OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305-9049
doi: 10.1111/obes.12426

Does the Left Spend More? An Econometric Survey of Partisan Politics*

GEORGIOS MAGKONIS,† KALLIOPI-MARIA ZEKENTE‡ and VASILIOS LOGOTHETIS§

†*Economics and Finance Subject Group, Portsmouth Business School, University of Portsmouth, Richmond Building, Portsmouth PO1 3DE, UK (e-mail: georgios.magkonis@port.ac.uk)*
‡*Economic Research Division, Alpha Bank, 11 Sophocleous St., Athens 10559, Greece (e-mail: kalliopi-maria.zekente@alpha.gr)*
§*Cardiff University, Cardiff, South Glamorgan, UK (e-mail: logothetisv@cardiff.ac.uk)*

Abstract

This study provides a quantitative review of the empirical literature on partisan politics. Given the voluminous work on this subject, we focus on the relationship between government ideology and public spending. By exploiting a dataset of 800 estimates from papers published between 1992 and 2018, we find no evidence of publication bias. Taking into account the differences in the various categories of spending, proxies of ideologies, estimations methods, as well as, data and publication characteristics, we find evidence of a small positive and significant effect.

I. Introduction

Partisan theory discusses the idea that a government's ideological position can influence its policy-making decisions. Politicians with ties to specific segments of the electorate adopt policies to enhance the well-being of their core constituencies when in office, with competing parties having different preferences over policy outcomes (Drazen, 2000; Franzese and Jusko, 2006). The literature has extensively focused on the different policies adopted from left-wing and right-wing governments. These differences are often perceived to be driven by ideological considerations. Hibbs (1992) argues that the support of the working class to left-wing parties stem from the fact that these parties are more likely to pursue policies that favour low unemployment. On the other hand, he argues that right-wing parties draw support from up-scale societal groups that have most probably invested in financial capital, favouring policies that promote low inflation.

JEL Classification numbers: D72, H00, C52.
*We thank the editor Prof. Anindya Banerjee and two anonymous referees for their constructive comments and helpful suggestions. In addition, we thank the participants of the 2019 MAER-Net Colloquium for their comments and suggestions. Any remaining errors are the responsibility of the authors.

Theoretical models introduce rational voters with uncertainty over the election outcomes resting as the crucial insight that allows for partisan effects on the economy (Alesina, 1987). This rational partisan theory (RPT) model predicts that these effects emerge only during the first half of a government's term after the elections. In the second half, the party's identity is already known when wage bargaining contracts are signed (in the first term). The main implication of the RPT model is that in the case of a left-wing party (right-wing party) winning the elections unemployment will be below (above) its natural rate.

The theoretical predictions of the models have been put to empirical test, with results providing a mixture of evidence conditional upon the set of countries under consideration and the specific policy instruments under examination. In this study, we focus on the effects of government ideology on government spending. The econometric framework is based on simple regression analysis. The typical model uses as dependent variable any measure of government expenditures and as main explanatory variable a proxy of government ideology. Similar to the other areas of applied economics that use regression analysis as the main empirical tool, several other variables are used as control set.

The evidence is far from being conclusive. Blais, Blake and Dion (1993) show that left-wing governments spend more than right-wing using aggregated government spending for 15 countries over 1960–87. The evidence outcome is supported by Cusack (1997). On the other hand, Bräuninger (2005) and Garrett and Mitchell (2001) do not find any evidence regarding the effect of ideology on the total amount of government spending. The results of the literature focusing on disaggregated level of expenditures remain mixed and inconclusive as well. For instance, Hicks and Swank (1992) support a positive relationship between more left-wing governments and social welfare expenditures. Similarly, Jensen (2011) argues that left-wing governments are associated with higher pensions and family services, but not with higher health care expenditures or higher unemployment protection. On the contrary, Kittel, Obinger and et al (2003) do not find any statistical significance to the ideology.

Since ideology is not an observable variable, there are several methods to proxy it. In order to avoid erroneous conclusions, the main focal variable of our meta-study is the partial correlation and not the direct estimates from the studies. We explain further this choice in the next section. Figure 1 gives a brief summary of the estimated partial correlations through time. One clear message from the figure is that the empirical findings considerably vary. One of the three targets of our present meta-analysis is to explain the factors of this variation. Another important question is to test whether researchers (including both the writers and the editors/referees) tend to accept more frequently a specific outcome. In our case, this question can be rephrased as follows: is there evidence of selective reporting in favour of a more positive relationship between left-wing governments and government spending. Finally, we explore the statistical and the economic effect (if any) of ideology on government spending.

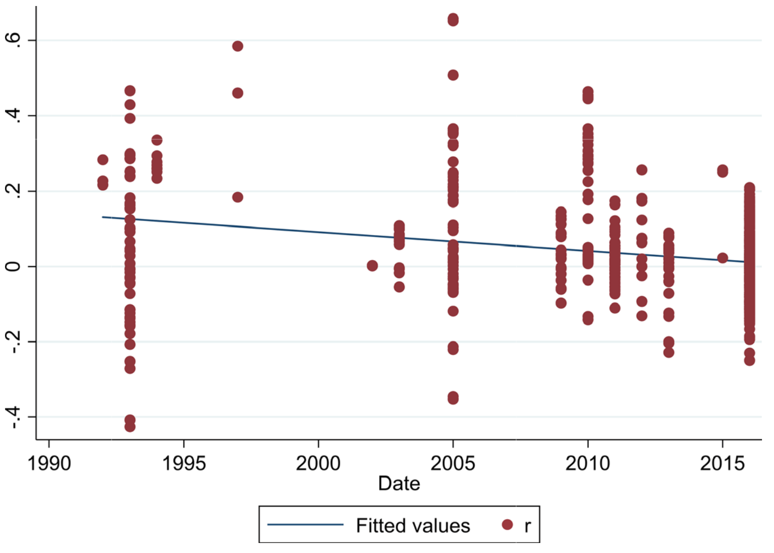

Figure 1. Estimated partial coefficient (r) per year of publication
Notes: The figure depicts the estimates (partial correlation coefficients, r) of the effect of ideology on government spending reported in the empirical literature over time. The horizontal axis shows the publication year of the examined studies.

While earlier narrative literature surveys (Potrafke, 2017) have provided important insights on the partisan politics literature, to the best of our knowledge the current paper is the first quantitative review that studies the ideology-spending nexus.[1] The partisan politics literature will be further benefited by additional future meta-analytic work focusing on other policy fields. For instance, it would be interesting to see a meta-analysis of the effects of ideology on tax, deficit/debt and privatization. In our study, we collect 28 empirical studies published over the last 26 years and define aspects related to the different categories of government spending, ideology proxies, model specification issues, econometric techniques used and data characteristics. We examine whether these study design issues can explain in a systematic way the reported estimates found in the collected papers. As discussed in the next sections, we propose a series of potential drivers that explain the variation of the reported estimates. We employ a series of model averaging techniques in order to deal with the problem of model uncertainty. This problem becomes quite significant when the number of drivers is large and there is no *a prior* knowledge whether certain factors are more important than others. In addition, our research examines the existence of publication bias.

The structure of the paper is as follows. Section II discusses the process followed in order to collect the meta-data sample. Section III examines the presence of publication bias. Section IV describes the moderator variables used in this study, while

[1]Imbeau, Pétry and Lamari (2001) review the effects of ideology on policy outcomes in general, without focusing on government expenditures.

section V presents the econometric model and the baseline results. Section VI presents a series of robustness checks and additional evidence. Section VII concludes.

II. Data collecting process

The first source of papers is the detailed survey of Potrafke (2017). In order to be as inclusive as possible, we also searched in Google Scholar using ‘partisan politics’, ‘government ideology’ and ‘government spending’ as keywords. This process produced 75 papers in total. Our inclusion strategy consists of three criteria. The first criterion for a study to be included in the meta-data sample is to report at least one estimated coefficient of the effect of government ideology on public spending. Therefore, we excluded papers that focus on other aspects of public policy outcomes, such as revenues, debt and unemployment as well as purely theoretical studies. The second inclusion criterion relates to the broader definition of government ideology. To ensure a minimum degree of comparability across studies, we include papers that measure the power of left-wing over right-wing governments. Specifically, we include studies that measure ideology by the number of seats in parliament by left-wing governments. Additionally, we include studies that use a variety of ideology indexes. Here, we focus on three main indexes.[2] The chosen proxies assign values in right-to-left scale, with the minimum value being 1 for extreme right-wing governments to 5 for extreme left-wing ones. Twenty-five studies use an ideology index in a left-to-right scale, giving higher value to right-wing government. These studies are excluded from the sample. Finally, we include papers that use a single dummy variable (1 for left, 0 otherwise) to distinguish between left and right. The third criterion for a study to be included is to report at least one measure of precision (either standard error or *t*-statistic or *p*-value) of the reported estimates. Following the guidelines of the Meta-Analysis of Economics Research Network (Havránek *et al*., 2020), we summarize the entire study collection process in a PRISMA chart (online Appendix A).

Using the above three criteria, we end up with 28 papers that empirically explore the impact of ideology on several categories of government spending. This process results in a total of 800 estimates which constitute our meta-dataset. All the included studies are published in journals. However, there was not any prior intention to focus only on published studies. We found only two working papers that did not satisfy the last two criteria. The first working paper uses an ideology index on the non-preferable left-to-right scale. The second one does not report any measure of precision as the authors indicate the statistical significance of their results using asterisks.[3] The full list of the studies included is provided in the online Appendix G.

As we stressed in the introduction, our analysis relies on partial correlation coefficients and not on the direct estimated effects reported by the studies or the corresponding *t*-statistics. In this way, we overcome the problem of incomparability of

[2]These indexes are developed by Budge, Keman and Woldendorp (1993), Woldendorp, Keman and Budge (1998), Potrafke (2009).

[3]We repeated the process of paper collection during the period of revision without being able to find any additional working papers.

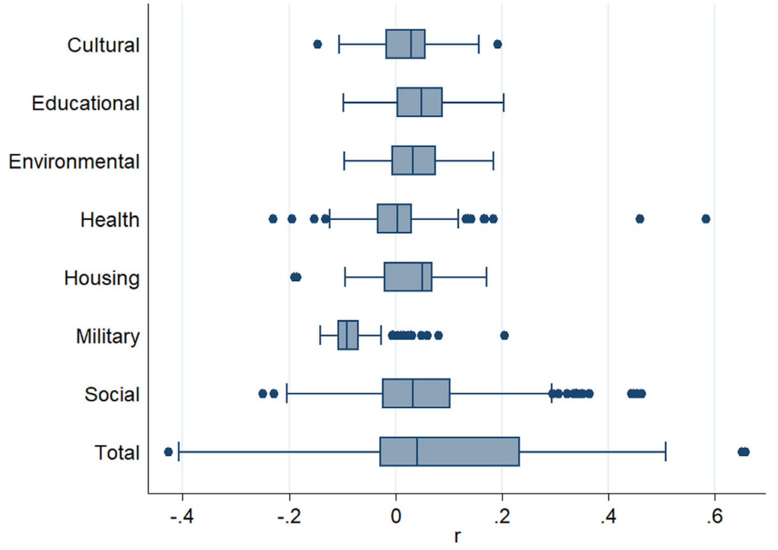

Figure 2. Boxplot of partial correlation (r) per spending category
Notes: The figure depicts the boxplot of the collected estimates (partial correlation coefficients, r). The estimates are sorted according to government spending categories.

the reported estimates across studies. The partial correlation coefficient is calculated as; $r_{ij} = t_{ij}/\sqrt{t_{ij}^2 + df_{ij}}$ where t and df are the t-statistics and the degrees of freedom of the reported estimates respectively, while i and j refer to the i observation from the j study. The corresponding standard errors are equal to $\sqrt{(1-r_{ij}^2)/df_{ij}}$. This approach renders all estimates comparable.[4] As Figure 1 shows the estimates (expressed as partial correlations) vary. This can also be observed when we split the estimates across different categories of spending. Figure 2 shows the dispersity of the collected estimates per spending category. Table 1 reports some additional statistics about the heterogeneity of the estimated partial correlations. Since there is a small number of outliers, in what follows we use winsorized data at 5% level. The results, however, remain robust even when the outliers are included in the dataset.

III. Analysing publication bias

Publication bias is a common feature of economics studies (Ioannidis, Stanley and Doucouliagos, 2017). This happens when the editors, reviewers and/or the authors are positively predisposed for a specific kind of result. More analytically, such a bias appears when different estimates have different probability of being reported depending on the statistical significance and/or the sign. In a bias-free literature, small imprecise estimates are equally reported as large imprecise estimates. Therefore, the reported estimates are not correlated with the standard errors or any measure of precision in

[4]The calculations follow the technical discussion in Doucouliagos, Haman and Stanley (2012).

TABLE 1

Partial correlations for different subsets

Variable	*Mean*	*SD*	*Min*	*Max*
Military	−0.073	0.005	−0.084	−0.062
Health	0.011	0.008	−0.005	0.028
Social	0.049	0.007	0.034	0.064
Educational	0.049	0.009	0.030	0.067
Environmental	0.035	0.009	0.015	0.055
Housing	0.025	0.013	−0.002	0.053
Cultural	0.026	0.012	0.001	0.050
Coalition	−0.018	0.003	−0.026	−0.011
Fiscal Pos	0.006	0.019	−0.034	0.046
GDP	−0.009	0.007	−0.023	0.004
GDPgr	0.085	0.014	0.057	0.113
Election	0.017	0.003	0.010	0.025
Inflation	0.040	0.025	−0.010	0.091
Unemployment	0.041	0.005	0.031	0.051
Population St	0.039	0.008	0.022	0.055
Population	0.025	0.004	0.017	0.033
Trade	0.087	0.012	0.062	0.113
Labour Power	0.182	0.036	0.108	0.256
Globalization	0.011	0.013	−0.015	0.038
Femate Part	0.049	0.022	0.001	0.096
GMM	0.011	0.005	0.001	0.021
Panel	−0.007	0.011	−0.029	0.013
DIV	0.022	0.004	0.013	0.031
Ideo Indexes	0.007	0.007	−0.007	0.022
Economic Journal	0.012	0.003	0.006	0.019

Notes: The table reports the mean, the standard deviation as well as the minimum and maximum values of the partial correlation coefficients for different subsets of data.

general. When large imprecise estimates are reported more frequently (as this increases the possibility of a study to be published), then the bias manifests itself with a correlation between the reported estimates and their corresponding standard errors.

A graphical illustration of the above analysis is traditionally explained through a funnel plot. Plotting the reported estimates against their inverse standard errors provides a visual inspection regarding the potential existence of publication bias. In the case of a bias-free literature, imprecise estimates, which are found in the bottom of the graph, are symmetrical around a mean that represents the true effect. This is because both large and small imprecise estimates are reported. In this case, no correlation between the estimates and their standard errors should be expected. On the contrary, if positive values are preferred for a certain reason, then the smaller imprecise estimates will be much less reported than the larger ones. This would produce an asymmetrical funnel plot and a positive correlation between the estimates and their standard errors. Figure 3 shows the funnel plot for our collected dataset. As expected, estimates with less precision (which appear towards the bottom) have more dispersion than estimates with more precision (which appear towards the top). Interestingly, the funnel plot reveals a slight asymmetry towards positive values. This visual indication suggests that

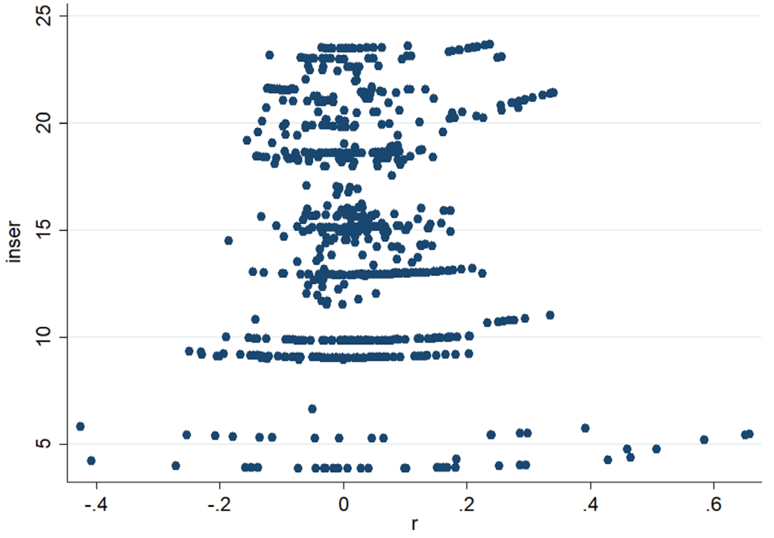

Figure 3. Funnel Plot
Notes: The figure depicts the scatter plot of the collected estimates (partial correlation coefficients, r) and the corresponding inverse standard errors.

the literature is characterized by a preference towards positive results, without being able to identify the cause of a potential bias. As we discuss in the next section, such preference might be due to heterogeneity.

A formal way of testing the existence of such a bias is the FAT test (Stanley, 2008), estimating the following model;

$$r_{ij} = \alpha + \beta SEr_{ij} + u_{ij} \tag{1}$$

where r_{ij} is the i-th partial coefficient extracted from the j-th study, SEr_{ij} is the corresponding standard error and u_{ij} is the error term. When bias exists, there is a statistically significant relationship between the estimated effect, r, and its standard error, SEr. More specifically, when positive values are over-preferred than negative ones, then the estimated beta should be positive and statistically significant. Regardless of the existence of publication bias, the most precise estimates are concentrated around the value that represents the true (genuine) effect. Therefore, the intercept, α, is interpreted as the average effect (here, the average partial correlation corrected for publication bias) conditional on the standard error approaching zero (Stanley, 2005). In our case, a positive and statistically significant β would be interpreted as an evidence of publication bias in favour of reporting more frequently positive values; that is, researchers are in favour of a positive relationship between ideology and government spending.

Table 2 presents the results of estimating equation (1). Firstly, we start with WLS using the whole sample. Secondly, we apply WLS in a subsample that consists of the baseline estimates of each study, excluding estimates from any reported robustness checks. We also apply an instrumental variable specification using as instrument the inverse of the square root of the number of observations. Havranek, Irsova and

TABLE 2

Funnel asymmetry testing (FAT)

	WLS	*Primary*	*IV*	*FE*	*BE*
β	0.056	0.146	0.313	0.399	0.838
publication bias	(0.199)	(0.354)	(0.264)	(0.584)	(0.536)
α	0.026	0.023	0.006	−0.001	0.002
effect beyond bias	(0.028)	(0.044)	(0.029)	(0.045)	(0.044)
Observations	800	524	800	800	800

Notes: The table reports the outcome from estimating equation (1). In all specifications the standard errors are clustered at study level and are reported in the parentheses. WLS refers to weighted least squares results. Primary shows the WLS output using a subset of the collected primary studies. IV refers to the instrumental variable estimation using as instrument for the standard error the inverse of the square root of the number of observations. FE refers to study-level fixed effects and BE refers to study-level between effects.

Zeynalova (2018) stress that some methods (used in the primary studies) may influence the estimates and their standard errors in the same direction. If this is the case, then the FAT results will be spurious. They propose the usage of inverse square root of the number of observations as instrument for the standard error; the root is correlated with standard error by definition but it is not likely to be correlated with a chosen method. As an additional robustness check, we also use fixed and between effects. In all specifications, the standard errors are clustered at study level. All the results are consistent with the view that the examined literature is not characterized by publication bias; β is found to be positive (as expected from the slightly asymmetric funnel plot) but not statistically significant. Interestingly, the effect beyond bias is very close to zero as the estimated intercept a is also positive but statistically insignificant.

The above analysis is based on the assumption of a linear relationship between the reported estimate (here partial correlation, r_{ij}) and the standard error SEr_{ij}. If, however, the relationship is characterized by nonlinearities then the inference could differ. We perform two recent methods that provide as outcome the mean effect conditional on different assumptions regarding the functional forms of publication bias. The first is developed by Andrews and Kasy (2019) who identify the conditional probability of publication as a function of a study's results. The publication bias is corrected given the identification of the conditional probability function (i.e. the probability of a study being published). Based on a theoretical framework, Furukawa (2019) proposes a method for correcting the bias that involves minimization of the trade-off between the bias and the variance. Contrary to Andrews and Kasy (2019), this method uses only the most precise estimates from each study. In a similar vein, Ioannidis *et al*. (2017) propose the usage of the estimates with adequate statistical power. Finally, Stanley, Jarrell and Doucouliagos (2010) suggest that the usage of the most 10% accurate estimates is sufficient to reveal the true effect beyond bias. Technical details are analysed in the online Appendixes B and C. The results are shown in Table 3. Based on Doucouliagos (2011) taxonomy, the estimated effect from all the methods is positive but quite small. This effect is expressed as partial correlation and therefore it does not provide any information about the economic significance. In the robustness section, we examine whether the evidence of a small partial correlation indicates a

TABLE 3

Genuine effect beyond publication bias

	Andrews & Kasy (2019)	*Furukawa (2019)*	*Ioannidis et al. (2017)*	*Stanley et al. (2010)*
Mean beyond bias	0.038	0.067	0.031	0.079
	(0.008)	(0.060)	(0.124)	(0.151)

Notes: The methods applied here are the ones developed by Andrews and Kasy (2019), Furukawa (2019), Ioannidis *et al*. (2017) and Stanley *et al*. (2010). Each entry shows the genuine effect when the publication bias is corrected. Standard errors are reported in the parentheses.

significant or an insignificant (statistically and economically) effect, when only a subsample of comparable studies is used. Before this step, we examine whether other features of this literature can affect the reported results. This task is addressed in the next section.

IV. Explaining the observed differences

The next step is to model the reported heterogeneity. We divide the existing literature into five broad groups of moderator variables, with each group capturing a specific feature of partisan politics. In particular, we consider the following dimensions: (1) the spending category, (2) the measures of ideology, (3) the econometric method, (4) the model specification and data characteristics and (5) the publication features. For each group, we define several moderator variables in order to capture in detail all the potential driving forces.

The first group of moderator variables takes into account various types of government spending. As Figure 2 shows, the literature focuses on both aggregated (total) and disaggregated amounts. Treating the total government expenditure as the reference category, we discern among seven broad disaggregated categories; (i) military, (ii) health, (iii) social protection (iv) educational, (v) environmental, (vi) housing and (vii) cultural.[5] This leads to seven separate dummy variables, with each assigned to the value of 1, when its corresponding spending category is used. For instance, the *military* dummy variable takes 1 if the paper focuses on military expenditures and so on.

The second major group of moderator variables is related to the different ideology proxies. Throughout the past three decades, a large number of variables have been used. As in other quantitative surveys in economics (Arestis, Chortareas and Magkonis, 2015), where it is rather impossible to capture all the variables employed, the best research strategy is to categorize the proxies into groups. We consider as reference category the cabinet seat shares of left-wing parties; that is, the number of seats taken by left-wing parties over the total number of seats in the government. The second moderator contains papers that have used indexes that measure the degree of

[5]Following Potrafke (2011), cultural expenditures are defined as the amount of money devoted to activities related to recreational, sporting and religious services and any other activity is related to the promotion of culture.

government ideology. More precisely, we focus on three broader indexes that measure the ideology in a similar way. These indexes include the ones developed by Budge *et al*. (1993), Woldendorp *et al*. (1998) and Potrafke (2009). Higher values indicate governments are considered to be more left-leaning, while lower values are used for more conservative governments. Our moderator (called *Ideo Indexes*) takes 1 when an index is used and 0 otherwise. Although each index has its own merits, it was not possible to create a dummy variable for each one of them. That would result in too many moderators that have only a limited number of 1s and, practically, it would create problems to the estimation process. However, in order to check the validity of the above strategy, we discuss an alternative grouping in the robustness subsection. The third moderator covers all the estimates from papers that use a dummy variable (1 for left, 0 for right-wing) in order to distinguish left from right-wing governments. This measure has been criticized on the grounds of its simplistic dichotomy which does not allow to capture intermediate cases like coalition governments (Potrafke, 2017). This moderator variable, named as *dichotomous ideological variable* (*DIV*, thereafter), assigns 1 when the estimate is drawn from a paper that uses a dummy variable to capture government ideology and 0 otherwise.

The third group captures the different estimation methods that have been used. As discussed in Potrafke (2017), the econometric technique may be an important factor in explaining the diversity of the reported estimates. Specifically, the problem of endogeneity is quite pronounced in the literature of partisan politics; when voters disagree with the implemented policies (measured by the dependent variable) then this may affect the government that is elected next time (and therefore its ideology measured by the independent variable). In order to deal with the reverse causality issue, researchers have started using instrumental variable methods. This leads us to separate the estimates from the collected studies into three different categories. The first, which acts as the reference category, consists of papers that use some form of least squares (e.g. OLS, pooled OLS or GLS). The second variable is a dummy variable that takes 1 when the observed estimated coefficient comes from a panel estimation method (fixed or random effects), while the third one includes the estimates from studies that use more advanced estimation techniques (GMM or 2SLS) that take into account the endogeneity problem.

The fourth group of moderator variables refers to the model specification, that is, the specific form of the estimated equation. Throughout the collected papers, this equation takes several forms; from a quite parsimonious model, containing four to eight variables in total, (Bräuninger, 2005; Gaston and Rajaguru, 2013) to a more extended specification with 15 to 18 covariates (Leibrecht, Klien and Onaran, 2011; Bove, Efthyvoulou and Navas, 2017). Following the suggestion of Philips (2016) on the importance of taking into account the specifications that have been used, we distinguish among 13 variables that are most commonly inserted as control group in the estimated equation. Furthermore, we account for the number of countries that are included in each empirical study (*ncountries*). Finally, we consider the sample period of each paper by adding the average year of each study (*Average Year*).

The last group of moderator variables accounts for publication characteristics. The empirical literature of partisan politics has attracted the research interest of both economists and political scientists. Therefore, our collected pool of papers contains studies published in journals of both fields. In order to capture this specific feature, we add a dummy variable (*Economics Journal*) that takes 1 when the estimate corresponds to a study published in an economics journal and 0 otherwise.[6] A second publication aspect is the journal quality and is captured by adding the impact factor as a separate variable (*ifactor*). Finally, we take into account the impact of each study by the number of citations (*citations*).[7] The Table in the online Appendix H summarizes the moderator variables and their definitions used in our analysis along with the summary statistics. Online Appendix D presents the VIF statistics.

V. Meta-regression analysis

The key purpose is to identify the main drivers that explain the variation of the reported estimates and, therefore, affect the ideology-government spending relationship. This section explores which of the factors analysed above systematically affect the reported estimates. Our meta regression model can be written as:

$$r_{ij} = c + \beta^k ser_{i,j} + \sum_{s=1}^{29} \gamma_S^k X_{S,ij} + e_{ij} \tag{2}$$

where r is the partial correlation, the X matrix contains the moderator variables, β is the coefficient of the standard errors of r, γ_s are the coefficients of each moderator, while $e \sim N(0, \sigma)$. Using the same notation with the previous section, i is an index for a regression estimate from the jth study. The uppercase k indicates that the above equation is valid under model M_k. Our empirical approach is based on Bayesian model averaging. The technical details are explained in the online Appendix E.

The benchmark findings are summarized in Table 4 that reports the estimated PIPs as well as the posterior means and standard deviations. To make the results more legible, we use a visual representation in Figure 4, where the models with the highest posterior inclusion probabilities are summarized. The horizontal axis measures the cumulative posterior model probabilities with the best models depicted on the left. As we move to the right, each model's posterior probability diminishes. In the vertical axis, the moderators are sorted by descending order according to their PIP. In other words, variables on top of the axis play a more significant role in explaining heterogeneity as compared to the ones in the bottom. The red colour (lighter grey) indicates that the variable is included, and its estimated sign is negative, while the blue colour (darker grey) indicates a positive sign.

Starting from the spending categories, the military dummy variable is found to be negative across almost all models. This becomes evident from the fact that the horizontal bar that corresponds to the *military* variable has no gaps and is continuously red, indicating that it appears always with a negative sign. Therefore, this variable proves to be a robust driver of the observed heterogeneity. Our findings support that

[6]We use the Association of Business Schools (ABS) list as a guide to the distinction between economics and political science journals.

[7]The cut-off date for the number of citations is July 2018, when we concluded the data collection process.

TABLE 4

BMA Baseline results

Variable	*PIP*	*post-Mean*	*post-SD*
ser	0.801[c]	0.399	0.256
ncountries	0.046	0.000	0.000
Average Year	0.169	0.000	0.000
Military	0.978[b]	−0.068	0.015
Health	0.033	−0.001	0.001
Social	0.049	−0.003	0.002
Educational	0.081	0.001	0.007
Environmental	0.040	0.005	0.004
Housing	0.028	−0.001	0.000
Cultural	0.027	−0.000	0.002
Coalition	0.189	0.009	0.023
Fiscal Pos	0.911[b]	−0.071	0.030
GDP	0.971[b]	−0.157	0.033
GDPgr	0.889[c]	−0.077	0.034
Election	0.033	0.001	0.003
Inflation	0.945[c]	−0.175	0.056
Unemployment	0.119	0.003	0.010
Population St	0.089	0.002	0.005
Population	0.036	−0.000	0.003
Trade	0.982[b]	0.073	0.023
Labour Power	0.254	0.012	0.031
Globalization	0.033	0.000	0.003
Femate Part	0.111	−0.003	0.023
GMM	0.981[b]	−0.099	0.038
Panel	0.999[a]	−0.092	0.019
DIV	0.911[c]	−0.091	0.041
Ideo Indexes	0.852[c]	0.071	0.039
Economics Journal	0.059	−0.003	0.009
ifactor	0.033	0.000	0.001
Citations	0.969[b]	0.000	0.000

Notes: PIP stands for posterior inclusion probability. We assume unit information prior as parameters' prior and uniform model prior. a/b/c denotes decisive/strong/positive evidence of a regressor having an effect, respectively, according to Kass and Raftery (1995).

studies that solely focus on military government spending estimate a weaker relationship between left-wing ideology and expenditures. This is in accordance to the evidence provided by Albalate, Bel and Elias (2012) and Bove *et al*. (2017), according to which military expenditures are higher during right-wing administrations.

The second important finding is that differences in ideology indexes matter. Both *ideo indexes* and *DIV* are found to be present in almost all of the estimated models and therefore, are having high PIPs. Specifically, the use of an ideology index tends to produce more positive results. In other words, studies using any of the three ideology indexes explained above tend to report a more positive relationship between government spending and left-wing governments. The opposite is true for the case of *DIV*; taking into account the ideology using only a dummy variable tends to give less positive estimates. This means that a simple dummy variable that takes 1 for left and 0

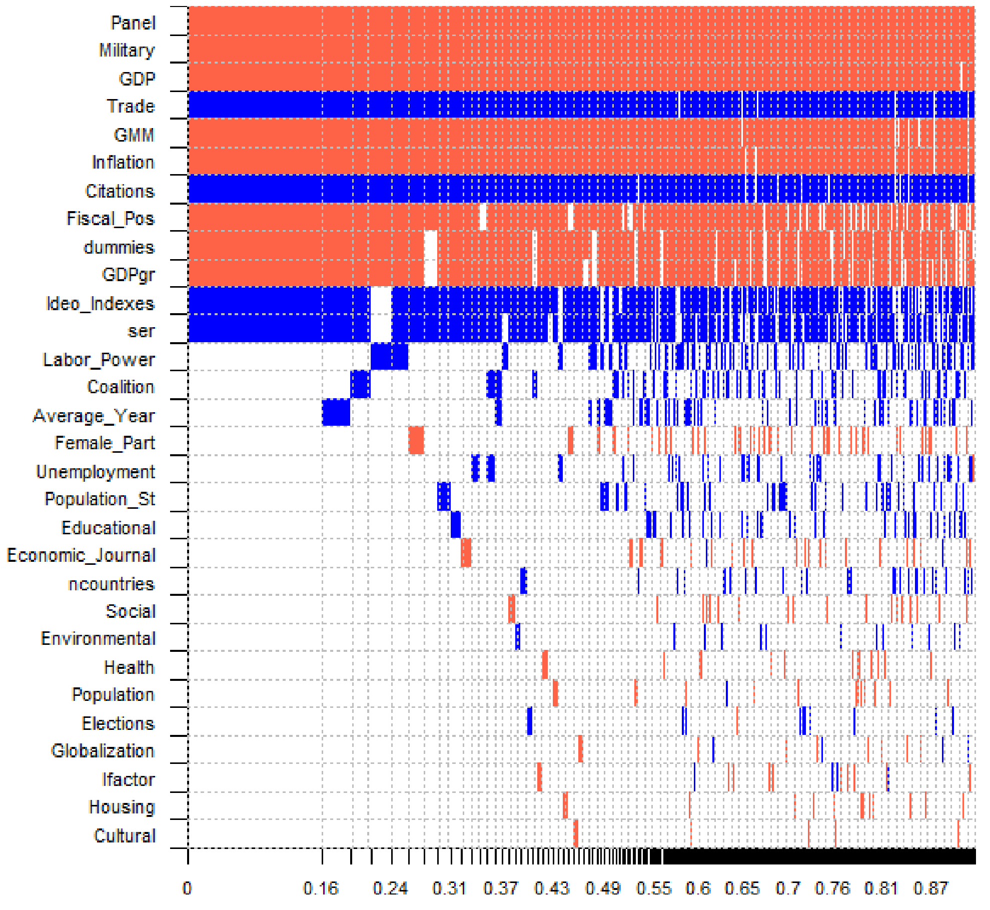

Figure 4. Bayesian Map-Baseline results
Notes: The horizontal axis measures the cumulative posterior model probabilities, while the vertical one depicts the moderator variables that are explained in the online Appendix H. Each column shows a different model. Each variable on the vertical axis is sorted according to its posterior inclusion probability in descending order meaning that variables on the top of the axis appear more frequently across different models than the ones at the bottom. Red colour (light grey) shows negative sign, while blue colour (dark grey) shows positive sign. Blank entries indicate that the variable is not included in the model. 3000 models with the highest posterior model probabilities are shown, while assuming unit information prior as parameters' prior and uniform model prior.

otherwise is not able to adequately capture the distinction among different ideological positions as detailed as an index. At least this is expected for cases where the distinction between left and right is not so clear (Beck *et al.*, 2001). On the contrary, in cases like United Kingdom and United States the distinction is easier. Overall, the measure of ideology employed is proved crucial to the final outcome. Given its significance as the main explanatory variable of interest, we explore further the categorization of ideology proxies. More precisely, we make an additional distinction between the more recent index developed by Potrafke (2009) and all the others. The results are discussed in the next section.

Another important outcome is the issue of the econometric methodology. Both moderators' (*Panel* and *GMM*) PIPs are found to be decisive and appear constantly in models with a negative sign. When a more advanced econometric technique is used, then the reported estimates tend to be less positive. This practically means that studies that do not take into account endogeneity tend to over-estimate the effect of ideology on government spending. This may also contribute in explaining the declining trend reported in Figure 1; more recent studies tend to use more modern techniques that take into account endogeneity.

The exact specification of the estimated model is also an important aspect of the literature. Studies that control for the level of economic activity (*GDP*), the growth rate (*GDPgr*), the inflation and the fiscal position tend to produce a smaller partial correlation, all else equal. Thus, basic macroeconomic fundamentals appear to be key factors in explaining the observed heterogeneity. This suggests that the level of development, the inflationary process, as well as, the fiscal conditions are factors that influence the reported results. The same appears to hold for trade openness. Specifically, studies that take into account the level of trade openness tend to produce larger partial correlations. The latter result reflects the so-called 'globalization effect' that has been extensively used as control in many branches of applied economics research (with growth econometrics and growth-finance literature being the most famous examples). On the other hand, other measures of globalization (like the globalization index developed by Dreher, 2006) that have been included in the control set are not found to be statistically significant. Moreover, other macroeconomic characteristics, like the unemployment rate or the population structure, do not have a statistically significant effect. This means that the partisan effect on government spending matters equally for low- and high-unemployment economies. The same is true for countries with ageing population and those with younger population, as the population-structure variable is not found to be significant. Furthermore, neither the *coalition* nor the *elections* variables appear to influence partial correlations, suggesting that the partisan effect on public spending is equally likely to occur under both coalition and non-coalition governments and regardless of the timing of elections.

Finally, some publication characteristics can also explain the heterogeneity in partial correlations. The most influential studies (i.e. those with higher number of citations) tend to report larger ideology-induced effects on government spending. The most cited study in our sample (Huber, Ragin and Stephens, 1993 with 1,244 citations) supports the view of a strong and positive relationship. On the other hand, studies (that are not so recent), as for instance, Kittel and Winner (2005) with much less citations (323), provide results in favour of a weak relationship between ideology and spending. It seems that some studies are used as references in the partisan-politics literature. However, this result should be cautiously interpreted as the estimated coefficient is almost zero. Finally, we do not find any difference between publishing in an economics or a political science journal. Therefore, the results of a study do not depend on whether the analysis is conducted by a political scientist or an economist.

TABLE 5

BMA results-Alternative Ideology categories

Variable	*PIP*	*post-Mean*	*post-SD*
ser	0.982[b]	0.492	0.141
ncountries	0.054	0.000	0.000
Average Year	0.169	0.000	0.001
Military	0.999[c]	−0.078	0.017
Health	0.033	−0.000	0.002
Social	0.056	−0.000	0.003
Educational	0.071	0.001	0.009
Environmental	0.041	0.000	0.004
Housing	0.039	−0.000	0.003
Cultural	0.024	−0.000	0.003
Coalition	0.067	0.001	0.008
Fiscal Pos	0.997[a]	−0.103	0.025
GDP	0.963[b]	−0.091	0.039
GDPgr	0.965[b]	−0.081	0.029
Election	0.031	0.000	0.003
Inflation	0.996[a]	−0.189	0.004
Unemployment	0.171	0.006	0.011
Population St	0.212	0.005	0.012
Population	0.029	−0.000	0.004
Trade	0.981[b]	0.069	0.011
Labour Power	0.031	−0.000	0.005
Globalization	0.033	0.000	0.002
Female Part	0.063	−0.003	0.011
GMM	0.986[b]	−0.052	0.011
Panel	0.997[a]	−0.110	0.031
DIV	0.997[a]	−0.121	0.030
Potrafke	0.051	−0.004	0.017
Other Indexes	0.998[a]	0.133	0.024
Economics Journal	0.031	0.371	0.007
ifactor	0.035	−0.000	0.002
Citations	0.943[c]	0.000	0.000

Notes: PIP stands for posterior inclusion probability. We assume unit information prior as parameters' prior and uniform model prior. a/b/c denotes decisive/strong/positive evidence of a regressor having an effect, respectively, according to Kass and Raftery (1995).

VI. Further evidence and robustness exercises

Alternative ideology categories

We start the robustness section by considering a supplementary dichotomy of ideology proxies. Potrafke (2017) discusses extensively the drawbacks of such measures and his motivation for the development of a new index discussed in Potrafke (2009). In this respect, we split the *Ideo Index* variable into two new candidate drivers; the first is assigned to 1 only when the Potrafke index is used (*Potrafke*), while the second takes 1 for all the remaining indexes (*Other Indexes*). Our results remain quantitatively and quantitatively the same under this alternative coding and are reported in Table 5. The

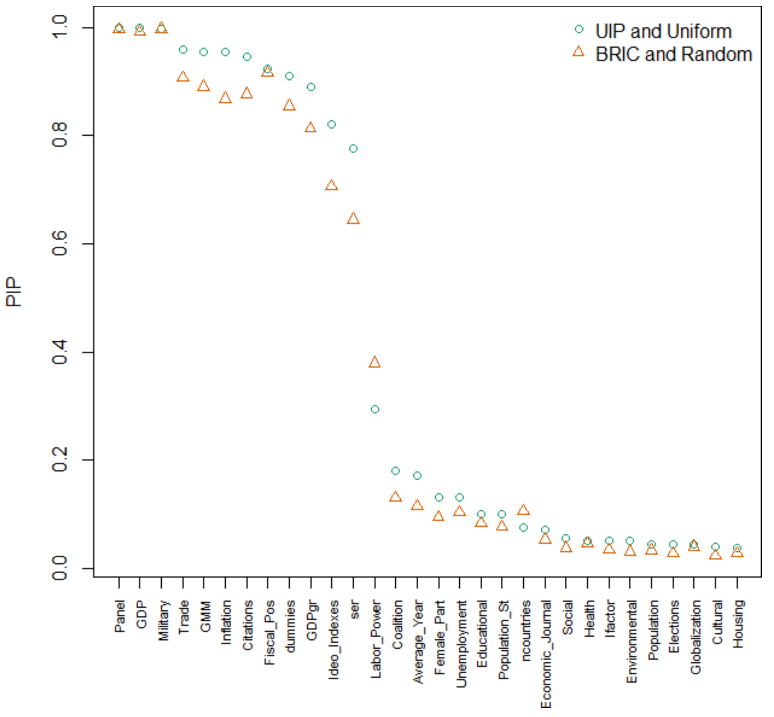

Figure 5. Comparison between two different sets of priors
Notes: The graph compares two BMA exercises that use two different sets of priors. The first model is the model estimated using UIP and uniform as parameters and model priors respectively. The second model uses hyper-g and beta-binomial as parameters and model priors respectively. The vertical axis depicts the estimated PIPs and the horizontal axis shows the variables. The cycles and the triangulars show the estimated PIPs of each variable for each set of priors.

use of indexes still tends to report a larger partial coefficient supporting the partisan theory. Also, the *DIV* continues to be an important driver in conditioning the ideology-spending nexus; studies that use a dummy variable as ideology proxy report a smaller effect.

Alternative priors and subsample analysis

The evidence presented so far is based on the usage of UIP and uniform as parameters and model priors, respectively. This choice is considered appropriate when there is not *a priori* knowledge. We check whether the result remain the same assuming hyper-g and beta-binomial respectively. To facilitate the comparison between the two estimations, we depict the two PIPs for each variable in Figure 5. Visual inspection suggests that the PIPs of each variable are almost identical regardless of the set of priors used. Furthermore, we investigate whether the results are driven by some specific publications or clusters of observations. For this purpose, we keep only the primary estimates reported from each study without including the reported robustness checks. The results are quantitatively and qualitatively the same (Table 6).

TABLE 6

BMA results-subsample analysis

Variable	*PIP*	*post-Mean*	*post-SD*
ser	0.704	0.245	0.211
ncountries	0.033	0.000	0.000
Average Year	0.174	0.000	0.000
Military	0.989[b]	−0.082	0.025
Health	0.058	−0.001	0.001
Social	0.017	−0.003	0.002
Educational	0.097	0.001	0.007
Environmental	0.028	0.000	0.000
Housing	0.037	−0.002	0.000
Cultural	0.014	−0.000	0.000
Coalition	0.187	0.007	0.013
Fiscal Pos	0.936[c]	−0.091	0.031
GDP	0.969[b]	−0.134	0.022
GDPgr	0.872[c]	−0.057	0.044
Election	0.022	0.000	0.000
Inflation	0.901[c]	−0.114	0.061
Unemployment	0.124	0.001	0.011
Population St	0.045	0.003	0.002
Population	0.041	−0.000	0.000
Trade	0.872[c]	0.023	0.023
Labour Power	0.287	0.010	0.012
Globalization	0.025	0.000	0.000
Femate Part	0.177	−0.001	0.002
GMM	0.874[c]	−0.091	0.087
Panel	0.990[a]	−0.075	0.023
DIV	0.923[c]	−0.083	0.087
Ideo Indexes	0.813[c]	0.062	0.044
Economics Journal	0.074	−0.001	0.003
ifactor	0.002	0.000	0.000
Citations	0.710	0.000	0.000

Notes: PIP stands for posterior inclusion probability. We assume unit information prior as parameters' prior and uniform model prior. a/b/c denotes decisive/strong/positive evidence of a regressor having an effect, respectively, according to Kass and Raftery (1995).

Alternative model averaging methods

As last robustness control, we estimate the model with alternative model averaging techniques. First, we apply frequentist model averaging (FMA). Apart from the generic differences between Bayesian and frequentist methodologies, the key difference between the two averaging methods is the construction of weights. While in the Bayesian case the posterior model probabilities are used as weights, in the frequentist approach the weights can vary depending on the researchers' preference. One choice is the usage of information criteria. Another choice that has been used in meta-analysis framework (Havranek, Rusnak and Sokolova, 2017) is the one proposed by Hansen (2007). The weights are selected by minimizing the Mallows criterion. The benefit is their asymptotic properties. The drawback (in all weighting schemes in the frequentist

TABLE 7

Frequentist model averaging and weighted-average least squares

	FMA		*WALS*	
Variable	*Coefficient*	*SD*	*Coefficient*	*SD*
ser	0.752*	0.420	0.200	0.147
ncountries	0.000	0.000	0.000	0.000
Average Year	0.003***	0.001	0.002***	0.001
Military	−0.064***	0.025	−0.060***	0.022
Health	−0.018	0.024	−0.017	0.021
Social	−0.007	0.020	−0.011	0.018
Educational	0.027	0.024	0.020	0.022
Environmental	0.022	0.026	0.010	0.023
Housing	−0.005	0.026	−0.009	0.023
Cultural	−0.004	0.026	−0.009	0.023
Coalition	0.098***	0.038	0.081***	0.031
Fiscal Pos	−0.092***	0.028	−0.078***	0.026
GDP	−0.164***	0.028	−0.141***	0.027
GDPgr	−0.089***	0.026	−0.083***	0.024
Election	−0.034	0.016	−0.002	0.014
Inflation	−0.138***	0.045	−0.104***	0.040
Unemployment	0.021	0.022	0.023	0.021
Population St	0.021	0.016	0.014	0.015
Population	0.017	0.020	0.015	0.017
Trade	0.091***	0.026	0.071***	0.021
Labour Power	0.062	0.041	0.063	0.037
Globalization	0.010	0.023	0.014	0.019
Female Part	−0.029	0.042	−0.011	0.037
GMM	−0.109***	0.026	−0.090***	0.025
Panel	−0.097***	0.018	−0.080***	0.016
DIV	−0.142***	0.033	−0.117***	0.032
Ideo Indexes	0.063***	0.027	0.046***	0.024
Economics Journal	0.021	0.031	0.011	0.023
ifactor	0.007	0.011	0.007	0.008
Citations	0.000	0.000	0.000	0.000

Notes: The WALS column show the results using the Weibull prior. *** and ** indicate statistical significant at 1% and 5%, respectively. For the case of WALS, the asterisks are used for illustrative purposes only and should be cautiously interpreted as this method is not purely a frequentist one.

methods) is that the MCMC algorithms are not anymore an option. This means that the model space has to be reduced. An appropriate way to do so has been proposed by Magnus, Powell and Prüfer (2010) and extended by Amini and Parmeter (2012). This method is based on the orthogonalization of the covariate space that leads to the significant reduction of the models that need to be estimated.

As an extra robustness check, we apply a closely related averaging scheme proposed by by Magnus *et al*. (2010) and Magnus and De Luca (2016). The weighted-average least squares (WALS) is a Bayesian combination of frequentist model averaging methods. More precisely, the parameters of each model are estimated in a frequentist way (least squares). Implementing a semiorthogonal transformation to the

TABLE 8

Best practice estimations

	Mean	*95% CI*	
Partial Correlations			
Using all estimates	0.092	0.079	0.104
Using only Indexes	0.081	0.071	0.092
Using only DIV	0.020	0.009	0.033
Using only Seats	0.067	0.056	0.077
Direct Estimates using log(gov)			
Using only Indexes	0.198	0.185	0.211
Using only DIV	0.063	0.052	0.073
Using only Seats	0.103	0.095	0.114
Direct Estimates using percentage of (gov) over GDP			
Using only Indexes	0.233	0.199	0.264
Using only DIV	0.098	0.089	0.101
Using only Seats	0.201	0.187	0.233

Notes: The implied estimates are based on the FMA estimations as BMA does not work with the concept of standard errors. CI stands for confidence interval. 'gov' stands for government expenditures and refers to the form of the dependent variable used in each case.

covariate space, the weighting scheme is based on Bayesian approach. The advantage of combining frequentist and Bayesian analysis is that it allows to reduce the computational burden. Applying both FMA and WALS leads to results that are quantitatively and qualitatively the similar to the BMA (Table 7). Overall, the moderator variables that were found to be robust drivers of the observed heterogeneity in the prior literature remain the same. Finally, we execute a pure frequentist exercise without using any weighting scheme. Taking into account the heterogeneity of the effect sizes, we run WLS. The results support the same evidence (online Appendix F).

Best practice estimations

So far we have explored the factors that explain the heterogeneity of our dataset. Additionally, we have shown that the partial correlation is small. A natural question is to ask 'how small'. The precise answer depends on the exact study design. It is a standard practice in meta-analysis to assume how the best study would be designed. Of course, there is not a unique research strategy. Using sample maxima for variables that are associated with the best practice, sample minima for the variables that depart from it and sample mean for variables that is not clear what it would be optimal, we create different designs and we get an estimate of the partial coefficient for each one.

The upper part of Table 8 shows four different 'best' scenarios. The first row of Table 8 shows the size of partial correlation when the best practice is based on the following design. Firstly, we do not have any prior preference regarding the category of government expenditures; whether a study focus on total spending or any other disaggregated category is equally preferable. Therefore, we use sample means for all the remaining spending moderator variables. Secondly, we assume that any ideology proxy is also equally preferable. So, we also use sample means for *share*, *DIV* and

Ideo Indexes. As far as the estimation method is concerned, we do prefer studies that take into account the endogeneity issue. Therefore, we align the sample maximum value of the variable *GMM* and the sample minima for *Panel* and *LS*. We also plug sample maxima for the impact factor. Given the lack of evidence regarding the existence of publication bias, it is legitimate to assign zero for the standard error. Regarding the average year, we use 2000 as the baseline year. As far as the number of citations is concerned, we use the sample average. Finally, for the variables related to the specification, we assign the maximum value (i.e. 1). In this way, we avoid any potential bias from omitting variables. This leads to an estimated effect of 0.092.

We repeat the same analysis for different choices about the ideology proxy. The second row shows the result when assuming that the best practice is the usage of an ideology index. Therefore, we add using the sample maximum of the *Ideo Indexes* and the sample minima to *DIV*. The resulting estimate is 0.081. We repeat the same exercise, treating the usage of *DIV* as the best practice design receiving 0.020. Lastly, the usage of parliament seats as an ideology proxy gives an estimate of 0.067. Based on Doucouliagos (2011) taxonomy, the estimated effects are small.

The above analysis provides an estimate of the implied partial correlation when the researcher follows a specific design. However, partial correlation is a measure of statistical, and not of economic strength of the examined relationship. In order to get an estimate of the economic strength, we have to restrict our sample to comparable studies only. For this reason we re-estimate the same model separately using two different subsets. Looking into the prior literature we discern that half of the papers use the logarithm of the government expenditure as dependent variable, while the remaining half of the papers use the percentage of government expenditure over GDP. Therefore, for each subgroup we evaluate the best practice estimate using the above-described design. The medium and the bottom panel of Table 8 show the corresponding estimates when the logarithm and the percentage over GDP is used respectively.[8] In all cases, the implied direct effect of government ideology on spending is found to be significant.

VII. Conclusions

The present paper is the first study that exploits the plethora of empirical studies on partisan politics by conducting a quantitative literature survey. The empirical research on partisan cycles has been extended to various policy fields. The meta-analytic approach herein focuses on the effects of government ideology on public spending in OECD economies. The collected papers, which have been published in both economics and political science journals, cover a period of 26 years providing a significant pool of reported estimates. The significant degree of heterogeneity of these estimates, both across and within studies as well as the conflicting results obtained, constitute the main motivation of our analysis; that is, to synthesize and evaluate the evidence published

[8]In the case of semilog form, the estimates can be interpreted as elasticities. Therefore, anything more than 0.1 can be reasonably evaluated as economically meaningful. We thank an anonymous referee for stressing this point.

so far, by exploring the factors that systematically affect the reported estimates. To this end, our analysis takes into account a series of factors that are related to different categories of spending, different proxies of government ideology, alternative model specifications as well as different methodologies in research design.

According to our findings, one important driver of the results is the adoption of more advanced econometric techniques that take into account the issue of endogeneity. Moreover, the government spending categories are found to be an additional important driver; in accordance to earlier evidence, more left-leaning administrations tend to spend less for military purposes. Not surprisingly, the proxy for measuring the concept of government ideology is also an important dimension that influences the reported results. The usage of ideology indexes tends to report a more positive relationship than simple measures based on the share of cabinet seats. Finally, the specification of the estimated equation plays a role in the final outcome, suggesting that the evidence of the literature hinge on the choice of the control variables which tend to influence the magnitude of the reported estimates. Interestingly, publication bias does not seem to be present in the examined literature. Our findings remain robust to a series of robustness checks that include different model averaging methods, different sets of priors as well as alternative coding of government ideology.

Overall, our quantitative analysis confirms the main arguments supported by Potrafke (2017); the relationship between ideology and government spending is positive and significant, although small. In this context, our findings provide implications for the study design of the partisan politics literature, especially amid the heterogeneous trends that emerge in the political landscapes and the complexities that arise for the measurement of government ideology as already stressed in the literature. For a more conclusive overview, additional quantitative surveys are needed though. The present paper is a first attempt to quantitatively analyse the empirical literature with main focus on a specific aspect of partisan politics. As future research project will enrich our understanding by examining other policy-making fields, such as taxation, debt and privatization.

Disclaimer

The views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of their respective institutions.

Authors contribution

Dr Georgios Magkonis was involved in conceptualization, writing, econometric execution and data collection. Dr Kallipi-Maria Zekente was involved in conceptualization, writing and data collection. V.Logothetis was involved in one third of data collection.

Final Manuscript Received: January 2021

References

Albalate, D., Bel, G. and Elias, F. (2012). 'Institutional determinants of military spending', *Journal of Comparative Economics*, Vol. 40, pp. 279–290.

Alesina, A. (1987). 'Macroeconomic policy in a two-party system as a repeated game', *The Quarterly Journal of Economics*, Vol. 102, pp. 651–678.

Amini, S. M. and Parmeter, C. F. (2012). 'Comparison of model averaging techniques: assessing growth determinants', *Journal of Applied Econometrics*, Vol. 27, pp. 870–876.

Andrews, I. and Kasy, M. (2019). 'Identification of and correction for publication bias', *American Economic Review*, Vol. 109, pp. 2766–2794.

Arestis, P., Chortareas, G. and Magkonis, G. (2015). 'The financial development and growth nexus: a meta-analysis', *Journal of Economic Surveys*, Vol. 29, pp. 549–565.

Beck, T., Clarke, G., Groff, A., Keefer, P. and Walsh, P. (2001). 'New tools in comparative political economy: the database of political institutions', *The World Bank Economic Review*, Vol. 15, pp. 165–176.

Blais, A., Blake, D. and Dion, S. (1993). 'Do parties make a difference? Parties and the size of government in liberal democracies', *American Journal of Political Science*, Vol. 37, pp. 40–62.

Bove, V., Efthyvoulou, G. and Navas, A. (2017). 'Political cycles in public expenditure: Butter vs. guns', *Journal of Comparative Economics*, Vol. 45, pp. 582–604.

Bräuninger, T. (2005). 'A partisan model of government expenditure', *Public Choice*, Vol. 125, pp. 409–429.

Budge, I., Keman, H. and Woldendorp, J. (1993). 'Political data 1945–1990. Party government in 20 democracies', *European Journal of Political Research*, Vol. 24, pp. 1–119.

Cusack, T. R. (1997). 'Partisan politics and public finance: changes in public spending in the industrialized democracies, 1955–1989', *Public Choice*, Vol. 91, pp. 375–395.

Doucouliagos, H., Haman, J. and Stanley, T. (2012). 'Pay for performance and corporate governance reform', *Industrial Relations: A Journal of Economy and Society*, Vol. 51, pp. 670–703.

Doucouliagos, H. C. (2011). How Large is Large? Preliminary and Relative Guidelines for Interpreting Partial Correlations in Economics. Working paper, Deakin University.

Drazen, A. (2000). 'The political business cycle after 25 years', *NBER Macroeconomics Annual*, Vol. 15, pp. 75–117.

Dreher, A. (2006). 'Does globalization affect growth? Evidence from a new index of globalization', *Applied Economics*, Vol. 38, pp. 1091–1110.

Franzese, R. and Jusko, K. L. (2006). 'Political-economic cycles', *Oxford Handbook of Political Economy*, Vol. VIII, pp. 545–564.

Furukawa, C. (2019). 'Publication bias under aggregation frictions: theory, evidence, and a new correction method', *Evidence, and a New Correction Method (March 29, 2019)*, Unpublished paper, MIT

Garrett, G. and Mitchell, D. (2001). 'Globalization, government spending and taxation in the oecd', *European Journal of Political Research*, Vol. 39, pp. 145–177.

Gaston, N. and Rajaguru, G. (2013). 'International migration and the welfare state revisited', *European Journal of Political Economy*, Vol. 29, pp. 90–101.

Hansen, B. E. (2007). 'Least squares model averaging', *Econometrica*, Vol. 75, pp. 1175–1189.

Havranek, T., Irsova, Z. and Zeynalova, O. (2018). 'Tuition fees and university enrolment: a meta-regression analysis', *Oxford Bulletin of Economics and Statistics*, Vol. 80, pp. 1145–1184.

Havranek, T., Rusnak, M. and Sokolova, A. (2017). 'Habit formation in consumption: a meta-analysis', *European Economic Review*, Vol. 95, pp. 142–167.

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*, Vol. 34, pp. 469–475.

Hibbs, D. A. (1992). 'Partisan theory after fifteen years', *European Journal of Political Economy* 8, pp. 361–373.

Hicks, A. M. and Swank, D. H. (1992). 'Politics, institutions, and welfare spending in industrialized democracies, 1960–82', *American Political Science Review*, Vol. 86, pp. 658–674.

Huber, E., Ragin, C. and Stephens, J. D. (1993). 'Social democracy, christian democracy, constitutional structure, and the welfare state', *American Journal of Sociology*, Vol. 99, pp. 711–749.
Imbeau L. M., Pétry, F. and Lamari, M. (2001). 'Left-right party ideology and government policies: a meta-analysis', *European Journal of Political Research*, Vol. 40, pp. 1–29.
Ioannidis, J., Stanley, T., and Doucouliagos, H. (2017). 'The power of bias in economics research', *Economic Journal*, vol. 127, pp. F236–F265.
Jensen, C. (2011). 'Two sides of the same coin? Left-wing governments and labour unions as determinants of public spending', *Socio-Economic Review*, Vol. 10, pp. 217–240.
Kass, R. E. and Raftery, A. E. (1995). 'Bayes factors', *Journal of the American Statistical Association*, Vol. 90, pp. 773–795.
Kittel, B., Obinger, H., et al. (2003). 'Political parties, institutions, and the dynamics of social expenditure in times of austerity', *Journal of European Public Policy*, Vol. 10, pp. 20–45.
Kittel, B. and Winner, H. (2005). 'How reliable is pooled analysis in political economy? The globalization-welfare state nexus revisited', *European Journal of Political Research*, Vol. 44, pp. 269–293.
Leibrecht, M., Klien, M. and Onaran, Ö. (2011). 'Globalization, welfare regimes and social protection expenditures in western and eastern european countries', *Public Choice*, Vol. 148, pp. 569–594.
Magnus, J. R. and De Luca, G. (2016). 'Weighted-average least squares (wals): a survey', *Journal of Economic Surveys*, Vol. 30, pp. 117–148.
Magnus, J. R., Powell, O. and Prüfer, P. (2010). 'A comparison of two model averaging techniques with an application to growth empirics', *Journal of Econometrics*, Vol. 154, pp. 139–153.
Philips, A. Q. (2016). 'Seeing the forest through the trees: a meta-analysis of political budget cycles', *Public Choice*, Vol. 168, pp. 313–341.
Potrafke, N. (2009). 'Did globalization restrict partisan politics? An empirical evaluation of social expenditures in a panel of oecd countries', *Public Choice*, Vol. 140, pp. 105.
Potrafke, N. (2011). 'Does government ideology influence budget composition? Empirical evidence from oecd countries', *Economics of Governance*, Vol. 12, pp. 101–134.
Potrafke, N. (2017). 'Partisan politics: the empirical evidence from oecd panel studies', *Journal of Comparative Economics*, Vol. 45, pp. 712–750.
Stanley, T. D. (2005). 'Beyond publication bias', *Journal of Economic Surveys*, Vol. 19, pp. 309–345.
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*, Vol. 70, pp. 103–127.
Stanley, T. D., Jarrell, S. B., and Doucouliagos, H. (2010). 'Could it be better to discard 90% of the data? A statistical paradox', *The American Statistician*, Vol. 64, pp. 70–77.
Woldendorp, J., Keman, H. and Budge, I. (1998). 'Party government in 20 democracies: An update (1990–1995)', *European Journal of Political Research*, Vol. 33, pp. 125–164.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix A. PRISMA chart
Appendix B. Andrew and Kasy (2019) method
Appendix C. Furukawa (2019) stem-based method
Appendix D. Multicollinearity Measures
Appendix E. Bayesian setup
Appendix F. Additional Results
Appendix G. Studies used
Appendix H. List of Moderator Variables