Incentive Effects of R&D Tax Incentives: A Meta-Analysis*

Carla Pöschel†

January 13, 2022

Abstract

Despite the growing literature on the effectiveness of tax incentives for research and development (R&D), little is known about the differing design aspects of the fiscal policies. This paper applies meta-regression analysis to assess how various tax incentives affect firms' R&D expenditures. Using 496 estimates from 37 studies exploiting direct approaches, the results indicate that the nature of implementing a tax incentive substantially drive the heterogeneity found in the literature. MetaForest, a novel machine learning algorithm, confirms these results. Furthermore, I find significant publication bias in favor of reporting positive effects of fiscal incentives, which is more prevalent among studies published in peer-reviewed journals

JEL classification codes: C81; H25; O31; O38

Keywords: R&D, tax incentives, meta-regression analysis, random forest

*The author would like to thank Jost Heckemeyer, Jochen Hundsdoerfer, Eva Matthaei, Caspar van Lissa as well as participants at the VHB Annual Conference 2020 for helpful comments. The results and their interpretation are solely the responsibility of the author. The author is grateful to Fulvio Castellacci and Christine Mee Lie for providing their data set. Data and replication codes are available upon request. A previous version of this manuscript was awarded the PwC arqus Thesis Prize 2019.

†Email: carla.poeschel@fu-berlin.de; Freie University Berlin.

1 Introduction

As of 2020, 33 out of 37 Organisation for Economic Co-operation and Development (OECD) countries implemented various tax policies to stimulate firms' research and development (R&D) spending (Organisation for Economic Co-operation and Development 2021).[1] As seen in Fig. 1, the number of countries providing R&D tax incentives has increased continuously over the last two decades. Yet, the nature of policy tools for R&D inputs varies substantially among countries, ranging from tax credits to enhanced tax allowances and accelerated depreciation. In some countries, immediate cash refunds, carry-backs, or carry-forwards of the unused tax reliefs are supposed to ensure an incentive effect even for firms without taxable profits. Furthermore, tax policies apply either to all qualifying R&D expenditures (volume based), to the additional amount of R&D expenditures exceeding a given base level (incremental), or to a mixture of both (hybrid).

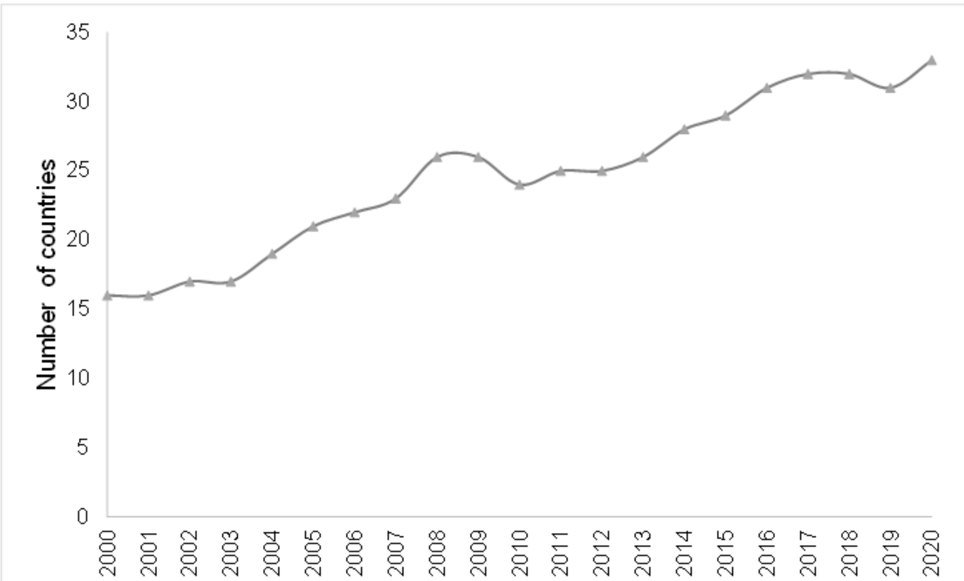

Figure 1: OECD countries providing R&D tax incentives, 2000–2020

Note: This figure presents the number of OECD countries providing R&D tax incentives.
Source: Own calculation based on the OECD R&D tax incentives database, https://www.oecd.org/sti/rd-tax-stats.htm.

A multitude of empirical studies corroborate evidence of positive effects—that is, tax incentives enhance private R&D outlays. This research has extensively investigated a number of factors associated with the effectiveness of R&D policies, such as the impact heterogeneity across sectors (e.g., Castellacci and Lie 2015; Bodas Freitas et al. 2017; Acconcia and Cantabene 2018; Chen and Li 2018) and firm size groups (e.g., Corchuelo

[1] Estonia, Finland, Latvia, and Luxembourg do not provide R&D tax provisions in 2020.

and Martínez-Ros 2010; Kobayashi 2014; Sterlacchini and Venturini 2019). Despite the advances made by the literature, little is known about the moderating effects of design aspects of schemes, even though this is a crucial topic in policy debates. Sterlacchini and Venturini (2019, p. 688) note that *"In the literature, there is a remarkable variation in the results, as these change with the policy measures under exam, the country coverage and the time span of the analysis"*. How do incentive effects of R&D tax incentives change with policy designs? This paper applies meta-regression analysis (MRA) to examine the effect heterogeneity of fiscal regimes for R&D. Beyond, the paper proposes the novel random forest algorithm MetaForest, by van Lissa (2017) to rank the moderators according to their relative importance in predicting the magnitude of estimates in new studies.

This paper extends earlier MRA by Castellacci and Lie (2015) and Gaillard-Ladinska et al. (2019) in several ways. Unlike former studies, I ensure comparability between estimates by converting initial (treatment) effect estimates into partial correlation coefficients (PCCs). In doing so, I set up a comprehensive meta-data set consisting of 496 estimates from 37 primary studies using direct approaches. With almost twice as much data than Castellacci and Lie (2015) and five-times more observations than Gaillard-Ladinska et al. (2019), I assess how the policy designs shape the reported results. I find, on average, a stronger statistical association between R&D expenditures and tax incentives for countries with hybrid scheme, while this effect is pronounced over time. In contrast, reported estimates for volume-based schemes have fallen over the evaluated time span. Given the modest number of observations in my sample, I apply the novel random forest algorithm MetaForest, by van Lissa (2017) to avoid over-fitting the data, that is, modeling random noise rather than true moderating effects. To my knowledge, I am the first to use MetaForest in a meta-analytical framework in the field of business and economics.[2] MetaForest confirms my MRA results: the design aspects are more important than firm characteristics in terms of predicting the estimates.

Furthermore, the results point to publication bias in favor of a R&D-enhancing effect of tax reliefs. This bias is more prevalent among studies published in peer-reviewed journals. Since the PCC is more of a statistical measure, I assess the average effect across more homogeneous subsets of estimates, allowing for an economically meaningful interpretation. Studies using nonparametric matching techniques to address selection

[2] I follow the meta-analyses of Bonapersona et al. (2019) and Curry et al. (2018) in behavioral science.

bias do not suffer from publication selectivity. In contrast to the weighted average of 0.90 identified by Gaillard-Ladinska et al. (2019), I find that the recipients of tax benefits have, on average, 36% higher R&D expenditures than non-recipients. Taken together, my findings are not only important for policy makers who are continuously improving R&D tax incentive schemes, but also give further guidance for interpreting empirical research on this topic.

The remainder of this paper is structured as follows. Section 2 discusses how R&D tax incentives have evolved over time among countries. Section 3 describes the data collection strategy and standardisation procedure and presents the meta-sample. Section 4 summarizes the variables and explains the methodology of MRA and MetaForest. Section 5 displays the estimation results. Finally, Section 6 concludes the paper by summarizing and discussing the implications of the results.

2 How R&D tax incentives have evolved

Within the OECD area, direct funding has been replaced by tax incentives as the main fiscal tool for promoting firm R&D due to lower administrative and compliance costs and the reduced scope for distortions in the selection process (Appelt et al. 2016; Dechezleprêtre et al. 2020).[3] While tax credits are the most common, some countries use a mix of instruments and, for example, combine a tax credit with an enhanced allowance (also known as super deduction) or accelerated depreciation to decrease both a firm's tax liability, as well as a firm's taxable income (Straathof et al. 2014). When eligible firms have no taxable profits, an incentive effect is often available through either immediate cash refunds or carry-backs or carry-forwards to future income years. Most countries explicitly target cash refunds to small and medium-sized enterprises (SMEs)—e.g., Australia, Canada, and France—since these firms are more likely to be liquidity constrained. Moreover, a tax scheme is (i) volume based if it applies to the whole amount of qualifying R&D expenditures, or (ii) it is incremental if the tax scheme covers only an increase in R&D spending compared to an initial level, which often corresponds to the average R&D expenditures of previous years (rolling average). However, this approach could distort a

[3]Among OECD countries, the share of indirect funding in total support increased from 36% in 2006 to around 57% in 2019 (Organisation for Economic Co-operation and Development 2021).

firm's R&D process, since the timing of R&D investment is determined by maximizing tax savings rather than optimizing R&D strategies (Straathof et al. 2014). This, in turn, may reduce the incentive effect among firms with high volumes of R&D expenditures.

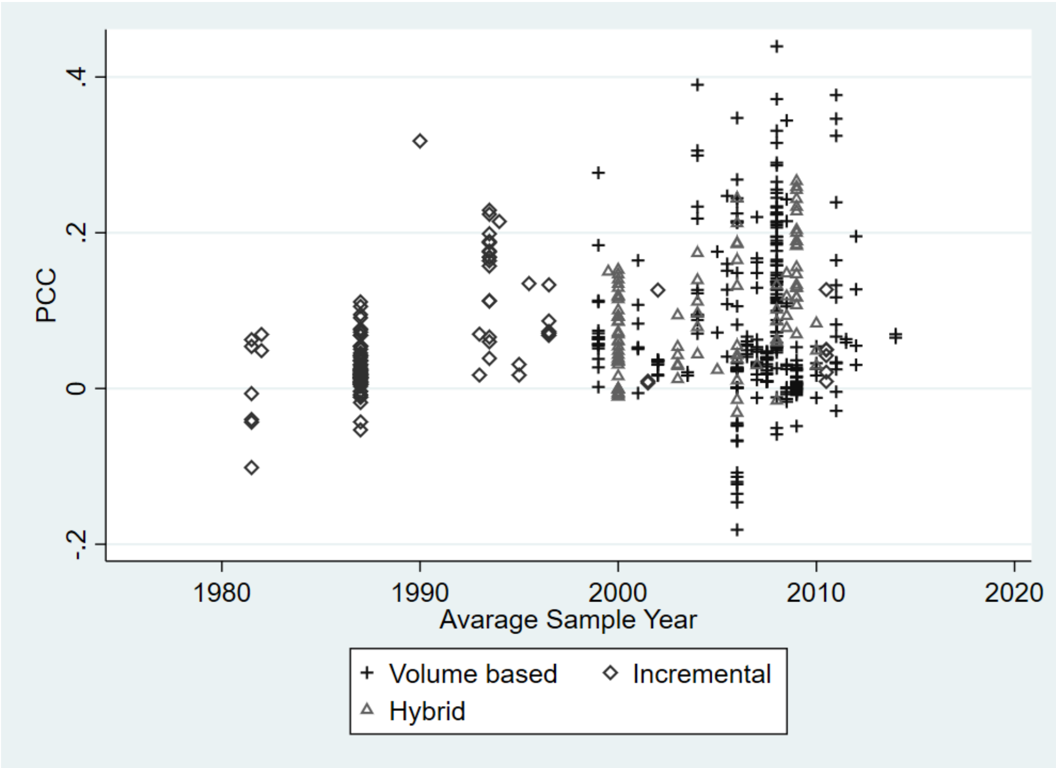

Figure 2: Evolution of tax incentives for R&D

Note: This plot presents the association between the PCCs and the average sample year, not normalized. The cross indicates estimates for volume-based schemes, the rhombus indicates estimates for countries with incremental schemes, and the triangle indicates estimates for hybrid schemes.

The widespread adoption of fiscal instruments among countries has encouraged scientists to evaluate their effectiveness, whereby a large share of studies assess the *input* additionality of tax reliefs on private R&D spending at the firm-level.[4] Fig. 2 plots the estimates found in the literature to the corresponding average sample year. While estimates of countries with incremental schemes primarily refer to early periods of investigation (1980s and 1990s), newer studies relate to countries with volume-based or hybrid incentives. As volume-based regimes are less complex for both firms and administrations and, thus, reach a larger target group, most governments have switched to volume-based schemes (Organisation for Economic Co-operation and Development 2021). France had an incremental scheme until 2003, a hybrid scheme between 2004 and 2007, and a pure

[4]The assessment of the input additionality of R&D focuses on innovation input, such as R&D spending, while the assessment of the output additionality of R&D relates to innovation output, like patents and product innovation.

volume-based scheme from 2008 onward. The Irish scheme evolved into a volume-based one over the years 2012 to 2014. Australia adopted a volume-based tax credit in 2012. However, Italy replaced its volume-based tax credit by an incremental incentive in 2015 to reduce public spending, but re-introduced a new volume-based R&D tax credit in 2020. Table B.2 in Appendix B provides further descriptions of the R&D tax policies among countries.

An incremental tax incentive can minimize the windfall gains due to R&D that would have been conducted, even in the absence of tax support. Lokshin and Mohnen (2012) evaluate the Dutch volume-based tax credit scheme during 1996–2004 and calculate a deadweight loss of 85% of total tax revenue loss saving due to the removal of the tax support on R&D. On the contrary, Lester and Warda (2014) argue that both incremental and volume-based incentives exhibit similar cost-effectiveness, that is, the growth of R&D spending in response to the tax incentive relative to the forgone tax revenue. Thus, some countries are implementing hybrid schemes by combining volume-based elements with incremental ones. Besides maintaining R&D spending levels, hybrid schemes promote strong R&D growth (Criscuolo et al. 2009). Japan and Spain, for example, combine a volume-based tax credit with an incremental one for firms with high R&D expansions.

3 Meta-data set

3.1 Assessing the effectiveness

Two key approaches for measuring the association between R&D expenditures and tax incentives exist: The *structural* and the *direct* approach. The structural approach builds on the neoclassical investment model, in which the R&D capital stock is explained by its user costs, that is, the after-tax price for R&D (Hall and Jorgenson 1967).[5] As structural approaches often suffer from simultaneity, most current empirical studies shifted towards direct approaches. This field of literature assess the impact on R&D expenditures through a variable indicating whether a firm is eligible for the tax incentive scheme:

$$RD_{it} = \beta_0 + \beta_1 \cdot E_{it} + \beta_2 \cdot X_{it} + \epsilon_{it} \tag{1}$$

[5] R&D user costs incorporate the tax reliefs for R&D, other tax provisions, and interest, inflation, and depreciation rates (Hall and van Reenen 2000).

where, for firm i at year t, RD_{it} is either the (logarithm of) R&D expenditures or the R&D intensity, that is, R&D expenditures scaled by total assets, sales, or employees, E_{it} is either a dummy variable indicating eligibility for the tax scheme or the amount of tax benefits, and the vector X_{it} contains various firm-level controls. The main coefficient of interest, β_1, captures the increase in R&D expenditures when claiming the tax benefit for R&D. A simple comparison of R&D expenditures between eligible firms and non-eligible firms through an simple ordinary least squares (OLS) regression can amount to an average treatment effect plus selection bias, because firms that use tax reliefs are likely to differ systematically from firms that do not. Selection into the treatment group is therefore not random, but could be driven by various confounding factors. To address the selection problem, many studies exploit an exogenous policy intervention in the tax scheme affecting only one (treatment) group of firms by adding an indicator variable $Reform_{it}$ (= 1 after the policy change):

$$RD_{it} = \beta_0 + \beta_1 \cdot E_{it} + \beta_2 \cdot Reform_{it} + \beta_3 \cdot E_{it} \cdot Reform_{it} + \beta_4 \cdot X_{it} + \epsilon_{it} \tag{2}$$

where the difference-in-differences (DiD) estimator is β_3 on the interaction variable that indicates investment in R&D due to the policy reform. A DiD estimate captures a causal treatment effect on R&D expenditures, assuming common trends between both groups (Angrist and Pischke 2009).

A widely applied nonparametric procedure is propensity score matching (PSM) which predicts a probability of receiving the treatment (eligibility) conditional on observable firm characteristics. Then, group identification is done by matching recipients and non-recipients with similar probability values. The parameter of interest is the average treatment effect on the treated (ATT), that is, the average difference in R&D expenditures between both groups. Besides the assumption of conditional independence, a causal interpretation requires that the difference between recipients and non-recipients manifest in observable variables (Pearl 2009).

3.2 Data collection

This paper follows the guidelines of Havránek et al. (2020) for conducting meta-analyses in economics. I primarily used the IDEAS database and Google Scholar to locate appropriate studies.[6] I employed the following keywords and combinations of them in the search process: *research and development*, *R&D*, *tax credit*, *tax incentive*, and *additionality*. Furthermore, I scanned the literature reviews and reference lists of identified studies. The final meta-sample is determined by two main selection criteria: First, the study must report results from assessing the impact of tax incentives on R&D expenditures or R&D intensity using the direct approach. Second, the study must provide standard errors or t-statistics, and the number of observations to derive PPCs. I just consider the latest version of a study to avoid autocorrelation among estimates. A detailed description of the data collection process is provided in Appendix A. Figure A.1 presents a PRISMA flow chart which illustrates the literature selection steps (Moher et al. 2009) and Table A.1 provides a list of the selected primary studies. In sum, 37 studies match the selection criteria. Since selecting a single estimate per primary study is quite subjective and results in less heterogeneity among estimates, I include multiple estimates from each primary study, as long as there is a considerable difference between the variables, estimation methods, model specifications, or samples.

3.3 Standardization procedure

MRA results are only meaningful if the estimates are comparable across primary studies (Stanley 2001). As measurement units (e.g., R&D expenditures, R&D growth, or R&D intensity), functional forms (e.g., log-log, log-level, or level-level), and estimation strategies (e.g., OLS, DiD, Matching) differ across effect estimates, I consistently transform a selected estimate into a partial correlation coefficient (PCC), as follows (Stanley and Doucouliagos 2012):

$$PCC_{is} = \frac{t_{is}}{\sqrt{t_{is}^2 + df_{is}}} \tag{3}$$

where t_{is} is the t-statistic of regression i of primary study s, and df_{is} is the regression's degrees of freedom. The corresponding standard error of the PCC is computed as $SEPCC_{is} = \sqrt{(1 - PCC_{is}^2)/df_{is}}$. The PCC captures the direction and significance level of the association between R&D spending and tax incentives. Since the PCC is more of a statistical measure, I use the initial (treatment) effect estimates as an alternative

[6] I completed the search process in December 2020.

Table 1: Distribution of estimates

	N	Mean	Median	Min	Max	Std. dev.
PCC (unweighted)	496	0.077	0.051	-0.120	0.347	0.088
PCC (weighted)	496	0.040	0.022	-0.120	0.347	0.056

Note: This table provides an overview of the distribution of the PCCs. The PCCs and corresponding standard errors are winsorized at the first and 95th percentiles.

dependent variable.

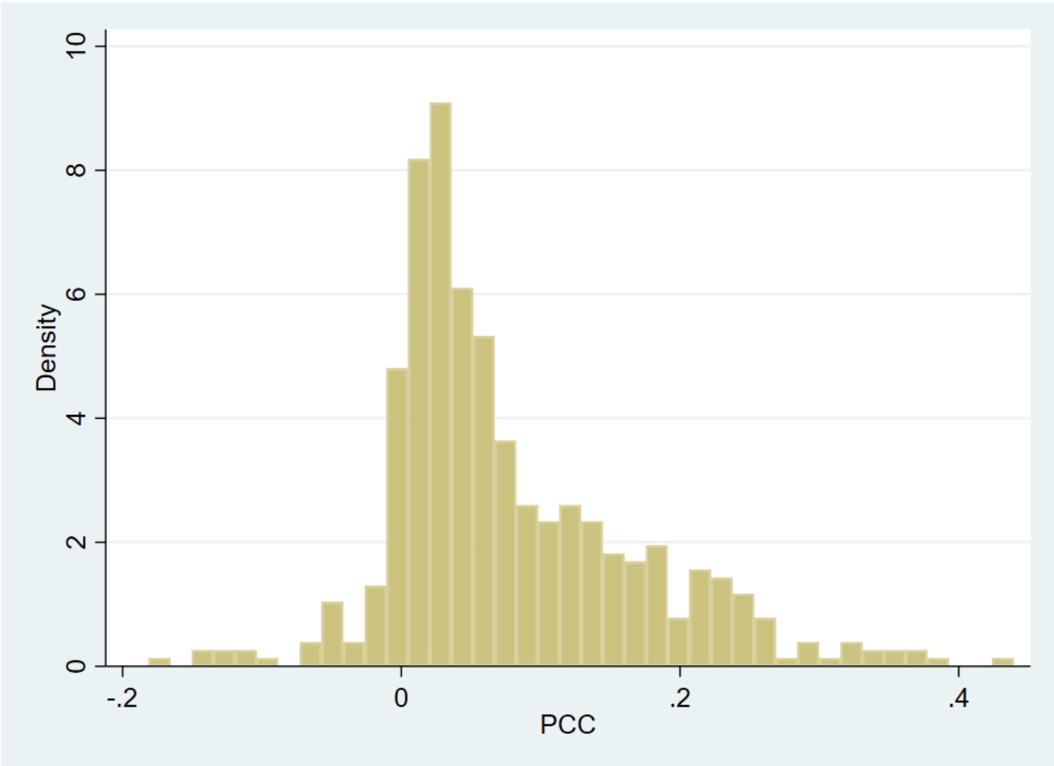

Figure 3: Density of the PPCs

Table 1 provides an overview of the distribution of the PCCs. PCCs and corresponding standard errors are winsorized at the first and 99th percentiles to remove the impact of outliers.[7] The total meta-sample comprises 496 estimates. The values of the PCCs vary far above -1.0 and below +1.0; the unweighted average is 0.077, with a standard deviation of 0.088. As the mean value exceeds 0.076, the total meta-sample exhibits a *large* correlation between private R&D expenditures and tax incentives, on average (see Doucouliagos 2011, Table 4, p. 14). The histogram in Fig. 3 illustrates a right-skewed distribution due to mostly positive values. When accounting for the precision of esti-

[7]The results are robust to not winsorizing estimates, see Tables C.5 and C.7 in Appendix C.

mates by weighting with the inverse variances, the average PCC diminishes to 0.040, with a standard deviation of 0.056.

4 Meta-analysis

4.1 Why reported estimates vary

The underlying theory or common practice—such as the reporting guidelines of Havránek et al. (2020) for conducting meta-analyses in economics—determine potential sources of heterogeneity among estimates. I code several variables, affecting the magnitude of the effect regarding the issue of publication bias, the definition of the outcome and explanatory variables, methodological choices (data and estimation strategy), and the design aspects of tax regimes for R&D among countries. Table 2 provides summary statistics for the full set of coded variables, along with descriptions. The depicted mean values of the variables can be interpreted in percentage terms.

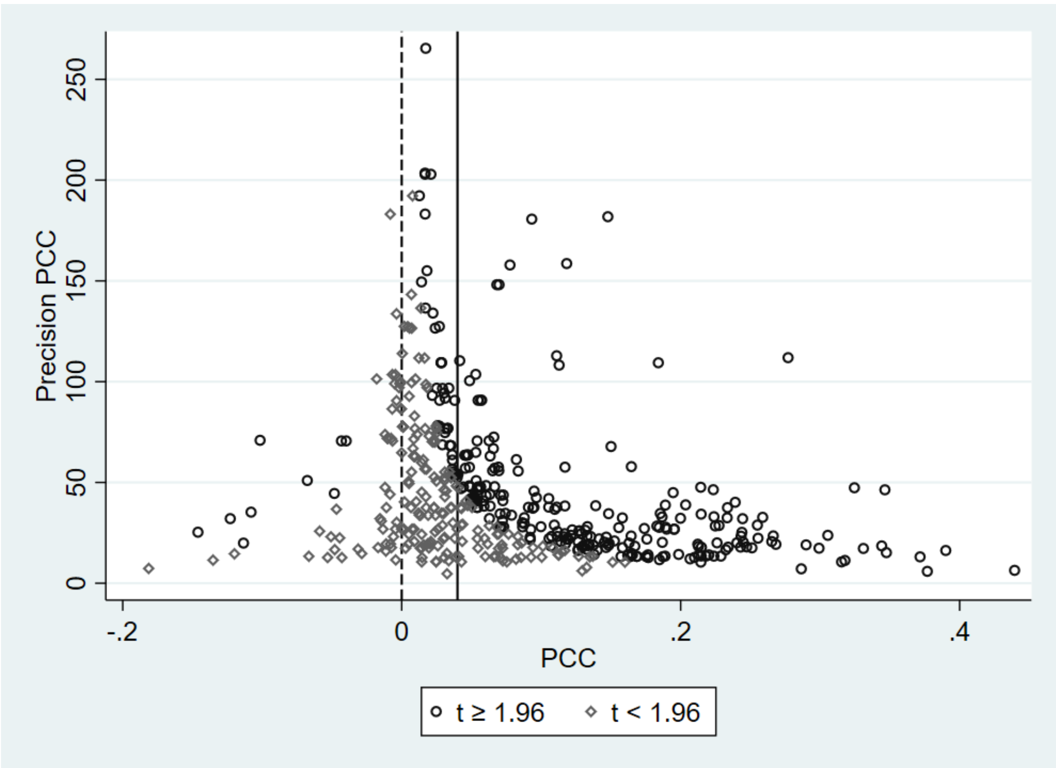

Figure 4: Funnel plot of the PPCs

Note: This funnel plot maps the PCC ($N = 496$) against its inverse of the standard error. Black circles indicate significant estimates at least at the 5% level ($t \geq 1.96$). Grey rhombs indicate nonsignificant ones ($t < 1.96$). The dotted vertical line marks the weighted *true* average (0.00). The solid vertical line marks the weighted average (0.040).

Publication bias. A publication bias arises if statistically nonsignificant or supposedly counterintuitive estimates are not published in a journal or do not even appear in a working paper due to certain preferences of authors, editors, and referees. To visually test for the presence of publication bias, the funnel plot in Fig. 4 maps the PCCs to their precision, that is, the inverse of the standard error (Egger et al. 1997). In absence of publication bias, the estimates should spread randomly around the average *true* effect (in my case a zero effect, see Table 5, column (2)). The plot shows a right tail missing the left side, since most estimates vary between zero and 0.2, while only a few studies report negative estimates. Moreover, 283 estimates are statistically significant at least at the 5% level ($t \geq 1.96$), while 213 PCCs are not significant. The peak is composed of the most precise estimates, scattered around zero and 0.03. I address the issue of publication bias more formally by including the standard error of the corresponding estimate as an explanatory variable. The funnel asymmetry test (FAT) for the coefficient on the standard error detects the presence of publication bias (Egger et al. 1997).[8]

Publication status. I code whether an estimate is published in a peer-reviewed journal (*Journal*, 64%), in a working paper (*WP*, 24%), or in a policy report (*Report*, 12%).

Outcome variable. Since the distribution of firms' private R&D expenditures is largely skewed, studies use the logarithm of R&D spending as the dependent variable, while often limiting the sample to observations with strictly positive values for R&D expenditures (*Log*, 57%). However, explicitly omitting observations with zero values can induce endogenous selection bias. Therefore, other studies use the R&D intensity as the outcome variable, that is, R&D expenditures scaled by assets, sales, or employees (*RDInt*, 36%).

Explanatory variable. As explained above, studies either use a dummy variable indicating eligibility for the tax scheme or the amount of tax benefits. I account for the difference by coding the variables *Dummy* (83%) and *Amount* (17%).

[8] A correlation between the PCC and the standard error appears to be due to publication bias, because the authors could be searching for the expected sign or statistical significance by testing various estimation methods or model specifications under the given conditions, such as a small sample size, resulting in increased (decreased) values of estimates (standard errors).

Table 2: Summary statistics of variables

Variables	Description	Summary statistics (N = 496)	
		Mean	Std. dev.
Publication bias			
SE	PCC standard error, winsorized at the first and 99th percentiles	0.036	0.026
Publication status			
Journal	=1 if published in a journal, and 0 otherwise	0.641	0.480
*WP**	=1 if published in a working paper, and 0 otherwise	0.236	0.425
*Report**	=1 if published in a policy report, and 0 otherwise	0.123	0.329
Outcome variable			
*Level**	=1 if the level of the outcome variable is used, and 0 otherwise	0.427	0.495
Log	=1 if the logarithm of the outcome variable is used, and 0 otherwise	0.573	0.495
*RDExp**	=1 if R&D expenditures as the outcome variable is used, and 0 otherwise	0.635	0.482
RDInt	=1 if R&D intensity as the outcome variable is used, and 0 otherwise	0.359	0.480
*RDGrow**	=1 if R&D growth as the outcome variable is used, and 0 otherwise	0.006	0.078
Explanatory variable			
*Dummy**	=1 if a dummy as the explanatory variable is used, and 0 otherwise	0.827	0.379
Amount	=1 if an amount as the explanatory variable is used, and 0 otherwise	0.173	0.379
Data			
*Panel**	=1 if panel data are used, and 0 otherwise	0.742	0.438
Cross	=1 if cross-sectional data are used, and 0 otherwise	0.258	0.438

continues on next page

*Admin**	=1 if administrative data are used, and 0 otherwise	0.726	0.447
Accounting	=1 if accounting data are used, and 0 otherwise	0.274	0.447
Average sample year	Average sample year, normalized between 0 and 1	0.637	0.252
Estimation strategy			
OLS	=1 if the OLS regression is used, and 0 otherwise	0.137	0.344
*FE**	=1 if the FE estimator is used, and 0 otherwise	0.093	0.290
Match	=1 if the Matching approach is used, and 0 otherwise	0.429	0.495
DiD	=1 if the DiD approach is used, and 0 otherwise	0.254	0.436
IV	=1 if the IV is used, and 0 otherwise	0.087	0.282
Tax policy			
*Credit**	=1 if a tax credit is implemented, and 0 otherwise	0.821	0.384
Deduction	=1 if a tax base deduction is implemented, and 0 otherwise	0.179	0.384
Vol	=1 if the implemented tax policy is volume based, and 0 otherwise	0.556	0.497
Hyb	=1 if the implemented tax policy is hybrid, and 0 otherwise	0.202	0.402
Inc	=1 if the implemented tax policy is incremental, and 0 otherwise	0.242	0.429
Firm characteristics			
All	=1 if a sample of all firms is used, and 0 otherwise	0.270	0.444
Manuf	=1 if a sample of manufacturing firms is used, and 0 otherwise	0.413	0.493
High	=1 if a sample of high-tech firms is used, and 0 otherwise	0.113	0.317
SME	=1 if a sample of SMEs is used, and 0 otherwise	0.099	0.299
Large	=1 if a sample of large firms is used, and 0 otherwise	0.210	0.407

Note: This table summarizes the full set of variables, along with a description and summary statistics for the PCCs. The mean of the variables can be interpreted in percentage terms. The superscript * marks the benchmark category of the respective study characteristics.

Data. I code whether a study used cross-sectional (*Cross*, 26%) instead of panel data (*Panel*, 74%). More recent studies often use administrative records, such as corporate tax return data to precisely infer which firms benefit or the amount at which firms received support (*Admin*, 73%). Not accounting for the actual support rate could underestimate the effectiveness of fiscal policies. Furthermore, to account for the underlying sample year, I code the variable *Average sample year*, normalized between zero and one, by setting the oldest average sample year (1981.5) to zero and the latest (2014) to one.

Estimation strategy. I further code dummy variables indicating the underlying estimation method. Since eligible firms could self-select into the treatment group, I expect a bias for simple OLS estimates (*OLS*, 14%). Studies correct for selection bias by using matching techniques (*Match*, 43%) or instrument variable (IV) estimations (*IV*, 9%).[9] Others apply a DiD identification strategy (*DiD*, 25%) or a fixed effects (FE) model (*FE*, 9%) to exploit changes over time within firms while controlling for unobserved but fixed heterogeneity between firms, such as managerial ability or attitude to risk.[10] A causal interpretation of the resulting estimates, however, depends heavily on whether the underlying key conditions are met (e.g., selection-on-observables, validity of IVs, common trends assumption). The common trends assumption implies that the growth in R&D expenditures would be the same between eligible and non-eligible firms in the absence of the treatment. An estimate is somewhat biased when time-varying confounding factors impact the group of user firms more strongly than non-user firms, or if spillover effects, such as competition or rising prices for R&D inputs affect R&D spending of firms not receiving the tax incentive.

Tax policy. The meta-data set shows a great geographical heterogeneity; it covers estimates for 17 countries over a vast time span of 40 years (1975 to 2016). I exploit these country and time variations in R&D policies to disentangle the incentive effects of the design attributes. I code whether a tax credit (*Credit*, 82%) or a tax allowance (*Deduction*, 18%) is implemented, and if its nature is volume based (*Vol*, 56%), incremental (*Inc*, 24%), or hybrid (*Hyb*, 20%). The information about the design features are taken from

[9] Generalized method of moments estimates, seemingly unrelated regression estimates, Heckman selection model estimates, and OLS random effects estimates are coded as *IV*. It should be noted that DiD can be combined with matching. I code the respective estimates as *Match*.

[10] A regression discontinuity (RD) design compares the difference in R&D expenditures between treated and untreated firms around the threshold to be eligible for the tax support. If a RD design or OLS balancing are used, I code the respective estimates as *FE*.

the primary studies as well as the OECD R&D tax incentive database (see Table B.2 in Appendix B for further policy descriptions).

An enhanced allowance allows firms to deduct more than their actual R&D expenditures from the taxable base rather than the tax liability to lower the after-tax cost of performing R&D. While the effect of tax credits on the after-tax price is not related to the tax rate, the impact of tax base deductions increases for higher marginal tax rates (see e.g., Elschner et al. 2011), rendering the actual tax advantage unpredictable in advance.

Furthermore, a tax scheme can be designed as a volume-based policy on the total amount of R&D expenditures, a incremental regime on the additional R&D spending relative to past expenditures, or a hybrid scheme by combining the previous two schemes. Assuming a positive link between firm responses and the generosity of a policy scheme, I expect a weaker effect for incremental than for volume-based schemes.[11] Moreover, these regimes tend to be more complex, resulting in higher compliance costs that can discourage firms from applying for such incentives at all if the (perceived) application costs exceed the expected tax benefits (Appelt et al. 2016). Unlike volume-based schemes, however, incremental ones promote R&D that, on average, would not have been conducted without a tax relief (Lokshin and Mohnen 2012). The disincentives of both fiscal measures can be mitigated by supplementing volume-based with incremental elements. Hybrid schemes could not only reduce business cycle effects and the degree of windfall gains, but could also prevent distortions in firm behavior, as firms do not have to increase their R&D expenditures to get access to the support regime. Whether hybrid regimes exhibit the incentives of both schemes is an empirical question.

Firm characteristics. Schumpeter's (1942) seminal work suggests the dominance of larger firms in the technological process through large-scale R&D in circumstances of greater market power, leading to better financial resources. Yet, studies find empirical evidence consistent with SMEs are less likely to innovate, since they face liquidity constraints (Czarnitzki and Hottenrott 2011). A tax incentive should provide additional cash flow when R&D is undertaken by reducing the firm's tax liability. The empirical findings on whether tax incentives are more effective for SMEs than large firms are mixed (e.g.,

[11] Elschner et al. (2011) find a stronger reduction in the firm's tax burden among European Union states with volume-based incentives than incremental schemes, which is particularly pronounced when stable R&D expenditures are assumed.

Koga 2003; Baghana and Mohnen 2009; Corchuelo and Martínez-Ros 2010; Kobayashi 2014; Sterlacchini and Venturini 2019). Without a refundability option of tax benefits, loss-making firms—e.g., innovative start-ups and SMEs—cannot exploit the reliefs, because they suffer from insufficient taxable income or tax liability (Stoneman 1991; Hall and van Reenen 2000; Elschner et al. 2011), which can explain the lower effectiveness for SMEs. On the other hand, limitations, such as upper ceilings and thresholds to eligible R&D expenditures, tax credits or allowances depress the incentive effect at the intensive margin among firms with high R&D intensity (Appelt et al. 2016). Therefore, I code variables for the estimates based on a sub-sample of SMEs (*SME*, 10%) and large firms (*Large*, 20%).

According to Castellacci and Lie (2015), the impact of tax credits varies among sectors, because firms across various industries differ regarding their innovation strategies and technological performance. Their results suggest, on average, smaller effects for high-tech firms. They conclude that low-tech firms react more strongly to R&D tax incentives, since these firms are more likely to be liquidity constrained due to lower technological and economic opportunities and less dynamic demand conditions. I categorized whether a effect estimate is related to a sub-sample of high-tech firms (*High*, 13%) or a sub-sample of manufacturing firms (*Manuf*, 41%).

4.2 Meta-regression model

My meta-analysis proceeds in two steps. First, I use MRA to explain the heterogeneity across estimates found. The meta-regression model is as follows:

$$PCC_{is} = \alpha + \beta \cdot X_{is} + \epsilon_{is} \tag{4}$$

where the outcome variable is the PCC of regression i of primary study s, X_{is} is a vector of moderator variables, α is the constant, and ϵ_{is} is the error term. To correct for heteroscedasticity, Eq. 4 is weighted by the inverse of the PCC's variance ($1/SE_{is}^2$), that is, weighted least squares (WLS) (Stanley and Doucouliagos 2012). Beyond correcting for heteroscedasticity, weighting by the inverse variance corrects for low-quality estimates, since imprecise coefficients are given less weight in the regression. I consider multiple estimates from each primary study in the meta-data set, which carries the risk of within-study dependency (i.e., autocorrelation). I allow for autocorrelation between the estimates per primary study due to unobserved study-level heterogeneity, and cluster standard errors at the study level (Stanley and Doucouliagos 2012).

4.3 MetaForest

Second, I use the nonparametric random forest algorithm MetaForest, by van Lissa (2017) to identify those variables contributing most to the model's predictive performance, that is, how well the model predicts the effect sizes of new studies.[12] To identify important variables, the random forest algorithm grows a number of decision trees (e.g., 1,000) using bootstrap samples of the initial meta-data set by separating the samples into two subgroups with most similar effect sizes; each sample split is based on a random subset of the moderators from the full set of variables, and the process of portioning is repeated until the subgroups contain a minimum number of cases. Then, the algorithm averages the predictions of the built decision trees and thus, is more robust to data over-fitting than MRA.

The application of a random forest to meta-analysis (MetaForest) applies a weighting scheme (uniform, fixed effects, or random effects weights) to the bootstrap sampling.[13] I use the R package `metaforest` (van Lissa 2020a) which comprises several tuning parameters, whose optimal values are chosen through k-fold cross-validation: the number of splitting variables, the minimum number of cases remaining in the subgroups, and the weighting scheme. A k-fold cross-validation involves (i) randomly portioning the data set into k equal portions, *k-1* training data sets, and a remaining testing (or out-of-bag) data set, (ii) fitting the model on the *k-1* training data sets, and (iii) validating the performance of the model on the testing data set by estimating the error rate, that is, the root mean squared error (RMSE); this process is repeated k times for all possible tuning parameter combinations. The combination of tuning parameters with the lowest averaged cross-validated error rate is chosen for the final model. I provide additional information on the random forest algorithm MetaForest in Appendix D.

[12] Unlike MRA, MetaForest calculates a predictive R-squared, showing how much variance the model would explain in *new* data.

[13] The meta-data set includes multiple estimates per primary study, which carries the risk of within-study dependency. The algorithm corrects for within-study dependency using clustered bootstrap sampling, meaning that multiple estimates are always included in the same portion of data (van Lissa 2020b).

5 Results

5.1 Sources of heterogeneity

Table 3 studies the degree to which PCCs vary with various methodological choices and tax policy designs by estimating Eq. 4. Instead of adding the full set of variables simultaneously, I include them step by step to address multicollinearity concerns. Column (1) contains the baseline model. Column (2) adds the variables for the applied estimation strategy. Columns (3) and (4) further include the variables *Deduction*, *Vol*, *Hyb*, and *Inc* to investigate whether these design features are associated with the reported estimates. Column (5) adds variables for the considered (sub-)sample of firms. To correct for the over- and underrepresentation of studies in the sample, column (6) uses the inverse of the number of estimates per study times the squared standard error as an alternative weighting factor.[14]

The positive and statistically significant estimates on the variable *SE* are consistent with the presence of publication bias in favor of a R&D-enhancing effect of tax incentives. It should be noted that the value of the constant in Table 3 is conditioned on the included variables when they take zero values (i.e., they depend on the reference categories) and thus cannot be interpreted as an average effect. The PCC captures the strength of the association in terms of statistical significance, but do not allow for an interpretation of the effect size. To account for this, I use the initial (treatment) effect estimate as an alternative dependent variable in Section 5.4.

Regarding the methodological choices, I obtain the following: The used outcome and explanatory variables as well as data characteristics do not explain the heterogeneity among estimates. Not accounting for endogeneity seems to lead to a weaker correlation, since the estimates on the variable *OLS* are significantly negative in some columns. Likewise, using DiD and IV produces, on average, less significant results.

[14]Table C.5 in Appendix C checks the sensitivity of the MRA results to changes in the meta-sample composition. My results stay stable when I remove Ho's (2006) as well as Bodas Freitas et al.'s (2017) estimates.

Table 3: Sources of heterogeneity

Variables	Baseline	+ Method	+ Policy	+ Policy	+ Sample	# · SEsq
	(1)	(2)	(3)	(4)	(5)	(6)
Publication bias						
SE	1.280***	0.752	1.105***	0.819*	1.301***	1.900***
	(0.389)	(0.457)	(0.378)	(0.418)	(0.425)	(0.313)
Publication status						
Journal	-0.014	0.003	-0.006	-0.003	0.001	-0.001
	(0.017)	(0.013)	(0.010)	(0.010)	(0.008)	(0.009)
Outcome variable						
Log	0.018	0.013	0.013	0.013	0.012	0.035
	(0.017)	(0.020)	(0.020)	(0.021)	(0.020)	(0.023)
RDInt	-0.002	-0.008	-0.006	-0.007	-0.004	0.014
	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)	(0.018)
Explanatory variable						
Amount	-0.020	-0.041	-0.029	-0.037	-0.027	-0.006
	(0.019)	(0.027)	(0.019)	(0.024)	(0.018)	(0.019)
Data						
Cross	0.059**	0.025	0.036	0.033	0.040	0.007
	(0.023)	(0.019)	(0.023)	(0.024)	(0.026)	(0.023)
Other	-0.011	-0.009	0.010	0.021	0.009	-0.003
	(0.022)	(0.015)	(0.032)	(0.024)	(0.029)	(0.023)
Estimation strategy						
OLS		-0.032	-0.039*	-0.032	-0.038*	-0.038*
		(0.026)	(0.019)	(0.025)	(0.020)	(0.020)

continues on next page

Match	-0.013	-0.029	-0.012	-0.029	-0.046
	(0.021)	(0.026)	(0.024)	(0.029)	(0.029)
DiD	-0.068***	-0.057***	-0.061***	-0.051**	-0.042
	(0.017)	(0.018)	(0.018)	(0.020)	(0.026)
IV	-0.041*	-0.055***	-0.033	-0.046***	-0.061**
	(0.023)	(0.018)	(0.021)	(0.015)	(0.023)
Tax policy					
Deduction		0.010	0.012	0.017	0.034*
		(0.015)	(0.016)	(0.019)	(0.018)
Vol		0.019		0.014	0.003
		(0.030)		(0.028)	(0.022)
Hyb		0.048*		0.047*	0.048**
		(0.028)		(0.026)	(0.022)
Inc			-0.030		
			(0.024)		
Firm characteristics					
All				0.002	-0.003
				(0.008)	(0.016)
Manuf				-0.022	-0.017
				(0.016)	(0.020)
High				-0.010	0.004
				(0.007)	(0.019)
SME				-0.004	-0.008
				(0.005)	(0.014)
Large				0.001	-0.013

continues on next page

					(0.006)	(0.014)
Constant	0.020	0.075**	0.042	0.066**	0.040	0.014
	(0.021)	(0.031)	(0.044)	(0.030)	(0.040)	(0.040)
Number of observations	496	496	496	496	496	496
Adj. R-squared	0.168	0.326	0.361	0.344	0.370	0.392
Variance inflation factor	1.45	1.96	3.04	2.63	3.19	3.08

Note: The dependent variable is the PCC, winsorized at the first and 99th percentiles. WLS with the inverse of the squared standard errors, winsorized at the first and 99th percentiles as weights, is used in columns (1)–(5). WLS with the inverse of the number of estimates per study times the squared standard errors, winsorized at the first and 99th percentiles as weights, is used in column (6). Standard errors are in parentheses and clustered at the study level. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively.

A further aim of the MRA is to assess the moderating effects of design aspects of tax incentives. Implementing tax allowances rather than tax credits does not yield statistically different results in most models. The parameter estimates on *Vol* and *Hyb* are both positive but only statistically significant on the latter moderator, while the coefficient on *Inc* is negative in column (4). Taken together, reported results for hybrid schemes are, on average, stronger correlated compared to volume-based ones, while the estimates in countries with incremental measures are weaker in the primary literature.[15]

I distinguish between firm size groups and industries by adding the moderators *All*, *High*, *High*, *SME* and *Large*. Interestingly, I find no evidence that the underlying sample of firms systematically drive the heterogeneity among estimates. This conclusion differs from the findings of Castellacci and Lie (2015), who find stronger effects of tax incentives for SMEs and smaller effects for high-tech firms. I further discuss this finding in Section 6.

5.2 MetaForest results

Next, I use the nonparametric machine learning algorithm MetaForest to estimate the relative importance of each moderator. I use the initial (treatment) effect estimates and include all variables as in Table 3, column (6), except for the variable *SE*. First, I apply *10*-fold cross-validation to select the optimal tuning parameters from the possible range of splitting moderators between 2 to 6, the minimum number of cases per subgroups between 2 to 6, and the three weighting schemes. The final model uses a set of 5 random candidate variables per split, a minimum of 4 cases per subgroup, and uniform weights. I construct 5,000 trees, since the model converges with approximately 5,000 trees (see Appendix Fig. D.2).

The algorithm MetaForest computes a permutation importance to assess the relative relevance of each moderator. This permutation importance reflects the strength of the relationship between the moderator and the outcome variable accounting for all linear, nonlinear, and interaction effects (van Lissa 2020b). Fig. 5 illustrates the MetaForest results for the importance of the variables. The moderators are ranked according to

[15] In a robustness check in Table C.5 in Appendix C, I account for unobserved heterogeneity by including study FE. As the meta-data set contains relatively few observations for some studies, using study FE runs the risk of over-fitting. Therefore, the results should be interpret with caution.

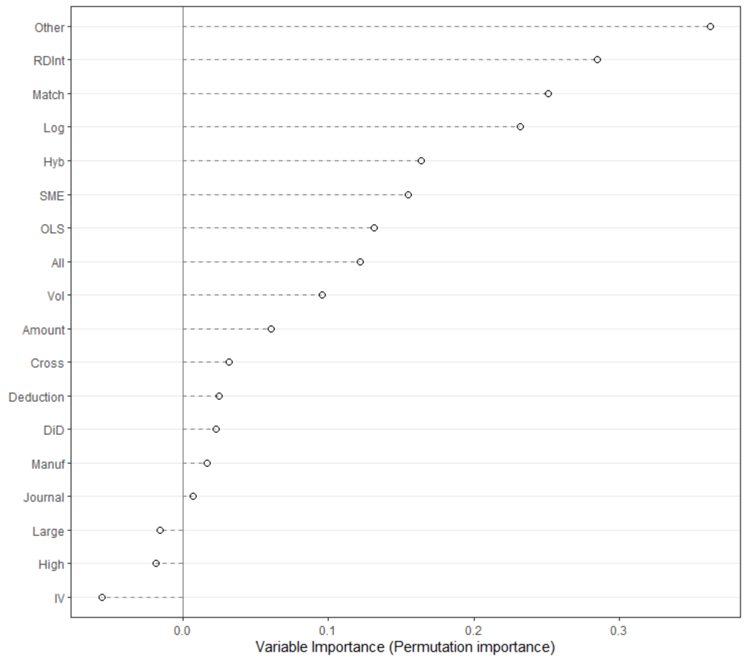

Figure 5: Variable importance plot

Note: This plot illustrates the MetaForest results for the importance of the moderator variables. The moderator variables are ranked according to their predictive power, from the highest, at the top, to the lowest, at the bottom. The dependent variable is the initial (treatment) effect estimate, winsorized at the first and 99th percentiles. The final model uses a set of 5 random candidate variables per split, a minimum of 4 cases per subgroup, and uniform weights.

their predictive power, from the highest, at the top, to the lowest, at the bottom. The algorithm determines a variable's importance by calculating the reduction of the model's performance, measured as the out-of-bag error rate, for a given moderator after randomly permuting the values of the moderator, averaged over all trees. When a moderator has no predictive power, randomly permuting coincidentally strengthens the relationship, and the variable importance score becomes negative. Fig. 5 shows that the variables on the design aspects of R&D tax incentives (e.g., *Vol*, *Hyb*, and *Deduction*) largely contribute to the overall prediction, while *Hyb* is more important than *Vol*. In contrast, the moderators on the industry and size classification have a low or even no predictive power, since the importance of the variables *Large* and *High* is negative.

5.3 R&D tax incentives over time

Table 4 investigates how PCCs depend on the evaluated time span. Column (1) tests whether the statistical relationship between R&D inputs and incentives is mitigated or pronounced over time by including the variable *Average sample year*.[16] The positive parameter estimate on the variable is consistent with fiscal policies have evolved towards more generous regimes over time, thereby raising the total volume of funding in relation to the gross domestic product (Organisation for Economic Co-operation and Development 2020). Next, I address that more recent empirical studies are more likely to be linked to countries with volume-based and hybrid support schemes rather than incremental ones, as observed in Fig. 2. Using an interaction term between the indicator variables *Vol*, *Hyb*, and *Inc*, as well as the variable *Average sample year*, I test if the computed estimate in Table 3 is driven by a time trend. I find a negative estimate on the interaction term in column (2) and a positive coefficient in column (3), both statistically significant. Interestingly, the association for hybrid schemes has become more prevalent over time, while the reverse is true for volume-based incentives. The PCCs for incremental schemes seem not related to the average sample year (column (4)), while this could be explained by estimates for incremental schemes are almost exclusively linked to earlier investigation periods. Taken together, this are important insights for interpreting prior empirical findings on R&D tax incentives.

5.4 Average effect

Table 5 tests for the existence of an average *true* effect of tax incentives on private R&D expenditures. Column (1) contains the weighted average PCC tested against zero, which serves as a reference point. Column (2) presents the result for the funnel asymmetry test precision effect test (FAT–PET) (Stanley 2008). Recall that the FAT for the coefficient on the standard error detects the presence of publication bias ($H_0 : \beta = 0$). Within this framework, the precision effect test (PET) on the constant tests whether an average *true* effect beyond publication bias exists ($H_0 : \alpha = 0$).

I find a *medium* to *large* statistical relationship in column (1). However, I cannot provide empirical evidence of a true effect, net of publication bias in column (2).[17] I test whether the extent of publication bias is associated with a study's publication status in

[16] Table C.6 in Appendix C uses the dummy variable *Post2000* which indicates PCCs with an average sample year after 2000. This robustness test fully confirms the main results.

[17] In Table C.7 in Appendix C, I provide evidence that their exist a true effect beyond publication bias when I use different estimation methods.

Table 4: R&D tax incentives over time

Variables	Year	Vol	Hyb	Inc
	(1)	(2)	(3)	(4)
Average sample year	0.061	0.131**	0.012	-0.002
	(0.041)	(0.057)	(0.054)	(0.076)
Vol		0.134		
		(0.082)		
Hyb			-0.157**	
			(0.072)	
Inc				-0.081
				(0.080)
Average sample year · Vol		-0.213*		
		(0.110)		
Average sample year · Hyb			0.261**	
			(0.112)	
Average sample year · Inc				0.129
				(0.119)
Constant	0.031	-0.016	0.053	0.073*
	(0.035)	(0.052)	(0.039)	(0.042)
Variables included	✓	✓	✓	✓
Number of observations	496	496	496	496
Adj. R-squared	0.348	0.391	0.387	0.355
Variance inflation factor	2.40	13.42	13.69	7.83

Note: The dependent variable is the PCC, winsorized at the first and 99th percentiles. WLS with the inverse of the squared standard errors, winsorized at the first and 99th percentiles as weights, is used. All variables as in Table 3, column (2) are included separately, but the coefficients are not reported. Standard errors are in parentheses and clustered at the study level. ** and * indicate significance levels of 0.01 and 0.05, respectively.

column (3). The estimate on the interaction term suggests that the publication bias is more pronounced among estimates published in peer-reviewed journals.

Table 5: Average effect

Variables	Avg.	FAT–PET	Journal	ATT	DiD	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
SE (publication bias)		1.631***	-0.084	0.985	1.984***	1.470
		(0.513)	(0.768)	(1.521)	(0.129)	(1.322)
$SE \cdot Journal$			2.714***			
			(0.834)			
Constant (average effect)	0.040***	0.018	0.038*	0.361***	0.004	0.490***
	(0.010)	(0.014)	(0.021)	(0.036)	(0.009)	(0.098)
Number of observations	496	496	496	53	91	35
Adj. R-squared	0.000	0.091	0.139	0.064	0.593	0.024

Note: The dependent variables are the PCC, winsorized at the first and 99th percentiles in columns (1)–(3), and the initial (treatment) effect estimate, winsorized at the first and 99th percentiles in columns (4)–(6). WLS with the inverse of the squared standard errors, winsorized at the first and 99th percentiles as weights, is used. *Journal* is included separately, but the coefficient is not reported. Standard errors are in parentheses and clustered at the study level. *** and * indicate significance levels of 0.01 and 0.05, respectively.

Since the PCC is more of a statistical measure, columns (4)–(6) use the initial (treatment) effect estimates as the dependent variable, allowing me to interpret the effect size. A complication of using the initial effect is that estimates are not directly comparable if the unit of measurement varies. I therefore assess the magnitude of the effect for more homogeneous subsets of estimates by focusing on semi-elasticity estimates, that is, studies using the logarithm of R&D expenditures as the dependent variable and an eligibility dummy as the explanatory variable. I include studies reporting ATTs in column (4), column (5) uses DiD estimates, and column (6) focuses on studies using OLS (FE) estimations. Studies using nonparametric matching techniques and OLS do not suffer from publication selectivity. The average effect for OLS estimates (0.490, column (6)) is relatively large in comparison to the two other effects, confirming that selection bias lead to larger estimates in the literature. The subset of estimates using DiD are, on average, much lower in absolute terms (0.004, column (6)), however, not statistically significant. The average true ATT is positive and equals 0.361 in column (4), indicating that the recipients of tax incentives appear to have, on average, 36% higher R&D expenditures than non-recipients. However, it should be noted that even though studies using matching and DiD claim to control for selection bias, the estimates cannot be automatically interpreted causally, since the underlying key assumptions have to be met (e.g., selection-on-observables and common trends assumption).

6 Concluding remarks

Yet, R&D tax incentives are a major policy tool for stimulating firms' private R&D expenditures. Despite the extensive literature on their effectiveness, little is known about the moderating effects of the design attributes. This paper applies meta-analytical techniques to a comprehensive meta-data set with 496 estimates from 37 studies exploiting direct approaches. While previous MRA of Castellacci and Lie (2015) and Gaillard-Ladinska et al. (2019) focus on heterogeneous effects among industrial sectors and firm size classes, this paper contributes to the literature by assessing how the fiscal policies among countries shape the reported results. Unlike former studies, beyond MRA, I additionally use the random forest algorithm MetaForest, by van Lissa (2017) to rank moderators according to their predictive power.

The findings contrast to previous MRA on this topic: I cannot confirm the role played by industries and firm size groups in explaining the heterogeneity found in the literature. In contrast, I observe robust evidence that differences in design features drive the variations in estimates. MetaForest supports this conclusion, suggesting that the differing tax incentives for R&D are more important moderators than firm characteristics. I find, on average, a stronger statistical association between R&D expenditures and tax incentives for countries with hybrid scheme, while this effect is pronounced over time. In contrast, volume-based regimes show a decreasing trend. While Castellacci and Lie (2015) use literature for a time span between 1975–2005, my meta-sample consists of estimates covering the years 1975–2016, with a median average sample year of 2006. Among OECD countries, the availability, simplicity of use, and the generosity of R&D tax incentives have been improved to enhance the stimulus effect (Appelt et al. 2016), resulting in a higher degree of diversity in the existing schemes across countries. Against this background, the observed difference in results is not surprising.

Contrary to prior MRA, I convert the initial coefficient into a PCC to ensure comparability between the estimates. I find publication bias in favor of reporting positive results of tax reliefs. This bias is more prevalent among studies published in journals. Since the PCC is more of a statistical measure, I assess the average effect across more comparable subsets of estimates, allowing me to interpret the association economically. In line with existing empirical evidence, the average true effect of tax incentives on R&D expenditures seems to be positive, but varies considerably from 0.004 to 0.490 with respect to the applied estimation method.

Appendices

A Included primary studies

A.1 Data collection

The inclusion criteria are as follows:

Association between R&D expenditures and tax incentives: A study must use R&D expenditures, R&D growth, or R&D intensity (= R&D expenditures scaled by total assets, sales, or employees) as the outcome variable (not components of R&D expenditures, such as in-house and outsourcing expenditures). A study must measure the association between R&D expenditures and tax incentives (not public grants or subsidies) through a dummy variable or the amount of tax savings/benefits.

Data: A study must use firm-level data for a single country.

Relevant information: A study must provide a correlation coefficient, the corresponding standard error or t-statistic, and the number of observations.

Language: A study must be written in English.

The exclusion criteria are as follows:

Method: All estimates for negative binomial regressions and limited information maximum likelihood are excluded.

Specification: All estimates that stem from interaction terms and lags of the dependent variable are excluded. All estimates for alternative control and treatment group definitions are excluded. All estimates for winsorized dependent variables are excluded. All estimates for samples including/excluding outliers are excluded.

Sample: All estimates for sub-samples beyond the scope of the paper are excluded, such as by region and skill intensity, for the culture industry, or for strategic groups of pharmaceutical firms etc.

Online Appendix: All estimates of the Online Appendix are excluded.

A.2 Standardization procedure

The calculation of PCCs requires a estimate's t-statistic, the number of observations, and the degrees of freedom. I compute t-statistics by dividing the reported regression coefficient by its standard error ($t_{is} = e_{is}/SE_{is}$). I calculate the degrees of freedom by subtracting the number of considered variables in a regression (e.g., controls as well as firm or industry dummies) from the number of observations minus one. I use the number of firms of Guceri (2018) for Guceri and Liu (2019). For Romero-Jordan et al. (2014), I approximate the number of observations by multiplying the number of firms and the number of sample years.

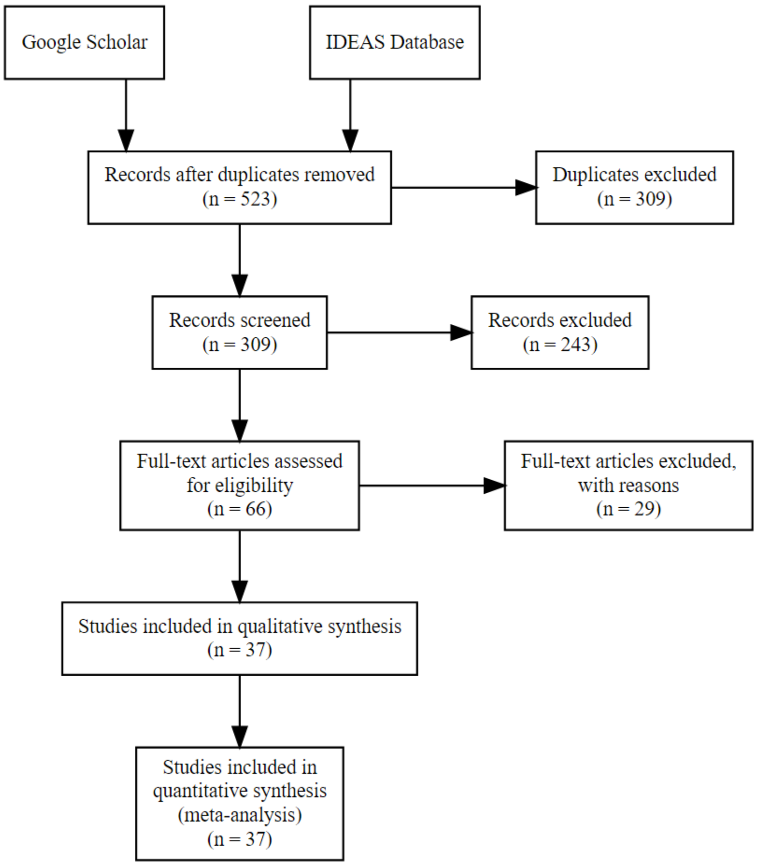

Figure A.1: PRISMA flow chart

Note: This PRISMA flow chart illustrates the selection steps of the literature (Moher et al. 2009).

Table A.1: Included primary studies

#	Authors and publication year	Country	Period
1	Acconcia and Cantabene (2018)	Italy	2008–2009
2	Acheson and Malone (2020)	Irland	2007–2014
3	Agrawal et al. (2020)	Canada	2000–2007
4	Álvarez-Ayuso et al. (2018)	Spain	1990–2009
5	Aralica and Botrić (2013)	Croatia	2006–2008
6	Aristei et al. (2015)	France, Spain	2007–2009
7	Balsmeier et al. (2018)	US	1977–1997, 1980–2006
8	Berger (1993)	US	1975–1989
9	Billings and Fried (1999)	US	1994
10	Billings et al. (2001)	US	1992–1998
11	Bodas Freitas et al. (2017)	France, Italy, Norway	2004, 2006, 2008, 2004–2006, 2004–2006, 2006–2008
12	Bøler et al. (2015)	Norway	1997–2005
13	Bozio et al. (2014)	France	2004–2010, 2005–2010, 2006–2009
14	Calderón-Madrid (2011)	Mexico	2004–2007
15	Cantabene and Nascia (2014)	Italy	2007–2009
16	Chang (2018)	US	1991–2006
17	Chen and Li (2018)	Taiwan	2006–2014
18	Chen and Yang (2019)	China	2010–2012
19	Corchuelo and Martínez-Ros (2010)	Spain	1998–2002
20	Dechezleprêtre et al. (2020)	UK	2009, 2010, 2011, 2006–2011, 2009–2011
21	Dumont (2017)	Belgium	2003–2011
22	Dumont (2019)	Belgium	2003–2015
23	Guceri (2015)	UK	1998–2006
24	Guceri (2018)	UK	1999–2013, 2003–2012
25	Guceri and Liu (2019)	UK	2002–2011
26	Hægeland and Møen (2007)	Norway	1993–2005
27	Ho (2006)	US	1975–1999
28	Holt et al. (2016)	Australia	2005–2012, 2011–2012, 2012
29	Kasahara et al. (2014)	Japan	2000–2003
30	Kobayashi (2014)	Japan	2009
31	Mercer-Blackman (2008)	Columbia	2000–2002
32	Paff (2005)	US	1994–1999
33	Ravšelj and Aristovnik (2020)	Slovenia	2012–2016
34	Romero-Jordan et al. (2014)	Spain	1995–2005
35	Sterlacchini and Venturini (2019)	France, Italy, Spain, UK	2007–2009
36	Swenson (1992)	US	1975–1988
37	Yang et al. (2012)	Taiwan	2001–2005

Note: This table outlines the total meta-data set of primary studies, with an overview of the underlying country and sample period(s).

B Supplementary statistics

Table B.2: R&D tax policies among countries

Country	**Period**	**Instrument**	**Type**	**Share**
Australia	1985–2001	Tax allowance	Volume based	0%
	2002–2011	Tax allowance	Hybrid	0.8%
	2012–	Tax credit	Volume based	1.2%
Belgium	2005–	Tax allowance	Volume based	0.4%
	2007–	Tax credit	Volume based	5.9%
Canada	1994–	Tax credit	Volume based	0.6%
China	2008–	Tax allowance	Volume based	2.8%
Colombia	1995–2015	Tax allowance	Volume based	0.6%
	2016–	Tax credit	Volume based	0%
Croatia	2003–	Tax allowance	Volume based	0.6%
France	1983–2003	Tax credit	Incremental	0%
	2004–2007	Tax credit	Hybrid	4%
	2008–	Tax credit	Volume based	6.7%
Ireland	2004–2011	Tax credit	Incremental	1.2%
	2012–2014	Tax credit	Hybrid	0%
	2015–	Tax credit	Volume based	0%
Italy	2007–2009	Tax credit	Volume based	12.3%
	2015–2019	Tax credit	Incremental	0%
	2020–	Tax credit	Volume based	0%
Japan	1967–2002	Tax credit	Incremental	0.4%
	2003–	Tax credit	Hybrid	4.2%
Mexico	1998–2001	Tax credit	Incremental	0%
	2002–2008	Tax credit	Volume based	1.21%
	2017–	Tax credit	Incremental	0%
Norway	2002–	Tax credit	Volume based	10.7%
Slovenia	2006–	Tax allowance	Volume based	0.4%
Spain	1995–	Tax credit	Hybrid	9.3%
Taiwan	1991–	Tax credit	Hybrid	1.8%
UK	2000–	Tax allowance	Volume based	12.3%
US	1981–	Tax credit	Incremental	22.6%

Note: This table provides descriptions of the R&D tax policies among countries. The last column indicates the share of the observations in the meta-data set.

Table B.3: Descriptive statistics for the PCCs

#	Author and publication date	PCCs				
		N	Mean	Min	Max	Std. dev.
1	Acconcia and Cantabene (2018)	13	0.079	-0.017	0.344	0.118
2	Acheson and Malone (2020)	6	0.050	0.009	0.127	0.041
3	Agrawal et al. (2020)	3	0.018	0.017	0.021	0.002
4	Álvarez-Ayuso et al. (2018)	1	0.150	0.150	0.150	.
5	Aralica and Botrić (2013)	3	0.171	0.129	0.220	0.046
6	Aristei et al. (2015)	3	0.052	-0.016	0.110	0.063
7	Balsmeier et al. (2018)	4	0.049	0.017	0.070	0.025
8	Berger (1993)	2	0.059	0.049	0.069	0.015
9	Billings and Fried (1999)	1	0.214	0.214	0.214	.
10	Billings et al. (2001)	2	0.024	0.017	0.031	0.010
11	Bodas Freitas et al. (2017)	96	0.134	-0.120	0.347	0.123
12	Bøler et al. (2015)	4	0.087	-0.006	0.164	0.071
13	Bozio et al. (2014)	16	0.039	0.018	0.148	0.031
14	Calderón-Madrid (2011)	6	0.139	0.041	0.247	0.068
15	Cantabene and Nascia (2014)	12	0.042	-0.059	0.088	0.047
16	Chang (2018)	19	0.158	0.039	0.318	0.065
17	Chen and Li (2018)	3	0.054	0.029	0.084	0.027
18	Chen and Yang (2019)	14	0.134	-0.029	0.347	0.131
19	Corchuelo and Martínez-Ros (2010)	24	0.091	0.016	0.153	0.038
20	Dechezleprêtre et al. (2020)	9	0.026	-0.012	0.054	0.017
21	Dumont (2017)	3	0.010	-0.012	0.032	0.022
22	Dumont (2019)	28	0.009	-0.048	0.053	0.018
23	Guceri (2015)	10	0.031	0.017	0.037	0.008
24	Guceri (2018)	30	0.024	0.001	0.048	0.016
25	Guceri and Liu (2019)	8	0.054	0.041	0.067	0.009
26	Hægeland and Møen (2007)	17	0.081	0.002	0.277	0.064
27	Ho (2006)	72	0.024	-0.053	0.111	0.030
28	Holt et al. (2016)	10	0.097	0.030	0.195	0.050
29	Kasahara et al. (2014)	2	0.008	0.007	0.010	0.002
30	Kobayashi (2014)	21	0.178	0.070	0.266	0.057
31	Mercer-Blackman (2008)	3	0.052	0.051	0.053	0.001
32	Paff (2005)	6	0.084	0.068	0.133	0.025
33	Ravšelj and Aristovnik (2020)	2	0.067	0.065	0.070	0.003
34	Romero-Jordan et al. (2014)	15	0.014	-0.010	0.060	0.025
35	Sterlacchini and Venturini (2019)	16	0.128	0.059	0.165	0.027
36	Swenson (1992)	6	-0.012	-0.102	0.063	0.063
37	Yang et al. (2012)	6	0.044	0.012	0.094	0.028
	Total meta-sample	496	0.077	-0.120	0.347	0.088

Note: TThis table provides descriptive statistics for the PCCs ($N = 496$, winsorized at the first and 99th percentiles) of the primary studies. The estimates of Chen and Li (2018) are multiplied by (-1) (reduction of tax credit rate).

Table B.4: Correlation matrix of variables

Variables	*SE*	*Journal*	*Log*	*RDInt*	*Amount*	*Cross*	*Other*	*OLS*	*Match*	*DiD*	*IV*	*Deduction*	*Vol*	*Hyb*	*Inc*	*All*	*Manuf*
SE	1.000																
Journal	0.307	1.000															
Log	-0.315	-0.205	1.000														
RDInt	0.299	0.525	-0.713	1.000													
Amount	-0.015	-0.013	0.374	-0.298	1.000												
Cross	0.263	0.412	-0.487	0.558	-0.270	1.000											
Other	0.101	-0.256	0.138	-0.403	0.148	-0.352	1.000										
OLS	-0.117	-0.105	0.084	-0.017	0.344	-0.048	-0.008	1.000									
Match	0.432	0.199	-0.543	0.455	-0.397	0.522	-0.040	-0.346	1.000								
DiD	-0.368	-0.056	0.326	-0.224	-0.267	-0.344	-0.047	-0.233	-0.506	1.000							
IV	0.025	0.096	0.208	-0.201	0.559	-0.182	0.068	-0.123	-0.267	-0.180	1.000						
Deduction	-0.144	0.043	0.181	-0.175	0.008	-0.240	-0.123	-0.095	-0.310	0.306	-0.051	1.000					
Vol	-0.051	0.026	-0.008	0.186	-0.031	0.156	-0.561	0.108	-0.234	0.167	-0.186	0.375	1.000				
Hyb	0.005	0.313	-0.104	0.179	-0.018	0.152	-0.196	-0.127	0.325	-0.270	0.131	-0.183	-0.563	1.000			
Inc	0.054	-0.323	0.107	-0.383	0.052	-0.322	0.835	-0.006	-0.034	0.060	0.094	-0.264	-0.633	-0.284	1.000		
All	-0.122	-0.226	0.241	-0.247	0.213	-0.276	0.054	0.114	-0.299	-0.021	0.184	-0.036	0.105	-0.159	0.027	1.000	
Manuf	0.072	0.508	-0.318	0.507	-0.179	0.385	-0.433	-0.227	0.240	0.056	0.003	0.184	0.172	0.221	-0.407	-0.511	1.000
High	-0.021	-0.278	-0.052	-0.174	-0.163	-0.123	0.480	-0.013	0.025	0.114	-0.110	-0.167	-0.310	-0.132	0.483	-0.217	-0.299
SME	0.069	-0.119	-0.055	-0.022	-0.045	-0.149	0.084	-0.053	0.163	-0.100	-0.030	-0.067	-0.194	0.137	0.097	-0.201	-0.127
Large	-0.148	-0.121	0.185	-0.189	-0.131	-0.270	0.094	-0.177	-0.097	0.382	-0.071	0.392	-0.059	-0.086	0.148	-0.313	0.161

Note: This matrix shows the correlation of the unweighted variables for the total meta-sample. Table 2 provides the descriptions of the variables.

C Robustness checks

Table C.5: Sources of heterogeneity

Variables	Pref.	No winsor	Study FE	Ho excl.	BF excl.
	(1)	(2)	(3)	(4)	(5)
Publication bias					
SE	1.900***	1.967***	0.300	2.116***	1.965***
	(0.313)	(0.319)	(0.675)	(0.315)	(0.311)
Publication status					
Journal	-0.001	0.000	0.023	0.005	0.000
	(0.009)	(0.009)	(0.036)	(0.010)	(0.009)
Outcome variable					
Log	0.035	0.034	-0.012***	0.060**	0.035
	(0.023)	(0.024)	(0.001)	(0.027)	(0.023)
RDInt	0.014	0.013	-0.011	0.029	0.011
	(0.018)	(0.018)	(0.007)	(0.021)	(0.018)
Explanatory variable					
Amount	-0.006	-0.008	0.040	-0.011	-0.006
	(0.019)	(0.019)	(0.026)	(0.017)	(0.019)
Data					
Cross	0.007	0.006	0.006***	0.010	-0.004
	(0.023)	(0.022)	(0.002)	(0.022)	(0.024)
Other	-0.003	-0.003	0.063**	-0.005	-0.003
	(0.023)	(0.023)	(0.026)	(0.020)	(0.022)
Estimation strategy					
OLS	-0.038*	-0.036*	0.001	-0.044**	-0.038*
	(0.020)	(0.020)	(0.030)	(0.019)	(0.020)
Match	-0.046	-0.046	-0.047	-0.056*	-0.054*
	(0.029)	(0.028)	(0.030)	(0.029)	(0.028)
DiD	-0.042	-0.041	-0.072*	-0.057**	-0.042
	(0.026)	(0.026)	(0.036)	(0.024)	(0.026)
IV	-0.061**	-0.059**	-0.018	-0.063**	-0.061**
	(0.023)	(0.023)	(0.049)	(0.024)	(0.023)
Tax policy					
Deduction	0.034*	0.033*	0.031*	0.042***	0.036**
	(0.018)	(0.018)	(0.017)	(0.015)	(0.018)
Vol	0.003	0.005	0.123***	0.006	0.002
	(0.022)	(0.022)	(0.037)	(0.020)	(0.021)
Hyb	0.048**	0.051**	0.103***	0.048**	0.051**
	(0.022)	(0.023)	(0.026)	(0.022)	(0.022)
Firm characteristics					
All	-0.003	-0.004	0.007	0.002	-0.003
	(0.016)	(0.016)	(0.007)	(0.016)	(0.016)
Manuf	-0.017	-0.018	0.007	-0.014	-0.021
	(0.020)	(0.020)	(0.025)	(0.020)	(0.020)
High	0.004	0.004	0.008	-0.001	0.005
	(0.019)	(0.019)	(0.017)	(0.027)	(0.019)
SME	-0.008	-0.008	-0.007	-0.008	-0.008
	(0.014)	(0.014)	(0.004)	(0.015)	(0.014)
Large	-0.013	-0.013	0.019	-0.023	-0.012
	(0.014)	(0.014)	(0.012)	(0.016)	(0.014)
Constant	0.014	0.012	-0.098	-0.012	0.014

continues on next page

	(0.040)	(0.040)	(0.073)	(0.038)	(0.040)
Number of observations	496	496	496	424	400
Adj. R-squared	0.392	0.393	0.660	0.405	0.400
Variance inflation factor	3.08	3.20	51.27	3.31	3.08

Note: The dependent variable is the PCC, winsorized at the first and 99th percentiles. WLS with the inverse of the number of estimates per study times the squared standard errors, winsorized at the first and 99th percentiles as weights, is used. Column (1) contains the preferred model from Table 3, column (6); PCCs and standard errors, not winsorized are used in column (2); study fixed effects are included in column (3); estimates of Ho (2006) are excluded in column (4); and estimates of Bodas Freitas et al. (2017) are excluded in column (5). Standard errors are in parentheses and clustered at the study level. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively.

[!h]

Table C.6: Robustness: R&D tax incentives over time

Variables	Year	Vol	Hyb	Inc
	(1)	(2)	(3)	(4)
Post2000	0.012	0.067**	-0.016	-0.010
	(0.020)	(0.028)	(0.034)	(0.021)
Vol		0.060		
		(0.040)		
Hyb			-0.027	
			(0.034)	
Inc				-0.068
				(0.046)
Post2000 · *Vol*		-0.111**		
		(0.042)		
Post2000 · *Hyb*			0.094*	
			(0.054)	
Post2000 · *Inc*				0.079
				(0.052)
Constant	0.069**	0.028	0.071***	0.076***
	(0.030)	(0.040)	(0.024)	(0.025)
Variables included	✓	✓	✓	✓
Number of observations	496	496	496	496
Adj. R-squared	0.329	0.385	0.382	0.346
Variance inflation factor	2.22	5.60	4.01	4.14

Note: The dependent variable is the PCC, winsorized at the first and 99th percentiles. WLS with the inverse of the squared standard errors, winsorized at the first and 99th percentiles as weights, is used. All variables as in Table 3, column (2) are included separately, but the coefficients are not reported. Standard errors are in parentheses and clustered at the study level. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively.

Table C.7: Robustness: Average effect

Variables	Pref.	No winsor	Cluster	IV	# · SEsq
	(1)	(2)	(3)	(4)	(5)
SE (publication bias)	1.629***	1.639***	1.295***	1.262***	1.370**
	(0.523)	(0.499)	(0.202)	(0.301)	(0.527)
Constant (average effect)	0.018	0.018	0.030**	0.031***	0.026**
	(0.014)	(0.013)	(0.014)	(0.012)	(0.013)
Number of observations	496	496	496	496	496
Adj. R-squared	0.091	0.094	0.141	0.132	0.061

Note: The dependent variable is the PCC, winsorized at the first and 99th percentiles. WLS with the inverse of the squared standard errors, winsorized at the first and 99th percentiles as weights, is used. Column (1) contains the preferred model from Table 5, column (2); PCCs and standard errors, not winsorized, are used in column (2); the inverse of the square root of the number of observations as instrument for the standard errors is used in column (4); and WLS with the inverse of the number of estimates per study times the squared standard errors, winsorized at the first and 99th percentiles as weights, is used in column (5). Standard errors are in parentheses and clustered at the study level and at the study and country level in column (3). ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively.

D Random forest algorithm MetaForest

Random forest algorithm MetaForest. The algorithm MetaForest applies a weighting scheme to the bootstrap sampling, that is, uniform, fixed effects, or random effects weights (van Lissa 2017). If uniform weights are applied, studies have the same probability of being included in the bootstrap sample, if fixed effects weights are used, studies with smaller standard errors have a greater probability of being selected, and if random effects weights are applied, studies with smaller standard errors have a greater probability of being selected, but this diminishes with increasing heterogeneity between the studies.

Predictive performance of a model. Machine learning techniques aim to improve the model's predictive performance in new data. MetaForest uses the root mean squared error (RMSE) to evaluate the predictive performance of a model on the out-of-bag portion of data. The out-of-bag RMSE is calculated by the difference between the observed outcome in the training data and the predicted outcome in the testing data. MetaForest chooses the model with the smallest out-of-bag RMSE.

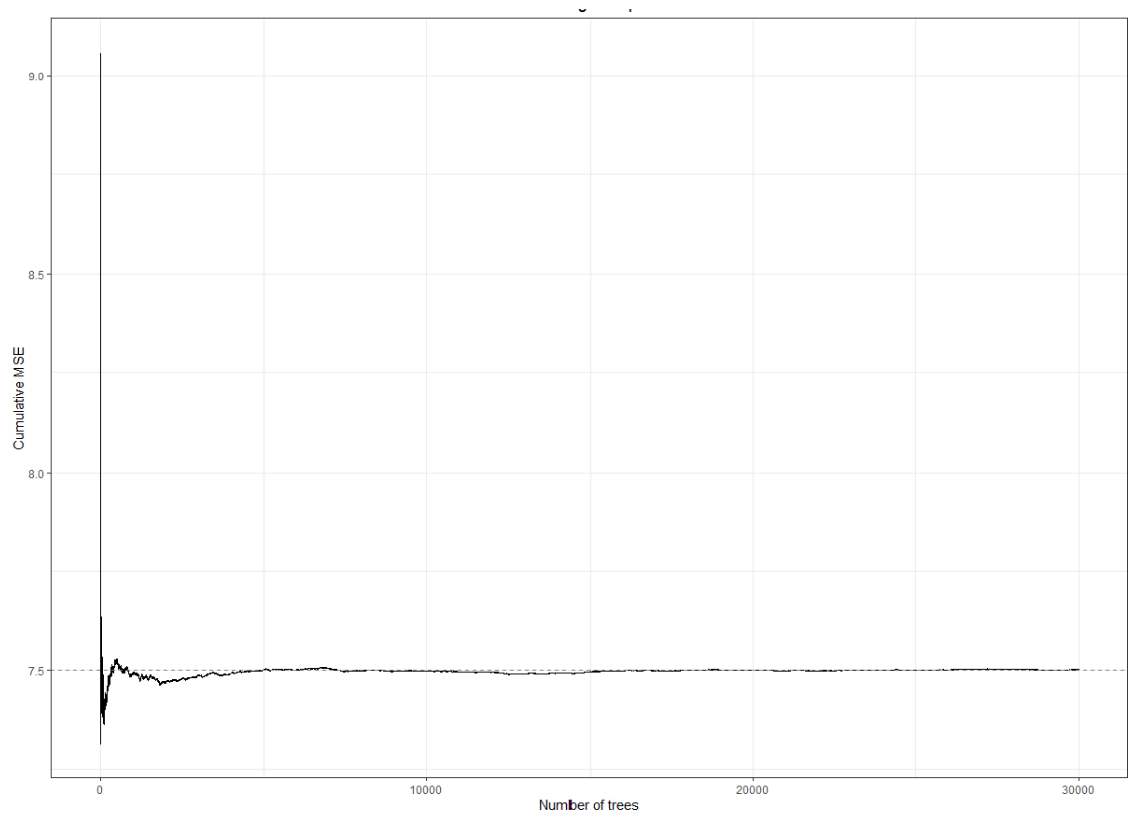

Figure D.2: Convergence plot

Note: This convergence plot depicts the convergence of the model with uniform weights by examining the cumulative mean squared out-of-bag prediction error (MSE) as a function of the number of trees in the model.

References

Acconcia, A. and Cantabene, C. (2018). Liquidity and firms' response to fiscal stimulus, *The Economic Journal* **128**(613): 1759–1785.

Acheson, J. and Malone, R. (2020). Respect your elders: Evidence from Ireland's R&D tax credit reform, *The Economic and Social Review* **51**(1): 105–131.

Agrawal, A., Rosell, C. and Simcoe, T. (2020). Tax credits and small firm R&D spending, *American Economic Journal: Economic Policy* **12**(2): 1–21.

Álvarez-Ayuso, I. C., Kao, C. and Romero-Jordán, D. (2018). Long run effect of public grants and tax credits on R&D investment: A non-stationary panel data approach, *Economic Modelling* **75**: 93–104.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Appelt, S., Bajgar, M., Criscuolo, C. and Galindo-Rueda, F. (2016). *R&D tax incentives: Evidence on design, incidence and impacts*, Organisation for Economic Co-operation and Development.

Aralica, Z. and Botrić, V. (2013). Evaluation of research and development tax incentives scheme in Croatia, *Economic Research-Ekonomska Istraživanja* **26**(3): 63–80.

Aristei, D., Sterlacchini, A. and Venturini, F. (2015). The effects of public supports on business R&D: Firm-level evidence across EU countries, number 64611, Munich Personal RePEc Archive.

Baghana, R. and Mohnen, P. (2009). Effectiveness of R&D tax incentives in small and large enterprises in Québec, *Small Business Economics* **33**(1): 91–107.

Balsmeier, B., Kurakina, M. and Fleming, L. (2018). R&D tax credits: Mechanisms of private and public value.

Berger, P. G. (1993). Explicit and implicit tax effects of the R&D tax credit, *Journal of Accounting Research* **31**(2): 131–171.

Billings, A., Glazunov, S. and Houston, M. (2001). The role of taxes in corporate research and development spending, *R&D Management* **31**(4): 465–477.

Billings, B. A. and Fried, Y. (1999). The effects of taxes and organizational variables on research and development intensity, *R&D Management* **29**(3): 289–302.

Bodas Freitas, I., Castellacci, F., Fontana, R., Malerba, F. and Vezzulli, A. (2017). Sectors and the additionality effects of R&D tax credits: A cross-country microeconometric analysis, *Research Policy* **46**(1): 57–72.

Bøler, E. A., Moxnes, A. and Ulltveit-Moe, K. H. (2015). R&D, international sourcing, and the joint impact on firm performance, *American Economic Review* **105**(12): 3704–3739.

Bonapersona, V., Kentrop, J., van Lissa, C., van der Veen, R., Joëls, M. and Sarabdjitsingh, R. (2019). The behavioral phenotype of early life adversity: A 3-level meta-analysis of rodent studies, *Neuroscience & Biobehavioral Reviews* **102**: 299–307.

Bozio, A., Irac, D., Py, L. et al. (2014). *Impact of research tax credit on R&D and innovation: Evidence from the 2008 French reform*, Banque de France.

Calderón-Madrid, A. (2011). A micro-econometric analysis of the impact of Mexico's R&D tax credit program on private R&D expenditure.

Cantabene, C. and Nascia, L. (2014). The race for R&D subsidies: Evaluating the effectiveness of tax credits in Italy, *Economia e Politica Industriale* **2014**(3): 133–158.

Castellacci, F. and Lie, C. M. (2015). Do the effects of R&D tax credits vary across industries? A meta-regression analysis, *Research Policy* **44**(4): 819–832.

Chang, A. C. (2018). Tax policy endogeneity: Evidence from R&D tax credits, *Economics of Innovation and New Technology* **27**(8): 809–833.

Chen, L. and Yang, W. (2019). R&D tax credits and firm innovation: Evidence from China, *Technological Forecasting and Social Change* **146**: 233–241.

Chen, M.-C. and Li, H.-Y. (2018). The effects and economic consequences of cutting R&D tax incentives, *China Journal of Accounting Research* **11**(4): 367–384.

Corchuelo, M. B. and Martínez-Ros, E. (2010). Who benefits from R&D tax policy?, *Cuadernos de Economía y Dirección de la Empresa* **13**(45): 145–170.

Criscuolo, C., Czarnitzki, D., Hambro, C. and Warda, J. (2009). *Design and Evaluation of Tax Incentives for Business Research and Development: Good Practice and Future Development*, Final report submitted by the Expert Group on Impacts of R&D Tax Incentives to the European Commission.

Curry, O. S., Rowland, L. A., van Lissa, C. J., Zlotowitz, S., McAlaney, J. and Whitehouse, H. (2018). Happy to help? A systematic review and meta-analysis of the effects of performing acts of kindness on the well-being of the actor, *Journal of Experimental Social Psychology* **76**: 320–329.

Czarnitzki, D. and Hottenrott, H. (2011). R&D investment and financing constraints of small and medium-sized firms, *Small Business Economics* **36**(1): 65–83.

Dechezleprêtre, A., Einiö, E., Martin, R., Nguyen, K.-T. and van Reenen, J. (2020). Do tax incentives increase firm innovation? An RD design for R&D.

Doucouliagos, H. C. (2011). *How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics*, Deakin University Faculty of Business and Law.

Dumont, M. (2017). Assessing the policy mix of public support to business R&D, *Research Policy* **46**(10): 1851–1862.

Dumont, M. (2019). *Tax incentives for business R&D in Belgium–Third evaluation*, Federal Planning Bureau.

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

Elschner, C., Ernst, C., Licht, G. and Spengel, C. (2011). What the design of an R&D tax incentive tells about its effectiveness: a simulation of R&D tax incentives in the European Union, *The Journal of technology transfer* **36**(3): 233–256.

Gaillard-Ladinska, E., Non, M. and Straathof, B. (2019). More R&D with tax incentives?: A meta-analysis, *International Financial Markets*, Routledge, pp. 316–335.

Guceri, I. (2015). Tax incentives and R&D: An evaluation of the 2002 UK reform using micro data, Oxford University Centre for Business Taxation.

Guceri, I. (2018). Will the real R&D employees please stand up? Effects of tax breaks on firm-level outcomes, *International Tax and Public Finance* **25**(1): 1–63.

Guceri, I. and Liu, L. (2019). Effectiveness of fiscal incentives for R&D: Quasi-experimental evidence, *American Economic Journal: Economic Policy* **11**(1): 266–91.

Hægeland, T. and Møen, J. (2007). *Input additionality in the Norwegian R&D tax credit scheme*, Statistisk Sentralbyrå.

Hall, B. and van Reenen, J. (2000). How effective are fiscal incentives for R&D? A review of the evidence, *Research policy* **29**(4-5): 449–469.

Hall, R. E. and Jorgenson, D. W. (1967). Tax policy and investment behavior, *The American Economic Review* **57**(3): 391–414.

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

Ho, Y. (2006). *Evaluating the effectiveness of state R&D tax credits*, PhD thesis, University of Pittsburgh.

Holt, J., Skali, A. and Thomson, R. (2016). *The additionality of R&D tax policy in Australia*, Centre for Transformative Innovation.

Kasahara, H., Shimotsu, K. and Suzuki, M. (2014). Does an R&D tax credit affect R&D expenditure? The Japanese R&D tax credit reform in 2003, *Journal of the Japanese and International Economies* **31**: 72–97.

Kobayashi, Y. (2014). Effect of R&D tax credits for SMEs in Japan: A microeconometric analysis focused on liquidity constraints, *Small Business Economics* **42**(2): 311–327.

Koga, T. (2003). Firm size and R&D tax incentives, *Technovation* **23**(7): 643–648.

Lester, J. and Warda, J. (2014). An international comparison of tax assistance for research and development: Estimates and policy implications, *The School of Public Policy Research Papers* **7**(36).

Lokshin, B. and Mohnen, P. (2012). How effective are level-based R&D tax credits? Evidence from the Netherlands, *Applied Economics* **44**(12): 1527–1538.

Mercer-Blackman, V. (2008). The impact of research and development tax incentives on Colombia's manufacturing sector: What difference do they make?, number 08/178.

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

Organisation for Economic Co-operation and Development (2020). Gross domestic spending on R&D (indicator), `https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm`. (Accessed December 22nd, 2021).

Organisation for Economic Co-operation and Development (2021). OECD R&D tax incentives database, 2021 edition, `https://www.oecd.org/sti/rd-tax-stats.htm`. (Accessed January 10th, 2022).

Paff, L. A. (2005). State-level R&D tax credits: A firm-level analysis, *The BE Journal of Economic Analysis & Policy* **5**(1).

Pearl, J. (2009). *Causality*, Cambridge University Press.

Ravšelj, D. and Aristovnik, A. (2020). The impact of public R&D subsidies and tax incentives on business R&D expenditures, *International Journal of Economics and Business Administration* **8**(1): 160–179.

Romero-Jordan, D., Delgado-Rodríguez, M. J., Alvarez-Ayuso, I. and de Lucas-Santos, S. (2014). Assessment of the public tools used to promote R&D investment in Spanish SMEs, *Small Business Economics* **43**(4): 959–976.

Schumpeter, J. A. (1942). *Socialism, capitalism and democracy*, Harper and Brothers.

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

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

Stanley, T. D. and Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*, Routledge.

Sterlacchini, A. and Venturini, F. (2019). R&D tax incentives in EU countries: Does the impact vary with firm size?, *Small Business Economics* **53**(3): 687–708.

Stoneman, P. (1991). The use of a levy/grant system as an alternative to tax based incentives to r&d, *Research Policy* **20**(3): 195–201.

Straathof, B., Gaillard-Ladinska, E., Kox, H., Mocking, R., Goldberg, I., Jensen, C., Lindholm, P., Sobolewksi, M., Berg-Andersson, B., Kaseva, H. et al. (2014). *A study on R&D tax incentives: Final report of CPB*, European Commission.

Swenson, C. W. (1992). Some tests of the incentive effects of the research and experimentation tax credit, *Journal of Public Economics* **49**(2): 203–218.

van Lissa, C. J. (2017). *MetaForest: Exploring heterogeneity in meta-analysis using random forests*, PsyArxiv.

van Lissa, C. J. (2020a). Package 'metaforest', `https://CRAN.R-project.org/package=metaforest`.

van Lissa, C. J. (2020b). Small sample meta-analyses: Exploring heterogeneity using metaforest, *Small Sample Size Solutions*, Routledge, pp. 186–202.

Yang, C.-H., Huang, C.-H. and Hou, T. C.-T. (2012). Tax incentives and R&D activity: Firm-level evidence from Taiwan, *Research Policy* **41**(9): 1578–1588.