JEL Classification: E62, C83.
Keywords: fiscal multiplier, meta-analysis, publication bias, model uncertainty.

Uncovering Publication Bias in Fiscal Multiplier Estimates

Michal HLAVÁČEK - Charles University, Prague

Ilgar ISMAYILOV - Charles University, Prague, Azerbaijan State Economic University (UNEC), Baku

Abstract

Our study provides an integrated overview of fiscal multiplier estimates, a key parameter assessing the economy's response to government interventions. Using a comprehensive database comprising 131 studies and over 3200 observations, we employ both linear and non-linear meta-analysis methods. Notably, our study marks the first application of Bayesian Model Averaging (BMA) in the context of meta-analysis for fiscal multipliers. Our results reveal a positive but moderate fiscal multiplier effect, ranging from 0.75 to 0.83 across different models. Significantly, our findings diverge from prior research by identifying a publication selection bias, largely attributed to our innovative use of BMA for heterogeneity investigation.

1. Introduction

Following the global financial crisis, the significance of fiscal policy heightened as interest rates hit the zero lower bound. Fiscal stimuli and subsequent budget deficit policies emerged as pivotal tools to prevent economic crises and stabilize declining economies. The fiscal multiplier, which measures the ratio of output change resulting from government actions to alterations in government spending or taxation, plays a key role in assessing the effectiveness of government interventions.

In our study, we contribute to the fiscal policy literature by assembling an extensive database consisting of 131 studies. Using this comprehensive dataset, we aim to provide an integrated overview of fiscal multiplier estimates. The primary innovation of our research lies in disentangling the presence of publication bias within estimates.

Fiscal policy literature is complex due to diverse approaches influenced by economic theories and modeling frameworks. This complexity is further compounded by the multifaceted nature of fiscal policy. Ioannidis (2005) argues that such complexity creates an environment susceptible to publication bias, where the pursuit of statistically significant results may distort findings to enhance publication likelihood. The recent increase in interest in fiscal policy research may further amplify the pressure to select statistically significant results. To address this, we employ meta-analysis, a method that utilizes rigorous quantitative survey techniques to unveil the existence of publication bias (Stanley, 2001; Stanley and Doucouliagos, 2012; Iwasaki and Tokunaga, 2016; Havranek et al., 2016a; Havranek et al., 2016b; Ioannidis

https://doi.org/10.32065/CJEF.2024.04.02

The authors reported no potential conflict of interest.
The authors wish to express their gratitude to the anonymous reviewers for their valuable feedback and recommendations.

et al., 2017; Zigraiova et al., 2021; Irsova et al., 2023).
Although several studies (Gechert, 2015; Gechert and Rannenberg, 2018; Ramey, 2019; Scandizzo and Pierleoni, 2020; Asatryan et al., 2020) have attempted to summarize the scientific research, a systematic analysis of fiscal multiplier estimates is still lacking in the literature. Recent advancements in meta-analysis, particularly regarding non-linear methods and the implementation of Bayesian Model Averaging (BMA), have not been adequately addressed in previous analyses of fiscal multipliers. Considering the existing gap in the literature, in this research, besides the traditional linear models, we also pioneer the application of advanced non-linear models, as well as BMA, marking the first implementation of these techniques within the context of fiscal multiplier estimates.
The findings of our research provide robust evidence supporting the existence of publication bias in the fiscal policy literature. This result contradicts previous studies (e.g., Gechert, 2015; Gechert and Rannenberg, 2018; and Asatryan et al., 2020), which either found no evidence or weak support for publication bias. Our study identifies that the discrepancy mainly stems from our use of BMA for heterogeneity investigation, a method not consistently applied in other studies. The methodology employed in this study surpasses its counterparts in handling model uncertainty and addressing omitted variable bias. Furthermore, our findings challenge the assumption that the number of observations is a reliable proxy for standard errors of fiscal multiplier estimates.

2. Data

The database utilized in this study constitutes an extension of the dataset originally developed and scrutinized by Gechert and Rannenberg (2018). Our study primarily focuses on empirical investigations analyzing fiscal policy issues using Vector Autoregression (VAR) and Single Equation Estimates (SEE). The initial dataset by Gechert and Rannenberg (2018) spanned the years 1992 to 2013, encompassing 98 studies and over 1900 observations. To enhance comprehensiveness, we extend the dataset until 2020, incorporating an additional 33 studies detailed in Table 1, thereby augmenting the dataset by more than 1300 observations. The combined dataset now encapsulates 131 papers and over 3200 observations. Additionally, we introduce several new control variables to augment the existing set.

Data collection employed Google Scholar due to its comprehensive coverage and powerful full-text search capabilities, distinguishing it as a superior resource in comparison to other databases. Google Scholar's unrestricted search capability is particularly advantageous as opposed to databases that limit searches to titles, keywords, and abstracts. We examine the first 500 studies yielded by Google Scholar. We scrutinized the abstracts of each study to identify papers potentially featuring empirical estimates of the fiscal multiplier. Subsequently, selected studies were subjected to detailed reading. Further refinement involved scrutinizing the reference lists of all selected studies to identify potentially crucial papers absent from the initial Google Scholar search. Figure A1 in the Appendix illustrates the PRISMA diagram, providing a detailed overview of the study inclusion process for this meta-analysis.

Table 1 New Studies Added to the Extended Database

Afonso and Leal (2018)	Koh (2017)
Alloza(2018)	Kuckuck and Westermann (2014)
Amendola et al. (2019)	Mencinger et al. (2017)
Auerbach and Gorodnichenko (2017)	Miyamoto et al. (2018)
Auerbach et al. (2018)	Mertens and Ravn (2010)
Ben Zeev and Pappa (2015)	Perotti (2014)
Boiciuc (2015)	Priftis and Zimic (2018)
Borg (2014)	Pyun and Rhee (2015)
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. (2013)
Dupor and Guerrero (2017)	Vlasov and Deryugina (2018)
Estevao and Samake (2013)	Yadav et al. (2012)
Forni and Gambetti(2016)	

Notes: The table lists new studies that have been added to the database originally compiled by Gechert and Rannenberg (2018).

The fiscal multiplier (μ) is conventionally defined as the ratio of the change in output (ΔY) to the change in government expenditure (ΔG), providing a metric for evaluating the efficacy of fiscal policy:

$$\mu = \frac{\Delta Y}{\Delta G} \tag{1}$$

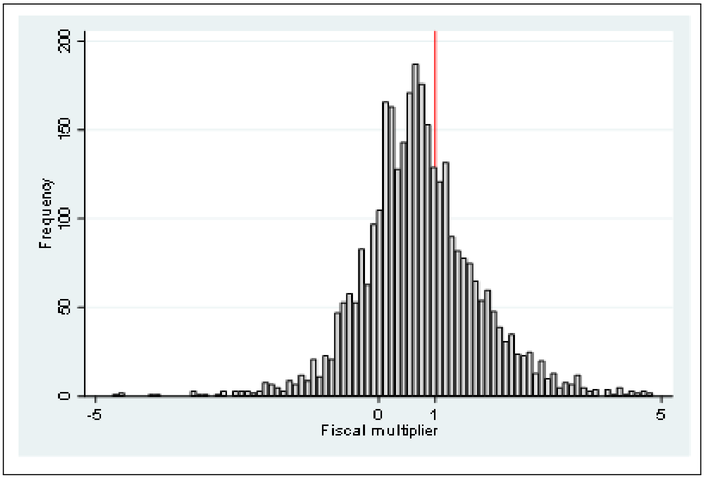
Empirical studies commonly calculate peak or cumulative multipliers as estimates of the fiscal multiplier:

$$\mu_{peak} = \frac{max_h \Delta Y_h}{\Delta G_1} \tag{2}$$

$$\mu_{cumulative} = \frac{\sum_t \Delta Y_t}{\sum_t \Delta G_t} \tag{3}$$

The peak multiplier captures the largest change for a given horizon *'h'* in the response variable concerning the government expenditure shock in period one. In contrast, the cumulative multiplier represents the ratio of the sum of all changes in output to the sum of all changes in government expenditure over a given period. Figure 1 depicts the distribution of collected observations, portraying an asymmetric, slightly right-skewed distribution of fiscal multipliers. Upon scrutinizing the data, a wide range of estimates emerged, ranging from -9.8 to 24.97 among the 3,279 collected estimates, with 2599 observations being positive. To solve the issue with the extreme values, we applied winsorization at 3.5%, the level at which our results stabilize. Post-winsorization, reported estimates range from -1.1 to 3.0, with a mean of 0.75 and a median of 0.68.

Figure 1 Distribution of Reported Estimates



Notes: Histogram illustrating the distribution of fiscal multipliers reported in individual studies. The vertical red line represents a multiplier value of 1.

Figure 2 illustrates the variation in multiplier estimates across different countries. Moreover, Figure A2 in the Appendix depicts the diversity of estimates within studies. Additionally, Table A1 in the Appendix provides summary statistics for various subsamples of the data, including a weighted mean that ensures equal weighting for all studies based on the number of estimates reported per study. Notably, both weighted and unweighted means of the fiscal multiplier are positive but less than one, indicating that, on average, economies expand less than the resources spent by governments in pursuit of fiscal policy objectives.

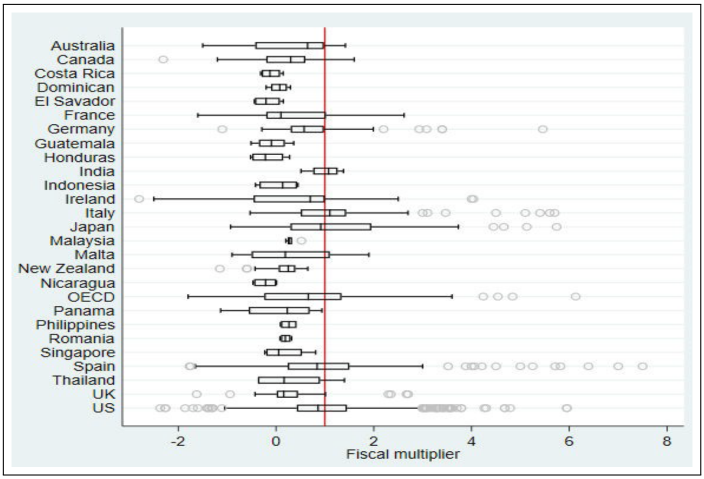
Further analysis reveals that the size of the fiscal multiplier is not uniformly stable across different subsamples. Data categorization based on models employed, data structure, identification methods, regimes, shock types, and frequency introduces significant variation. Approximately three-quarters of studies investigating fiscal policy topics utilize VAR models, while the remaining studies rely on SEE. Estimates obtained under both models exhibit remarkable similarity, as visualized in Figure 3(a).

Identification methods and shock types are central to debates regarding the accurate measurement of the fiscal multiplier, as evidenced by studies such as Forni and Gambetti (2015), Tang et al. (2013), and Broner et al. (2019). Figure 3(c) summarizes the role of identification methods in generating a fiscal multiplier in VAR models, while Figure 3(b) depicts the same relationship for SEE models.

Analyzing causal effects over a multi-year framework necessitates exogenous variation in policy variables. The issue of reverse causality between government expenditure and output, as highlighted by Ramey (2019), is addressed through multiple equation models utilizing five different methods. Figure 3(c) also reveals that narrative (VARNAR) and sign-restriction (VARSR) approaches tend to yield larger multipliers compared to other approaches. For SEE models, larger estimates are, on average, obtained using the instrumental variable method (SEEIV).

Another important factor to think about is the condition of the economy, which is considered a speculative element affecting the size of the multiplier. Figure 3(d) compares the fiscal multiplier under linear and multiple states models, illustrating the relationship between the multiplier's size and the states of the economy. This becomes particularly significant as contrasting results emerge in the literature; some argue for the essential role of states in estimating responses to government intervention (Auerbach and Gorodnichenko, 2012a, 2012b, 2017; Bachman and Sims, 2012), while others find either small or no impact in favor of non-linearity (Ramey and Zubairy, 2018; Afonso et al., 2018). The literature overview in Figure 3(d) aligns with the former group, indicating that multipliers under single-state models may be interpreted as the average of estimates obtained under different regimes of multiple-state models, with the multiplier being higher during economic downturns.

Figure 2 Cross-Country Heterogeneity



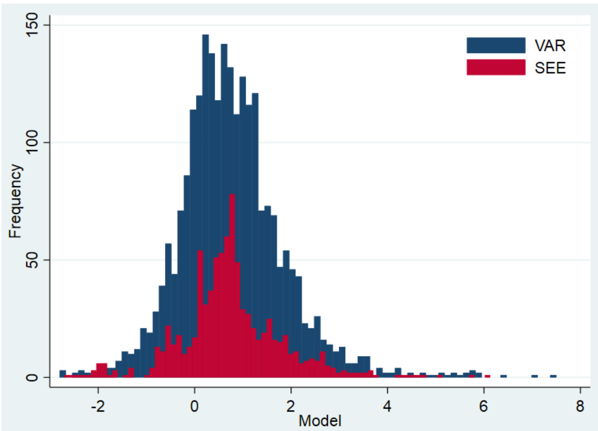
Notes: Each box represents the interquartile range (P25-P75), with the dividing line inside indicating the median value. Whiskers extend to the highest and lowest data points within 1.5 times the interquartile range. The vertical red line represents a multiplier value of 1. Outliers are excluded from the figure but included in all statistical tests.

3. Publication Bias

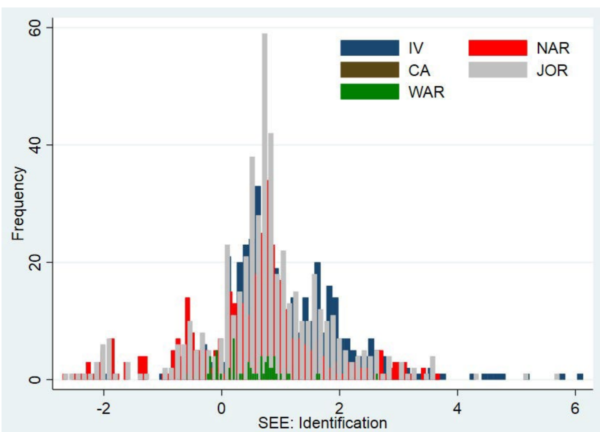
The size of the fiscal multiplier is a key focus in economic research. Particularly, when monetary policy interest rates hit the zero lower bound, fiscal policy became the sole tool for many nations to manage their economies and combat crises. This situation has heightened the significance of fiscal policy and increased interest in its effective implementation. In addition to the theoretical debates between New Keynesian and Neoclassical economic schools about the magnitude of the fiscal multiplier, a key question arises: How does the increased importance of fiscal policy, given the mentioned circumstances, affect research findings in this field? According to Stanley (2008), publication bias, characterized by a preference for statistically significant and theory-compliant results, can be identified and alleviated through meta-analyses. Publication bias can be understood in two ways: narrowly, it involves cases where studies with insignificant or counterintuitive results go unpublished or unreported; broadly, it encompasses all situations, including p-hacking. P-hacking introduces an additional layer of complexity, entailing the selective reporting or manipulation of statistical tests and significance levels to achieve statistically significant results. This practice may lead to the publication of findings aligning with the researcher's expectations, introducing a potential bias not immediately apparent in reported results (Irsova et al., 2023b). In our study, we adopt a broader definition of publication selection bias, encompassing all instances of conscious or unconscious manipulations to attain desired results.

Figure 3 Patterns in the Data

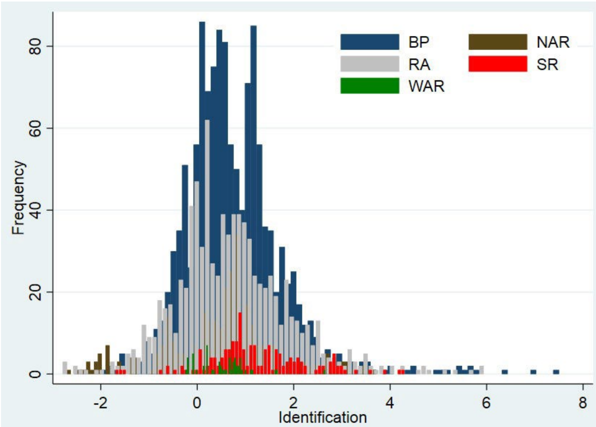
(a) Model types



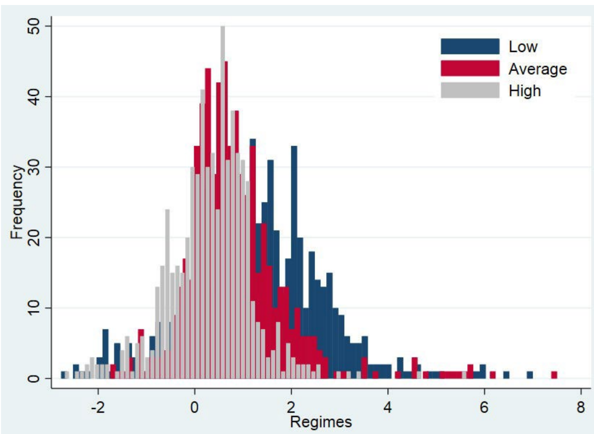
(b) SEE: Identification strategies



(c) VAR: Identification strategies



(d) Regimes



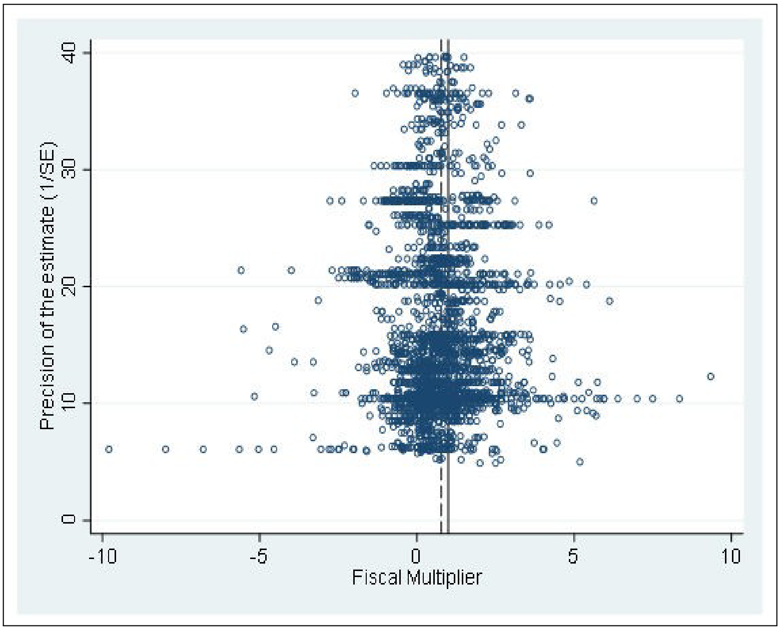
Notes: Figure 4(a): VAR – Vector Autoregression, SEE – Single Equation Estimates; Figure 4(b): IV – Instrumental Variable, CA – Cyclically Adjusted, WAR – War Episodes, NAR – Narrative Approach, JOR - Jorda Method; Figure 4(c): BP - Blanchard and Perotti, RA - Recursive Approach, WAR – War Episodes, NAR – Narrative Approach, SR - Sign-Restriction Method; Figure 4(d): Low Regime – Recession, High Regime-Booms, Average Regime - Linear or Non-Specified Regime.

The standard idea in the meta-analysis is that, in the absence of publication selection bias in the literature, the precision (reciprocal of standard errors) should not contribute to any variance in fiscal multiplier estimates. Despite the various methods available in meta-analysis literature, the underlying rationale remains consistent. However, a challenge inherent in studies exploring fiscal policy issues is the prevalent reporting of impulse-response functions. Moreover, numerous papers lack the comprehensive information necessary to compute comparable standard errors; for example, they lack the level of confidence bounds, or they present uncentered confidence bounds (Gechert and Rannenberg, 2018).

Havranek (2015) suggests that the number of observations can serve as either a proxy or an instrumental variable for standard errors in any study.

Similarly, Stanley and Doucouliagos (2012) contend that the number of observations can function as a second-best proxy in cases where standard errors are unavailable. Given the prevalent issue of poor reporting of standard errors, we employ the number of observations in our analysis and estimations.

Figure 4 Funnel Plot



Notes: In the absence of publication bias, the funnel should be symmetrical around the most precise estimates. A solid vertical line indicates the case when the multiplier is equal to one and the dashed line is the (unweighted) sample average. The most precise estimates (those with the smallest standard errors) should cluster closely around the true mean value. As precision decreases, estimates scatter more widely around the true mean value.

We initiate the publication bias investigation with a funnel plot (Figure 4), where a solid vertical line represents a multiplier equal to one, and the dashed line denotes the sample (unweighted) average. A lack of publication bias should result in estimates forming a symmetric funnel around the most precise estimations of the true value. The funnel plot depicted in Figure 4 displays an approximately symmetric pattern.

We then proceed to formally test the bias using the funnel-asymmetry test (FAT). To conduct FAT, it is necessary to have reported multipliers and their corresponding standard errors. However, due to missing standard error information in many studies, the number of observations serves as a proxy for standard errors in our estimations. The model takes the following form:

$$\mu_{ij} = \alpha + \gamma\left(\frac{1}{\sqrt{n}}\right)_{ij} + \varepsilon_{ij} \tag{4}$$

where, μ_{ij} represents the i-th multiplier reported in research j, n is the number of observations corresponding to the multiplier i from paper j, and ε_{ij} is the error term.

If the estimation results indicate that the coefficient, γ, significantly differs from zero, it would document the asymmetry in the funnel plot, revealing the presence of selection publication bias in the estimations. In any case, the intercept, α, represents the true effect corrected for potential publication bias.

Meta-analyses frequently rely on empirical findings from overlapping samples, where aggregated observational data is often shared across multiple studies. Overlapping samples occur when the same dataset is used in different investigations or when multiple estimates originate from a single study, as is the case in our dataset. Neglecting to properly address overlapping samples can result in inflated false positives, particularly in scenarios with large meta-sample sizes (Bom and Rachinger, 2020; Irsova et al., 2023a). Various strategies have been proposed to alleviate the issue of estimate dependence induced by sample overlap in meta-studies. These include incorporating one estimate per study or averaging estimates (Van der Sluis et al., 2005; Stanley, 1998), employing a generalized-weights meta-estimator that explicitly models the variance-covariance matrix (Bom and Rachinger, 2020), and integrating fixed-study effects into the meta-analysis model (Stanley and Doucouliagos, 2012). To efficiently tackle this problem, we adopt a two-way clustering at the level of studies and countries, the approach inspired by Havranek and Irsova (2017).

To ensure the robustness of our findings, we apply three different estimation techniques, covering various linear and non-linear models detailed in Table 2. The results across nearly all models in this phase consistently support the conclusion regarding the magnitude and significance of both the intercept, α, and the coefficient on standard errors, γ. Nearly all models yield consistent results, showing a uniform effect size beyond bias ranging from 0.75 to 0.83 and indicating an absence of publication bias in the fiscal multiplier literature.

Table 2 depicts the results of all three techniques, whereas Panel A contains the results of linear models. Pooled OLS indicates an actual effect of less than one with a non-significant coefficient on standard errors. The second column contains estimates derived from the panel fixed-effect model, controlling for unobserved heterogeneity across studies through study-level fixed effects. The third column presents between estimators, capturing variations between studies in panel data. The last column in Panel A introduces the hierarchical Bayes model, a multi-level estimation applying weights through partial pooling at the study level and utilizing within-study variations. All four linear models share similar results.

Part B of Table 2, presents results using two weighting schemes. In the first scheme (inverse of) the number of observations per study uniformly weights all studies, irrespective of the number of estimates. The second scheme employs the inverse of (proxy of) standard errors, precision, as the weighting criterion. While the weighted models generally mirror previous estimates, the model with study weights reveals a genuine effect of the multiplier corrected for bias (α) slightly larger than one. Despite conflicting results on the intercept, under both weighting schemes, the coefficient on standard errors does not reach statistical significance.

Table 2 Results of Funnel Asymmetry Test

PANEL A	*OLS*	*FE*	*BE*	*Hierarchical*	
γ	-0.53	0.01	1.56	0.97	
	(1.47)	(1.55)	(1.45)	(1.76)	
	[-3.79, 2.82]	-	-	-	
α	0.79***	0.75***	0.99***	0.82***	
	(0.15)	(0.11)	(0.12)	(0.13)	
	[0.40, 1.16]	-	-	-	
PANEL B		*Study - weighted*		*Precision - weighted*	
γ		0.51		-0.65	
		(2.2)		(1.88)	
		[-7.55, 6.76]		[-5.48, 3.93]	
α		1.03***		0.79***	
		(0.18)		(0.13)	
		[0.49 1.61]		[0.39, 1.06]	
PANEL C	*Ioannidis et al.(2017)*	*Bom and Rachinger (2019)*	*Furukawa (2021)*	*Andrews and Kasy (2019)*	*van Aert and van Assen (2023)*
α	0.75***	0.78***	0.83	0.80***	0.78***
	(0.15)	(0.04)	(0.65)	(0.03)	(0.37)

Notes: The table contains the results of the regression provided in Equation (4). Standard errors are reported in parentheses. OLS refers to ordinary least squares, FE denotes study fixed effects, and BE indicates between effects. In square brackets, 95% confidence intervals from wild bootstrap clustering are reported; the implementation follows Roodman et al. (2019), using Rademacher weights with 9999 replications. In the regression equation, α represents the constant, while γ represents the coefficient on standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, with standard errors in parentheses. The number of observations in all regressions presented in Table 2 is 3279.

Traditional methods are effective when the relationship between estimates and standard errors is linear. However, to account for more complex, non-linear connections, we employ recently developed techniques. The results of these non-linear models are summarized in Table 2, Part C. Non-linear methods for bias correction fall into two groups. The first group includes selection models (van Aert and van Assen, 2023; Andrews and Kasy, 2019), which assume that the likelihood of a study being published depends on the significance of its findings. Essentially, more significant results are given higher weight through re-weighting estimates using the inverse publication probability. The second group relies on the funnel plot method (Egger et al., 1997; Stanley, 2008; Stanley and Doucouliagos, 2012; Stanley and Doucouliagos, 2014; Ioannidis et al., 2017. Bom and Rachinger, 2019; Furukawa, 2021). In contrast to selection models, these techniques assume that selective reporting is based on the size of the reported result, not the p-value. Funnel-based methods calculate the impact size that would be reported in a perfect, highly precise study (Irsova et al., 2023a). To enhance the robustness of our results and align with theoretical considerations, we incorporate models from both categories. This approach is in line with recent meta-analysis studies (e.g., Bajzik et al., 2020; Kocenda and Iwasaki, 2022; Zigraiova et al., 2021; Gechert et al., 2022; Havranek et al., 2024).

The first approach we use is the weighted average of the adequately powered (WAAP) method 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). If the standard error is less than this threshold, it means that the estimate is adequately powered to detect the unknown actual effect.

Bom and Rachinger (2019) model introduces non-linearity due to publication selection, assuming selection occurs only when the standard error exceeds a threshold. The model uses a piecewise linear meta-regression with a kink at the standard error cutoff, determined endogenously based on an initial estimate of the true effect and a significance threshold. Above this threshold, the model is linear in standard error. Below it, estimates are considered sufficiently significant, and increased standard errors are not linked to publication selection.

Furukawa (2021) model addresses bias problems by prioritizing precise estimates, as higher precision is associated with reduced bias severity. While this focus on precision may result in larger estimate variances, the stem-based method ensures coherence between assumed and implied variances. The model achieves coherence by internally choosing the optimal ratio of the most precise studies, effectively balancing the trade-off between bias and efficiency.

The model developed by Andrews and Kasy (2019) falls under the category of selection models. Selection models typically employ a step function to introduce non-linearity in the publication probability for each research. The distinction among selection models lies in the weight functions used to capture dissimilarity across different groups. Andrews and Kasy (2019) model assumes a significant change in publication probability beyond conventional t-statistics thresholds. Unlike linear dependencies, this model estimates the true effect under a multiple regime model, where standard p-value cut-offs (1%, 5%, 10%) are linked to jumps in the distribution of reported estimates.

The final model in Table 2, labeled p-uniform*, was developed by van Aert and van Assen (2023). It also falls within the selection models category. This model operates on the statistical principle that p-values should follow a uniform distribution at the true effect size, thereby predicting the true effect based on this principle. The technique is robust to heterogeneity and also to the endogeneity of the standard errors.

Summarizing the findings of non-linear models, all five models indicate a corrected effect below one, ranging from 0.75 to 0.83. With the exception of the Furukawa (2021) model, all estimates are statistically significant, and their results are detailed in Part C of Table 2.

Comparing our FAT results with the two most recent studies in this field, Asatryan et al. (2020) and Gechert and Rannenberg (2018), both our study and Asatryan et al. (2020) do not find a statistically significant coefficient, whereas only Gechert and Rannenberg (2018) shows a significant coefficient indicating publication bias. It is important to note that all three studies use a similar dataset; however, Gechert and Rannenberg (2018) results are primarily influenced by nonclustered errors, as described in Appendix Table A2.

4. Heterogeneity

In the previous section, we primarily analyze the variation in estimates through standard errors (inverse of the number of observations). However, beyond standard errors, critical distinctions in data, methods, models, and other factors across studies can significantly contribute to the variability of estimates. In this section, we investigate the sources of heterogeneity beyond standard errors. The variables in our dataset can be categorized into nine groups, each representing pivotal aspects of the dataset. A summary of these variables is provided in Table 3.

We explore heterogeneity issues by using Bayesian techniques. Bayesian techniques offer a robust way to handle issues of model uncertainty and identify a parsimonious model from a list of potential variables. To conduct our meta-analysis, we expand upon the database developed by Gechert and Rannenberg (2018), incorporating new variables that account for (new) identification strategies, data types, and specific publication characteristics. Our comprehensive dataset comprises 39 variables, conditionally grouped into nine categories: Model, Dataset Type, Identification, Regime, Frequency, Impulse Response Type, Shock Type, Publication Characteristics, and Others.

To address variations across studies, it is important to consider diverse key characteristics in study design. One notable difference lies in the models employed, which can either be SEE or VAR. The shock type, regime, and identification subgroups within the dataset play a crucial role in understanding the variation in fiscal multiplier estimates, providing insights into prominent discussions in the literature (e.g., Ilzetzki et al., 2013; Caggiano et al., 2015; Ramey and Zubairy, 2018; Fatas and Mihov, 2001; Batini et al., 2014).

The issue of endogeneity, arising from the reverse causality between GDP and government expenses, is prevalent in economic models studying fiscal multipliers. Consequently, shock types and identification are crucial tools to address endogeneity. Various methods have been developed for both SEE and VAR models. The summary of the 131 papers provides an overview, indicating that SEE models rely on one of the five following techniques: SEEWAR relies on war episodes and subsequent defense spending increases as exogenous shocks; SEENAR (narrative method) incorporates both war episodes and exogenous tax changes; SEEIV employs instrumental variables for government expenses; SEECA relies on event studies using cyclically adjusted time series, and SEEJOR relies on the local projection method by Jorda (2005). The local projections method has gained popularity in the last several years (Auerbach and Gorodnichenko, 2012a, 2012b; Ramey and Zubairy, 2018; Riera-Crichton et al., 2015; Miyamoto et al., 2018; Broner et al., 2019) due to its simplicity, non-parametric calculation of impulse response functions, robustness to misspecifications, and flexibility in capturing non-linear relations (Ramey, 2016).

The VAR models, on the other hand, employ five identification strategies to mitigate endogeneity: VARWAR and VARNAR consider war episodes and the narrative method, similar to the SEE case; VARRA is the recursive approach, ordering variables in a Cholesky decomposition to eliminate contemporaneous impact of GDP on Government; VARBP, based on the classic paper by Blanchard and Perotti (2002), imposes elasticity for automatic stabilizers; and VARSR imposes sign restrictions on impulse responses while calculating them.

Table 3 Description and Summary Statistics of Variables

Variable	*Description*	*Mean*	*STD. Dev.*	*Weighted mean*
Multiplier	Fiscal multiplier, indicator measuring the effectiveness of fiscal policy, response variable	0.75	0.96	0.87
Standard error (SE)	Standard error of the fiscal multiplier	0.07	0.03	0.08
Model				
SEE	= 1 if single equation models	0.26	0.44	0.32
VAR	= 1 if Vector Autoregression models	0.74	0.44	0.68
Data				
Panel	= 1 if the dataset type is panel	0.39	0.49	0.29
Time series	= 1 if the dataset type is time series	0.61	0.49	0.71
Identification				
VARNAR	= 1 if VAR model is developed based on a narrative action-based approach	0.03	0.18	0.06
VARBP	= 1 if VAR model is developed based on the Blanchard – Perotti approach	0.37	0.48	0.30
VARRA	= 1 if VAR model is developed on a recursive approach	0.27	0.45	0.23
VARSR	= 1 if VAR model is developed based on a sign-restriction approach	0.05	0.22	0.07
VARWAR	= 1 if VAR model is developed based on a war episode approach	0.01	0.09	0.03
SEEIV	= 1 if SEE with the instrumental variable approach	0.06	0.25	0.14
SEENAR	= 1 if SEE with a narrative action-based approach	0.03	0.17	0.05
SEECA	= 1 if SEE with prior cyclical adjustment of the public budget	0.01	0.12	0.04
SEEWAR	= 1 if SEE with war episode-based approach	0.01	0.09	0.04
SEEJOR	= 1 if Jorda method is used to calculate IRFs	0.16	0.36	0.07
Regime				
RAV	= 1 if average or unspecified regime	0.62	0.49	0.80
RLO	= 1 if downturn or crises regime	0.19	0.39	0.10
RUP	= 1 if recovery or expansion regime	0.19	0.39	0.10
Frequency				
Annual	=1 if data frequency is annual	0.22	0.42	0.27
Semiannual	=1 if data frequency is semi-annual	0.02	0.15	0.03
Biannual	=1 if data frequency is biannual	0.02	0.14	0.02
Quarterly	=1 if data frequency is quarterly	0.73	0.44	0.67
Monthly	=1 if data frequency is monthly	0.00	0.06	0.02
IRF Type				
Cumulative	=1 if calculated as a cumulative multiplier	0.77	0.42	0.73
Peak	=1 if calculated as a peak multiplier	0.21	0.41	0.25

Variable	Description	Mean	STD. Dev.	Weighted mean
Publication				
TOP5 journal	= 1 if the estimate was published in the top five journal	0.07	0.25	0.08
Journal	= 1 if the estimate is from a published study	0.36	0.48	0.42
Working paper	= 1 if the estimate is from a non-published study	0.57	0.49	0.49
Citations	The logarithm of yearly citations according to Google Scholar	4.37	1.50	4.73
Published year	The logarithm of publication year	03.6	0.19	2.92
Shock				
Spending	= 1 if public spending is unspecified	0.41	0.49	0.44
Consumption	= 1 if public consumption	0.20	0.40	0.11
Investment	= 1 if public investment	0.07	0.25	0.06
Military	= 1 if public military spending	0.07	0.25	0.11
Transfers	= 1 if transfer to the private sector	0.02	0.15	0.02
Tax	= 1 if tax reliefs to the private sector	0.17	0.37	0.17
Deficit	= 1 if unspecified tax relief or spending increase	0.05	0.22	0.06
other factors				
Horizon	The horizon of multiplier calculation	8.42	09.11	8.32
MGDP	Import-to-GDP ratio of surveyed country sample	29.36	21.19	22.24

Notes: The table provides the summary of estimates for different subsets of the data. Weighted means are calculated using weights based on the inverse of the number of estimates reported per study.

Apart from empirical strategies, shock type emerges as another crucial factor influencing the size of the fiscal multiplier. Governments generally utilize six different channels to finance activities. Another key group of data characteristics relates to economic regimes. Estimates obtained under regime-dependent models correspond to periods of economic upturns (RUP) or downturns (RLO). The remaining estimates, derived from either linear models or unspecified regimes, fall under the average regime (RAV).

In addition to study design factors, this study also considers publication-quality characteristics of the research that publishes these estimates. We introduce study-level variables to control for study quality that might not be captured by observable differences in study design, methods, and other similar factors. Our dataset includes five publication characteristic variables, three of which are new factors not covered in Gechert and Rannenberg (2018): TOP5[2], a dummy variable indicating if the research was published in the top 5 economic journals; the logarithm of the number of citations per published year; and the logarithm of the publication year.

Other variables to be included in the multivariable meta-regression analysis to control for differences belong to the following subgroups: frequency of data, type of

[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 Economic Studies.

impulse response function, type of data set, and others (horizon of impulse responses and import to GDP ratio[3]).

4.1 Estimation

We run the model with the extra variables outlined in the preceding section:

$$\mu_{ij} = \alpha + \gamma \left(\frac{1}{\sqrt{n}}\right)_{ij} + \sum\nolimits_{m} \beta_m X_{m,ij} + \varepsilon_{ij} \tag{5}$$

The distinction between equation (4) and equation (5) lies in the additional term $\sum_m \beta_m X_{m,ij}$, where $X_{m,ij}$ represents the m-th control variable corresponding to the i-th observation from research j, and β_m is the corresponding coefficient. The rest of the parameters remain as described in equation (4). The intercept, α, despite being adjusted for publication bias, should not be construed as the true multiplier, as its interpretation depends on the reference specification. In alignment with recent literature (see Elminejad et al., 2023; Kocenda and Iwasaki, 2022; Gechert et al., 2022; Havranek et al., 2024), we employ BMA as a practical approach to address model uncertainty. BMA helps manage model uncertainty by preventing overfitting. The method includes only crucial variables and alleviates the impact of potential misspecifications. BMA evaluates various models by calculating posterior model probabilities (PMP), similar to the information criterion in frequentist econometrics, and assigns weights based on PMP to construct a weighted average for each coefficient across models. The weighted sum of PMPs yields the posterior inclusion probability (PIP), which determines the variables included in the multivariable model. For a more detailed explanation, see Raftery et al. (1997) and Eicher et al. (2011), among others. The algorithm involves estimating 2^k regressions initially, where k is the number of control variables. After excluding nine variables to avoid a dummy variable trap, as we have mutually exclusive and collectively exhaustive 9 subgroups of dummy variables, the task of estimating 2^{30} regressions in our study becomes impractical. To achieve feasibility, we employ a Markov Chain Monte Carlo algorithm (Madigan and York, 1995), which prioritizes models with a higher PMP. In the second step, a 'frequentist check' tests the robustness of findings from the first step. It selects only variables with a PIP above 85% (according to BMA results) and conducts multivariable regression with two-way clustering at the level of studies and countries.

4.2 Results

The result presented in Figure 5 summarizes the outcomes of BMA. Each column corresponds to an individual regression model, and the column width indicates the PMP. Columns are arranged from left to right based on descending PMP values. Additionally, each row represents a variable, ordered from top to

---

[3] The data for the share of imports on GDP were downloaded from the World Bank World Development Indicators, the indicator code: NE.IMP.GNFS.ZS; https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS

bottom according to descending PIP values. Blue cells indicate a positive posterior mean of the coefficient, red cells indicate a negative value and blank cells imply that the variable was not included in the model. Examination of Figure 5 reveals that 18 variables significantly contribute to the variation in multiplier estimates. In our baseline model, we choose a uniform distribution as the model prior and the Unit Information Prior (UIP) for coefficients (g-prior). We also test alternative priors for models and coefficients, comparisons available in Figure 6. Results with alternative distributions can be found in Appendix, Figures A4 and A5.

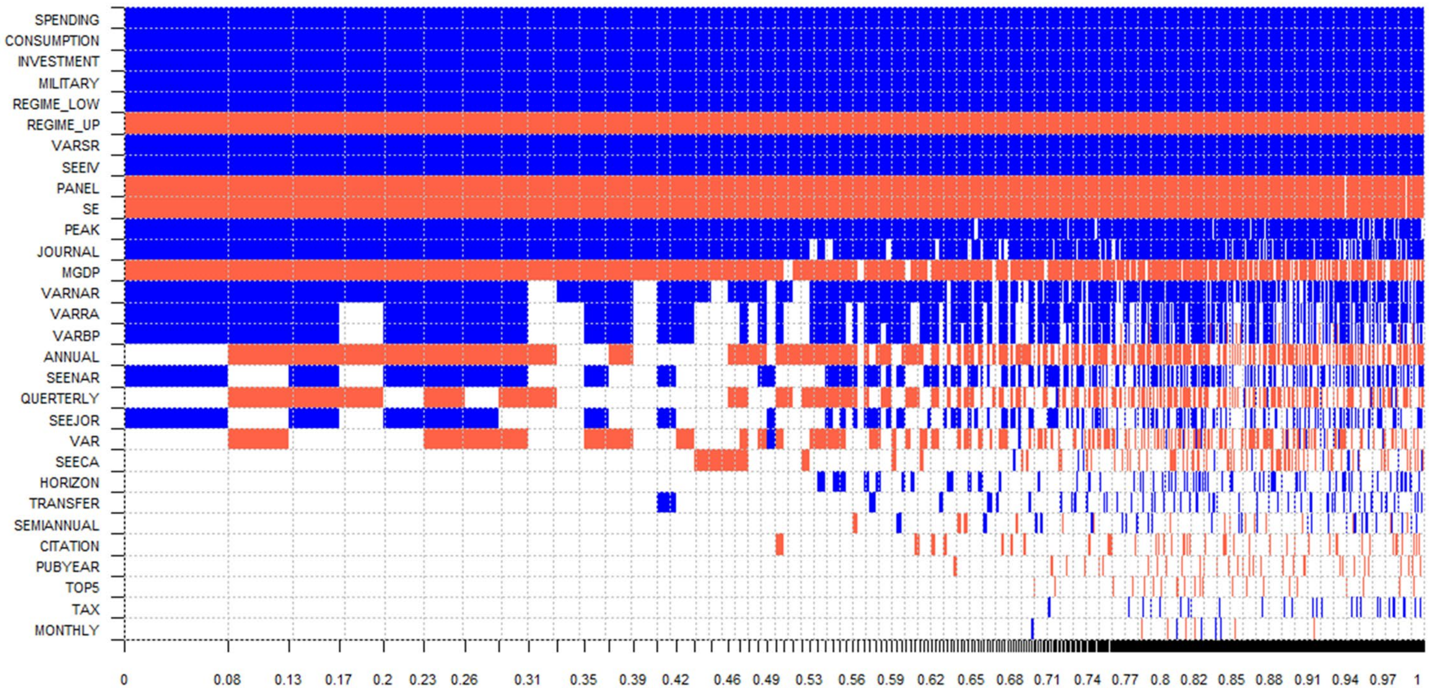
Upon closer inspection of Figure 5, more than half of the variables included in BMA prove to be significant in explaining the heterogeneity of the fiscal multiplier, with consistent sign impacts across models. Table 4 Part A provides a numerical representation of Figure 6, presenting values of PMP, posterior mean, and posterior standard deviation. Following Jeffreys (1961), variables are classified into four groups based on PIP: 'decisive' (PIP $>$ 0.99), 'strong' (0.9 $<$ PIP $\leq$ 0.99), 'substantial' (0.75 $<$ PIP $\leq$ 0.95), and 'weak' (0.50 $<$ PIP $\leq$ 0.75). Table 4 indicates that 18 variables significantly influence the fiscal multiplier estimates, with 11 being decisive, 3 strong, and 4 weak.

Table 4 Part B presents the results of the 'frequentist check,' a hybrid model of BMA and the frequentist approach, supporting the findings of BMA. Additionally, Table A6 in the Appendix includes results of Frequentist Model Averaging (FMA) using Mallow's criteria as weights, aligning with BMA outcomes. Contrary to the FATs discussed in the previous chapter, all three approaches highlight the presence of publication bias in fiscal policy literature. The discrepancy, where FATs show a statistically insignificant coefficient, can be attributed to omitted variable bias in the single-variable regression. According to Irsova et al. (2023a), the multivariable model holds an advantage over FATs that rely on a single variable, as it explicitly addresses omitted variable bias in both observational primary studies and multivariable regression analysis.

The initial assessment of BMA suggests that three categories of variables — shock types, regimes and identification strategies — play a crucial role in determining the fiscal multiplier's size. These variables align with the general theory on fiscal policy issues and represent attempts to address endogeneity problems. The meta-analysis also supports the idea that the fiscal multiplier is generally less than one, with investment activities being the most effective tool for stimulating the economy.

When comparing the findings of our study with two recent papers on the same topic, Gechert and Rannenberg (2018) and Asatryan et al. (2020), notable differences emerge. Asatryan et al. (2020) do not conduct heterogeneity analysis due to differences in their research design. On the other hand, Gechert and Rannenberg (2018) utilize subset analysis for their heterogeneity analysis. The disparity in results can be attributed to our use of heterogeneity analysis based on BMA, which is a robust tool compared to subset analysis. As discussed earlier, subset analysis similar to FATs is susceptible to omitted variable bias (Irsova et al., 2023a).

Figure 5 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. The blue color indicates if the estimated parameter of a corresponding explanatory variable is positive. The 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 Table 4. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals. The definition of all variables is available in Table 3.

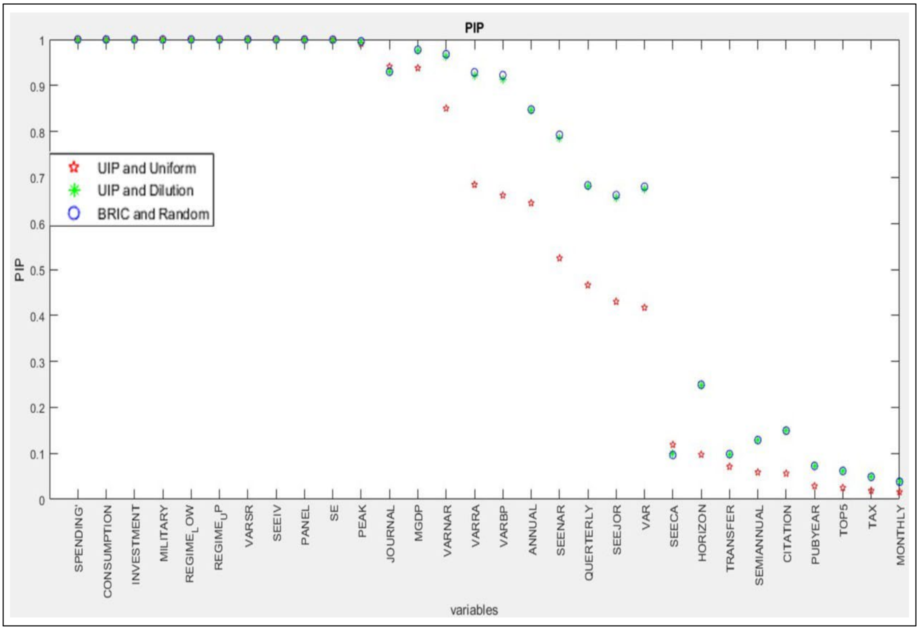
Shock types. Shock types play a critical role in explaining the variations observed across fiscal multipliers. Our analysis reveals that most shock types are considered 'decisive' based on their PIP, except for tax reliefs and transfers to the private sector. Notably, all significant variables possess positive coefficients, indicating their larger impact relative to the baseline case. However, it's important to note that the magnitudes of these coefficients vary, providing policymakers with valuable insights into the effectiveness of different fiscal tools. Specifically, investments emerge as the most influential factor, supported by findings from studies such as Leeper, Walker and Yang (2010), Auerbach and Gorodnichenko (2012a), Mourougane et al. (2016), Scandizzo and Pierleoni (2020). This result suggests that investing in productive assets can enhance the multiplier effect in the economy. In terms of impact, consumption ranks as the second most effective factor, followed by government spending, and then military expenditure.
Identification. Identification strategies are crucial in fiscal policy estimation due to the challenge of reverse causality between government expenditure and output (Ramey, 2019). Multiple equation models employ five distinct methods to address this issue, while SEE models utilize their own set of five methods. The variations in estimated sizes based on these strategies have been interpreted either as differences in their effectiveness in mitigating attenuation bias or in capturing exogenous variation in policy variables (Ramey and Zubairy, 2018; Ben Zeev and Pappa, 2017). In SEE models, the instrumental variable (SEEIV) method emerges as significant, whereas the Jorda method, despite recent popularity, ranks lower in importance. In VAR models, all identification strategies are important but to varying degrees, with the sign restriction method being decisive. Notably, both sign restriction and instrumental variable methods contribute similarly to the coefficient, adding approximately 0.47-0.43 to the baseline specification, while the narrative approach in VAR models (VARNAR) contributes only 0.23.

Table 4 Results of Bayesian Model Averaging and Frequentist Check

	PART A: Bayesian Model Averaging			*PART B: Frequentist check*		
variables	***PIP***	***Post. mean***	***Post. SD***	***Coefficient***	***S.E.***	***p-value***
SE	1.00	-3.98	0.83	-4.23	1.45	0.00
VAR	0.42	-0.25	0.35			
PEAK	0.99	0.15	0.04	0.16	0.11	0.17
HORIZON	0.10	0.00	0.001			
SPENDING	1.00	0.59	0.04	0.58	0.10	0.00
CONSUMPTION	1.00	0.78	0.05	0.77	0.11	0.00
INVESTMENT	1.00	1.15	0.07	1.13	0.27	0.00
MILITARY	1.00	0.77	0.09	0.69	0.26	0.01
TRANSFER	0.07	0.01	0.06			
TAX	0.02	0.001	0.01			
REGIME LOW	1.00	0.43	0.04	0.44	0.07	0.00
REGIME UP	1.00	-0.48	0.04	-0.47	0.07	0.00
SEEIV	1.00	0.54	0.20	0.43	0.09	0.00
SEENAR	0.52	0.24	0.26			
SEEJOR	0.43	0.14	0.18			
SEECA	0.12	-0.03	0.11			
VARNAR	0.85	0.56	0.38	0.24	0.11	0.02
VARBP	0.66	0.39	0.34			
VARRA	0.68	0.44	0.37			
VARSR	1.00	0.89	0.36	0.47	0.15	0.00
PANEL	1.00	-0.37	0.07	-0.38	0.15	0.01
ANNUAL	0.64	-0.18	0.17			
SEMIANNUAL	0.06	-0.00	0.06			
QUARTERLY	0.47	-0.12	0.15			
MONTHLY	0.02	0.001	0.04			
JOURNAL	0.94	0.12	0.05	0.15	0.07	0.03
TOP5	0.03	-0.002	0.02			
MGDP	0.94	-0.003	0.001	-0.003	0.002	0.13
CITATION	0.06	-0.01	0.01			
PUBLICATION YEAR	0.03	-0.01	0.06			
INTERCEPT	1.00	0.60	NA	0.62	0.18	0.00

Notes: The dependent variable is the fiscal multiplier. ‘Post. mean’ refers to the posterior mean, ‘Post. SD’ denotes the posterior standard deviation, ‘PIP’ represents the posterior inclusion probability, and ‘S.E.’ stands for standard error. Part A contains numerical results of BMA based on the UIP g-prior and a prior uniform distribution for the model. Part B reports the results of the frequentist check, which includes substantial variables with PIP higher than 85% obtained from the baseline BMA specification. Standard errors in the frequentist check are clustered at the level of studies and countries using two-way clustering. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals.The definition of all variables is available in Table 3.

Figure 6 Posterior Inclusion Probabilities Across Different Prior Settings



Notes: UIP (unit information prior) and Uniform model priors are based on Eicher et al. (2011). The UIP and Dilution prior recommended by Eicher et al. (2011) and George (2010), respectively, are utilized. BRIC and Random represent the benchmark g-prior by Fernandez et al. (2001) for parameters with the beta-binomial model prior for the model space, indicating that each model size has equal prior probability. The results remain robust regardless of model and g-prior choices. The results of BMA with alternative distributions are provided in the appendix, Figures A4 and A5. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals. The definition of all variables is available in Table 3.

Regimes. Another crucial factor to consider is the state of the economy, which significantly influences the size of the multiplier. While various studies provide differing perspectives on this matter, Auerbach and Gorodnichenko (2012a, 2012b, 2017), along with Bachman and Sims (2012), emphasize the pivotal role of economic conditions in estimating responses to government intervention. Conversely, Ramey and Zubairy (2018) and Afonso et al. (2018) suggest minimal or no impact, favoring non-linearity. The results of our multivariable model reject the assumption of linearity in fiscal multipliers, supporting the notion of state dependency. Both variables indicating economic regimes are deemed 'decisive' in our BMA analysis. Furthermore, the estimated coefficients highlight a substantial 0.9 difference between multipliers estimated during prosperous and challenging economic periods.

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 (Journal) 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 academia; on average, estimates reported in published works are 0.12 higher than their non-published counterparts.

Furthermore, among other subgroups (**Frequency, Data, Model, IRF Type, Others**) only two variables, panel type databases (PANEL) and the import-to-GDP ratio (MGDP) emerge as other significant variables that notably influence the magnitude of the multiplier. However, other factors, particularly those related to frequency, do not exhibit any significant impact.

The meta-analysis of the fiscal multiplier, besides testing for publication bias, also seeks to answer the question about the size of the true effect. This question is especially important in the scope of the interest of policymakers to precisely evaluate the effectiveness of implemented fiscal policies. As the bottom line of the current analysis, a 'best practice' multiplier was calculated. Simply, this 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. We use sample maxima for weights plugged into the regression when the variable is preferred, the sample means when there is no preference, and sample minima when the variable is far from the 'best practice'. As the benchmark case, we refer to the 'best practice' multiplier identified by Gechert and Rannenberg (2018). Additionally, we prefer quarterly frequency, published studies, and frequently cited studies.

Table 5 Best Practice: Alternative Specifications

	Multiplier	*95% Confidence Interval*
Best practice	0.73	(0.38, 1.07)
Crises	1.52	(0.99, 2.05)
Boom	0.61	(0.22, 1.10)
Higher import	0.57	(-0.30, 1.44)
Annual	0.99	(0.36, 1.63)
Investment	1.62	(0.93, 2.31)
Tax	0.48	(-0.07, 1.03)
Military	1.28	(0.66, 1.89)

Notes: This table presents mean estimates of the fiscal multiplier conditional on various models, identification, shock type, multiplier type, and publication characteristics. The analysis places more weight on selected aspects of the study design, following the definition of best practice by Gechert and Rannenberg (2018). Each row represents the implied multiplier when changing one aspect of the best practice definition. The 95% confidence intervals, reported in the second column, are constructed using standard errors, with two-way clustering at the level of study and countries.

Table 5 contains the multiplier calculated according to the definition of best practice. For the benchmark case, the multiplier is equal to 0.73, which means the preferred features almost offset the impact of each other. The second column contains 95% confidence interval borders. From the table, it is evident that if investment data is preferred the 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. For alternative specifications, the multiplier varies from 0.5 to 1.6; for comparison in Gechert and Rannenberg (2018), a 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 0.6 to 1.7 and is the largest for the high-import samples.

4.3 Robustness Check

We conduct a robustness check to confirm our main findings, particularly focusing on the precision coefficient discrepancy. To address this concern, we examine a smaller subset of papers that provided standard errors or relevant data to calculate them. Only 13 papers are suitable for this analysis, contributing 525 observations, which comprise 16% of the complete dataset. The papers are listed in Table A4 in the Appendix. Using this new subsample, we apply a similar test as described in Equation (4) in the 'Publication Bias' section, but this time incorporating standard errors instead of the inverse of the square root of the number of observations.

$$\mu_{ij} = \alpha + \gamma * (SE)_{ij} + \varepsilon_{ij} \tag{6}$$

The results, detailed in Table 6, indicate that the coefficient of standard errors is statistically significant, unlike the results presented in Table 2. Additionally, we observe that the intercepts of the models, representing the true effect, are lower than the estimates reported in Table 2. When interpreting the findings from Table 6, it is essential to note that all observations are from the new subsample of the extended dataset, covering papers written or published between 2013 and 2020. Moreover, the discrepancy in numbers may also be attributed to the majority of papers covered by Gechert and Rannenberg (2018) being published or written immediately after the Great Recession, suggesting that multipliers tend to be larger in economically stressed periods.

Moreover, a recent paper by Irsova et al., (2023b) addresses the issue of spurious precision, which arises when standard errors are manipulated by authors to attain statistical significance. The paper introduces a novel approach to tackle the potential endogeneity problem between estimates and standard errors. This new method, known as the meta-analysis instrumental variable estimator (MAIVE), proposes a two-step estimation process. In the first step, variances are regressed on the inverse of the number of observations. In the second step, the fitted values obtained from the first step are used in the FAT.

Due to limitations in our database, we are unable to implement MAIVE in the general model. However, we can apply it to a sub-sample, as we possess all the necessary information. Unfortunately, as shown in Appendix Table A8, the coefficients obtained from the first step of the MAIVE method, where variances are regressed on the inverse of the number of observations, yield statistically insignificant results. This further reinforces our suspicion that the number of observations may be a poor instrument or proxy in our case.

Table 6 Robustness Check: Results of Funnel Asymmetry Test

PANEL A	*OLS*	*FE*	*Hierarchical*
γ	0.30***	0.27***	0.97***
	(0.03)	(0.02)	(0.48)
	[-2.92, 1.85]	-	-
α	0.63***	0.64***	0.23
	(0.12)	(0.01)	(0.21)
	[0.22, 0.864]	-	-
PANEL B	*BE*	*Study - Weighted*	*Precision - Weighted*
γ	0.49***	0.30***	1.8***
	(0.14)	(0.02)	(0.44)
	-	[-3.7, 2.27]	[0.67, 2.88]
α	0.39***	0.55***	0.01***
	(0.16)	(0.14)	(0.00)
	-	[0.08 0.81]	[-0.11, 0.20]

Notes: The table contains the results of the regression provided in Equation (6). Standard errors are reported in parentheses. OLS refers to ordinary least squares, FE refers to study fixed effects, and BE refers to Between effects. In square brackets, 95% confidence intervals are reported using wild bootstrap clustering; the implementation follows Roodman et al. (2019), and Rademacher weights with 9999 replications are used. α represents the constant, while γ represents the coefficient on standard errors. Significance levels are denoted by *, **, and ***, indicating significance at the 10%, 5%, and 1% levels, respectively. The number of observations in all regressions presented in Table 6 is 525.

Our research provides robust evidence supporting the presence of publication bias in the fiscal policy literature, indicating a systematic distortion in reported estimates. Based on these findings, we conclude that the literature suffers from publication bias.

In addition to the drawbacks of the models discussed earlier, the estimation with standard errors suggests that the number of observations, in our case, does not appear to be a reliable replacement for standard errors. Therefore, similar attempts, whether in this study or in other works employing similar methodologies and data, such as Gechert and Rannenberg (2018) or Asatryan et al. (2020), could not find significant evidence supporting publication bias.

5. Concluding Remarks

Our study presents an integrated overview of the fiscal multiplier estimates, the key parameter that indicates how large the economy's response to government intervention is. In this study, we use a large database comprising 131 studies and over 3200 observations. We employ both linear and non-linear meta-analysis methods, many of which are applied to fiscal multipliers for the first time. Due to poor reporting of standard errors, we choose to use the number of observations per study as the proxy of standard errors. Our results indicate that the genuine effect of the fiscal multiplier is positive but less than one, ranging from 0.75 to 0.83 under different models. Moreover, our study marks the first implementation of BMA in the meta-analysis of fiscal multipliers. BMA plays a central role in all analyses. The multivariable model developed according to the results of BMA indicates the presence of publication selection bias in the literature.

Our findings diverge from earlier research (e.g., Gechert, 2015; Gechert and Rannenberg, 2018; Asatryan et al., 2020), which either found limited support or no evidence of publication bias. This difference is mainly due to use of BMA for heterogeneity investigation, a neglected method in prior studies. The methodology we employ effectively addresses model uncertainty and omit- ted variable bias, while also challenging the assumption that the number of observations reliably represents standard errors in fiscal multiplier estimates.

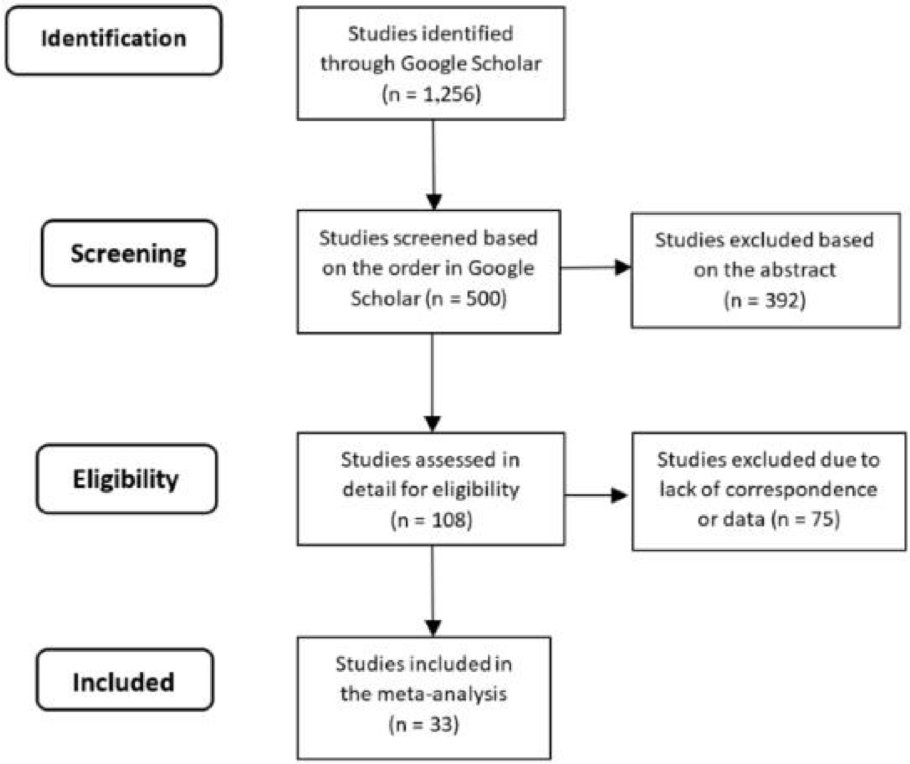
Specific shock types and data characteristics contribute significantly to the size of the fiscal multiplier, providing insight into which government tools may yield better results. We find that government intervention through public investment is associated with larger multipliers. Additionally, our overall assessment of fiscal policy literature supports the state dependency of the fiscal multiplier.

We provide robust evidence supporting the presence of publication bias in fiscal policy literature, indicating a systematic distortion in reported estimates. This outcome is the primary contribution of our paper and challenges the conclusions of previous studies that do not support publication bias in fiscal multiplier estimates.

While our study provides insights into fiscal multipliers and publication bias, it also highlights limitations present in both our research and previous studies. Specifically, relying on the number of observations as a proxy for standard errors may introduce inaccuracies. Future research addressing these limitations could significantly enhance our understanding of the true determinants and nature of fiscal multipliers.

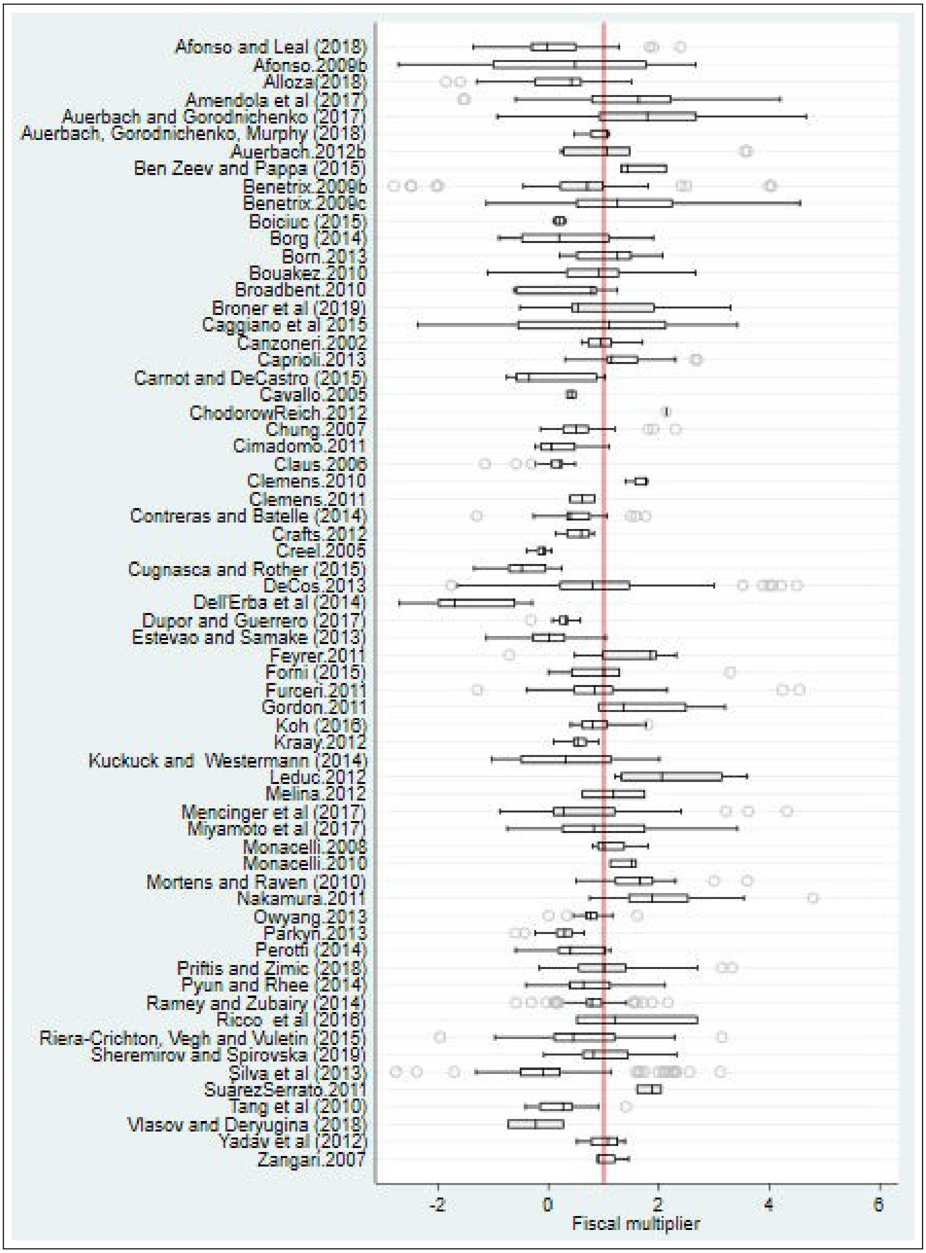
APPENDIX

Figure A1 PRISMA Flow Diagram

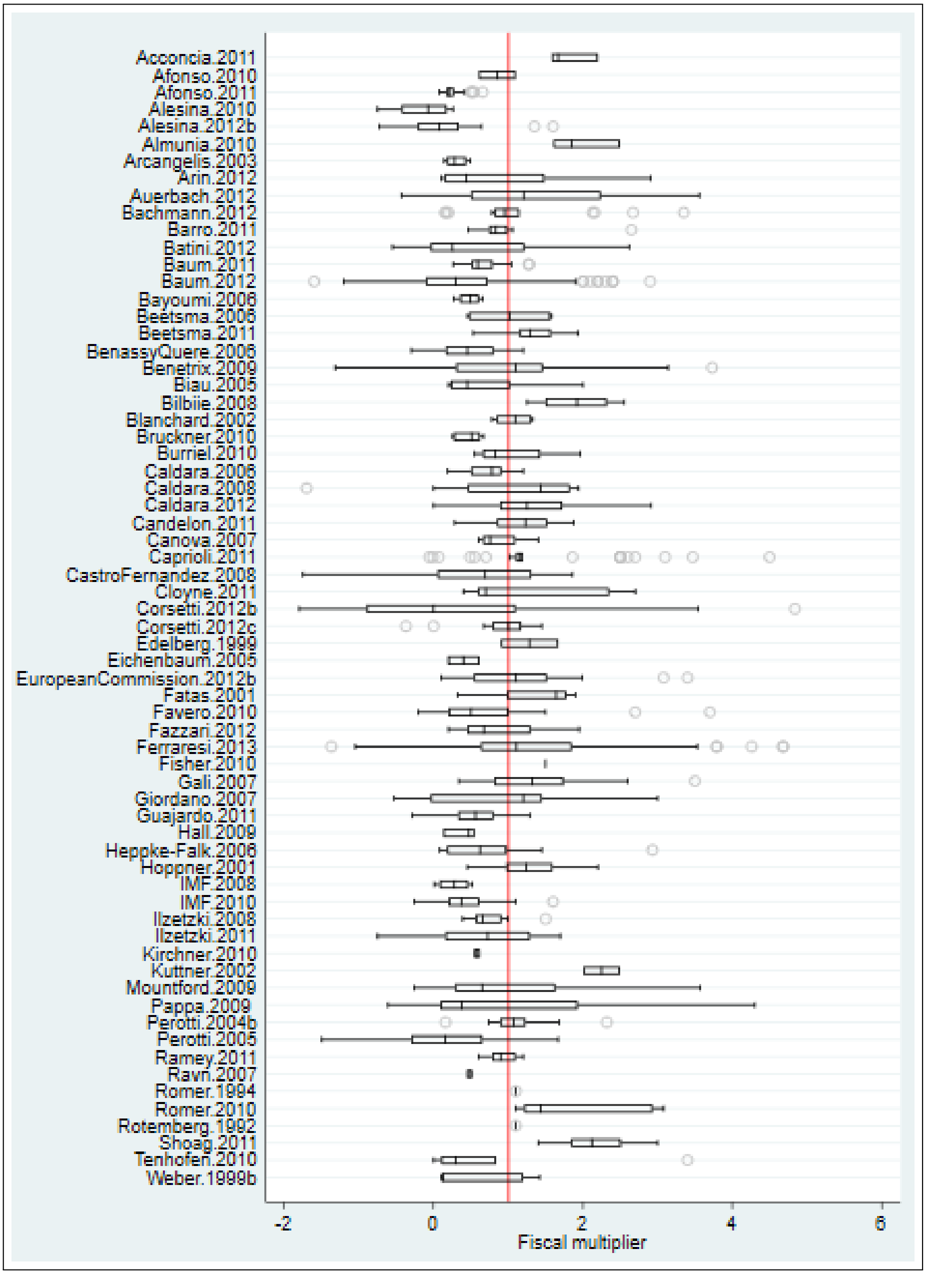


Notes: We use the following query in Google Scholar: ("fiscal multiplier" and "fiscal policy"). Note that Google Scholar provides full-text search, not only the search of the title, abstract and keywords; consequently, our query is very general. The search for studies was terminated on January 31, 2021. The list of the 33 studies included in the meta-analysis is available in Table 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is an evidence-based set of items for reporting in systematic reviews and meta-analyses. More details on PRISMA and the reporting standard of meta-analysis in general are provided by Havranek et al. (2020).

Figure A2 Estimates of Multiplier Across Studies



(a)



(b)

Notes: The length of each box represents the interquartile range (P25-P75), with the dividing line inside the box, indicating 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, but they are included in all statistical tests. The vertical red line represents a multiplier value of 1.

Table A1 Fiscal Multiplier for Different Subsets

Variable	*Obs.*	*Unweighted*			*Weighted*		
		Mean	*95 % CI*		*Mean*	*95 % CI*	
MULTIPLIER	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
SEEJOR	514	0.79	0.71	0.86	0.79	0.70	0.87
Regime							
AVERAGE	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
SEMIANNUAL	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	2394	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							
CUMULATIVE	2515	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							
TOP5	226	0.95	0.87	1.03	1.15	1.07	1.24
JOURNAL	1179	0.92	0.86	0.97	0.92	0.87	0.98
WORKING PAPER	1874	0.62	0.58	0.68	0.77	0.73	0.83
Shock							
SPENDING	1336	0.88	0.83	0.92	0.94	0.89	0.98
CONSUMPTION	665	0.95	0.89	1.02	0.88	0.82	0.94
INVESTMENT	218	1.26	1.11	1.41	1.43	1.30	1.56
MILITARY	222	0.93	0.83	1.03	1.00	0.91	1.10
TRANSFERS	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
DEFICIT	160	-0.4	-0.14	0.06	0.31	0.19	0.43

Notes: The table provides a summary of estimates for different subsets. 'Obs.' denotes the observation number, '95% CI' represents 95% confidence intervals. The definition of the variables is available in Table 3. Weighted estimates are calculated by weighting each estimate by the inverse of the number of estimates reported per study.

Table A2 Replication of the Results by Gechert and Rannenberg (2018)

PART A	***Estimates reported in GR(2018)***		
FAT-PET	**Estimates**	**STD Err**	**p-value**
γ	-3.13**	1.51	0.04
α	1.10***	0.12	0.00
FAT-PEESE			
γ	-18**	8.61	0.04
α	1.00***	0.07	0.000
PANEL B	**Replication**		
FAT-PET	**Estimates**	**STD Err**	**p-value**
γ	-2.98**	1.51	0.05
α	1.10***	0.12	0.00
FAT-PEESE			
γ	-17.71**	8.74	0.04
α	1.00***	0.07	0.00
PANEL C	**With clustered errors**		
FAT-PET	**Estimates**	**STD Err**	**p-value**
γ	-2.98	3.00	0.32
α	1.10***	0.24	0.00
FAT-PEESE			
γ	-17.71	18.69	0.35
α	1.00***	0.15	0.00

Notes: ‘STD Err’ - standard error; ‘FAT’ - ‘funnel-asymmetry tests’, ‘PET’ - ‘precision effect test’, ‘PEESE’ - ‘precision effect estimate with standard error’. This table presents estimations using data from Gechert and Rannenberg (2018). The results from Gechert and Rannenberg (2018) were extracted from Table 2 on page 1165. In the estimations, α represents the constant, indicating the mean beyond bias, while γ represents the coefficient on standard errors. For more details, you may refer to Stanley and Doucouliagos, 2012, 2014. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The number of observations in all regressions presented in Table A2 is 1914.

Table A3 Summary of the Benchmark BMA Estimation

Mean no. regressors 17.9132	Draws $3 * 10^6$	Burn-ins $1 * 10^6$	Time 6.393208 mins	No. models visited 409823
Modelspace $1.1 * 10^9$	Models visited 0.038%	Topmodels 100	Corr PMP 0.9981	No. Obs. 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 4. Considering Eicher et al. (2011), a uniform distribution was used as a model prior, and the Unit Information Prior (UIP) was employed for the g-prior.

Figure A3 Model Size and Convergence for the Benchmark BMA Model

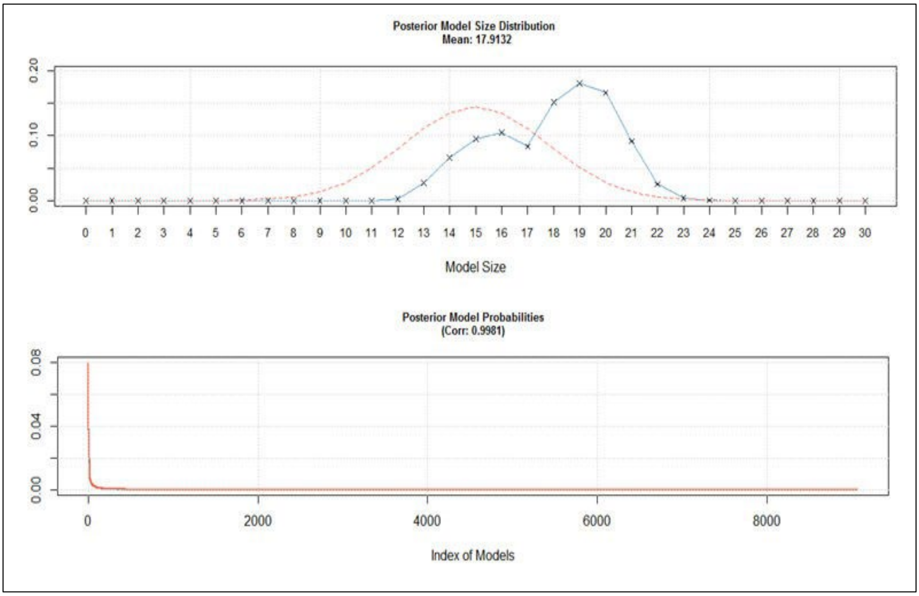
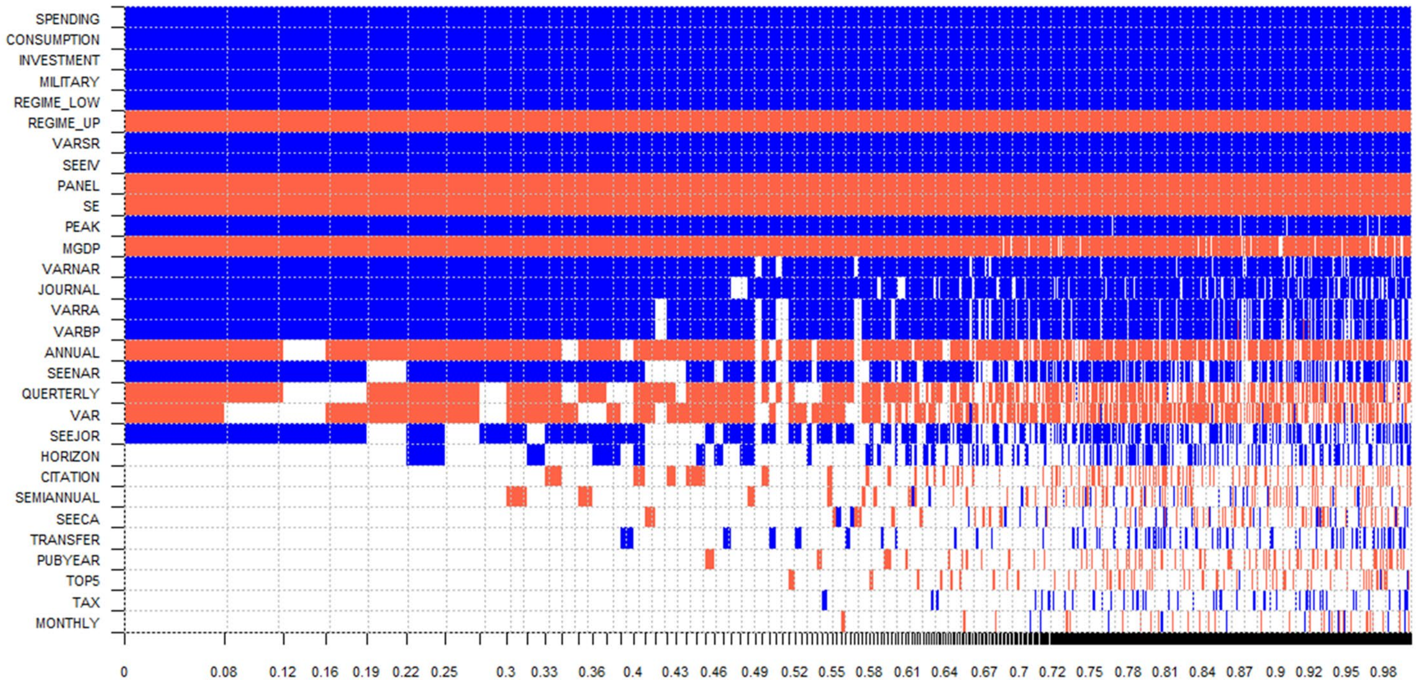
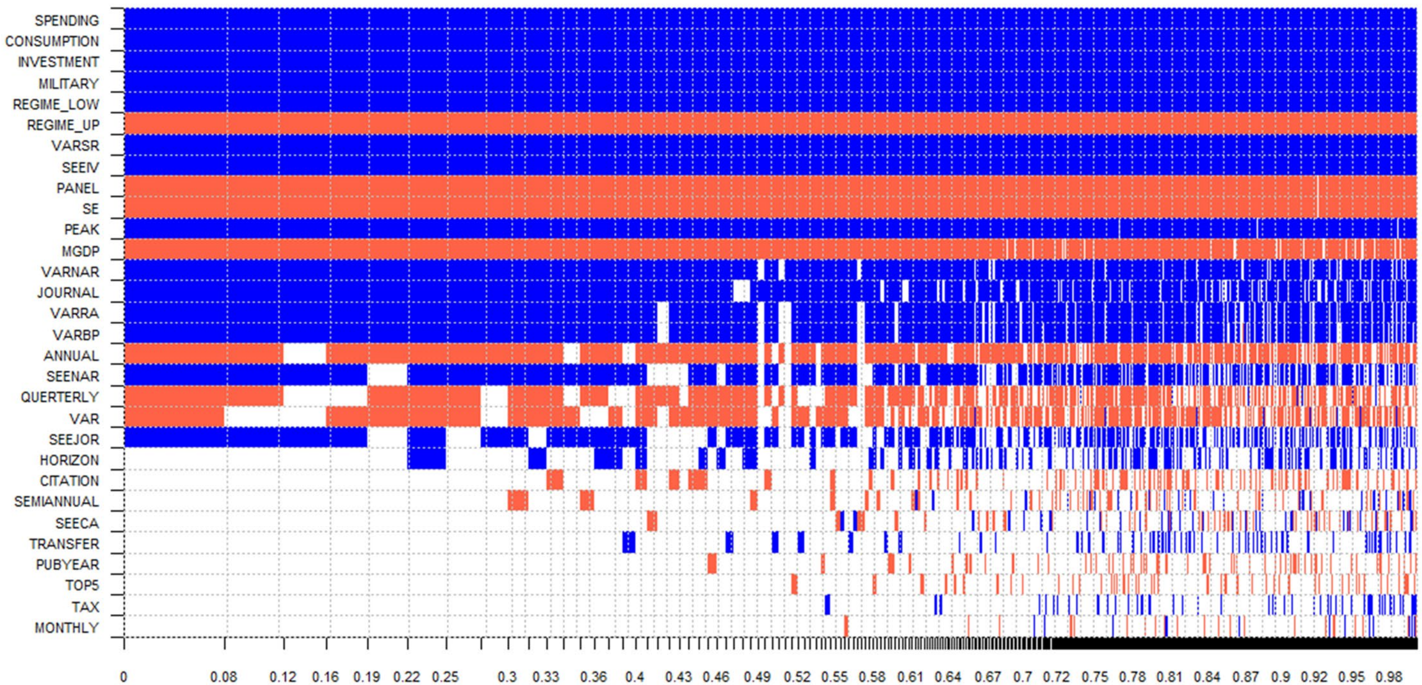


Figure A4 Model Inclusion in BMA (g-prior UIP; m-prior - Dilution)



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. The blue color indicates if the estimated parameter of a corresponding explanatory variable is positive. The 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 Table A4. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals. The definition of the variables is available in Table 3.

Figure A5 Model Inclusion in BMA (g-prior – BRIC; m-prior - Random)



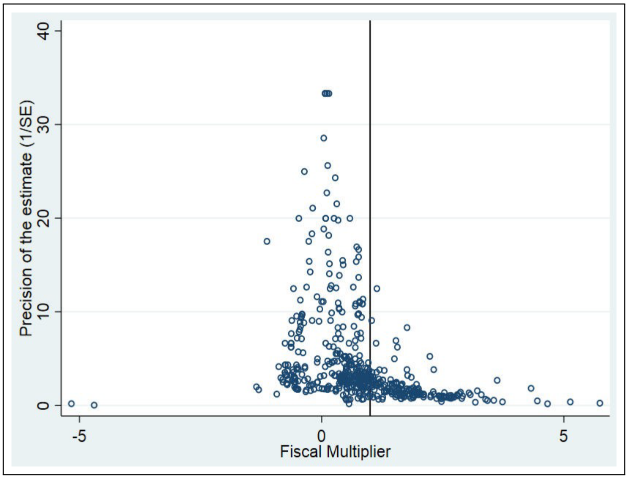
Notes: Explanatory variables are ranked according to their posterior inclusion probabilities along the vertical axis from the highest at the top to the lowest at the bottom. The horizontal axis depicts the values of cumulative posterior model probability. The blue color indicates if the estimated parameter of a corresponding explanatory variable is positive. The 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 Table A4. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals. The definition of the variables is available in Table 3.

Table A4 Alternative BMA Priors

	g-prior=UIP,	m-prior=Dilution		g-prior =BRIC, m-prior=Random		
Variables	***PIP***	***Post. mean***	***Post. SD***	***PIP***	***Post. mean***	***Post. SD***
SE	1.00	-3.83	0.82	1.00	-3.83	0.82
VAR	0.67	-0.39	0.34	0.67	-0.39	0.33
PEAK	1.00	0.16	0.04	1.00	0.16	0.04
HORIZON	0.25	0.001	0.002	0.25	0.001	0.001
SPENDING	1.00	0.60	0.04	1.00	0.60	0.04
CONSUMPTION	1.00	0.79	0.05	1.00	0.79	0.05
INVESTMENT	1.00	1.16	0.07	1.00	1.16	0.06
MILITARY	1.00	0.81	0.08	1.00	0.81	0.08
TRANSFERS	0.10	0.02	0.07	0.10	0.02	0.07
TAX	0.05	0.005	0.02	0.05	0.002	0.02
REGIME-LOW	1.00	0.43	0.04	1.00	0.43	0.04
REGIME-UP	1.00	-0.48	0.04	1.00	-0.49	0.04
SEEIV	1.00	0.58	0.18	1.00	0.58	0.18
SEENAR	0.78	0.35	0.23	0.78	0.35	0.23
SEEJOR	0.65	0.20	0.17	0.65	0.20	0.17
SEECA	0.10	-0.01	0.09	0.10	-0.01	0.09
VARNAR	0.96	0.77	0.32	0.96	0.77	0.32
VARBP	0.91	0.59	0.28	0.91	0.59	0.28
VARRA	0.92	0.65	0.30	0.92	0.65	0.30
VARSR	1.00	1.09	0.30	1.00	1.09	0.30
PANEL	1.00	-0.36	0.07	1.00	-0.36	0.07
ANNUAL	0.84	-0.27	0.17	0.84	-0.27	0.17
SEMIANNUAL	0.13	-0.03	0.10	0.13	-0.02	0.10
QUARTERLY	0.68	-0.19	0.16	0.68	-0.19	0.16
MONTHLY	0.04	-0.001	0.06	0.04	-0.001	0.06
JOURNAL	0.93	0.11	0.05	0.93	0.11	0.05
TOP5	0.07	-0.01	0.03	0.07	-0.004	0.08
MGDP	0.97	-0.003	0.001	0.97	-0.003	0.001
CITATIONS	0.15	-0.01	0.02	0.15	-0.08	0.02
PUBLICATION YEAR	0.08	-0.02	0.09	0.08	-0.02	0.09
INTERCEPT	1.00	0.62	NA	1.00	0.62	NA

Notes: Dependent variable is fiscal multiplier, Post. mean - posterior mean, Post. SD - posterior standard deviation, PIP - posterior inclusion probability. Part A contains numerical results of BMA based on the UIP g-prior and prior Dilution distribution for the model. Part B contains numerical results of BMA based on the BRIC g-prior and prior Random distribution for the model. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals. The definition of the variables is available in Table 3.

Figure A6 Funnel Plot: with Standard Errors



Notes: In the absence of publication bias the funnel should be symmetrical around the most precise estimates. A solid vertical line indicates the case when the multiplier is equal to one. The most precise estimates (those with the smallest standard errors) should cluster closely around the true mean value. As precision decreases, estimates scatter more widely around the true mean value. The graph shows severe asymmetry in the under-representation of negative numbers.

Table A5 List of Studies: Robustness Check

Arin et al. (2015)	Sheremirov and Spirovska (2019)
Auerbach and Gorodnichenko (2017)	Broner et al. (2019)
Auerbach et al. (2018)	Mencinger et al. (2017)
Estevao and Samake (2013)	Carnot and DeCastro (2015)
Ramey and Zubairy (2018)	Cugnasca and Rother (2015)
Miyamoto et al. (2018)	Dupor and Guerrero (2017)
Kuckuck and Westermann (2014)	

Table A6 Results of Frequentist Model Averaging

Variables	*Coefficient*	*STD Err*	*p-value*
SE	-3.73	0.86	0.00
VAR	**-0.50**	**0.21**	**0.01**
PEAK	0.18	0.04	0.00
HORIZON	**0.004**	**0.002**	**0.03**
SPENDING	0.71	0.08	0.00
CONSUMPTION	0.90	0.08	0.00
INVESTMENT	1.25	0.09	0.00
MILITARY	0.91	0.10	0.00
TRANSFERS	0.12	0.14	0.40
TAX	0.12	0.08	0.11
REGIME-LOW	0.44	0.05	0.00
REGIME-UP	-0.46	0.05	0.00
SEEIV	0.56	0.14	0.00
SEENAR	0.48	0.15	0.00
SEEJOR	**0.27**	**0.12**	**0.02**
SEECA	0.06	0.17	0.75
VARNAR	0.91	0.20	0.00
VARBP	0.73	0.18	0.00
VARRA	0.81	0.18	0.00
VARSR	1.20	0.20	0.00
PANEL	-0.37	0.08	0.00
ANNUAL	-0.50	0.18	0.01
SEMIANNUAL	-0.28	0.19	0.14
QUARTERLY	**-0.42**	**0.18**	**0.02**
MONTHLY	-0.13	0.26	0.62
JOURNAL	0.09	0.04	0.03
TOP5	-0.04	0.08	0.59
MGDP	-0.003	0.001	0.00
CITATIONS	-0.05	0.04	0.20
PUBLICATION YEAR	-0.25	0.25	0.36
INTERCEPT	0.99	0.45	0.03

Notes: 'STD Err' - standard error; We use Mallow's weights according to Hansen (2007), and the orthogonalization of the covariate space suggested by Amini and Parmeter (2012) to conduct FMA exercise. Bold lines show variables important in FMA but not in the benchmark BMA. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals. The definition of the variables is available in Table 3.

Table A7 Frequentist Check: OLS and WLS

	PART A: Frequentist check: OLS			*PART B: Frequentist check WLS*		
variables	***Coefficient***	***S.E.***	***p-value***	***Coefficient***	***S.E.***	***p-value***
SE	-4.23	1.45	0.00	-7.63	1.35	0.00
PEAK	0.16	0.11	0.17	-0.02	0.18	0.91
SPENDING	0.58	0.10	0.00	0.80	0.11	0.00
CONSUMPTION	0.77	0.11	0.00	0.97	0.18	0.00
INVESTMENT	1.13	0.27	0.00	1.90	0.16	0.00
MILITARY	0.69	0.27	0.01	1.00	0.27	0.00
REGIME LOW	0.44	0.07	0.00	0.46	0.16	0.01
REGIME UP	-0.47	0.07	0.00	-0.42	0.08	0.00
SEEIV	0.43	0.09	0.00	**0.41**	**0.25**	**0.10**
VARNAR	0.24	0.11	0.02	0.31	0.11	0.01
VARSR	0.47	0.15	0.00	**0.49**	**0.27**	**0.08**
PANEL	-0.37	0.15	0.01	-0.60	0.11	0.00
JOURNAL	0.15	0.07	0.03	0.26	0.09	0.00
MGDP	-0.003	0.002	0.13	**0.003**	**0.001**	**0.01**
INTERCEPT	0.62	0.18	0.00	0.45	0.13	0.00

Notes: The dependent variable is the fiscal multiplier, 'S.E.' refers to the standard error. Part A contains results from the frequentist check using OLS. Part B reports frequentist check results using WLS, with the number of observations serving as the weight. The variables included are substantial, with PIPs higher than 85%, obtained from the baseline BMA specification. Standard errors in the frequentist check are two-way clustered at the study and country levels. Bold lines show variables important in WLS, but not in the OLS. For VAR (Vector Autoregression) models: VARSR - Sign-Restriction; VARNAR - Narrative Approach; VARRA - Recursive Approach; VARBP - Blanchard and Perotti. For single equation estimates (SEE): SEEIV - Instrumental Variable; SEENAR - Narrative Approach; SEEJOR - Jorda method; SEECA - Cyclically adjusted series. SE stands for standard error. MGDP represents import-to-GDP ratio. TOP5 refers to the top 5 Economics journals. The definition of the variables is available in Table 3. The number of observations in the regressions presented in Table A7 is 3279.

Table A8 Meta-analysis Instrumental Variable Estimator, First Step Regression Results

First step	***Estimates***	***STD Err***	***p-value***
ψ_0	7.02	7.41	0.36
ψ_1	33.72	274.53	0.90

Notes: The Meta-analysis Instrumental Variable Estimator (MAIVE) is a two-step procedure designed to exclude spurious elements from reported standard errors related to p-hacking. STD. Err. refers to standard errors. We run the following regression: $SE(\mu)_i^2 = \psi_0 + \psi_1 * (1/N_i) + v_i$, where $SE(\mu)_i^2$ is the variance, μ_i is the effect size reported in the primary study, N_i is the sample size in the primary study, v_i is the error term, and ψ_0 and ψ_1 re the constant and coefficient, respectively. For further information, please refer to Irsova et al. (2023b). Due to a non-significant coefficient, we do not implement the second step of the procedure. The number of observations in the regression presented in Table A8 is 525.

REFERENCES

Afonso A, Baxa J, Slavık M (2018): Fiscal Developments and Financial Stress: A Threshold VAR Analysis. *Empirical Economics*, 54(2):395-423.

Afonso A, Silva Leal F (2018): Fiscal Multipliers in the Eurozone: A SVAR Analysis. *REM Working Paper*, 047-2018.

van Aert RC, van Assen MA (2023): Correcting for Publication Bias in a Meta-Analysis with the P-uniform* Method. Working paper. Tilburg University and Utrecht University.

Alloza M (2018): Is Fiscal Policy More Effective in Uncertain Times or During Recessions?. Banco de Espana Working Paper No. 1730, Available at SSRN: https://ssrn.com/abstract=3024538

Amendola A, di Serio M, Fragetta M, Melina G (2019): The Euro-Area Government Spending Multiplier at the Effective Lower Bound. (No WP/19/133). IMF Working paper.

Amini SM, Parmeter CF (2012): Comparison of Model Averaging Techniques: Assessing Growth Determinants. *Journal of Applied Econometrics*, 27(5):870-876.

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 AJ, Gorodnichenko Y (2017): Fiscal Multipliers in Japan. *Research in Economics*, 71(3):411-421.

Auerbach AJ, Gorodnichenko Y (2012a): Measuring the Output Responses to Fiscal Policy.

American Economic Journal: Economic Policy, 4(2):1-27.

Auerbach AJ, Gorodnichenko Y (2012b): Fiscal Multipliers in Recession and Expansion.

Fiscal Policy after the Financial Crisis, 63:98.

Auerbach AJ, Gorodnichenko Y, Murphy D (2019): *Local Fiscal Multipliers and Fiscal Spillovers in the United States* (No. w25457): National Bureau of Economic Research.

Bachmann R, Sims ER (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, Weber A (2014): A Simple Method to Compute Fiscal 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 Defense 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.

Boiciuc I (2015): The Effects of Fiscal Policy Shocks in Romania. A SVAR Approach. *Procedia Economics and Finance*, *32:*1131-1139.

Bom PR, Rachinger H (2019): A Kinked Meta-Regression Model for Publication Bias Correction. *Research Synthesis Methods*, *10* (4):497-514.

Bom PR, Rachinger H (2020): A Generalized-Weights Solution to Sample Overlap in Meta-Analysis. *Research Synthesis Methods*, *11* (6):812-832.

Borg I (2014): *Fiscal multipliers in Malta* (No. WP/06/2014). CBM Working Papers.

Broner F, Clancy D, Erce A, Martin A (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.

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.

Contreras J, Battelle H (2014): Fiscal Multipliers in a Panel of Countries. (No. 2014-15). Working Papers.

Cugnasca A, Rother P (2015): Fiscal Multipliers During Consolidation: Evidence from the European Union (No. 1863). *ECB Working Paper.*

Dell'Erba MS, Poplawski-Ribeiro MM, Koloskova K (2014): Medium-Term Fiscal Multipliers During Protracted Recessions. (No. WP/14/213). International Monetary Fund.

Dupor B, Guerrero R (2017): Local and Aggregate Fiscal Policy Multipliers. *Journal of Monetary Economics*, *92:*16-30.

Egger M, Smith GD, Schneider M, Minder C (1997): Bias in Meta-Analysis Detected by a Simple, Graphical Test. *BMJ*, *315* (7109):629-634.

Eicher TS, Papageorgiou C, Raftery AE (2011): Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants. *Journal of Applied Econometrics*, 26(1):30-55.

Elminejad A, Havranek T, Horvath R, Irsova Z (2023): Intertemporal Substitution in Labor Supply: A Meta-Analysis. Review of Economic Dynamics, 51:1095-1113.

Estevao M, Samake I (2013): The Economic Effects of Fiscal Consolidation with Debt Feedback. *International Monetary Fund*.

Fatas 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 MF (2001): Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, 100(2):381-427.

Forni M, Gambetti L (2016): Government Spending Shocks in Open Economy VARs. *Journal of International Economics*, 99:68-84.

Furukawa C (2021): 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 (2022): Measuring Capital-Labor Substitution: The Importance of Method Choices and Publication Bias. *Review of Economic Dynamics.* 45:55-82.

George EI (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.

Hansen BE (2007): Least Squares Model Averaging. *Econometrica*, 75(4):1175-1189.

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, Irsova Z (2017): Do Borders Really Slash Trade? A Meta-Analysis. *IMF Economic Review*, *65:*365-396.

Havránek T, Stanley TD, Doucouliagos H, Bom P, Geyer-Klingeberg J, Iwasaki I ... & van Aert RC (2020): Reporting Guidelines for Meta-Analysis in Economics. *Journal of Economic Surveys*, *34* (3):469-475.

Havranek T, Irsova Z, Laslopova L, Zeynalova O (2024): Publication and Attenuation Biases in Measuring Skill Substitution. *Review of Economics and Statistics,* 1-14.

Havranek T, Horvath R, Zeynalov A (2016a): Natural Resources and Economic Growth: A Meta-

Analysis. *World Development*, 88:134-151.

Havranek T, Irsova Z, Lesanovska J (2016b): Bank Efficiency and Interest Rate Pass-Through: Evidence from Czech Loan Products. *Economic Modelling*, 54:153-169.

Ilzetzki E, Mendoza EG, Vegh CA (2013): How Big (Small?) Are Fiscal Multipliers?

Journal of Monetary Economics, 60(2):239-254.

Ioannidis JP (2005): Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8):e124.

Ioannidis JP, Stanley TD, Doucouliagos H (2017): The Power of Bias in Economics Research. The Economic Journal, 127(605): F236-F265.

Irsova Z, Doucouliagos H, Havranek T, Stanley TD (2023a): Meta-Analysis of Social Science Research: A Practitioner's Guide. *Journal of Economic Surveys*.

Irsova Z, Bom PR, Havranek T, Rachinger H (2023b): Spurious Precision in Meta-Analysis of Observational Research. (No. 2023/05). Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Iwasaki I, Tokunaga M (2016): Technology Transfer and Spillovers from FDI in Transition Economies: A Meta-Analysis. *Journal of Comparative Economics*, 44(4), 1086-1114.

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

Jorda O (2005): Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161-182.

Kocenda E, Iwasaki I (2022): Bank Survival around the World: A Meta-Analytic Review.

Journal of Economic Surveys, 36(1):108-156.

Chian Koh W (2017): Fiscal Multipliers: New Evidence from a Large Panel of Countries. *Oxford Economic Papers*, *69* (3):569-590.

Kuckuck J, Westermann F (2014): On the Size of Fiscal Multipliers: A Counterfactual Analysis. *Economics letters*, *123* (1):26-32.

Leeper EM, Walker TB, Yang SCS (2010): Government Investment and Fiscal Stimulus. *Journal of Monetary Economics*, 57(8):1000-1012.

Madigan D, York J, Allard D (1995): Bayesian Graphical Models for Discrete Data. *International Statistical Review/Revue Internationale de Statistique*, 215-232.

Mencinger J, Aristovnik A, Verbič M (2017): Asymmetric Effects of Fiscal Policy in EU and OECD Countries. *Economic Modelling*, *61:*448-461.

Mertens K, Ravn M (2010): Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks (No. w16289). *National Bureau of Economic Research.*

Miyamoto W, Nguyen TL, Sergeyev D (2018): Government Spending Multipliers under the Zero Lower Bound: Evidence from Japan. *American Economic Journal: Macroeconomics*, 10(3):247-77.

Mourougane A, Botev J, Fournier JM, Pain N, Rusticelli E (2016): Can an Increase in Public Investment Sustainably Lift Economic Growth? (No. 1351). OECD Publishing.

Perotti R (2014): Defense Government Spending is Contractionary, Civilian Government Spending Is Expansionary (No. w20179). *National Bureau of Economic Research.*

Priftis R, Zimic S (2018): Sources of Borrowing and Fiscal Multipliers (No. 2209). *ECB Working Paper.*

Pyun JH, Rhee DE (2015): Fiscal Multipliers during the Global Financial Crisis: Fiscal and Monetary Interaction Matters. *Contemporary Economic Policy*, *33* (1):207-220.

Raftery AE, Madigan D, Hoeting JA (1997): Bayesian Model Averaging for Linear 24regression Models. *Journal of the American Statistical Association*, 92(437):179-191.

Ramey V (2016): Macroeconomic Shocks and Their Propagation. *Handbook of macroeconomics,* 2:71-162.

Ramey VA, 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 VA (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.

Ricco G, Callegari G, Cimadomo J (2016): Signals from the Government: Policy Disagreement and the Transmission of Fiscal Shocks. *Journal of Monetary Economics*, *82*:107-118.

Riera-Crichton D, Vegh CA, Vuletin G (2015): Procyclical and Countercyclical Fiscal Multipliers: Evidence from OECD Countries. *Journal of International Money and Finance*, 52:15-31.

Roodman D, Nielsen MØ, MacKinnon JG, Webb MD (2019): Fast and Wild: Bootstrap Inference in Stata Using Boottest. *The Stata Journal*, 19(1):4-60.

Scandizzo PL, Pierleoni MR (2020): Short and Long-Run Effects of Public Investment: Theoretical Premises and Empirical Evidence. *Theoretical Economics Letters*, 10(04):834.

Sheremirov V, Spirovska S (2019): Fiscal Multipliers in Advanced and Developing Countries: Evidence from Military Spending (No. 19-3). *Working Papers.*

Silva R, Carvalho VM, Ribeiro AP (2013): How Large are Fiscal Multipliers? A Panel Data VAR Approach for the Euro Area. FEP Working Papers, Issue 500, FEP-UP. Porto, Portugal: School of Economics and Management, University of Porto, pp. 1–33

Stanley TD (1998): New Wine in Old Bottles: A Meta-Analysis of Ricardian Equivalence. *Southern Economic Journal*, *64* (3):713-727.

Stanley TD (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 TD, Doucouliagos H (2012): Meta-Regression Analysis in Economics and Business. Oxford: Routledge.

Stanley TD, Doucouliagos H (2014): Meta-Regression Approximations to Reduce Publication Selection Bias. *Research Synthesis Methods*, *5* (1):60-78.

Stanley TD (2001): Wheat from Chaff: Meta-analysis as Quantitative Literature Review. *Jour- nal of Economic Perspectives*, 15(3):131-150.

Tang HC, Liu P, Cheung EC (2013): Changing Impact of Fiscal Policy on Selected ASEAN Countries. *Journal of Asian Economics*, 24:103-116.

Van der Sluis J, Van Praag M, Vijverberg W (2005): Entrepreneurship Selection and Performance: A Meta-Analysis of the Impact of Education in Developing Economies. *The World Bank Economic Review*, *19* (2):225-261.

Vlasov SA, Deryugina EB (2018): Fiscal Multipliers in Russia. *Journal of Economic Theory Studies of the Russian Economy Issues of Economic Policy Hot Topic*, *2* (38):83-95.

Yadav S, Upadhyay V, Sharma S (2012): Impact of Fiscal Policy Shocks on the Indian Economy. *Margin: The Journal of Applied Economic Research*, *6* (4):415-444.

Zigraiova D, Havranek T, Irsova Z, Novak J (2021): How Puzzling is the Forward Premium Puzzle? A Meta-Analysis. *European Economic Review*, 134:103714.