INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

META-ANALYSIS: FISCAL MULTIPLIER

Michal Hlaváček
Ilgar Ismayilov

IES Working Paper 7/2022

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Hlaváček M. & Ismayilov I. (2022): "Meta-analysis: Fiscal Multiplier" IES Working Papers 7/2022. IES FSV. Charles University.

Meta-analysis: Fiscal Multiplier

Michal Hlaváček[a]
Ilgar Ismayilov[a]

[a]Charles University, Prague

May 2022

Abstract:
Interest in fiscal policy has been dynamically improved in the last two decades; the number of research conducted on this topic has significantly increased in recent years. One of the key areas in fiscal policy investigations is the size of the fiscal multiplier, and most studies find contradictory results. In the current study, the unique dataset of 132 studies and more than 3200 observations were used to conduct a meta-analysis on multiplier effects, and several linear and non-linear models were involved in implementing this exercise. Additionally, Bayesian Model Averaging was first implemented to investigate heterogeneity effects in the meta-analysis of the fiscal multiplier. The results show that the fiscal multiplier is significantly less than one in the range of 0.75- 0.82. Moreover, the main contribution of the current study to the fiscal policy literature is disentangling the existence of selection publication bias in the literature.

JEL: E62
Keywords: fiscal multiplier, meta-analysis, publication bias

1. Introduction

Interest in fiscal multipliers has been dynamically improved in the last two decades, especially after the global financial crises. This was due to the increased role of fiscal policy after the monetary policy interest rates hit the zero lower bound, and fiscal stimulus and later budget deficit policies were one of the primary tools to prevent economies from freefalling into crises. The term ‘fiscal multiplier’ was first introduced in the classic paper by Blanchard and Perotti (2002), and it is defined as the ratio of the change in output induced by government intervention actions to the government spending or tax changes. Despite there being no consensus on the methodology, surprisingly bulk of empirical studies provide a narrow range of 0.6 to 1 for the estimated multiplier (Ramey, 2019). In the current study, an extensive database from 132 studies was developed, and systematic statistical analysis was conducted to obtain an integrated overview of the fiscal multiplier estimates. In particular, the main research question is to check whether publication bias exists in fiscal policy literature.

Literature of empirical analysis conducted on fiscal policy topics possesses high complexity. Several alternative approaches are the result, partially caused by the development and perception of economic theory and modeling framework and partially by the rich nature of the fiscal policy and the availability of a list of tools and instruments to reach the strategic goals. The large set of options that can be the focus of research provides excellent flexibility in study design and analytical models. According to Ioannidis (2005), such flexibility opens the way for publication bias that searches for statistically significant results to maximize the probability of publications. Meta-analysis is the method that provides rigorous quantitative survey techniques shedding light on the existence of publication bias (Stanley, 2001; Havranek, Horvath, and Zeynalov, 2014). In the current study, estimates of fiscal multiplier were tested for publication bias using linear and non-linear models, and by applying Bayesian techniques, systematic differences were investigated. The detailed analysis provides that the literature suffers from publication bias.

It should be noted that results supporting the existence of publication bias in fiscal policy literature might seem as findings contradicting the previous studies written on this topic, i.e., Gechert (2015), Gechert and Rannenberg (2018), Asatryan, Havlik, Heineman and Nover (2020) that either find no evidence or week support for publication bias. The current analysis partially shares the results of the stated works; however, the main difference comes from heterogeneity investigation built on Bayesian Model Averaging (BMA) that either missing or relied on different techniques than the one employed in the current work.

The contribution of this paper is threefold. The first contribution is constructing a new database on fiscal multiplier estimates. The database developed and used in the current paper is an extended version of Gechert and Rannenberg (2018). By adding 33 new studies and more than 1300 observations, the extended database contains 132 studies and more than 3200 observations. Additionally, 39 variables were considered in our analysis, whereas only 24 variables were used in the previous database. The number of potential variables is almost unlimited and to cover all differences across studies is an unfeasible task. However, it is believed that the new series is controlling for essential features missed by previous studies that might be crucial in determining heterogeneity impacts, i.e., publication-quality variables. The second contribution is the implementation under formal testing of both linear (OLS, fixed-effect, hierarchical) and non-linear (selection, the weighted average of adequately powered (WAAP), stem-based, p-uniform*) models to find actual multiplier size and to check for publication bias. To the best of our knowledge, the current paper is the first work implementing the stated advanced methods in the meta-analysis of the fiscal multiplier. In addition, most of the models confirm that the multiplier corrected for bias ranges from 0.751 to 0.827. The third main contribution is the implementation of Bayesian model averaging (BMA), "frequentist check," a hybrid of frequentist and Bayesian model and frequentist model averaging in the heterogeneity analysis that allows us to handle model uncertainty and precision of estimates and prevent excluding important variables from multi-variable regression. Finally, the model developed employing Bayesian techniques supports the hypothesis that fiscal policy literature suffers from publication bias, which is the current study's leading and most important contribution.

The rest of the paper proceeds as follows. Chapter 2 reviews data samples, variables, and patterns in the series. Formal tests for publication bias are provided in Chapter 3. Heterogeneity effects, multi-variable model estimation and results are discussed in Chapter 4, while Chapter 5 concludes.

2. Dataset

In the current research, we are using a dataset constructed by Gechert and Rannenberg (2018) (*further referred as GR(2018)*). The data set of GR (2018) covers only empirical studies that either vector autoregressive (VAR) models, or single equation estimates (SEEs), leaving out all multipliers from structural model simulations. Most papers in their sample have been published after the Great Recession and policy actions following the crises (GR 2018, page 1161). In addition to the publication bias investigations, one of the key questions that GR (2018) focused on was the regime dependency of the multiplier among the other sources of heterogeneity in the reported estimates.

The database used in the current paper is the extension of the one developed and analyzed by GR(2018). For the new portion of the dataset, observations were collected closely following the data structure in the GR(2018). Our primary focus is on empirical studies that used VAR and SEE in fiscal policy investigations. GR(2018) covered the studies belonging to the period 1992 - 2013, with 99 studies and more than 1900 observations. We extend the database by adding the period till 2020 and more than 1300 observations from 33 studies listed in Table 1. We cover 132 papers and more than 3200 observations in the combined dataset. Additionally, we include several new control variables into the existing set of factors.

We use Google Scholar for collecting data as it is superior to all other databases. Google Scholar has a powerful full-text search and covers all papers. However, other databases might limit their search with the title, keyword, and abstract. We examine the first 500 studies returned by the search. Initially, the abstract of each study was examined to identify those that may potentially include empirical estimates of the fiscal multiplier. After determining the shortlist of studies with possible empirical estimates, only those were downloaded and read in detail. Furthermore, we inspected the lists of references of all these studies to find potentially important papers omitted by our Google Scholar search; the literature search was terminated on January 31, 2021.

Generally, the fiscal multiplier (μ) is defined as the ratio of a change in output (ΔY) to a change in government expenditure (ΔG). Thus, the fiscal multiplier measures the effectiveness of the fiscal policy by providing how large is the economy's response to the government intervention. If the ratio is larger than one, then it shows the value created in the economy by the multiplication effect exceeds the resources spent by the government.

$$\mu = \frac{\Delta Y}{\Delta G} \quad (1)$$

Table1: List of studies used in meta-analysis

Afonso and Leal (2018)	Koh (2016)
Alloza(2018)	Kuckuck and Westermann (2014)
Amendola et al (2017)	Mencinger et al (2017)
Auerbach and Gorodnichenko (2017)	Miyamoto et al (2018)
Auerbach et al (2018)	Mortens and Raven (2010)
Ben Zeev and Pappa (2015)	Perotti (2014)
Boiciuc (2015)	Priftis and Zimic (2018)
Borg (2014)	Pyun and Rhee (2014)
Broner et al (2019)	Ramey and Zubairy (2018)
Caggiano et al (2015)	Ricco et al (2016)
Carnot and DeCastro (2015)	Riera-Crichton et al (2015)
Contreras and Batelle (2014)	Sheremirov and Spirovska (2019)
Cugnasca and Rother (2015)	Silva et al (2013)
Dell'Erba et al (2014)	Tang et al (2010)
Dupor and Guerrero (2017)	Vlasov and Deryugina (2018)
Estevao and Samake (2013)	Yadav et al (2012)
Forni (2015)	

As the estimate of the fiscal multiplier, empirical studies mainly calculate two types of it which are peak and cumulative multipliers:

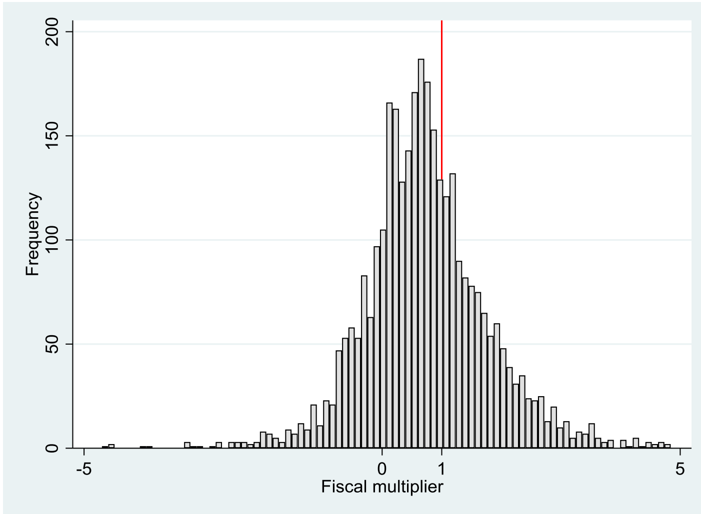
$$\mu_{peak} = \frac{max_h \Delta Y_h}{\Delta G_1} \quad (2) \qquad\qquad \mu_{cumulative} = \frac{\sum_t \Delta Y_t}{\sum_t \Delta G_t} \quad (3)$$

Peak multiplier considers relation of the largest change for the given horizon *'h'* in response variable to the government expenditure shock in the period one, however, cumulative multiplier is the ratio of the sum of all changes in output to the sum of all changes in government expansion for given period.

The distribution of collected observations is depicted in the Figure 1. The histogram of the collected estimates provides asymmetric, slightly skewed to the right distribution of fiscal multipliers. Inspection of the data discloses large range of estimates, between -9.8 and 24.97.

Of the 3,279 estimates we have, 2599 observations are positive, 1154 of them support the hypothesis that fiscal multiplier is greater than 1, further 356 observations provide cases when studies found estimates greater than 2, merely 1243 observations (or 37.9%) fall into the (0; 1) interval. Although we see a sharp drop around zero, the distribution of estimates around 1 seems relatively smooth. It might result from tending not to report negative estimates in the literature.

Figure 1: Distribution of the reported estimates



Notes: Histogram of the multipliers collected from individual studies.
The vertical red line denotes the value of multiplier equal to 1.

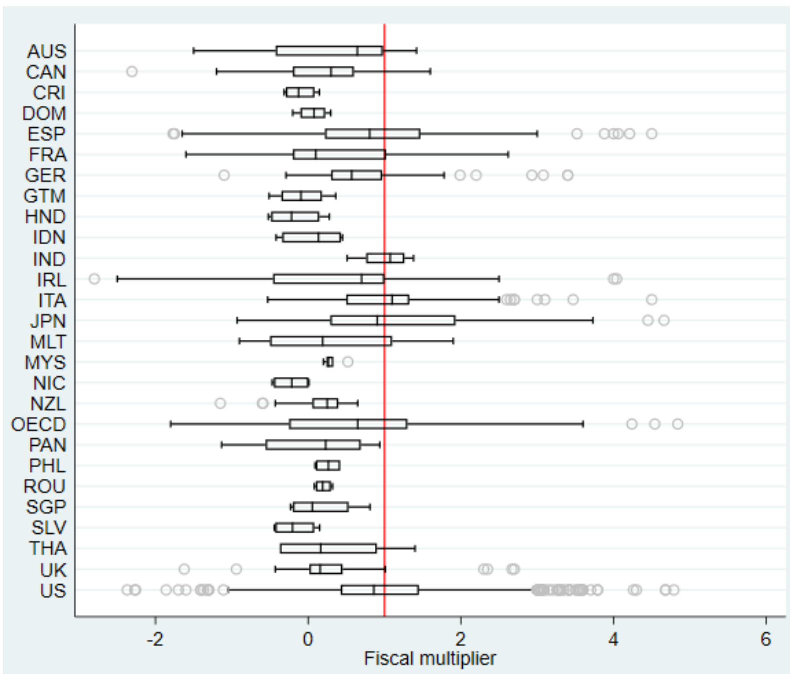
Theory mainly supports a positive fiscal multiplier, and most of the studies argue that it is approaching either one or zero. Therefore, negative estimates might be excluded due to the perception of them as miscalculated.

Some observations provide too side values, which are an apparent outlier in the sample. To prevent these outliers from driving our results, we winsorize the sample at 3.5% (the level at which our results stabilize and hold irrespective of further winsorization) and work with the winsorized sample from now on. After winsorization, the reported estimates range from -1.1 to 3.0 and are characterized by a mean of 0.75 and a median of 0.68. Figure 3 shows how estimates of multiplier vary around theoretically desired value in different studies. Moreover, from Figure 2, it is also obvious that variation in estimates exists for within-study observations, and the same picture holds for individual countries as well.

Before starting a detailed analysis, first, we introduce a brief overview of heterogeneity in the data. Table 2 provides summary statistics for various subsamples of the data. The table also contains a weighted mean: weighting by the inverse of the number of estimates reported per study ensures that all studies get the same weight. The exact definitions of the listed groups can be found later in Table 5.

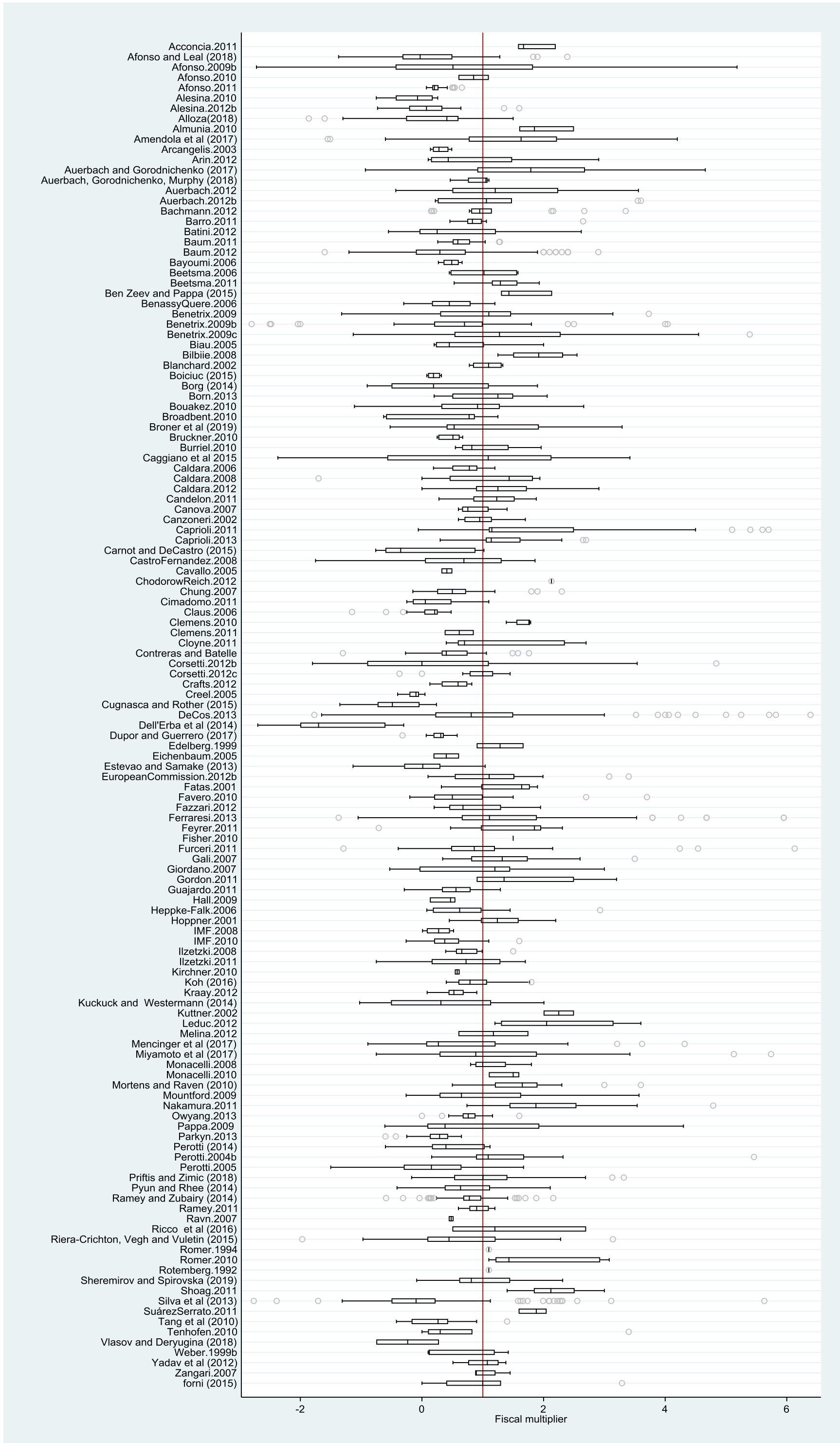
First, we see that both weighted and unweighted means of the fiscal multiplier are positive; however, both are less than one, which means that, on average, economies expand less than resources spent by governments while following fiscal policy aims. However, the size of the fiscal multiplier is not uniformly stable across different subsamples, as can be detected at first glance. Data can be conditionally grouped based on models employed, data structure, identification methods, regimes, shock types, frequency and many other characteristics. We see that almost 3/4 of the studies investigating fiscal policy topics employ VAR models; the rest of the works use SEE. However, estimates obtained under both models are similar. Figure 4.A visualizes this distribution of estimates under different model specifications. The identification techniques and financing schemes stand at the center of huge debates about correctly measuring the size of the fiscal multiplier. Figure 4.C summarizes the role of the identification methods in generating a fiscal multiplier.

Figure 2: Cross-country heterogeneity



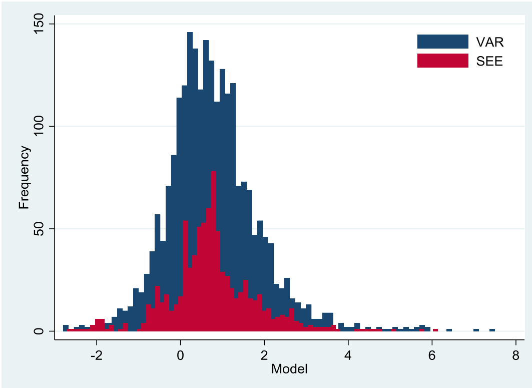
Notes: The length of each box represents the interquartile range (P25-P75), and the dividing line inside the box is the median value. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. Outliers are excluded from the figure. The vertical red line denotes value of multiplier equal to 1. For ease of exposition, outliers are excluded from the figure but included in all statistical tests.

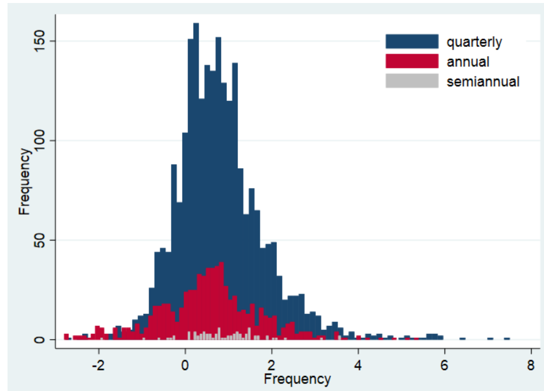
Figure 3: Estimates of multiplier vary both within and across studies



Notes: The length of each box represents the interquartile range (P25-P75), and the dividing line inside the box is the median value. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. Outliers are excluded from the figure. The vertical line denotes unitary elasticity. For ease of exposition, outliers are excluded from the figure but included in all statistical tests.

Figure 4: Patterns in the data

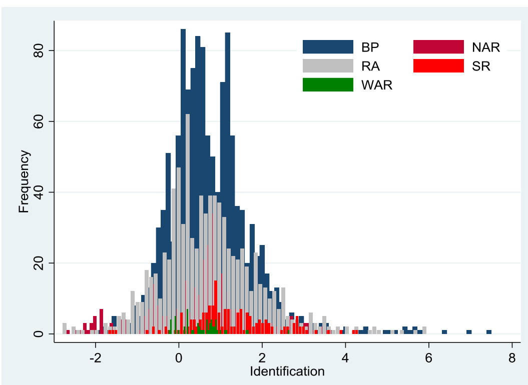




Notes: VAR – vector autoregression
SEE – single equation estimates

Figure 4.A Model types

Figure 4.B Frequency patterns



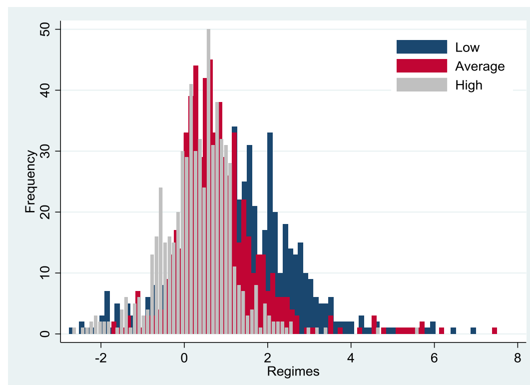


Figure 4.C: VAR: Identification strategies

Figure 4.D Regimes

Notes: BP - Blanchard and Perotti, RA- recursive approach
WAR – war episodes, NAR – Narrative approach;
SR - sign-restriction method

Notes: Low regime – recession, High regime - booms
Average regime- linear or non-specisfied regime

To measure the causal effect in a multi-year framework requires exogenous variation in policy variables. Any analysis built on time series of variables confronts identification issues and the estimation of the effects of fiscal policy is also not an exception. The problem that arises with the fiscal policy estimation is reverse causality between government expenditure and output (Ramey, 2019). In order to solve endogeneity problem stemming from reverse causality, as described in GR (2018), multiple equation models rely on 5 different methods. The same problem in terms of SEE models involves using 4 different approaches (complete description of methods are provided in chapter 4. Heterogeneity). As depicted in Figure 4.C, Narrative (VARNAR) and sign-restriction (VARSR) approaches are able to generate larger multipliers compared to others; for SEE models, on average larger estimates are obtained using the instrumental variable method (SEEIV). The difference in the size of estimates based on identification strategies should be interpreted differently than simple heterogeneity. The variation of the estimates based on the identification methods could be considered either as the effectiveness, suitability of the method in fighting against attenuation bias (Ramey and Zubairy, 2018, page 869), or drawback in capturing all exogenous variation in policy variables (Ben Zeev and Pappa, 2017, page 11). In both cases, the existence of bias is consistent.

The following important point is the state of the economy, which is considered a speculative factor affecting the size of the multiplier. A comparison of fiscal multiplier under linear and multiple states models is provided in Figure 4.D. In addition, Figure 4.D depicts the relation between the size of the multiplier and states of the model. This point is also important as different results contradict each other, i.e., Auerbach and Gorodnichenko (2012a, 2012b, 2017), Bachman and Sims (2012) argue that the role of states is essential in estimating responses to government intervention. However, Ramey and Zubairy (2018) and Afonso et al. (2018) find either small or no impact in favor of nonlinearity. Overview of the literature summarized in Table 2 and Figure 4.D supports the position of the former group of studies, describing that simply models not distinguishing between states of the economy is linear averaging of estimates from the multiple regime models, with the multiplier being higher during the economic downturn.

TABLE 2: Fiscal multiplier for different subsets

Variable	Obs	Unweighted			Weighted		
		Mean	95% Conf. interval		Mean	95% Conf. interval	
Mult	3279	0.75	0.72	0.79	0.86	0.84	0.89
Model							
SEE	859	0.75	0.69	0.82	0.84	0.78	0.90
VAR	2420	0.75	0.71	0.79	0.87	0.84	0.91
Data							
PANEL	1271	0.63	0.58	0.68	0.64	0.59	0.69
TIME SERIES	2008	0.84	0.80	0.88	0.96	0.93	1.00
Identification							
VARNAR	110	0.97	0 .78	1.16	1.11	0.94	1.28

VARBP	1215	0.72	0.68	0.77	0.80	0.75	0.84
VARRA	893	0.69	0.62	0.74	0.92	0.87	0.98
VARSR	175	1.20	1.07	1.33	0.92	0.79	1.05
VARWAR	28	0.35	0.15	0.54	0.62	0.39	0.84
SEEIV	481	1.02	0.93	1.08	0.47	0.36	0.58
SEENAR	305	0.44	0.37	0.56	0.33	0.15	0.52
SEECA	44	0.03	-0.12	0.18	1.16	1.08	1.23
SEEWAR	29	0.72	0.54	0.91	0.60	0.44	0.77
JORDA	514	0.79	0.71	0.86	0.79	0.70	0.87
Regime							
AV	2031	0.72	0.68	0.76	0.87	0.83	0.90
LOW	620	1.24	1.16	1.32	1.27	1.19	1.36
UP	628	0.37	0.31	0.43	0.43	0.38	0.49
Frequency							
ANNUAL	732	0.57	0.49	0.64	0.78	0.71	0.86
SEMI-ANN	78	0.89	0.71	1.07	0.98	0.80	1.16
BIANNUAL	63	1.21	0.95	1.47	1.42	1.16	1.67
QUARTERLY	2,394	0.79	0.75	0.83	0.87	0.84	0.90
MONTHLY	12	1.24	0.51	1.96	1.64	1.03	2.26
Type							
CUM	2,515	0.74	0.70	0.78	0.87	0.84	0.90
PEAK	702	0.86	0.79	0.92	0.92	0.86	0.98
Publication							
TOP 5	226	0.95	0.87	1.03	1.15	1.07	1.24
JOURNAL	1,179	0.92	0.86	0.97	0.92	0.87	0.98
W/PAPER	1,874	0.62	0.58	0.68	0.77	0.73	0.82
Shock							
SPEND	1,336	0.88	0.83	0.92	0.94	0.89	0.98
CONS	665	0.95	0.89	1.02	0.88	.82	0.94
INVEST	218	1.26	1.11	1.41	1.43	1.30	1.56
MILIT	222	0.93	0.83	1.03	1.00	0.91	1.10
TRANS	77	0.54	0.35	0.74	0.62	0.45	0.79
TAX	543	0.22	0.16	0.28	0.59	0.52	0.66
DEF	160	-0.04	-0.14	0.06	0.31	0.19	0.43

Notes: The table provides the summary of estimates for different subsets of the data. The exact definition of the variables is available in Table 5. Weighted estimates: are weighted by the inverse of the number of estimates reported per study.

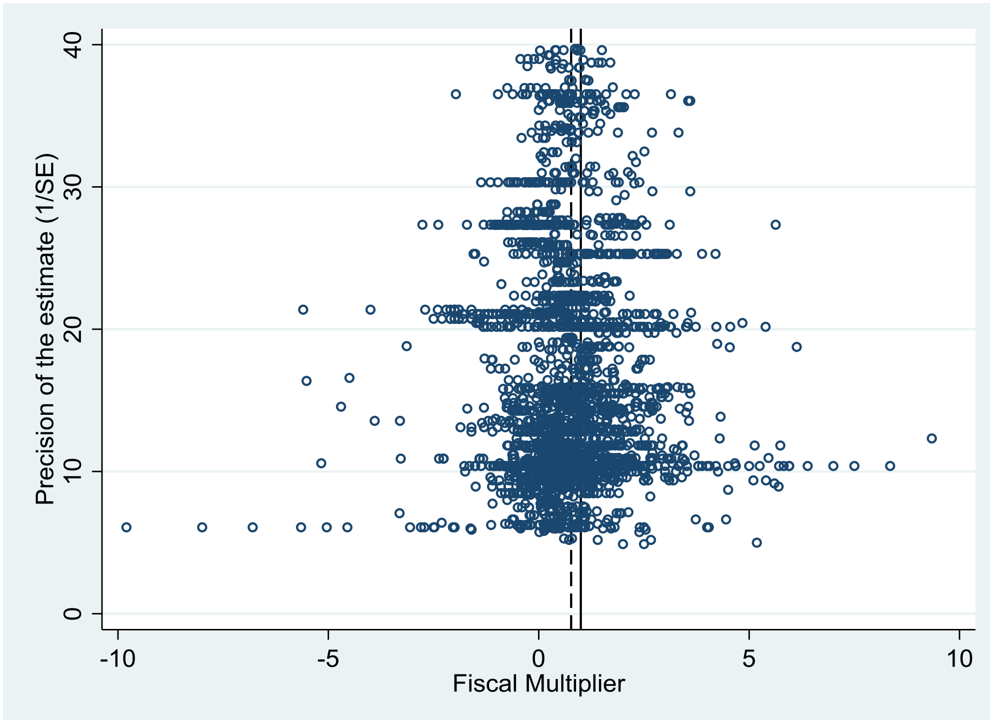
3. Publication bias

The size of the fiscal multiplier is one the central topics in economics. Additionally, after European Monetary Union emerged and the Great Recession hit the global economy, fiscal policy was the only tool to control the economy and fight back against crises for many national countries. Therefore, the importance of it and the interest to the fiscal policy and its effective implementation has been increased. In addition to long lasting and irreconcilable theoretical debate between New Keynesian and Neoclassical economic schools on the size of the fiscal multiplier, the question arises of how the situation with the increased importance of fiscal policy as described above may affect the research and its results on this topic. One should note that publication selection bias does not necessarily involve any ulterior motives on the side of authors, editors, referees; the existence of it could be quite natural (Havranek et al., 2021). And it is the task for reviewers to highlight such trend in the literature. In general publication bias is inevitable in economic studies and it is the role of reviews to check and correct the literature. According to Stanley (2008), the preference for statistically significant and theory-compliant results in publication selection, that is publication bias, can be controlled and identified by the meta-analyses.

The standard idea is that if there is no publication selection bias in the literature, then precision (reciprocal of standard errors) does not have an important impact on the effect size. In the meta-analysis literature, many ways, including linear and non-linear methods, have been developed. However, the logic behind all of them remains the same: if publication bias is not present in the literature, then standard errors of the estimates cannot explain any variation in the estimates of the variable of interest.

As a rule, a meta-analysis of any topic starts with the investigation of the funnel plot, which helps to detect possible bias in the literature. If there is no publication bias on the coefficient of interest, then estimates should draw a symmetric funnel around the most precise estimations of the true value. Figure 5 depicts a funnel plot where a solid vertical line indicates the case when the multiplier is equal to one and the dashed line is the sample average. It is easy to see that graph does not follow the symmetric pattern, which may indicate the presence of bias in the literature. Next, we switch to the formal testing of the bias, which is the implementation of the funnel-asymmetry test (FAT). To apply FAT we need to collect reported multipliers and their

Figure 5: Funnel plot



standard errors. However, studies investigating fiscal policy issues mainly report impulse response analysis. Moreover, many papers lack full information to calculate comparable standard errors, like the level of confidence bounds, or they display uncentered confidence bounds (GR, 2018). Standard errors are directly linked to study observation numbers, $\sigma_x = \frac{\sigma}{\sqrt{n}}$, where σ_x is standard error of i.i.d. random variable 'x', 'n' is the number of draws and σ population standard deviation. Standard error of a regression parameter is proportional to the inverse of the square root of the observations numbers; to see this relation from the formula is straightforward. Therefore, the inverse square root of the observations number represents a natural link for the standard error of the estimate, and according to Havranek (2015), they can be used either as a proxy or instrumental variable of standard errors for any studies. Stanley and Doucouliagos (2012) also argue that the number of observations can be used as a second-best proxy if standard errors are not available. Considering the situation with poor reporting of standard errors number of observations enters as the instrumental variable of standard errors in our estimations.

Conducting FAT means running an econometric model in the following form and checking if the coefficient of precision is statistically significant than zero:

$$\mu_{ij} = \alpha + \gamma(1/\sqrt{n})_{ij} + \varepsilon_{ij} \quad (4)$$

where μ_{ij} is the i-th multiplier reported in research j, n is the number of observation taken from paper j, and ε_{ij} is the error term. If estimation results reveal that the coefficient, γ, significantly different from zero, it will be the documentation of the asymmetry in the funnel plot and existence of selection publication bias in the estimations. However, in any case, the intercept, α, represents the true effect corrected for potential publication bias. In the formal analysis, we start by replicating the results of the GR (2018). Table 3 summarizes the results of this task, where one can see that both results are identical. In general, α is equal to 1.097, γ is around -3.0 and statistically different from zero. Such results support negative publication selection bias in the fiscal multiplier literature.

As described in GR(2018) page 1164, estimates are based on weighted least squares (WLS); however, clustering of errors was not applied. Considering the panel structure of the dataset, implementation of study-based clustering seems quite natural, and the majority of the recent meta-analyses follow this approach. Panel C of Table 3 shows the estimates calculated with standard errors clustered at the study level. With clustered errors, the coefficient on standard error becomes insignificant. This result might doubt the findings of the GR (2018) that finds negative selection bias in the literature. The next step is to run the same model with the extended database.

To check the robustness of the results, three different types of estimation techniques have been employed. The complete list of linear and non-linear techniques that have been used to check publication bias is provided in Table 4. The results from almost all models share the view on the size and the significance of both - the intercept, α, and the coefficient on standard errors, γ.

The first column of Table 4 contains the benchmark case, which is pooled OLS model with clustered errors. OLS finds true effect less than one and a nonsignificant coefficient on standard errors. The second column contains estimates obtained under the panel fixed-effect model, where unobserved heterogeneity across studies is controlled by study level-fixed effects. The estimates in the third column are obtained with the hierarchical Bayes model. This technique is a multi-level estimation that applies weights by implementing partial pooling at the study level and using with-in study variations. All linear models provide similar results, the size of the effect beyond bias matches across all linear approaches; it varies from 0.75 to 0.82. Additionally, all models reveal that there is no sign of the presence of publication bias in the literature on the fiscal multiplier. The first Colum of part B of Table 4 provides estimates of intercept and coefficient on standard errors using between the variations of studies. Among the linear models, only the last model provides the higher corrected effect of government intervention, yet it also found a fiscal multiplier less than one and no sign of publication bias.

TABLE 3: Replication the results of Gechert and Rannenberg (2018)

Part A	Estimates reported in GR (2018)		
FAT-PET	Estimates	STD Err	p- value
γ	-3.138**	1.509	0.037
α	1.099***	0.117	0.000
FAT-PEESE			
γ	-18**	8.61	0.037
α	0.9952***	0.0695	0.000
Part B	Replication		
FAT-PET			
γ	-2.976**	1.512	0.049
α	1.097 ***	0.118	0.000
FAT-PEESE			
γ	-17.708**	8.743	0.043
α	1.000***	0.070	0.000
Part C	With clustered errors		
FAT-PET			
γ	-2.976	2.996	0.323
α	1.097 ***	0.238	0.000
FAT-PEESE			
γ	-17.708	18.694	0.346
α	1.000 ***	0.154	0.000

Notes: In the estimations only data of GR(2018) were used. GR(2018) results were taken from the Table 2 at the page 1165. *, **, *** indicate significance at 10%, 5%, 1% levels.

The next sub-section of Table 4, columns 2 and 3 of Part B, contains the results of the estimates using two weighting schemes. In column 2 (inverse of) the number of observations per study was used to weight studies; thus, all studies included in the dataset possess the same weight, not depending on the number of estimates were taken from them. In the second scheme, the inverse of the standard errors, precision was used to weight the studies. It should be noted that GR(2018) results should be compared with columns 2 and 3 of Part B, as they were also obtained under WLS models. Estimates from WLS models are the closest ones in terms of the model to the estimates reported in GR(2018); however, as it was described above, the main contradictory result is an insignificant coefficient on standard errors, which is the direct result of using clustered standard errors. Weighted models also provide more or less similar estimates to the previous results; however, results obtained under study weights find the genuine effect of multiplier corrected for bias, α, larger than one. Despite contradictory results on intercept, under both weighting schemes, coefficient on standard errors is not statistically different from zero, which means there is no need to be cautionary about the presence of publication bias in the economic literature.

Previous techniques can provide reliable results if the relation between estimates and standard errors evolves as a linear dependency. Next, to cover the gap related to non-linear models, we switch to exploiting recently developed techniques to investigate the research questions. The summary of the results of non-linear models can be found in Table 4, Part C. In particular, used models are WAAP, Stem-based method (Furukawa, 2020), Selection model (Andrews and Kasy, 2019), and P-uniform*, which consider the more realistic relation of series and these models have been repeatedly employed in many meta-analysis works, i.e., Bajzik et al. (2020), Kočenda and Iwasak (2021), Havranek et al. (2021), Gechert et al. (2021), Zigraiova et al. (2021).

The first approach we use is the WAAP developed by Ioannidis et al. (2017). This approach focuses on estimations in the literature with adequate statistical power. For an estimate to have adequate power, its standard error should be less than the value of the estimate in absolute terms divided by 2.8. The value of 2.8 is the sum of the usual 1.96 for a significance level of 5% and 0.84, the standard normal value that makes a 20/80% split in its cumulative distribution (Ioannidis et al. (2017), page 239). If the standard error is less than this threshold, it means that estimate is adequately powered to detect the unknown actual effect.

The model proposed by Furukawa (2020) relies on the stem-based method in the meta-analysis of estimates. If by any means one can determine the subset of data that exhibits less bias, by focusing on the less biased subsample, bias can be mitigated even without estimating the publication selection process. According to Furukawa (2020), an overview of theory reveals that precise estimates are less biased compared to imprecise ones. Therefore, including precise estimates into the data problem related to the selection bias can be significantly improved. The method, in essence, is to optimize the trade-off between efficiency and bias to select the most relevant estimates.

Table 4: Results of funnel asymmetry test

Panel A	OLS	FE	Hierarchical
γ	-0.529	0.010	0.97
	(1.164)	(1.55)	(1.76)
	[-3.81 2.75]	-	-
α	0.790***	0.751 ***	0.82***
	(0. 152)	(0.110)	(0.13)
	[0.49 1.09]	-	-

Panel B	BE	Study-weighted	Precision-weighted
γ	-1.559	0.514	-0.652
	(1.449)	(3.235)	1.873
	-	[-7.50 7.91]	[-5.01 3.87]
α	0.993 ***	1.025 ***	0.785***
	(0.124)	(0.236)	(0 .132)
	-	[0.48 1.62]	[0.39 1.05]

Panel C	Stem-based method	WAAP	Selection model	p-uniform*
α	0.827*	0.752***	0.802***	0.7844***
	(0.446)	(0.152)	(0.027)	(0.3696)

Notes: The table contains results of the regression provided at equation (4). Standard errors are reported in the brackets. In square brackets were reported 95% confidence intervals from wild bootstrap clustering; implementation follows Roodman (2019) and we use Rademacher weights with 9999 replications.*, **, *** indicate significance at 10%, 5%, 1% levels, s.e. in the parantheses

The following model is based on Andrews and Kasy (2019), which belongs to the class of selection models. In general, selection models use a step function to provide the non-linearity in the publication probability of each research and based on the heterogeneity re-weights estimates in each bracket. The significant difference between the selection models is the weight functions they implement to obtain dissimilarity among the various groups. The model developed by

Andrews and Kasy (2019) is based on the assumption that publication probability changes noticeably after passing conventional t-statistics thresholds.

The last model given in Table 4 is p-uniform* developed by van Aert and van Assen (2021). This model also belongs to the group of selection models where the random-effects model the effect size. This model relies on the statistical principle that p-values should reflect uniform distribution at the correct effect size and, according to this rule, foresees the true effect. If we summarize the results of the nonlinear models, we see that all four models find corrected effect less than one. They range between 0.75 and 0.83 and all of them are statistically significant. Part C of Table 4 provides the results of the nonlinear models.

4. Heterogeneity

In the previous section, standard errors (inverse of the number of observations) were the main factor in analyzing variation in the size of estimates. However, the critical differences in terms of data, methods, models and many other key factors in each work might be as crucial as standard errors in the variation of estimates. Moreover, many additional characteristics of studies can undoubtedly cause systematic differences among the reported values. In this section, other factors than the standard error that can impact the magnitude of the multiplier will be an object of the attempt to statistically explain the source of the heterogeneity. The variables can be sorted into nine groups that each capture important features of the dataset. The following subsections explain each group and variable explicitly. Additionally, a summary can be found in Table 5. In addition to the traditional least-squares approach, we will employ Bayesian techniques also to conduct the exercise related to the heterogeneity. Bayesian techniques enable us to properly handle the model uncertainty issues and find the parsimonious model for the given list of potential variables.

Meta-regression analysis is mostly predetermined using the dataset developed by GR(2018). However, it is true that to cover all differences across studies is an unfeasible task. Therefore, the current study extends the initial dataset by including several additional variables that are believed could be useful in heterogeneity analysis. These new variables control for (new) identification strategy, type of data, and some publication characteristics not covered in GR(2018). In total, our complete dataset covers 39 variables that can be conditionally grouped into nine categories: Model, Data set type, Identification, Regime, Frequency, Impulse response type, Shock type or Financing, Publication characteristics, and Others.

The source of variations could be controlled by adding further key characteristics varying across studies. Many different aspects of study design should be considered in this context. First of all, the most obvious difference is the models employed, which are either single equation estimations (SEE), or vector autoregression (VAR). It should be noted that 73.8% of total observations are from VAR models. Additionally, three subgroups, financing, regimes, and identification, are important factors that investigation of their roles in the variation of fiscal multiplier might shed light on many discussions, i.e., Ilzetzki et al. (2013), Caggiano et al. (2015), Ramey and Zubairy (2018), Fatas and Mihov (2001), Batini et al. (2014).

TABLE 5: Description and summary statistics of variables

Variable	Description	Mean	STD. Dev.	Weighted mean
Multiplier	Fiscal multiplier, indicator measuring effectiveness of the fiscal policy, response variable	0.752	0.963	0.870
Standard error (SE)	Standard error of the fiscal multiplier	0.071	0.035	0.079
Model				
SEE	Single equation models	0.262	0.44	0.315
VAR	Vector Autoregression models	0.738	0.44	0.685
Data				
Panel	= 1 if dataset type is panel	0.388	0.487	0.285
Time series	= 1 if dataset type is time series	0.612	0.487	0.715
Identification				
VARNAR	=1 if VAR model developed based on narrative action based approach	0.034	0.18	0.063
VARBP	= 1 if VAR model developed based on Blanchard – Perotti approach	0.371	0.483	0.298
VARRA	= 1 if VAR model developed on recursive approach	0.272	0.445	0.234
VARSR	= 1 if VAR model developed based on sign restriction based approach	0.053	0.225	0.067
VARWAR	= 1 if VAR model developed based on war episode based approach	0.009	0.092	0.025
SEEIV	= 1 if SEE with instrumental variable approach	0.147	0.354	0.172
SEENAR	= 1 if SEE with narrative action based approach	0.093	0.29	0.069
SEECA	= 1 if SEE with prior cyclical adjustment of public budget	0.013	0.115	0.039
SEEWAR	= 1 if SEE with war episode based approach	0.009	0.094	0.035

Jorda	= 1 if Jorda method used to calculate IRF	0.157	0.364	0.072
Regime				
Av	= 1 if average or unspecified regime	0.619	0.486	0.804
Low	= 1 if downturn or crises regime	0.189	0.392	0.098
Up	= 1 if recovery or expansion regime	0.192	0.394	0.098
Frequency				
Annual	=1 if data frequency is annual	0.223	0.416	0.268
semi-annual	=1 if data frequency is semi-annual	0.024	0.152	0.032
biannual	=1 if data frequency is biannual	0.019	0.137	0.015
quarterly	=1 if data frequency is quarterly	0.730	0.444	0.670
monthly	=1 if data frequency is monthly	0.004	0.060	0.015
Type				
Cum	=1 if calculated as cumulative multiplier	0.767	0.423	0.732
Peak	=1 if calculated as peak multiplier	0.214	0.410	0.247
publication				
Top journals	= 1 if estimate was published in the top five journal	0.946	0.040	1.155
Journal	= 1 if the estimate is in a published study	0.262	0.440	0.361
Working paper	= 1 if the estimate is in a non-published study	0.738	0.440	0.638
Citations	The logarithm of per year citations, according to Google scholar	4.371	0.026	4.734
Published year	The logarithm of publication year	3.059	0.003	2.916
Shock				
Spend	= 1if public spending is unspecified	0.407	0.491	0.439
Cons	= 1if spending is public consumption	0.203	0.402	0.113
Invest	= 1if spending is public investment	0.066	0.249	0.061
Milit	= 1if public military spending	0.068	0.251	0.109
Trans	= 1if transfer to privet sector	0.023	0.151	0.021
Tax	= 1if tax reliefs to private sector	0.166	0.372	0.173
Def	= unspecified tax relief or spending increase	0.049	0.215	0.063
other factors				
HOR	Horizon of multiplier calculation	8.416	9.106	8.318
MGDP[1]	Import-to-GDP ratio of surveyed country sample	22.933	10.836	18.731

Endogeneity is present in the economic models developed to study fiscal multiplier topics due to reverse causality between GDP and Government expense. Therefore, financing and identification

[1]Data for share of import on GDP downloaded from World Bank World Development Indicators, indicator code: NE.IMP.GNFS.ZS; https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS

are seen as the main tools to solve the endogeneity issue. Related to this, several methods were developed for both SEE and VAR models. Summary of the 132 papers provides an overview that SEE models rely on one of the four following techniques: SEEWAR using war episodes (mainly the US) increase in defense spending as exogenous shocks; SEENAR using the same logic, however not limiting with war episodes, also considering exogenous tax changes; SEEIV using instrumental variables for Government expenses; SEECA identification relies on event studies using cyclically adjusted time series. Additionally, the influential paper by Auerbach and Gorodnichenko (2012) used the local projections method developed by Jorda (2005), which was highly widespread in fiscal multiplier investigations recently, Ramey and Zubairy (2018), Riera-Crichton et al. (2015), Miyamoto et al. (2018), Broner et al. (2019) and many others. The popularity of the local projections method results from its relative simplicity, non-parametric way of calculating impulse response functions, its robustness to misspecifications, and flexibility in capturing non-linear relations (Ramey, 2016).

VAR models choose among five identifications strategies to mitigate the endogeneity problem: VARWAR and VARNAR are identical to the SEE cases; VARRA is the recursive approach which consists of ordering variables such that in Cholesky decomposition, no contemporaneous impact of GDP on Government; VARBP the model based on the classic paper by Blanchard and Perotti (2002) that imposes elasticity for automatic stabilizers; and VARSR is the final approach that puts sign restriction on impulse responses while generating them. The first two approaches use additional historical information. However, the last three take all the information directly from the times series.

The reason why empirical strategies are complex and many, is the direct result of the perception of how they are effective in fighting against bias caused by endogeneity. One of the practical problems related to investigations of publication bias on reported fiscal multipliers is coexistence bias sourced by endogeneity additional to the bias as a result of publication selections, which further worsens the situation and makes it more complicated.

In addition to the crucial differences in empirical strategies, financing is also one of the key factors that might play the leading role in the size of the fiscal multiplier. In general, six different financing channels took the role in government actions.

The next important group of data characteristics is economic regimes. Estimates obtained under regime-dependent models correspond to the periods either the state of the economy is good (RUP) or bad (RLO). The rest of the estimates, from linear models or non-specified regimes, belong to the average regime (RAV). It should be noted that starting from 2012, many studies investigated the response of the economy to government intervention under different states of the economy; 47.3% of total observations starting with the first work with a multiple-regime model are reported referring to the state of the economy.

Other variables to be included in the multivariable meta-regression analysis to control for differences belong to the following subgroups: frequency of data, type impulse response function, type of data set, and others (horizon of impulse responses and import to GDP ratio).

In addition to study design factors, the current study also considers publication-quality characteristics of the research that estimates were published. The role of study level variables is to control for study quality that cannot be captured by study design, methods and many other similar factors. Our dataset covers five publication characteristic variables. Three of them are new factors and not covered in GR(2018): TOP5 dummy variable if research were published in the top 5 economic journals[2], (logarithm of) number of citations per published year, (logarithm of) publication year.

4.1. Estimation

The next task is to run the multi-variable regression using the control variables described in the previous section. With the additional control variables, the model will take the following form:

$$\mu_{ij} = \alpha + \gamma(1/\sqrt{n})_{ij} + \sum_{m} \beta_m * X_{m,ij} + \varepsilon_{ij} \quad (5)$$

As described in equation (1) μ_{ij} is the i-th multiplier reported in research j, n is the number of observation taken from paper j, the coefficient, γ, represents the intensity of publication bias and ε_{ij} is the error term. $X_{m,ij}$ is the m-th control variable corresponding to the i-th observation from research j, and β_m is the corresponding coefficient. The intercept α despite corrected for

[2] Papers published in one of the following journals are included to Top5: American Economic Review, Quarterly Journal of Economics, Journal of Political Economy, Econometrica, Review of Economics and Statistics, Science.

publication bias, unlike the single variable model, should not be interpreted as true multiplier since it depends on reference specification.

One can run regression including all variables; however, this approach would be problematic due to the underlying assumption that all variables are equally essential and ignoring model uncertainty. Additionally, including all variables will substantially decrease the precision of the parameters. An alternative way would be limiting the number of factors to a small set of them or running stepwise regression to exclude variables one by one. The drawback of these approaches would be the possibility of excluding important controls by chance and not covering all dissimilarities across studies. Therefore, following the literature of recent years, Aliminejad et al. (2021), Kočenda and Iwasak (2021), Havranek et al. (2021), Gechert et al. (2021) Bayesian model averaging (BMA) was applied as an optimal solution to address the above-stated problems. And then, the "frequentist check," which is a hybrid of the frequentist and Bayesian model, was applied to test the robustness of our estimates.

BMA follows the logic described below to assess the inclusion of each factor in the multivariable model. Firstly, using all possible combinations of explanatory variables, BMA runs numerous models and, for each model, calculates posterior model probability (PMP), which is the equivalent of the information criterion in frequentist econometrics. PMP is the performance indicator of the model compared to other ones. Then the procedure is to assign weights based on PMP to each model and construct, according to these PMPs, the weighted average for each coefficient across all models. The weighted sum of PMPs is the posterior inclusion probability (PIP) that determines which variables will be included in the model. For a more detailed explanation, one can refer to Raftery et al. (1997) and Eicher et al. (2011), among others. The total number of the regression that needs to be estimated in the first step of this algorithm is 2^k, where *'k'* is the number of control variables. After dropping nine variables not to cause a dummy trap, the number of estimations in our exercise would be 2^{30}, which is not a feasible task. Therefore, to make this task feasible Markov Chain Monte Carlo algorithm (Madigan and York, 1995) was implemented, where only models with the highest PMP were considered. The second step is to apply a 'frequentist check' to test the robustness of findings in step one. In the 'frequentist check,' only the variables with PIP above 85%, according to BMA results, were chosen and based on these variables, multivariable regression was run using OLS with study level clustered standard errors.

4.2. Results

Figure 6 contains the results of the BMA, where each column represents an individual regression model and the width of each column indicates the PMP of the corresponding model. According to descending PMP values, columns are ordered from the left to the right. On the other hand, each row represents an individual variable included in the analysis; they are also ordered according to descending PMP values from top to bottom. Blue cells indicate the positive value of the posterior mean of the coefficient on the variable in the model; in contrast, the red color reveals the negative value of the coefficient and blank cells indicate that the variable was not included in the model. The results show that 18 variables are important in the variation of the multiplier estimates.

To implement BMA priors for models and coefficients (g-priors) should be specified. In our baseline model, uniform distribution as model prior and Unit Information Prior (UIP) for coefficients were chosen. However, other alternative priors for model and coefficients (model prior, Dilution; g-prior, UIP and model prior, Random; g-prior BRIC) were also tested, results are robust irrespective of model and g-prior choices. The results with alternative distributions can be found in the Appendix, Figures AA2 and AA3.

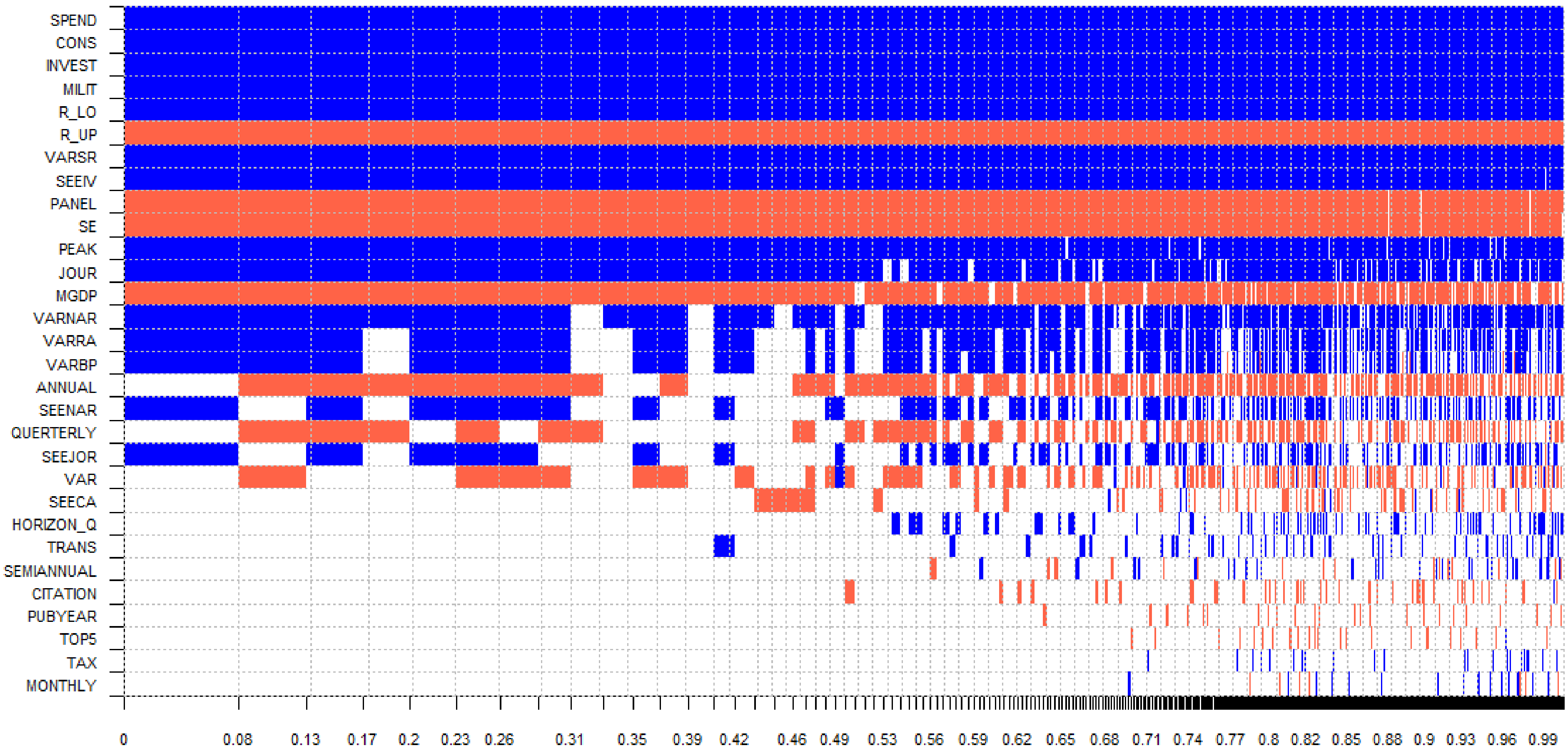
After inspection of Table 6 Part A, one can easily see that more than half of the variables included in BMA are important in explaining the heterogeneity of the fiscal multiplier and the sign of the impact is not changing across models. Table 6 is the numerical representation of Figure 6, where one can find values of PMP, posterior mean, and posterior SD. According to Jeffreys (1961), variables with respect to PIP could be classified into four groups: 'decisive' variables with PIP larger than 0.99, 'strong' variables with PIP between 0.95 and 0.99, 'substantial' variables with PIP from 0.75 to 0.95 and 'weak' variables labeled those that PMP does not reach to previous groups, but it is not less than 0.50. Table 6 provides that the 18 series has an important impact on the magnitude of the fiscal multiplier estimates where 11 out of the total 18 factors have decisive and 3 factors have strong and the remaining 4 factors have a weak effect. Table 6 part B provides the 'frequentist check' results which is the hybrid model of Bayesian model averaging and frequentist approach, where you can find that results support the finding of BMA. Additionally, in Table 8, results of the Frequentist model averaging (FMA) implemented using Mallow's criteria as weights are provided, which coincides with the results of

BMA. In contrast to the FAT tests described in the previous chapters, the most important finding of all three approaches is publication bias in fiscal policy literature; the multi-variable model confirms a statistically significant coefficient on standard errors. This result might seem confusing; however, it could be explained in terms of pure technical means. The FAT is based on regression with only a single variable, i.e., SE, and as BMA, 'frequentist check,' and FMA suggest other important variables, we may conclude that the FAT seriously suffers from omitted variable bias and the discrepancy in the results is caused by it. If we delve even further into details, a not statistically significant coefficient results from a small value of coefficient (in absolute terms) rather than the inflated standard deviation. The comparison of parameters makes reveals the existence of the positive omitted variable bias in the single variable regression.

The initial assessment and rough summary of the BMA is that three categories of variables play a crucial role in determining the size of the multiplier. These categories are financing schemes, regimes and identification strategies, and their role in determining the size of the multiplier coincides with the general lines of the theory developed on fiscal policy issues. These variables represent the attempts to mitigate the problems driven by endogeneity. The first line of struggle against the endogeneity would be finding pure exogenous series, i.e., military spending, tax changes. Additionally, on top of pure exogenous series, further measures might be developed or adjusted to obtain unbiased estimates.

Data characteristics. As it was described in the previous passage, standard errors (SE), consumption (CONS), investment (INVEST), states of the economy (RLO and RUP), panel type databases (PANEL) have a systematic impact on the magnitude of the multiplier, at least they belong to the group of strong variables if they are not the decisive variables. However, frequency variables do not seem to possess a substantial effect, only data with annual frequency (ANNUAL) has PIP above 50%, yet it does not qualify even for substantial variables and remains in the weak effect group. This result does not coincide with the literature because data with higher frequency are considered more reliable in fiscal policy investigations. Additionally, we do not see any clear pattern in terms of PIP along with the decreasing or increasing frequency of used databases.

Figure 6: Model inclusion in Bayesian model averaging



Notes: Explanatory variables are ranked according to their posterior inclusion probabilities along the vertical axis from the highest at the top to the lowest at the bottom. The horizontal axis depicts the values of cumulative posterior model probability. Blue color indicates if the estimated parameter of a corresponding explanatory variable is positive. Red color indicates if the estimated parameter of a corresponding explanatory variable is negative. No color means the corresponding explanatory variable is not included in the model. Numerical results are reported in Table6.

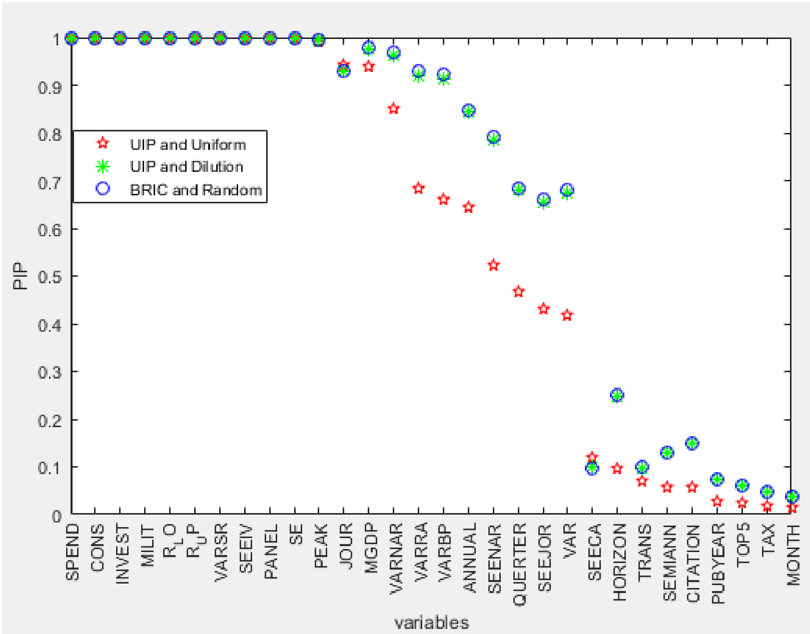
Table 6: Results of Bayesian Model Averaging and Frequentist check

	PART A: Bayesian Model Averaging			PART B: Frequentist check		
	PIP	Post Mean	Post SD	Coefficient	S.E.	p-value
SE	0.9992	-3.9773	0.8311	-4.231	1.452	0.004
VAR	0.4170	-0.2474	0.3492			
PEAK	0.9908	0.1532	0.0404	0.158	0.095	0.100
HORIZON	0.0974	0.0004	0.0012			
SPEND	1.0000	0.5875	0.0408	0.580	0.096	0.000
CONS	1.0000	0.7789	0.0474	0.770	0.117	0.000
INVEST	1.0000	1.1466	0.0647	1.131	0.228	0.000
MILIT	1.0000	0.7712	0.0875	0.686	0.219	0.002
TRANS	0.0711	0.0143	0.0601			
TAX	0.0188	0.0009	0.0116			
R_LO	1.0000	0.4308	0.0406	0.444	0.123	0.000
R_UP	1.0000	-0.4811	0.0403	-0.469	0.089	0.000
SEEIV	0.9999	0.5414	0.2008	0.431	0.144	0.003
SEENAR	0.5245	0.2377	0.2562			
SEEJOR	0.4300	0.1425	0.1781			
SEECA	0.1188	-0.0321	0.1085			
VARNAR	0.8503	0.5612	0.3821	0.237	0.233	0.309
VARBP	0.6612	0.3905	0.3373			
VARRA	0.6844	0.4360	0.3637			
VARSR	1.0000	0.8896	0.3604	0.472	0.190	0.014
PANEL	0.9995	-0.3694	0.0721	-0.377	0.127	0.004
ANNUAL	0.6444	-0.1832	0.1705			
SEMIANN	0.0588	-0.0019	0.0616			
QUERTER	0.4659	-0.1243	0.1492			
MONTH	0.0160	0.0012	0.0354			
JOUR	0.9412	0.1236	0.0467	0.149	0.081	0.068
TOP5	0.0252	-0.0019	0.0171			
MGDP	0.9381	-0.0030	0.0012	-0.003	0.002	0.111
CITATION	0.0563	-0.0025	0.0124			
PUBYEAR	0.0290	-0.0072	0.0567			
Intercept	1.0000	0.6007	NA	0.623	0.164	0.000

Notes: Dependent variable is fiscal multiplier, Post. mean - posterior mean, Post. SD - posterior standard deviation, PIP - posterior inclusion probability, SE - standard error. Part A, contains numerical results BMA based on the UIP g-prior and prior uniform distribution for model. Part B reports frequentist check results, which includes substantial variables with PIPs higher than 80% obtained from the baseline BMA specification. Standard errors in the frequentist check are clustered at the study level.

The next important result is the systematic differences in the size of the multiplier across the states of the economy, which excludes linearity of the fiscal multiplier, contributing to the discussion on the state dependency nature of multipliers, Ramey and Zubairy (2018), Auerbach and Gorodnichenko (2012), etc. Furthermore, most of the identification strategies appear as important factors in the result of BMA, but not all of them have the same highest importance level. For example, sign restriction (VARSR) for VAR models and instrumental variable approach for SEE models (SEEIV) have a decisive impact compared to other identification strategies. Additionally, among *the other factors*, the import-to-GDP ratio (MGDP) substantially affects the magnitude of the multiplier.

Figure 7: Posterior inclusion probabilities across different prior settings



Notes: UIP (unit information prior) and Uniform model priors according to Eicher et al. (2011). UIP and Dilution prior recommended by Eicher et al. (2011) and George(2010), respectively; BRIC and Random - the benchmark g-prior by Fernandez et al. (2001) for parameters with the beta-binomial model prior for the model space, which means that each model size has equal prior probability.

Publication characteristics. The majority of publication characteristics do not possess a substantial effect on the estimates of the fiscal multiplier. For example, only the variable indicating publication in journals (JOUR) has a meaningful impact; however, the number of citations per year, publication year, and publication in the top 5 journals do not provide any substantial impact. Nevertheless, the statistically significant impact of publication on journals with the positive coefficient might support the hypothesis about the selective nature of the academia; on average, estimates reported in published works are 0.12 higher than non-published counterparts.

As the bottom line of the current analysis 'best practice' multiplier was calculated considering all reported estimates and BMA results presented previously. Simply it is an exercise in which different weights are given to various data characteristics according to the author's preferences to compute the multiplier that might represent some key features of the database. Generally, we calculate fitted values for given conditions that represent essential features of the data. For weights plugged into the regression to calculate the fitted value, were used sample maxima if the variable is preferred, the sample mean if there is no preference and sample minima if the variable is far from the best practice. As the benchmark case, we refer to GR(2018)'s 'best practice' multiplier (2018). Additionally, we prefer quarterly frequency, published studies and frequently cited studies. In the benchmark case, the multiplier represents cumulative estimates from the VAR model with Blanchard and Perotti identification method, with quarterly frequency and panel data type, with mean import-to-GDP ratio, publication year and horizon length. Additionally, frequently cited and published papers also have larger weights. Table 7 contains the multiplier calculated according to the definition of best practice. For the benchmark case, the multiplier is equal to 0.73, which means preferred features almost offset the impact of each other.

Table 7: Best practice: Alternative specifications

	Multiplier	95% CI
Best practice	0.728	(0.384, 1.072)
Crises	1.519	(0.989, 2.050)
Boom	0.610	(0.220, 1.100)
Higher import	0.570	(-0.298, 1.439)
Annual	0.991	(0.358, 1.625)
Investment	1.620	(0.934, 2.306)
Tax	0.480	(-0.070, 1.030)
Military	1.275	(0.661, 1.890)

Notes: The table shows mean estimates of the fiscal multiplier conditional on model, identification, financing, multiplier type and publication characteristics. The exercise uses all information used in the literature but puts more weight on selected aspects of study design. Best practice definition is based on Gechert and Rannenberg (2018). The remaining rows report implied multiplier when we change one aspect in the definition of best practice. The 95% confidence intervals are reported in the second column. They are constructed using the standard errors estimated by OLS that are clustered at the study level.

The second column contains 95% Confidence intervals borders. From the table, it is evident that if investment data preferred fiscal multiplier would have the highest value; on the other hand, tax data generates the lowest value, almost half of the multiplier in the benchmark case. Moreover, bad states create higher multipliers compared to average regimes or economic booms.

TABLE 8: Results of frequentist model averaging

	Coefficient	Sd.Err.	p-value
SE	-3.7345	0.8571	0.000
VAR	-0.5049	0.2064	0.014
PEAK	0.1779	0.0390	0.000
HORIZON	**0.0040**	**0.0018**	**0.026**
SPEND	0.7123	0.0756	0.000
CONS	0.8968	0.0788	0.000
INVEST	1.2510	0.0921	0.000
MILIT	0.9051	0.0951	0.000
TRANS	0.1175	0.1390	0.398
TAX	0.1225	0.0764	0.109
R_LO	0.4444	0.0475	0.000
R_UP	-0.4575	0.0473	0.000
SEEIV	0.5632	0.1448	0.000
SEENAR	0.4765	0.1523	0.002
SEEJOR	**0.2727**	**0.1165**	**0.019**
SEECA	0.0564	0.1731	0.745
VARNAR	0.9079	0.2006	0.000
VARBP	0.7285	0.1790	0.000
VARRA	0.8061	0.1823	0.000
VARSR	1.2006	0.2041	0.000
PANEL	-0.3712	0.0767	0.000
ANNUAL	-0.4993	0.1785	0.005
SEMIANN	-0.2829	0.1901	0.137
QUERTER	**-0.4244**	**0.1801**	**0.018**
MONTHLY	-0.1282	0.2578	0.619
JOUR	0.0911	0.0415	0.028
TOP5	-0.0431	0.0789	0.585
MGDP	-0.0032	0.0011	0.004
CITATION	-0.0481	0.0376	0.201
PUBYEAR	-0.2485	0.2524	0.325
Intercept	0.9864	0.4473	0.027

Notes: I use Mallow's weights Hansen (2007), and the orthogonalization of the covariate space suggested by Amini and Parmeter (2012) to conduct frequentist model averaging (FMA) exercise. Bold black lines show variables important in FMA but not in the benchmark BMA.

For alternative specifications, the multiplier varies from 0.4 to 1.6; for comparison in GR(2018), the similar interval is from 0.5 to 1.3. Meanwhile, a high degree of uncertainty indicated by the broad boundaries of the confidence interval should also be noted. The interval of the estimates varies from around 1.0 to 1.6 and is the largest for the high import samples.

5. Concluding remarks

Using the large database, the current study presents an integrated overview of the fiscal multiplier estimates, the key parameter that provides how large is the economy's response to government intervention. The magnitude of the multiplier possesses high importance for researchers and policymakers in assessing the effectiveness of fiscal policy. Due to poor reporting of standard errors number of observations per study was chosen as the instrument of standard errors. The Bayesian and frequentist model averaging is the heart of all the analysis that solves the model uncertainty inherent in meta-analysis research. The reason behind the variation of fiscal multiplier estimates could be addressed by systematic analysis of the economic conditions, study design characteristics, and publication quality features.

In the current study, using the large database consisting of 132 studies and more than 3200 observations, linear and nonlinear meta-analysis methods were employed to quantify the impact of factors on the estimates. The results support that the genuine effect of the fiscal multiplier is positive but less than one ranging from 0.75 to 0.82 under different models. Moreover, the results of BMA suggest that one of the most important variables for explaining the variation in the reported multipliers is (the instrument) of standard errors. The results show that, despite it is not large in magnitude, fiscal policy literature suffers from publication bias. Furthermore, the heterogeneity analysis reveals the group of factors crucial in explaining the variation of the fiscal multiplier. The analysis supports that different financing schemes, the states of the economy, and empirical identification strategies have an important role in determining the size of the multiplier.

6. References

Afonso, A., Baxa, J., & Slavík, M. (2018). Fiscal developments and financial stress: a threshold VAR analysis. *Empirical Economics*, *54*(2), 395-423.

van Aert, R. C. M., & van Assen, M. A. L. M.. Correcting for Publication Bias in a Meta-Analysis with the P-uniform* Method. https://doi.org/10.31222/osf.io/zqjr9

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*(8), 2766-94.

Asatryan, Z., Havlik, A., Heinemann, F., & Nover, J. (2020). Biases in fiscal multiplier estimates. *European Journal of Political Economy*, *63*, 101861.

Auerbach, A. J., & Gorodnichenko, Y. (2017). Fiscal multipliers in Japan. *Research in Economics*, *71*(3), 411-421.

Auerbach, A. J., & Gorodnichenko, Y. (2012a). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, *4*(2), 1-27.

Auerbach, A. J., & Gorodnichenko, Y. (2012b). Fiscal multipliers in recession and expansion. *Fiscal policy after the financial crisis*, *63*, 98.

Bachmann, R., & Sims, E. R. (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics*, *59*(3), 235-249.

Bajzik, J., Havranek, T., Irsova, Z., & Schwarz, J. (2020). Estimating the Armington elasticity: The importance of study design and publication bias. *Journal of International Economics*, *127*, 103383.

Batini, N., Eyraud, L., and Weber, A. (2014). A simple method to compute _scal multipliers. IMF Working Papers 14/93, International Monetary Fund.

Ben Zeev, N., & Pappa, E. (2017). Chronicle of a war foretold: The macroeconomic effects of anticipated defence spending shocks. *The Economic Journal*, *127*(603), 1568-1597.

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of economics*, *117*(4), 1329-1368.

Broner, Fernando; Clancy, Daragh; Erce, Aitor; Martin, Alberto (2019).Fiscal multipliers and foreign holdings of public debt, ECB Working Paper, No. 2255, ISBN 978-92-899-3517-3, European Central Bank (ECB), Frankfurt a. M.,http://dx.doi.org/10.2866/109192

Caggiano, G., Castelnuovo, E., Colombo, V., & Nodari, G. (2015). Estimating fiscal multipliers: News from a non-linear world. *The Economic Journal*, *125*(584), 746-776.

Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, *26*(1), 30-55.

Elminejad, M. A., Havranek, T., & Horvath, R. (2021). Publication and Identification Biases in Measuring the Intertemporal Substitution of Labor Supply.

Fatás, A., & Mihov, I. (2001). The effects of fiscal policy on consumption and employment: theory and evidence. *Available at SSRN 267281*.

Fernandez, C., Ley, E., & Steel, M. F. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*(2), 381-427.

Furukawa, C. (2020). *Publication bias under aggregation frictions: from communication model to new correction method*. Working paper, MIT, mimeo.

Gechert, S. (2015). What fiscal policy is most effective? A meta-regression analysis. *Oxford Economic Papers*, *67*(3), 553-580.

Gechert, S., & Rannenberg, A. (2018). Which fiscal multipliers are regime-dependent? A Meta-regression analysis. *Journal of Economic Surveys*, *32*(4), 1160-1182.

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

George, E. I. (2010). Dilution priors: Compensating for model space redundancy. In *Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown* (pp. 158-165). Institute of Mathematical Statistics.

Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choices and selective reporting. *Journal of the European Economic Association*, *13*(6), 1180-1204.

Havranek, T., Horvath, R., & Zeynalov, A. (2016). Natural resources and economic growth: A meta-analysis. *World Development*, *88*, 134-151.

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

Ilzetzki, E., Mendoza, E. G., & Végh, C. A. (2013). How big (small?) are fiscal multipliers?. *Journal of monetary economics*, *60*(2), 239-254.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*(8), e124.

Ioannidis, J. P., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research.

Jeffreys, H. (1961): Theory of Probability. Oxford Classic Texts in the Physical Sciences. Oxford: Oxford University Press, third edition.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, *95*(1), 161-182.

Kočenda, E., & Iwasaki, I. (2021). Bank survival around the world: A meta-analytic review. *Journal of Economic Surveys.*

Madigan, D., York, J., & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, 215-232.

Miyamoto, W., Nguyen, T. L., & Sergeyev, D. (2018). Government spending multipliers under the zero lower bound: Evidence from Japan. *American Economic Journal: Macroeconomics*, *10*(3), 247-77.

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

Ramey, V. (2016). Macroeconomic Shocks and Their Propagation, volume 2. Elsevier.

Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of political economy*, *126*(2), 850-901.

Ramey, V. A. (2019). Ten years after the financial crisis: What have we learned from the renaissance in fiscal research?. *Journal of Economic Perspectives*, *33*(2), 89-114.

Riera-Crichton, D., Vegh, C. A., & Vuletin, G. (2015). Procyclical and countercyclical fiscal multipliers: Evidence from OECD countries. *Journal of International Money and Finance*, *52*, 15-31.

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., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. routledge.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of economic perspectives*, *15*(3), 131-150.

Zigraiova, D., Havranek, T., Irsova, Z., & Novak, J. (2021). How puzzling is the forward premium puzzle? A meta-analysis. *European Economic Review*, *134*, 103714.

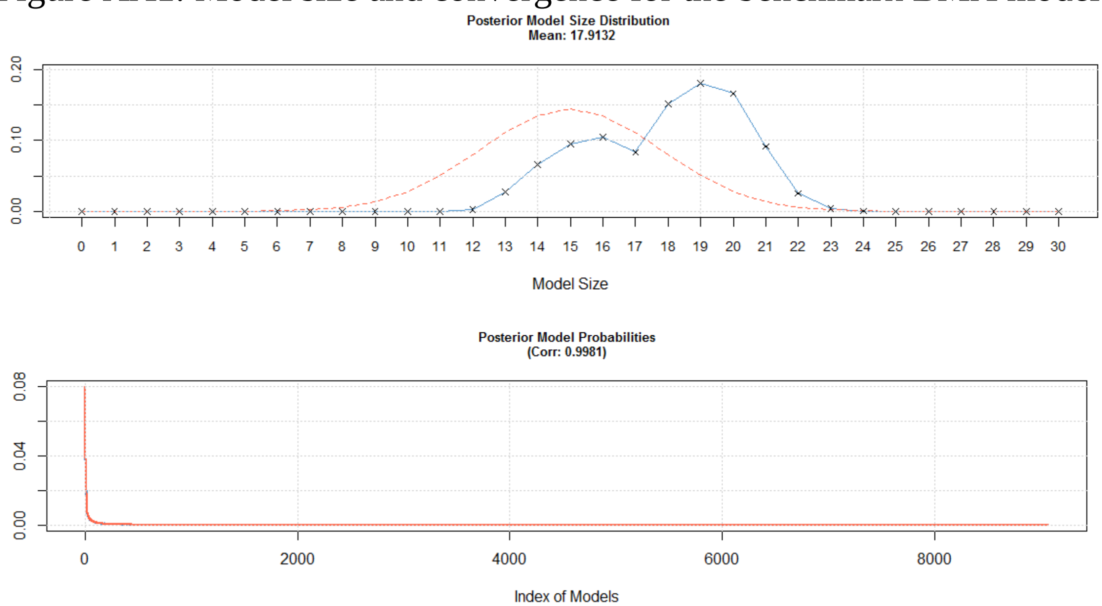
Appendix

Table AA1: Summary of the benchmark BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
17.9132	$3*10^6$	$1*10^6$	6.393208 mins	409823
Modelspace	*Models visited*	*Topmodels*	*Corr PMP*	*No. Obs.*
$1.1*10^9$	*0.038%*	*100*	0.9981	3279
Model prior	*g-prior*	*Shrinkage-stat*		
uniform / 15	UIP	*Av=0.9997*		

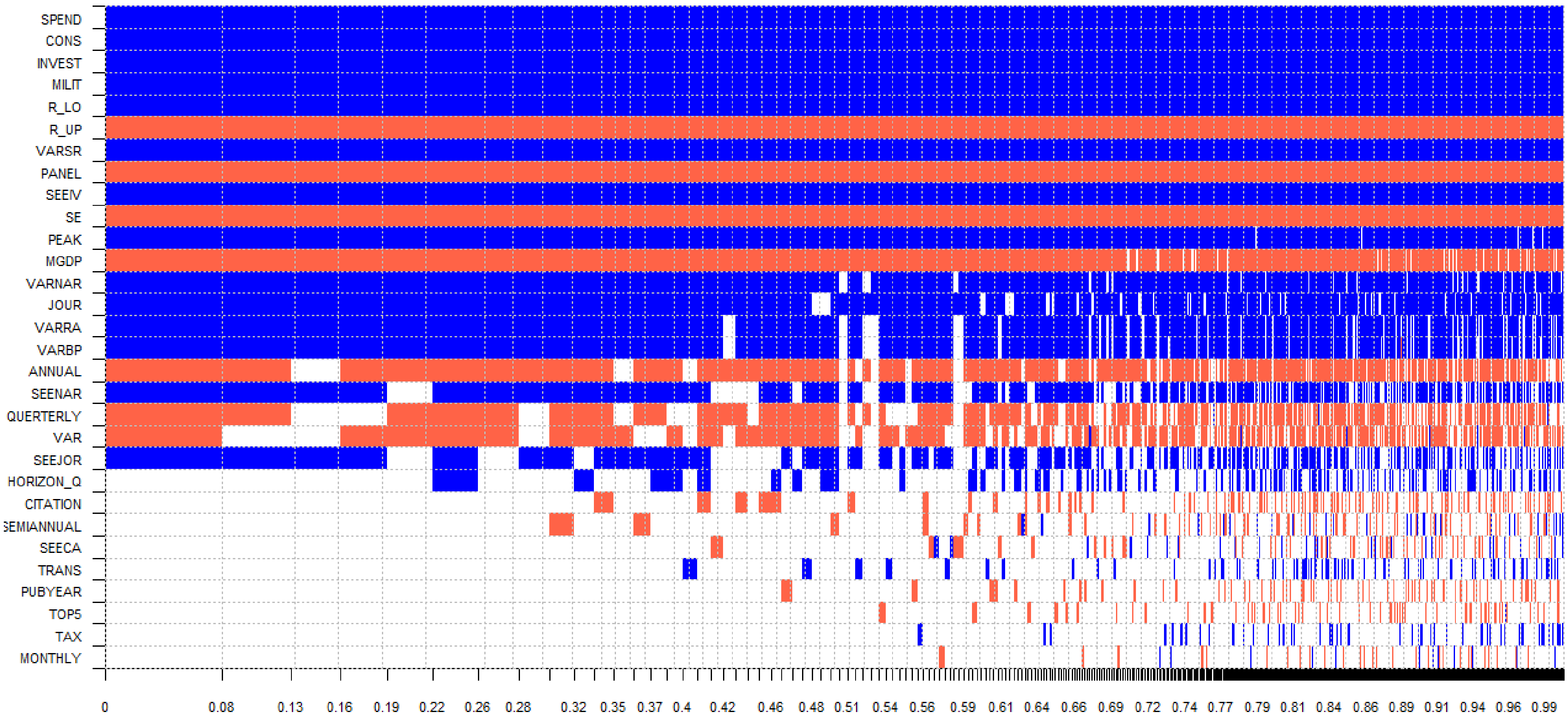
Notes: The corresponding results of this BMA specification are reported in Table 6. Considering Eicher et al. (2011), uniform distribution as model prior and Unit Information Prior (UIP) for g-prior were employed.

Figure AA1: Model size and convergence for the benchmark BMA model



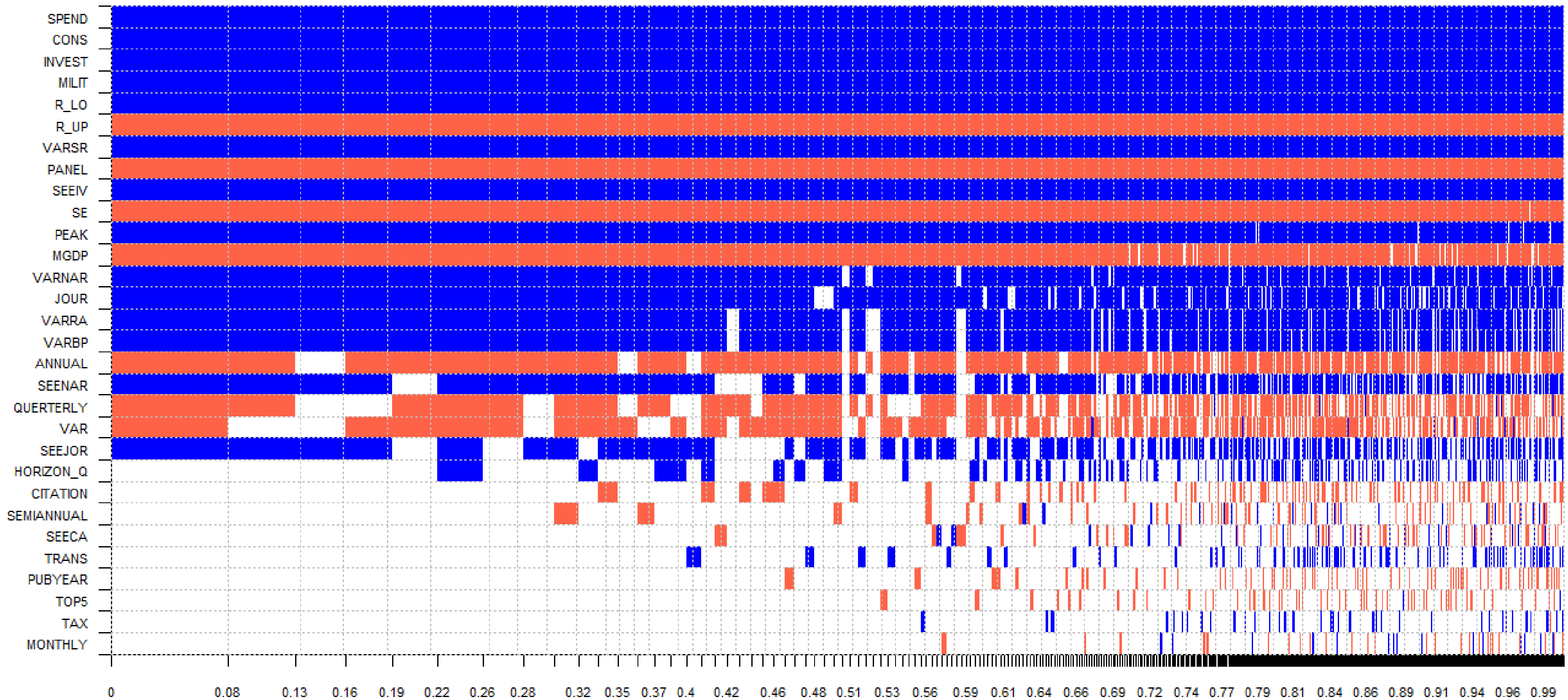
Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA exercise reported in Figure 5 and Table 6 part A.

Figure AA2: Model inclusion in BMA (g-prior – UIP; m-prior - Dilution)



Notes: Explanatory variables are ranked according to their posterior inclusion probabilities along the vertical axis from the highest at the top to the lowest at the bottom. The horizontal axis depicts the values of cumulative posterior model probability. Blue color indicates if the estimated parameter of a corresponding explanatory variable is positive. Red color indicates if the estimated parameter of a corresponding explanatory variable is negative. No color means the corresponding explanatory variable is not included in the model.

Figure AA3: Model inclusion in BMA (g-prior – BRIC; m-prior - Random)



Notes: Explanatory variables are ranked according to their posterior inclusion probabilities along the vertical axis from the highest at the top to the lowest at the bottom. The horizontal axis depicts the values of cumulative posterior model probability. Blue color indicates if the estimated parameter of a corresponding explanatory variable is positive. Red color indicates if the estimated parameter of a corresponding explanatory variable is negative. No color means the corresponding explanatory variable is not included in the model.

Table AA 2 Alternative BMA priors

	g-prior=UIP, mprior=Dilution			g-prior =BRIC, mprior=random		
	PIP	Post Mean	Post SD	PIP	Post Mean	Post SD
SE	0.999	-3.831	0.823	0.999	-3.831	0.822
VAR	0.671	-0.394	0.335	0.671	-0.394	0.334
PEAK	0.995	0.156	0.040	0.995	0.156	0.040
HORIZON	0.250	0.001	0.002	0.251	0.001	0.001
SPEND	1.000	0.596	0.043	1.000	0.596	0.043
CONS	1.000	0.790	0.049	1.000	0.789	0.049
INVEST	1.000	1.157	0.066	1.000	1.158	0.066
MILIT	1.000	0.809	0.083	1.000	0.809	0.083
TRANS	0.103	0.017	0.065	0.103	0.017	0.065
TAX	0.053	0.003	0.021	0.053	0.002	0.020
R_LO	1.000	0.426	0.041	1.000	0.426	0.041
R_UP	1.000	-0.484	0.040	1.000	-0.485	0 .040
SEEIV	0.999	0.578	0.184	0.999	0.578	0.184
SEENAR	0.781	0.349	0.230	0.781	0.348	0.230
SEEJOR	0.650	0.196	0.167	0.650	0.196	0.167
SEECA	0.104	-0.014	0.089	0.104	-0.014	0.089
VARNAR	0.960	0.770	0.317	0.960	0.769	0.317
VARBP	0.907	0.586	0.283	0.907	0.586	0.283
VARRA	0.915	0.649	0.299	0.915	0.649	0.299
VARSR	1.000	1.094	0.300	1.000	1.0944	0.301
PANEL	0.999	-0.364	0.071	0.999	-0.364	0.071
ANNUAL	0.841	-0.270	0.171	0.841	-0.270	0.171
SEMIANN	0.133	-0.025	0.104	0.133	-0.024	0.104
QUERTER	0.677	-0.188	0.160	0.677	-0.188	0.160
MONTH	0.041	-0.001	0.056	0.041	-0.001	0.056
JOUR	0.927	0.111	0.046	0.927	0.111	0.046
TOP5	0.065	-0.005	0.028	0.065	-0.004	0.077
MGDP	0.972	-0.003	0.001	0.972	-0.003	0.001
CITATION	0.152	-0.007	0.021	0.152	-0.007	0.021
PUBYEAR	0.077	-0.020	0.094	0.077	-0.020	0.094
Intercept	1.000	0.617	NA	1.000	0.617	NA

Notes: Dependent variable is fiscal multiplier, Post. mean - posterior mean, Post. SD - posterior standard deviation, PIP - posterior inclusion probability, SE - standard error. Part A, contains numerical results BMA based on the UIP g-prior and prior Dilution distribution for model. Part B contains numerical results BMA based on the BRIC g-prior and prior Random distribution for model.

IES Working Paper Series

2022

1. Klara Kantova: *Parental Involvement and Education Outcomes of Their Children*
2. Gabriel Nasser, Doile de Doyle, Paulo Rotella Junior, Luiz Célio Souza Rocha, Priscila França Gonzaga Carneiro, Rogério Santana Peruchi, Karel Janda, Giancarlo Aquila: *Impact of Regulatory Changes on Economic Feasibility of Distributed Generation Solar Units*
3. Paulo Rotella Junior, Luiz Célio Souza Rocha, Rogério Santana Peruchi, Giancarlo Aquila, Karel Janda, Edson de Oliveira Pamplona: *Robust Portfolio Optimization: A Stochastic Evaluation of Worst-Case Scenarios*
4. Adam Kučera, Evžen Kočenda, Aleš Maršál: *Yield Curve Dynamics and Fiscal Policy Shocks*
5. Karel Janda, Ladislav Kristoufek, Barbora Schererova, David Zilberman: *Price Transmission and Policies in Biofuels-Related Global Networks*
6. Daniel Kolář: *Wealth Survey Calibration: Imposing Consistency with Income Tax Data*
7. Michal Hlaváček, Ilgar Ismayilov: *Meta-analysis: Fiscal Multiplier*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz