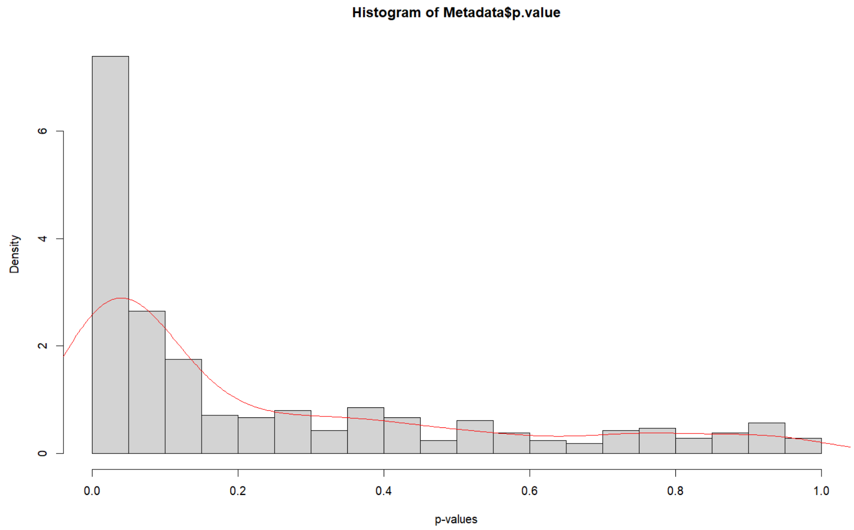

Figure 13: Histogram of p-values

This Figure shows the histogram of the p-values, represented by the gray bars with the approximate distribution of the variable (the red line).

We analyze whether publication bias is affected by the statistical significance of the results. First, we plot the histogram of the coded p-values in Figure 13, with the density approximation represented by the red line. We note that the p-values tend to concentrate around values less than or close to the 0.05 statistical significance threshold.
Then, we perform a three-parameter selection test, following McShane et al. (2016). Table 13 shows the results of this test. The estimated effect size is not different from the RE and the multi-level estimations in Table 1. The estimate in the third row represents the probability of being selected for publication when the p-value is less than 0.05. According to our analysis, all results with p-values less than the threshold are published, while (in the fourth row) this probability is reduced to 0.62 when p-values are greater than the threshold. This result provides evidence of publication bias, given the statistical significance of the published results.

Table 13: Results of the Three-Parameter Selection Model

Intervals	K	Estimate	SE	Lower Bound 95% C.I.	Upper Bound 95% C.I.	p-value
Effect Size		-0.2394***	0.0356	-0.3091	-0.1697	0.0000
0 < p <= 0.025	50	1.0000	—	—	—	—
0.025 < p <= 1	372	0.6195**	0.1579	0.3099	0.9290	0.0160

This Table shows the results of the three-parameter selection model. It reports the number of observations whose p-value is smaller or equal, or grater than 0.05 (K), the coefficient estimates, the standard error, the extreme bounds of the 95% confidence interval, and the p-value.
*** p<0.01, ** p<0.05, * p<0.1

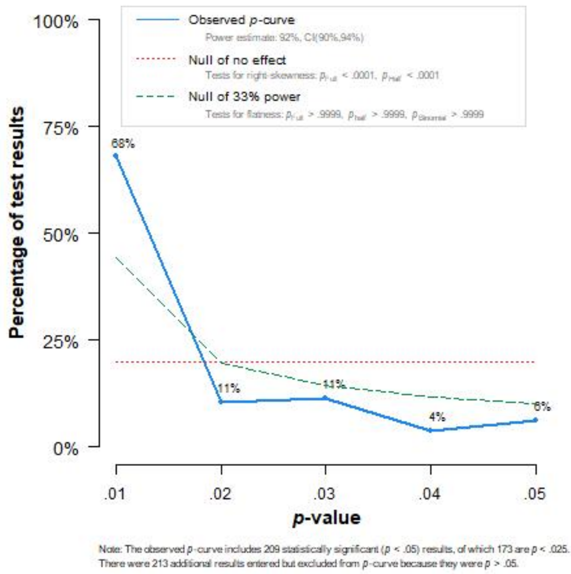

Figure 14: P-curve Analysis

This Figure shows the result of the p-curve analysis. The dashed red line represents the ideal distribution in the no-effect case, and the dashed gray line in the 33% power case. The solid blue line instead shows the observed p-curve from the data.

Finally, we perform the p-curve analysis. Figure 14 shows the percentage of p-values ranging between 0.01 and 0.05 in the observed studies. In the absence of publication bias, the ditribution of p-values ranging between 0.01 and 0.05 would be uniform, as in the dashed red line. Since the curve lies above the two dotted lines of no publication bias for p-values greater than 0.01 and below them for p-values between 0.02 and 0.05, we can state that too high a percentage of studies report significant statistics at the $\alpha = 0.01$ level. This confirms the tendency to publish certain results when they are statistically significant.
In conclusion, this meta-analysis shows a strong publication bias, both driven by the results of the PCCs (tendency to publish positive results) and by the statistical significance of the results.

6 Conclusion

In this study, we performed a meta-analysis on the link between public debt and economic growth, using a dataset of 32 studies, with a total of 422 observations. The selected studies estimate linear cross-country or panel regressions and report at least the Partial Correlation Coefficient, or the t-stat (or standard error and coefficient, or p-value, so the t-stat can be calculated), the degrees of freedom, and the number of observations. The coded effect size is the PCC and 7 continuous moderators and 11 categorical moderators were included in the meta-dataset.
The estimated average effect size is negative (about -0.2) in the multilevel analysis. Looking at the entire sample, it ranges from about -0.4 (with fixed effects) and -0.2 (with random effects). On the

other hand, if we analyze the mean effect size for each study, the estimate leads to an average PCC ranging from about 0.1 to -0.1. As we rely more on multi-level analysis, we conclude that on average the linear effect of public debt on economic growth in the sample of studies analyzed is negative, with an average PCC of -0.2. This meta-analysis investigates only the linear relationship between the two variables, although we do not exclude that in reality, the relationship may be nonlinear. We postpone this question as an in-depth study of a second meta-analysis, with a different sample and with the debt turning point threshold as a possible effect size to be analyzed. These new results could be compared to Heimberger et al. (2021).

Then, we found the presence of substantial heterogeneity, mainly due to within-studies variability rather than between-studies variability. Heterogeneity is also strongly present when analyzing moderators. Among these, the region of the sample, the income level, the development level, the variables used as proxies, the estimation method, the first year considered, the number of countries, and the journal ranking allow us to slightly lower the level of heterogeneity.

It is interesting to stress the following results. The first one concerns the development level of the countries considered: in fact, the link between public debt and growth has, on average, a negative direction when studying samples of developing countries, while showing debt neutrality for developed countries. This indicates a better capacity to use public debt to finance investments with higher returns to economic growth in developed countries than in developing countries.

The second one concerns income level: the link is positive for high-income countries and negative for low-income countries, although the results are not statistically significant. It is worth noting, however, that a middle-income trap arises, as the link is negative for middle-income countries, and this is also confirmed by the estimates from the BMA analysis and predictor importance. This result is in line with a stylized fact discussed by Aghion et al. (2021).

Also, it is interesting to note that as the journal ranking in which the results are published increases, the coefficient linking public debt and growth tends to decrease, while this tends to increase in unpublished papers.

As regards the moderator analysis, the journal ranking, the publication type, the estimation method used, the region, the income and development level of the sample, the debt type, the proxy used for the independent variable, and the use of inflation as control variable stand out in terms of predictive importance, according to the multi-model inference. The BMA and stepwise regression analysis reveal that the estimation of the PCC between public debt and economic growth is influenced by the proxies used for the dependent and the independent variables, the region, the income level, the sample size in time, the publication status, the presence of inflation as a control variable, the journal ranking, and the estimation method used.

Finally, with this analysis, we have identified the presence of publication bias in this area of the economic literature. In particular, the funnel plot is asymmetrical, with a prevalence of studies reporting positive effect sizes. Several tests confirm this result. The average corrected effect size is in a range of -0.9 to -0.5, highlighting a tendency to publish average positive results for this link.

The study of the three-parameter selection model also allows us to state that there is a tendency to publish studies with statistically significant values at the $\alpha = 0.05$ level. Furthermore, the

p-curve analysis shows a massive presence of estimates with a statistical significance level lower than $\alpha = 0.01$. Therefore, also in this sense, the presence of publication bias arises.
To confirm the scientific validity of the meta-analysis developed, we propose some research directions. Firstly, other moderators could be coded for the studies considered, such as if the paper directly assesses endogeneity issues. In addition, meta-analyses could be conducted considering a subset of the already coded data, for example answering research questions such as: what is the link between foreign debt and economic growth? How does this link change when only studies on the non-linearity of the relationship are considered? A further extension might involve comparing the same studies in the working paper version and the published version, observing whether the estimates have been changed, and/or analyzing working papers that were later not published.
We conclude this study with a call to all authors of coded papers, which can also be extended to authors of papers that meet the inclusion and eligibility criteria listed above, to contribute to the rectification and extension of the meta-dataset. It would be interesting to set up a public dataset in which everyone can contribute their estimates, perhaps on an online platform where the results of the meta-analysis are continuously updated. These would certainly be more complete, correct, and accurate than those presented here. We hope that this study will pave the way for an ever-increasing sharing of data and estimates in the academy and with institutions on economic issues of relevance to today and posterity.

References

Abbas, S. A. and Christensen, J. E. (2010). The role of domestic debt markets in economic growth: An empirical investigation for low-income countries and emerging markets. IMF Staff Papers, 57(1):209–255.

Afonso, A. and Alves, J. (2014). The role of government debt in economic growth.

Afonso, A. and Jalles, J. T. (2011). Economic performance and government size.

Aghion, P., Antonin, C., and Bunel, S. (2021). The power of creative destruction. In The Power of Creative Destruction. Harvard University Press.

Ahlborn, M. and Schweickert, R. (2018). Public debt and economic growth–economic systems matter. International Economics and Economic Policy, 15(2):373–403.

An, P. T. H., Duong, M. B., Toan, N. T., et al. (2020). The impact of public debt on economic growth of asean+ 3 countries. International Journal of Economics & Business Administration (IJEBA), 8(4):87–100.

Barro, R. J. (1989). The ricardian approach to budget deficits. Journal of Economic perspectives, 3(2):37–54.

Baum, A., Checherita-Westphal, C., and Rother, P. (2013). Debt and growth: New evidence for the euro area. Journal of international money and finance, 32:809–821.

Bökemeier, B. and Clemens, C. (2016). Does it pay to fulfill the maastricht convergence criteria?–reflections on the public debt and growth nexus for selected european economies.

Bökemeier, B. and Greiner, A. (2015). On the relation between public debt and economic growth: an empirical investigation. Economics and Business Letters, 4(4):137–150.

Bouchrara, K., Rachdi, H., Guesmi, K., et al. (2020). The non-linear relationship between economic growth and public debt. Economics Bulletin, 40(3):2336–2343.

Buchanan, J. M. (1976). Barro on the ricardian equivalence theorem. Journal of political economy, 84(2):337–342.

Buchanan, J. M. and Buchanan, J. M. (1958). Public principles of public debt: a defense and restatement, volume 2. RD Irwin Homewood, IL.

Caner, M., Grennes, T., and Koehler-Geib, F. (2010). Finding the tipping point–when sovereign debt turns bad.

Cecchetti, S. G., Mohanty, M., Zampolli, F., et al. (2011). Achieving growth amid fiscal imbalances: the real effects of debt. In Economic Symposium Conference Proceedings, volume 352, pages 145–96. Federal Reserve Bank of Kansas City.

Checherita-Westphal, C. and Rother, P. (2011). The impact of government debt on growth. Revue économique, 62(6):1015–1029.

Clements, B., Bhattacharya, R., and Nguyen, T. Q. (2003). External debt, public investment, and growth in low-income countries.

Cochran, W. G. (1954). Some methods for strengthening the common χ 2 tests. Biometrics, 10(4):417–451.

Cordella, T., Ricci, L. A., and Ruiz-Arranz, M. (2010). Debt overhang or debt irrelevance? IMF Staff Papers, 57(1):1–24.

Corsetti, G. et al. (2012). Austerity: Too much of a good thing. Centre for Economic Policy Research, London.

Dafermos, Y. (2015). The 'other half'of the public debt–economic growth relationship: a note on reinhart and rogoff. European Journal of Economics and Economic Policies: Intervention, 12(1):20–28.

Dao, M. Q. et al. (2011). Debt and growth in developing countries-debito e crescita nei paesi in via di sviluppo. Economia Internazionale/International Economics, 64(2):173–193.

Diamond, P. A. (1965). National debt in a neoclassical growth model. The American Economic Review, 55(5):1126–1150.

Duval, S. and Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56(2):455–463.

Égert, B. (2015). Public debt, economic growth and nonlinear effects: Myth or reality? Journal of Macroeconomics, 43:226–238.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. Bmj, 315(7109):629–634.

Elmendorf, D. W. and Mankiw, N. G. (1999). Government debt. Handbook of macroeconomics, 1:1615–1669.

Gaies, B. and Nabi, M.-S. (2019). Financial openness and growth in developing countries. Journal of Economic Integration, 34(3):426–464.

Gani, A. (1999). The burden of external debt in the south pacific island countries/le fardeau de la dette externe dans des îles du sud pacifique. Savings and development, pages 29–45.

Gechert, S., Havranek, T., Irsova, Z., and Kolcunova, D. (2021). Measuring capital-labor substitution: The importance of method choices and publication bias. Review of Economic Dynamics.

Gómez-Puig, M. and Sosvilla-Rivero, S. (2018). Public debt and economic growth: Further evidence for the euro area. Acta Oeconomica, 68(2):209–229.

Greenidge, K., Craigwell, R., Thomas, M. C., and Drakes, M. L. (2012). Threshold effects of sovereign debt: Evidence from the Caribbean. International Monetary Fund.

Greiner, A. (2011). Economic growth, public debt and welfare: Comparing three budgetary rules. German Economic Review, 12(2):205–222.

Harrer, M., Cuijpers, P., A, F. T., and Ebert, D. D. (2021). Doing Meta-Analysis With R: A Hands-On Guide. Chapman Hall/CRC Press, Boca Raton, FL and London, 1st edition.

Havránek, T., Irsova, Z., Laslopova, L., and Zeynalova, O. (2021). Skilled and unskilled labor are less substitutable than commonly thought.

Havránek, T., Stanley, T., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W. R., Rost, K., and Van Aert, R. (2020). Reporting guidelines for meta-analysis in economics. Journal of Economic Surveys, 34(3):469–475.

Heimberger, P. et al. (2021). Do higher public debt levels reduce economic growth? Technical report, The Vienna Institute for International Economic Studies, wiiw.

Herndon, T., Ash, M., and Pollin, R. (2014). Does high public debt consistently stifle economic growth? a critique of reinhart and rogoff. Cambridge journal of economics, 38(2):257–279.

Higgins, J. P. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. Statistics in medicine, 21(11):1539–1558.

Karadam, D. Y. (2018). An investigation of nonlinear effects of debt on growth. The Journal of Economic Asymmetries, 18:e00097.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. Journal of the american statistical association, 90(431):928–934.

Khanfir, W. et al. (2019). Threshold effect of public debt on economic growth: An empirical investigation for selected north african countries. Economic Alternatives, 3:429–436.

Koroglu, M. (2019). Growth and debt: An endogenous smooth coefficient approach. Journal of Risk and Financial Management, 12(1):23.

Kourtellos, A., Stengos, T., and Tan, C. M. (2013). The effect of public debt on growth in multiple regimes. Journal of Macroeconomics, 38:35–43.

Lartey, E. Y., Musah, A., Okyere, B., and Yusif, N. (2018). Public debt and economic growth: Evidence from africa. International Journal of Economics and Financial Issues, 8(6):35.

Lee, S., Park, H., Seo, M. H., and Shin, Y. (2017). Testing for a debt-threshold effect on output growth. Fiscal studies, 38(4):701–717.

McShane, B. B., Böckenholt, U., and Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. Perspectives on Psychological Science, 11(5):730–749.

Megersa, K. (2014). The laffer curve and the debt-growth link in low-income sub-saharan african economies. Technical report, University Library of Munich, Germany.

Megersa, K. and Cassimon, D. (2015). Public debt, economic growth, and public sector management in developing countries: Is there a link? Public Administration and Development, 35(5):329–346.

Mencinger, J., Aristovnik, A., and Verbic, M. (2015). Revisiting the role of public debt in economic growth: The case of oecd countries. Engineering Economics, 26(1):61–66.

Minea, A. and Parent, A. (2012). Is high public debt always harmful to economic growth? reinhart and rogoff and some complex nonlinearities.

Modigliani, F. (1961). Long-run implications of alternative fiscal policies and the burden of the national debt. The Economic Journal, 71(284):730–755.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P., et al. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. PLoS medicine, 6(7):e1000097.

Moore, W. and Thomas, C. (2010). A meta-analysis of the relationship between debt and growth. International Journal of Development Issues.

Myers, S. C. (1977). Determinants of corporate borrowing. Journal of financial economics, 5(2):147–175.

Padoan, P. C., Sila, U., and Van den Noord, P. (2012). Avoiding debt traps: financial backstops and structural reforms.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. Bmj, 372.

Panizza, U. and Presbitero, A. F. (2012). Public debt and economic growth: Is there a causal effect?

Panizza, U. and Presbitero, A. F. (2014). Public debt and economic growth: is there a causal effect? Journal of Macroeconomics, 41:21–41.

Pattillo, C., Poirson, H., Ricci, L., Kraay, A., and Rigobon, R. (2003). Through what channels does external debt affect growth?[with comments and discussion]. In Brookings Trade Forum, pages 229–277. JSTOR.

Próchniak, M. (2011). Determinants of economic growth in central and eastern europe: the global crisis perspective. Post-communist economies, 23(4):449–468.

Raftery, A. E. (1995). Bayesian model selection in social research. Sociological methodology, pages 111–163.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. Journal of the American Statistical Association, 92(437):179–191.

Reinhart, C. and Rogoff, K. (2010a). Debt and growth revisited.

Reinhart, C. M. and Rogoff, K. S. (2010b). Growth in a time of debt. American economic review, 100(2):573–78.

Ricardo, D. (1888). Essay on the funding system. The Works of David Ricardo, pages 513–548.

Rücker, G., Schwarzer, G., Carpenter, J. R., and Schumacher, M. (2008). Undue reliance on i 2 in assessing heterogeneity may mislead. BMC medical research methodology, 8(1):1–9.

Saungweme, T. and Odhiambo, N. M. (2018). The impact of public debt on economic growth: A review of contemporary literature. The Review of Black Political Economy, 45(4):339–357.

Schclarek, A. et al. (2005). Debt and economic growth in developing and industrial countries. Department of Economics.

Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, pages 461–464.

Shkolnyk, I. and Koilo, V. (2018). The relationship between external debt and economic growth: Empirical evidence from ukraine and other emerging economies. Investment Management and Financial Innovations, 15(1).

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations: Volume one. London: printed for W. Strahan; and T. Cadell, 1776.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. Oxford Bulletin of Economics and statistics, 70(1):103–127.

Stanley, T. D. and Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods, 5(1):60–78.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G., et al. (2013). Meta-analysis of economics research reporting guidelines. Journal of economic surveys, 27(2):390–394.

Stanley, T. D. and Jarrell, S. B. (2005). Meta-regression analysis: a quantitative method of literature surveys. Journal of economic surveys, 19(3):299–308.

Szabó, Z. (2013). The effect of sovereign debt on economic growth and economic development. Public Finance Quarterly, 58(3):251.

Uzun, A., Karakoy, C., Kabadayi, B., and Emsen, O. S. (2012). The impacts of external debt on economic growth in transition economies. Chinese business review, 11(5).

Wamboye, E. and Tochkov, K. (2016). External debt relief initiatives and economicgrowth in least developed countries. The Journal of Developing Areas, pages 213–229.

Woo, J. and Kumar, M. S. (2015). Public debt and growth. Economica, 82(328):705–739.

Xue, X., Cheng, M., and Zhang, W. (2021). Does education really improve health? a meta-analysis. Journal of Economic Surveys, 35(1):71–105.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. Bayesian inference and decision techniques.

Zeugner, S. (2011). Bayesian model averaging with bms. Tutorial to the R-package BMS 1e30.

Appendix

A Included Studies

Table 14: List of Coded Studies

Study	Number of observations
Abbas and Christensen (2010)	33
Afonso and Jalles (2011)	22
Ahlborn and Schweickert (2018)	12
An et al. (2020)	8
Baum et al. (2013)	37
Bökemeier and Clemens (2016)	12
Bouchrara et al. (2020)	6
Caner et al. (2010)	6
Cordella et al. (2010)	10
Dao et al. (2011)	6
Gaies and Nabi (2019)	9
Gani (1999)	8
Greenidge et al. (2012)	4
Karadam (2018)	13
Khanfir et al. (2019)	2
Koroglu (2019)	10
Kourtellos et al. (2013)	6
Lartey et al. (2018)	2
Lee et al. (2017)	17
Megersa (2014)	1
Megersa and Cassimon (2015)	34
Minea and Parent (2012)	2
Padoan et al. (2012)	6
Panizza and Presbitero (2014)	34
Pattillo et al. (2003)	11
Próchniak (2011)	1
Schclarek et al. (2005)	40
Shkolnyk and Koilo (2018)	2
Szabó (2013)	4
Uzun et al. (2012)	1
Wamboye and Tochkov (2016)	2
Woo and Kumar (2015)	61
32	422

This Table shows the list of coded studies with the respective number of coded PCCs.

Table 15: Additional Papers

Source	Study
Moore and Thomas (2010)	Abbas and Christensen (2010)
Saungweme and Odhiambo (2018)	Afonso and Jalles (2011)
	Ahlborn and Schweickert (2018)
	Baum et al. (2013)
	Panizza and Presbitero (2014)
	Schclarek et al. (2005)
	Uzun et al. (2012)
Szabó (2013)	Caner et al. (2010)
	Greenidge et al. (2012)
	Minea and Parent (2012)
	Padoan et al. (2012)

This Table shows the list of additional papers and the respective sources.

B Coded Variables

Table 16: List and Definitions of Coded Variables

Variable name	Definition	Type of moderator
Author(s)	Author(s) surname(s)	-
Publication year	Year in which the study was published	Continuous
Journal	Name of the journal in which the study is published	-
T-stat	T-statistics associated with the coefficient estimate	-
Degrees of Freedom	Number of degrees of freedom associated with the coefficient estimate	-
PCC	Partial Correlation Coefficient, own estimation	-
SE(PCC)	Standard errors of the PCC, own estimation	-
N	Sample size	-
P-value	P- value associated with the coefficient estimate	-
Beta	Beta coefficient estimate (used when necessary)	-
SE(Beta)	Standard errors of the Beta coefficient	-
Dependent variable	Proxy used for the dependent variable	Categorical
Independent variable	Proxy used for the independent variable	Categorical
First year	First year of the sample	Continuous
Last year	Last year of the sample	Continuous
Number of years	Number of years included in the sample	Continuous
Number of countries	Number of countries included in the sample	Continuous
Debt lag	Time lag of the independent variable to the dependent, expressed in years	Continuous
Focus on non linearities	Dummy variable, if the research question/methodology/results focus on non-linearities	Categorical
Published	Dummy variable, if studies are published or not	Categorical
Publication type	Type of publication in which the study is published	Categorical
Type of countries	Type of countries according to their level of development	Categorical
Inflation	Dummy variable, if inflation is used as a regressor	Categorical
Estimation method	Econometric method used for parameter estimation	Categorical
Region	Region of the world to which the sample countries belong	Categorical
Income level	Income level to which the sample countries belong, according to the WB classification[15]	Categorical
Debt type	Type of debt used as a proxy for the dependent variable	Categorical
Simple impact factor	Impact factor from RePEc (all years)[16]	Continuous
Number of t-stat	Total of t-statistics coded for each study	-

This Table shows the variable name, definition, and whether the moderator is continuous or categorical, for each variable encoded in the meta-dataset.

[15] The World Bank classification is available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

[16] When not present, we impute 0.

C Fixed and Random Effects Theoretical Model

For the fixed effects model, we assume that the effect sizes reported in each study ($\hat{\theta}_k$) are generated by the same true effect size (θ), but differ due to a sampling error typical of the k-th study (ϵ_k). Thus, each study represents an estimator of the true effect size, as follows:

$$\hat{\theta}_k = \theta + \epsilon_k. \tag{13}$$

In the random effects model, it is assumed that there is another source of variability besides sampling error. There is not only a true effect size but a distribution of effect sizes is referred to, the mean of which is to be estimated. Each study is independently drawn from a universe of populations. Therefore, the observed effect size is assumed to be composed as follows:

$$\hat{\theta}_k = \theta_k + \epsilon_k, \tag{14}$$

in which θ_k is the true effect size of the k-th study and in turn is generated by the following process:

$$\theta_k = \mu + \zeta_k. \tag{15}$$

Combining (14) and (15) gives the following formula:

$$\hat{\theta}_k = \mu + \zeta_k + \epsilon_k. \tag{16}$$

According to this model, the effect sizes of the individual study deviate from the true value of the individual study by a component of sampling error, and each true effect size of the individual study is sampled from a universe of true effect sizes, centered around μ.

D Heterogeneity Analysis Theoretical Model

The statistics analyzed to assess the degree of heterogeneity are the following:

1. Cochran (1954)'s Q is a weighted sum of the squares of the variances of each effect size observed in the studies $\hat{\theta_k}$ from the summary effect $\hat{\theta}$ (the pooled effect according to the fixed effects model), weighted by the inverse of the study's variance w_k. Mathematically, the formula is given by:
$$Q = \sum_{k=1}^{K} w_k(\hat{\theta_k} - \hat{\theta})^2. \tag{17}$$
2. The I^2 statistic, developed by Higgins and Thompson (2002), is based on the assumption that Q, under the null hypothesis of non-heterogeneity, is distributed as a χ^2 with $K-1$ degrees

of freedom. It is calculated according to the following formula:

$$I^2 = \frac{Q - (K-1)}{Q}. \tag{18}$$

Therefore, the values assumed by this statistic range from 0% (no heterogeneity) to 100% (maximum level of heterogeneity). According to the rule of thumb by Higgins and Thompson (2002), when $I^2 = 25\%$, there is low heterogeneity; when it is equal to 50%, there is moderate heterogeneity; substantial heterogeneity is assessed by $I^2 = 75\%$.

3. Since, when the value of Q is less than $K-1$, I^2 is corrected to present a null value, Higgins and Thompson (2002) develop a second statistic, called H. Mathematically, it is calculated as follows:

$$H = \frac{Q}{K-1}. \tag{19}$$

When there is no heterogeneity, the value of this statistic is 1 (or smaller).

4. Finally, τ^2 represents the variance of the true effect size, whose standard deviation is the measure τ.

E Variance Components

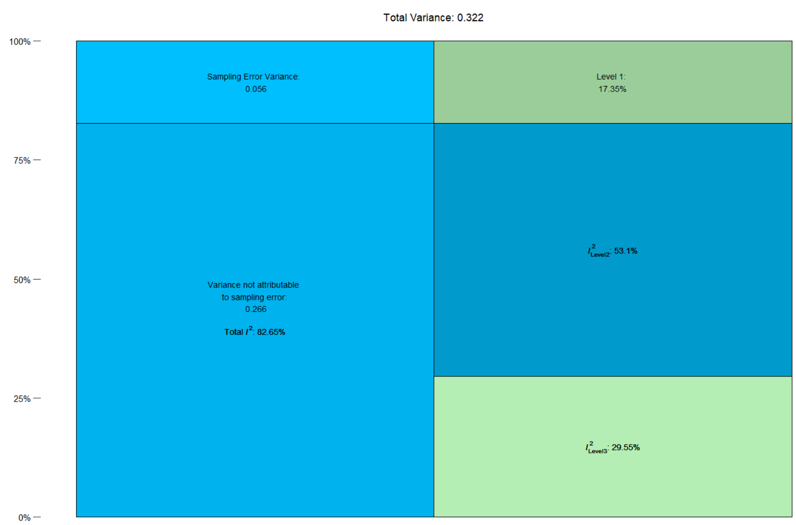

Figure 15: Variance Decomposition

This Figure shows the decomposition of variance and heterogeneity.

F Categorical Moderators

Regarding the proxy of the dependent variable, the most used is the GDP per capita growth rate (about 300 observations), the GDP growth rate is used in about 100 observations, while the remaining proxies do not reach 50 observations.

Regarding the independent variable, the debt-to-GDP ratio is used in about 300 observations, followed by the logarithm of the debt-to-GDP ratio (about 100); the remaining proxies are almost negligible.
Almost 85% of the observations analyzed are from studies that focus on nonlinearities, and the same percentage of observations are extracted from published papers. In addition, about 300 observations come from papers published in peer-reviewed journals, about 100 from institutional staff papers, and the remainder from working papers.
The distribution of the type of countries considered is more balanced: each category (developed, developing, both) counts for about one-third of the sample.
More than 150 observations are estimated with GMM, about 150 with OLS, and the remainder with the other methods (2SLS, FE, IV, GLS, etc.).
More than 75% of the observations are extracted from samples considering more than one region and about 20% from European samples.
One-third of the observations are from samples of high-income countries, followed by middle, lower-middle, and upper-middle-income. Fewer than 25 observations consider low-income levels.
Finally, most observations (more than 300) are of government debt, while less than 25% consider external debt.

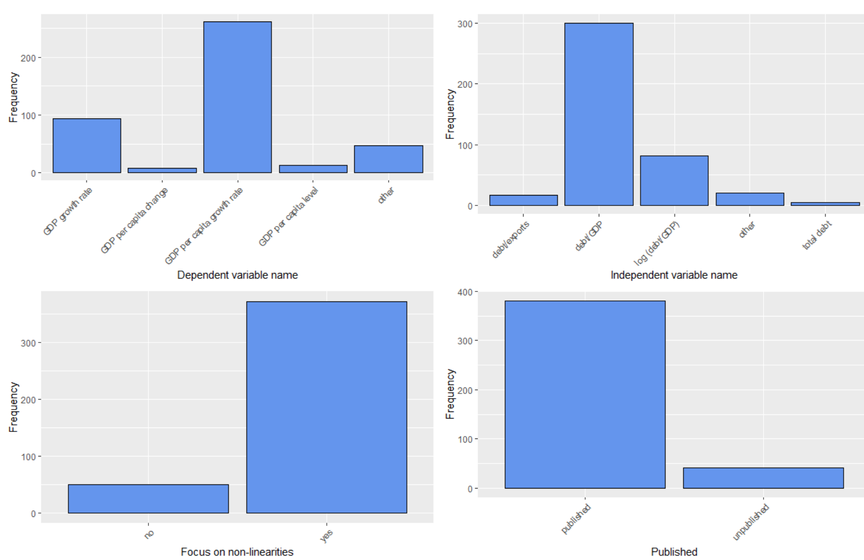

Figure 16: Barplots for Categorical Moderators (1/3)

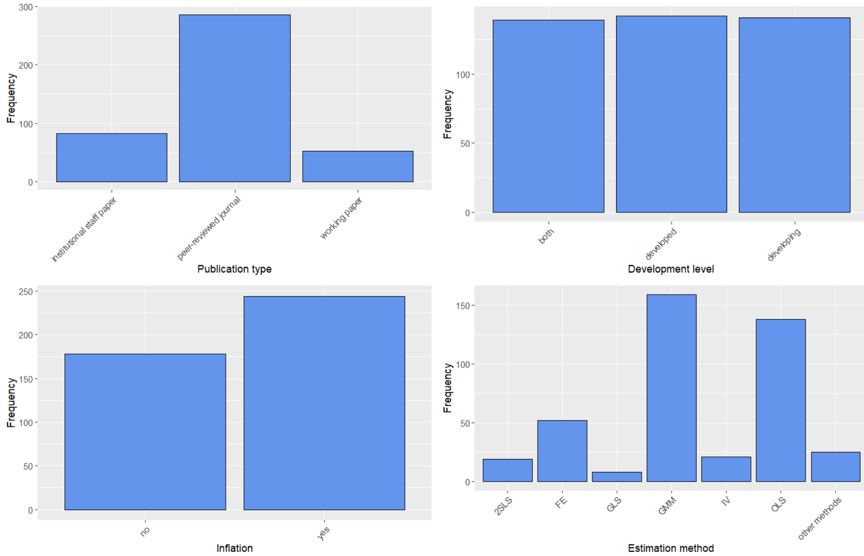

Figure 17: Barplots for Categorical Moderators (2/3)

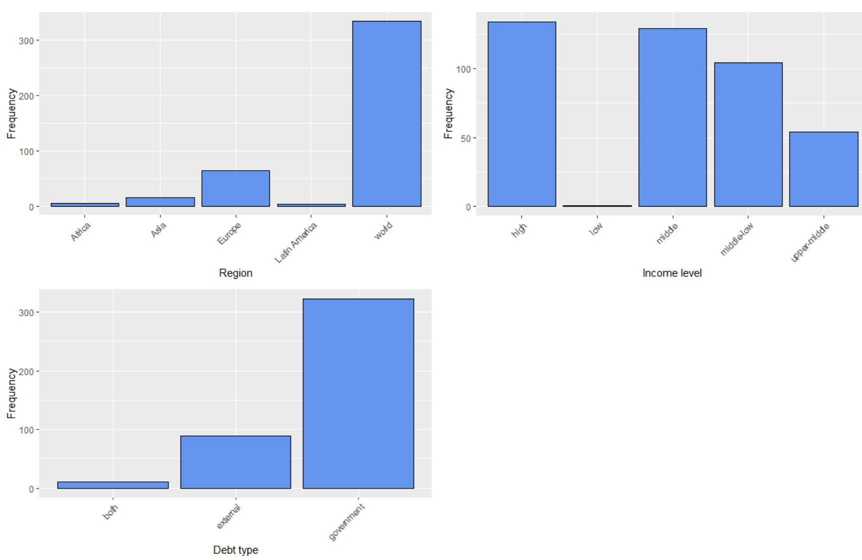

Figure 18: Barplots for Categorical Moderators (3/3)

These Figures show the barplots for each category of each categorical moderator. Every category within the same moderator has a color. In particular, Figure 16 shows the barplots for the moderators "dependent variable name", "independent variable name", "focus on non linearities", and "published". Figure 17 shows the barplots for the moderators "publication type", "ddevelopment level", "inflation", and "estimation method". Figure 18 shows the barplots for the moderators "region", "income level", and "debt type".

G Continuous Moderators

Table 17: Continuous Moderators-Descriptive Statistics

Variable name	Mean	SD	Min	Max
Publication year	2013	4.56	1999	2020
First year	1978	13.39	1880	2008
Last year	2007	4.74	1979	2016
Number of years	30.01	13.26	2	130
Number of countries	44	33.47	3	135
Debt lag	2.30	3.19	0	33
Simple impact factor	3.88	1.73	1.02	21.07

This Table shows the main descriptive statistics for the continuous moderators. It reports the mean, the standard deviation, the minimum, and the maximum value.

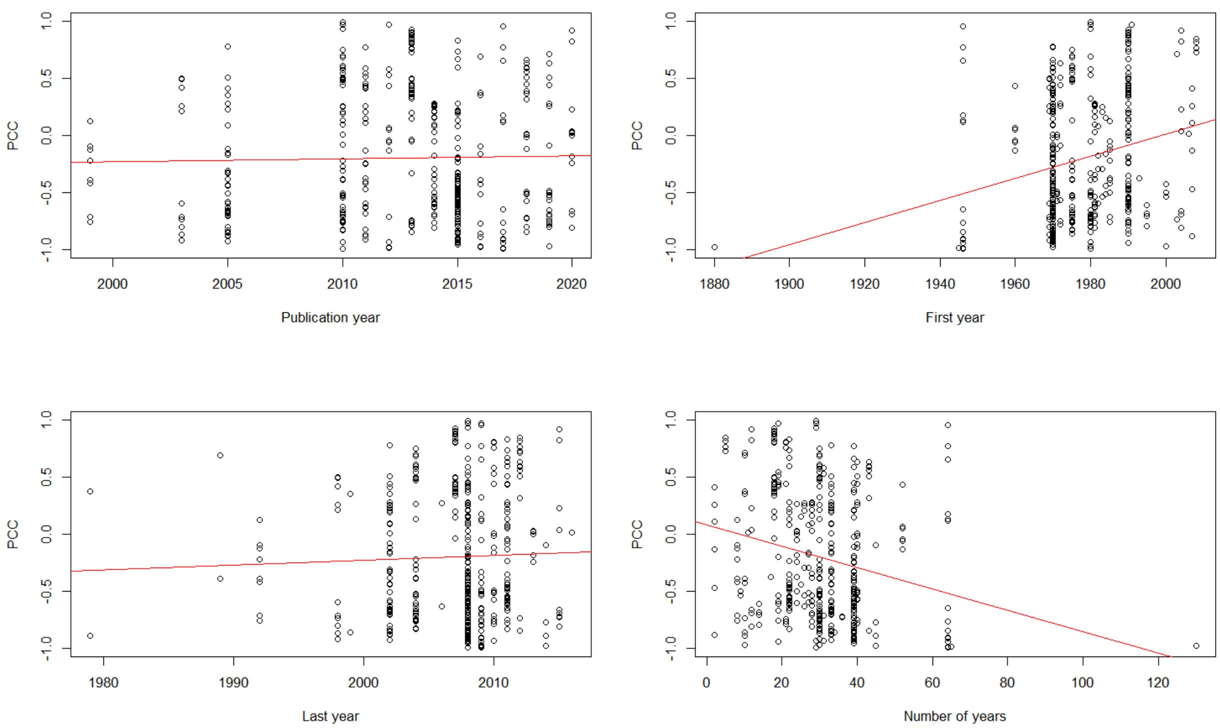

Figure 19: Single Regressions of Continuous Moderators (1/2)

This Figure shows the scatterplots (the empty circles) between each moderator and the PCC and the red linear regression line.

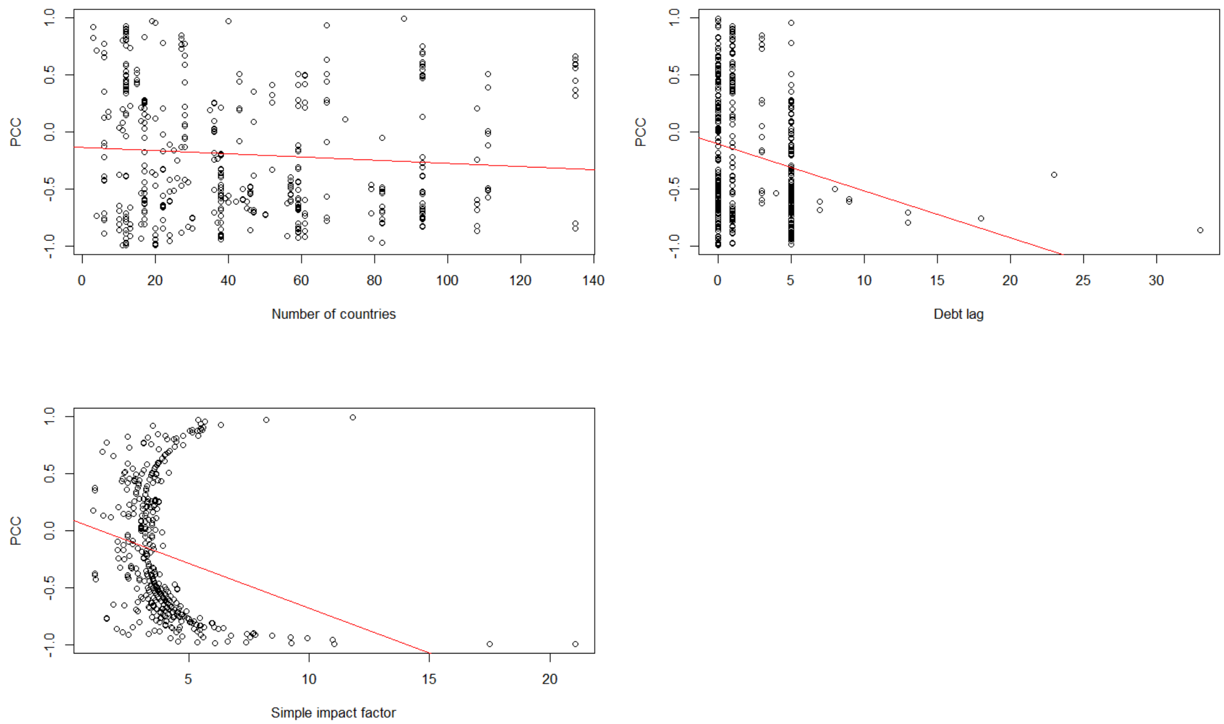

Figure 20: Single Regressions of Continuous Moderators (2/2)

This Figure shows the scatterplots (the empty circles) between each moderator and the PCC and the red linear regression line.

H BMA Theoretical Model

We assume a linear regression model with a constant of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_Z + \epsilon, \tag{20}$$

where y is the dependent variable and we have Z potential explanatory variables $x_1, x_2, ..., x_Z$. We define the set of all possible models $M = \{M_1, ..., M_J\}$, with $J = 2^Z$ representing all the regressors combinations.

The posterior distribution of the quantity of interest Δ is estimated through the BMA. It is a weighted average of the posterior distribution under each model, with weights equal to the posterior model probabilities

$$P(\Delta|D) = \sum_{j=1}^{J} P(\Delta|M_j, D)P(M_j, D). \tag{21}$$

By applying the Bayes rule, we calculate the weights

$$P(M_j, D) = \frac{P(D|M_j)P(M_j)}{P(D)}, \tag{22}$$

where $P(M_j)$ represents the model prior,

$$P(D) = \sum_{i=1}^{J} P(D|M_i)P(M_i) \tag{23}$$

is the weighted average of all the marginal likelihoods and the marginal likelihood of model j is

$$P(D|M_j) = \int P(D|\theta_j, M_j)P(\theta_j|M_j)d\theta_j, \tag{24}$$

where θ_j is the vector of parameters of model j. If we apply this procedure to (20), the quantity we want to estimate is the set of the beta coefficients $\Delta = \{\beta_1, ..., \beta_Z\}$. The posterior distribution of the general β_z coefficient is

$$P(\beta_z|D) = \sum_{j=\beta_z \in M_j} P(\beta_z|M_j)P(M_j|D). \tag{25}$$

The expected value of the β_z coefficient is

$$E[\beta_z|D] = \sum_{j=\beta_z \in M_j} \hat{\beta}_z P(M_j|D) \tag{26}$$

and the variance is

$$V[\beta_z|D] = \sum_{j=\beta_z \in M_j} (Var[\beta_z|D, M_j] + \hat{\beta}_z^2)P(M_j|D) - E[\beta_z|D]^2. \tag{27}$$

We are interested in estimating the Posterior Inclusion Probability (PIP) of every variable, obtained by summing up the posterior probabilities of the set of all the models that include β_z (i.e. in which $\beta_z \neq 0$). Formally, this quantity is:

$$P(\beta_z \neq 0|D) = \sum_{j=\beta_z \in M_j} P(M_j|D). \tag{28}$$

Table 18: BMA Diagnostics

Mean no. regressors	Draws	Burnins	Time
11.5394	3e+06	1e+06	5.240668 mins
No. models visited	Modelspace	% visited	% Topmodels
812,018	1.4e+11	0.00059	41
Corr PMP	No. Obs.	Model Prior	g-Prior
0.9893	465	uniform / 18.5	UIP
Shrinkage-Stats			
Av=0.9979			

This Table shows the main statistics for the BMA model diagnostics.

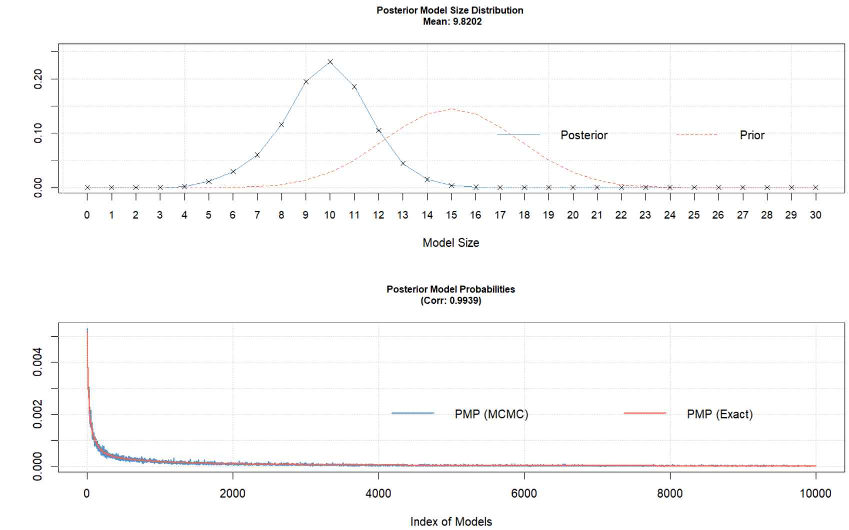

Figure 21: BMA Model Size and Convergence (UIP and Uniform Model Prior)

This Figure, in the top panel, shows the distribution of prior (dashed red line) and posterior (dashed blue line) model size in the BMA. The bottom box represents the Posterior Model Probability for each estimated model, sorted in descending order.

I Trim-n-Fill Analysis on Fisher's Z

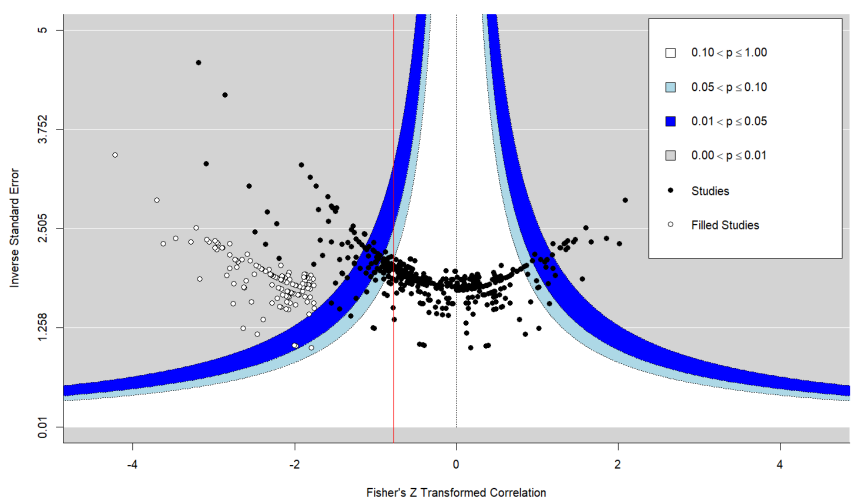

Figure 22: Contour Enhanced Funnel Plot (Trim-n-Fill) on Fisher's Z

This Figure shows the contour-enhanced funnel plot of the Trim-n-Fill test on Fisher's z. Thus, on the x-axis is represented the z, while on the y-axis is the precision of the estimates. Black dots represent coded studies, while white dots represent added studies. This funnel plot is symmetrical to the estimated average correct effect size (-0.78), represented by the red vertical line.

In Figure 22, the Fisher's z is represented on the x-axis. This PCC transformation is calculated as follows:

$$z_i = \frac{1}{2} ln \frac{(1 + PCC_i)}{(1 - PCC_i)}. \tag{29}$$

On the y-axis is represented the precision of the estimates, i.e., the inverse of the standard errors. This Trim-n-fill analysis adds 132 studies to the 422 reviewed, for a total of 554 observations. The new average effect size is -0.78 and is statistically significant at the $\alpha = 0.01$ level.

The dashed black line represents the standard mean effect size, while the adjusted effect size is represented by the solid red line. Added studies are represented by empty circles. Confirming the results discussed above, the added studies all have effect sizes smaller than both the mean and the lowest coded values, indicating the lack of studies showing a negative relationship between debt and growth.