Arbeitskreis Quantitative Steuerlehre
Quantitative Research in Taxation – Discussion Papers

Carla Pöschel

Incentive Effects of R&D Tax Incentives: A Meta-Analysis Focusing on R&D Tax Policy Designs

arqus Discussion Paper No. 243

August 2019
revised and renamed August 2020

www.arqus.info
ISSN 1861-8944

Incentive Effects of R&D Tax Incentives: A Meta-Analysis Focusing on R&D Tax Policy Designs

Carla Pöschel*

Freie University Berlin

August 2020

Abstract

Despite the growing literature on the effectiveness of research and development (R&D) tax incentives, little is known about the differing design aspects of the underlying tax policies. In this paper, I apply meta-regression analysis (MRA) to separate the distinct provisions through which various tax schemes affect firms' R&D expenditures. Using 192 estimates from 19 studies exploiting the direct approach, the results indicate, on average, greater input additionality effects of hybrid regimes in comparison to volume-based and incremental ones. MetaForest, a novel machine learning algorithm, confirms these results: the moderator for hybrid schemes is the most important variable in explaining the heterogeneity among estimates. Unlike previous MRA, I find only weak evidence for publication bias in this stream of literature. Overall, the relation between tax incentives and R&D expenditures is positive, on average, but the strength varies with methodological variations across studies.

Keywords: R&D, tax incentives, additionality effects, direct approach, meta-regression analysis, random forest

*Corresponding author contact e-mail: carla.poeschel@fu-berlin.de. I thank Jost Heckemeyer, Jochen Hundsdoerfer, and Eva Matthaei for helpful comments. I am grateful to Fulvio Castellacci and Christine Mee Lie for providing their meta-sample. The paper was presented at the VHB Annual Conference 2020. A previous version of this paper was awarded the PwC arqus Thesis Prize 2019.

1. Introduction

A country's economic growth is largely driven by research and development (R&D) (Schumpeter, 1942; Solow, 1956; Romer, 1986). Due to market failure, however, from a society point of view, the incentives to invest in R&D are too small, since, overall, economic returns significantly exceed private returns (Arrow, 1962). In response, 30 of 35 Organisation for Economic Co-operation and Development (OECD) member states have already implemented various tax instruments to stimulate firms' private R&D spending over the past decades (OECD, 2018).

The design of tax incentives for R&D varies substantially among countries, ranging from tax credits to enhanced tax allowances and accelerated depreciations (Straathof et al., 2014). Hereby, tax instruments 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). Additionally, in some countries, immediate cash refunds, carrybacks, or carryforwards of the unused tax benefits are supposed to ensure an incentive effect for loss-making firms.

While a large body of empirical literature provides evidence of positive effects of tax incentives on R&D spending, little is known about the design aspects of the underlying incentive schemes, although this is a crucial issue in many policy debates. In addition, individual studies apply a wide set of estimation strategies to evaluate the additionality of R&D tax incentives, which limits the comparability of their results. Therefore, my paper contributes to the literature by examining the impact heterogeneity of diverse tax provisions for R&D and methodological choices between studies using meta-regression analysis (MRA). Beyond MRA, I apply the random forest algorithm MetaForest (Van Lissa, 2020a) to rank moderators according to their relative importance in predicting the size of the estimates.

Previous MRA by Castellacci and Lie (2015) and Gaillard-Ladinska et al. (2015) assesses the effectiveness of R&D tax credits across industrial sectors and firm size groups. I limit my MRA to literature exploiting the direct approach and investigate sources of heterogeneous results among studies and the size of the average true effect after correcting for publication bias. Thus, this paper extends earlier MRA in several ways. I update the meta-samples of Castellacci and Lie (2015) and Gaillard-Ladinska et al. (2015) and collect estimates published between 1993 and 2019. While Gaillard-Ladinska et al. restrict their sample to similar estimates consisting of only 95 observations, I ensure comparability between estimates by converting each additionality estimate into a partial correlation coefficient (PCC), thus enhancing the variability of my sample.

I broaden the scope of my study by including additional moderator variables. Taking advantage of great cross-country variation in the underlying primary literature, I explain the heterogeneity in estimates through differences in design provisions (e.g., incremental, volume-based, and hybrid schemes). My findings reveal that diverse design choices indeed result in heterogeneous incentive effects for firms. More precisely, I find, on average, stronger input additionality effects in countries with hybrid schemes, followed by volume-based tax regimes, while countries with incremental designs seem to provide the poorest incentives for firms. Additional MRA results, however, show that the lesser additionality of incremental schemes is likely to be driven by a time effect, because many countries have adapted their tax regime toward more generous incentives in recent years.

Due to the modest number of observations (192) in my sample, I apply the novel random forest algorithm MetaForest by Van Lissa (2020a) to avoid overfitting the data (e.g., modeling random noise rather than true moderating effects). To my knowledge, I am the first to apply MetaForest in a meta-analytical framework in the field of business and economics. The algorithm ranks moderators by their importance in terms of predicting the size of the underlying estimates. MetaForest confirms my MRA results: the design aspects of tax incentives for R&D are very important moderators, and a hybrid scheme is by far the most important.

Unlike former MRA, I find only weak evidence for publication bias after controlling for heterogeneity in the estimates and, when this is the case, the bias is larger in studies using a difference-in-differences (DiD) strategy than in those applying ordinary least squares (OLS) estimators or matching. The average true association between tax incentives and R&D expenditures is positive, but its strength is also largely driven by methodological variations across studies; the average true effects range between 0.04 for DiD and 0.486 for OLS, comparable to previous findings. Overall, these findings are not only important for policy makers who are continuously improving the design aspects of R&D tax incentive schemes, but also give further guidance for future empirical research on this topic.

The remainder of this study is structured as follows. Section 2 reviews the literature on the additional effects of R&D tax incentives. Section 3 briefly presents the underlying sample, summarizes the moderator variables, and discusses the methodology of MRA and MetaForest. The results of my MRA and MetaForest are displayed in Section 4. Finally, Section 5 concludes the study.

2. Review of the Literature

2.1 Input Additionality of R&D Tax Policy

Tax incentives have become an increasingly popular instrument available to policy makers to stimulate firms' private R&D expenditures. Their implementation among countries has encouraged researchers to evaluate their effectiveness, whereby the vast majority of studies assess the input additionality effects of tax incentives on private R&D spending using firm-level datasets.

In their comprehensive review on the effectiveness of R&D tax incentives, Hall and Van Reenen (2000) distinguish between two key approaches for measuring the (causal) effect of tax instruments on R&D expenditures: the structural and the direct approach. The structural approach is based on the neoclassical investment model formalized by Hall and Jorgenson (1967), in which R&D capital stock is explained by its user costs. R&D user costs incorporate the tax incentives for R&D, other tax provisions, and interest, inflation, and depreciation rates (Hall and Van Reenen, 2000).

Since several papers relying on the structural approach are plagued by reverse causality between user costs and tax incentives, there has been a shift away from the structural approach toward the direct approach during the past few years. Thus, I restrict my MRA to studies that are based on the direct approach, thereby setting focusing on more recent literature on R&D tax policies.

Generally, these studies explore the impact of various tax instruments on R&D expenditures through a dummy variable indicating whether a firm is eligible for a R&D tax incentive scheme. A simple comparison of R&D expenditures between eligible firms (treatment group) and non-eligible firms (counterfactual or control group) through an OLS regression amounts to an average treatment effect plus selection bias, because firms that use tax benefits are likely to differ systematically from firms that do not. Therefore, selection into the treatment group is not random, but could be due to confounding variables (e.g., firm characteristics). Many underlying studies address the selection problem by applying matching or by exploiting a policy intervention in the tax scheme. DiD estimates are causal effects of the tax incentives on R&D expenditure, assuming common trends between both groups (Angrist and Pischke, 2009). Besides the assumption of conditional independence, matching requires that the difference between recipients and non-recipients manifest in observable variables (Pearl, 2009).

2.2 Design Aspects of R&D Tax Policy

The design aspects of current R&D tax incentives differ considerably among countries (Straathof et al., 2014). R&D tax incentives either relate to R&D expenditures (input based) or the income generated by intellectual property (output based). In this MRA, I focus on the former, because most research analyzes input-based provisions.

Input-based incentives are considered within a wide set of instruments: for example, R&D tax credits, enhanced allowances, and accelerated depreciations. While tax credits are the most common, some countries use a mix of several tax measures and combine tax credits with enhanced allowances or accelerated depreciations to decrease both firms' taxes due, as well as firms' tax base. The studies underlying my MRA mostly evaluate the effects through tax credits; however, four primary studies refer to enhanced allowances (e.g., Guceri, 2015, 2018; Holt et al., 2016; Guceri and Liu, 2019). Additionally, an incentive effect of R&D tax instruments is ensured, even for firms in a loss position, through either immediate cash refunds or carrybacks or carryforwards to future income years. Many countries explicitly target cash refunds to small and medium-sized enterprises (SMEs)—for example, Australia, Canada, France, and the United Kingdom—since these firms are more likely to be liquidity constrained.

An R&D tax policy can be further classified according to its type. The implementation of tax schemes is volume based if they apply to all qualifying R&D expenditures, or it is incremental if the tax schemes cover only an increase in R&D expenditures compared to an initial level, which mostly corresponds to the average R&D expenditure of previous years (moving average). However, this approach could distort firms' R&D process, since the timing of R&D investment is determined by maximizing tax advantages rather than optimizing R&D strategies (Straathof et al., 2014). Since volume-based regimes are less complex for both firms and administrations and, thus, reach a larger target group, most governments have recently switched to volume-based schemes. For example, France had an incremental scheme until 2003, a hybrid scheme between 2004 and 2007, and a volume-based scheme from 2008 onward. The Irish scheme evolved into a volume-based one over the years 2012 to 2015. Australia adopted a volume-based incentive in 2011. However, Italy replaced its volume-based scheme by an incremental one in 2011 to reduce costs. An incremental R&D tax policy can minimize the deadweight loss through forgone tax revenue due to R&D that would have been conducted, even in the absence of tax incentives. Lokshin and Mohnen (2012) evaluate the Dutch volume-based tax credit scheme during 1996 and 2004 and measure a deadweight loss of 85 percent of forgone tax revenue. Consequently, a preference for incremental schemes over volume-based ones supports firms with high R&D growth (Appelt et al., 2016). On the contrary, Lester and Warda (2014) argue that both incremental and volume-based incentives exhibit similar cost-effectiveness. Thus, more and more countries are implementing hybrid schemes by combining volume-based elements with incremental ones. Besides maintaining R&D spending levels, hybrid schemes promote high R&D growth (Appelt et al., 2016). Japan and Spain combine volume-based tax credits with incremental ones for firms with high R&D growth.

3. Meta-Analysis Methodology

Studies vary considerably with respect to the underlying methodologies and design aspects of tax incentives for R&D. Therefore, it is not surprisingly that the findings in the literature are inconsistent or even contradictory. Meta-analysis is a methodology for quantitatively combining a strand of literature and investigating heterogeneity in estimates across primary studies (Stanley, 2001).

3.1 Measuring Input Additionality

As explained in Section 2.1, the literature on the direct approach uses different estimation methods. The baseline model is

$$RD_{it} = \beta_0 + \beta_1 Eligibility_{it} + \beta X_{it} + \varepsilon_{it}$$

where, for firm i at time t, RD_{it} is either the (logarithm of) R&D expenditures or R&D intensity (R&D expenditures scaled by assets or sales), $Eligibility_{it}$ is a dummy variable indicating eligibility for a tax incentive, and the vector X_{it} contains various firm-level controls. The main coefficient β_1 captures the increase in R&D expenditures when claiming the tax benefit for R&D.

Another approach is to exploit an exogenous reform of the incentive scheme affecting only one (treatment) group of firms by adding an indicator variable $Reform_{it}$ (which equals one after the policy change):

$$RD_{it} = \beta_0 + \beta_1 Eligibility_{it} + \beta_2 Reform_{it} + \beta_3 Eligibility_{it} \cdot Reform_{it} + \beta X_{it} + \varepsilon_{it}$$

where the DiD estimator is β_3 on the interaction variable that measures the rise in R&D expenditures due to the policy reform.

Moreover, a matching strategy predicts a propensity score of receiving the treatment (eligibility in this case) conditional on observable firm characteristics. Then, group identification is carried out by matching observations with equal values of propensity scores. The parameter of interest is the average treatment effect on the treated, which is the average difference in R&D expenditures between both groups.

I collect the various coefficients of interest, which are the dependent variables in my MRA. The MRA results, however, are only meaningful if the estimates are comparable across primary studies (Stanley, 2001). Due to varying measurement units (e.g., R&D expenditures and R&D

intensity) and functional forms (e.g., level or logarithmic level) between the underlying studies, I transform each estimate into a PCC, as follows:

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

where t_{is} is the *t*-statistic of the regression $i = 1, \dots, I$ from primary study $s = 1, \dots, S$, and df_{is} refers to the *t*-statistics' degrees of freedom (Stanley and Doucouliagos, 2012). The corresponding standard error of the PCC is given by $SEPCC_{is} = \sqrt{\frac{(1-PCC_{is}^2)}{df_{is}}}$.

(insert Table 1 about here)

The PCCs measure the direction and size of the association between R&D expenditures and tax incentives while holding all other variables constant and is thus a statistical measure, rather than an economic one. As shown in Table 1, the values of the PCCs scatter much less compared to the additionality effects. Moreover, the values range far above and below the limits of -1.0 and +1.0, and thus truncation is not a concern in my MRA.

3.2 Selection of the Literature

To locate appropriate studies, I used Google Scholar and the IDEAS database. In the literature research, I employed the following keywords: *research and development, R&D, Tax policy, tax credits, tax incentives,* and *additionality*. Additionally, I scanned literature reviews (e.g., Straathof et al., 2014; Spengel et al., 2017) and references of the identified studies. The literature search was completed in March 2019.

I selected papers relying on the direct approach (n = 57), which estimate the effect of tax incentives on R&D expenditures or R&D intensity (inclusion criterion 1) using one of the three identification strategies explained in Section 3.1 (inclusion criterion 2). Moreover, these studies use firm-level within-country datasets (inclusion criterion 3) and provide information concerning the standard error or t-statistics and the number of observations (inclusion criterion 4). I only consider studies with estimates for subsamples within the scope of my study, such as for large firms, SMEs, high-tech firms, low-tech firms, manufacturing firms, and non-manufacturing firms, and only choose the latest version of a study to avoid autocorrelation among the estimates.

Since selecting a single effect estimate per primary study is quite subjective and results in a small sample size and less heterogeneity among the effect estimates, I include multiple estimates from each primary study, as long as there is a substantial difference regarding the specification, estimation strategy, or sample. I do not sample the estimates in the robustness analysis.

(insert Table 2 about here)

Table 2 outlines the primary studies in my meta-sample and provides an overview of the underlying countries and PCCs (i.e., number, mean, median, minimum, maximum, and standard deviation).[1] While Castellacci and Lie (2015) and Gaillard-Ladinska et al. (2015) cover studies published from 1992 through 2013 and 2006 through 2014, respectively, I consider papers published between 1993 and 2019 in my meta-sample. Studies are marked by a superscript *n* if they were not subject to the meta-analyses of Castellacci and Lie (2015) or Gaillard-Ladinska et al. (2015). As shown in Table 2, my total meta-sample contains 19 primary studies and 192 PCCs. The number of effect estimates from each primary study ranges from one for Billings and Fried (1999) and Acheson and Malone (2016) to 72 for Ho (2006), showing an enormous range. The paper by Ho (2006) reports significantly more coefficients than the average primary study (34 estimates). If I exclude this study, the number of observations and consequently the variability is greatly reduced. To address this concern, however, I drop the effect estimates of Ho (2006) from my meta-sample as a robustness test.

(insert Figure 1 about here)

The histogram in Figure 1 illustrates the right-skewed distribution of the PCCs. The mean value equals 0.061, with a standard deviation of 0.060, and the median corresponds to 0.046. The most extreme values (minimum and maximum) are -0.067 and 0.277 for Hægeland and Møen (2007). The means of the PCCs per primary study are positive and range from 0.019 to 0.214, suggesting substantial heterogeneity among the underlying studies. The estimates seem to vary far less within primary studies, since the standard deviations in the last column of Table 1 range below 0.060, except for Hægeland and Møen (2007). In addition, my meta-sample shows great geographical variation, since it covers estimates for 11 different countries over a long sample period of 40 years (1975 to 2014). I exploit this country heterogeneity to separate the incentive effects of various policy design aspects.

3.3 Moderator Variables

The underlying theory or common practice—such as the reporting guidelines of Havránek et al. 2020 for meta-analysis in economics—generally determine potential sources of heterogeneity among estimates. I coded 16 moderators, potentially affecting the magnitude of the effects regarding the issue of publication selection, the definition of the dependent variable, methodological choices, firm characteristics, and the design choices of tax regimes among countries. Table 3 summarizes the full set of moderator variables, along with their summary statistics. At first glance, the summary statistics point to heterogeneous estimates, since the means vary greatly between the various moderators.

(insert Table 3 about here)

3.3.1 Publication Bias

A publication bias arises if statistically nonsignificant and unexpected estimates are not published in a paper. The theory in the R&D literature undoubtedly predicts a positive relation between R&D tax incentives and R&D expenditures. To visually test the presence for publication bias, a funnel plot is used to map the estimates (in my case, PCCs) against their precision, as measured by the inverse of the estimates' standard error (Egger et al., 1997). All estimates that rely on a small sample size and a large standard error (i.e., low precision) scatter widely at the bottom of the diagram, whereas estimates based on a larger sample size with, thus, higher precision are distributed at the top of the diagram. In the absence of publication bias, the estimates are randomly spread around the average true effect.

(insert Figure 2 about here)

The funnel plot in Figure 2 shows a right tail with a missing left side, since the vast majority of the estimates vary between zero and 0.2, while negative values are almost absent. The peak of the funnel plot is composed of the most precise estimates, scattered around zero and 0.03. The asymmetrical funnel plot is consistent with the presence of a publication bias against positive estimates. Therefore, estimates with positive values are more likely to be selected for publication in the primary literature for the direct approach.

To address the issue of publication selection more formally, I include the PCC (squared) standard error (*PCC (Squared) Standard Error*) of the corresponding estimate as an explanatory variable. The funnel asymmetry test of the coefficient on *PCC (Squared) Standard Error* indicates the presence of publication bias (Stanley, 2008). The underlying intuition is simple: authors with small samples have to search longer for statistical significance by testing various specifications and samples, resulting in larger effects than those of authors with larger samples. Moreover, the precision effect test on the constant assesses whether there is an average true effect of tax incentives on R&D expenditures beyond publication bias (Stanley, 2008).

Additionally, I collect both published studies and working papers and code the moderator variable *Published Study*. The majority of the estimates (66 percent) are obtained from unpublished working papers.

3.3.2 Definition of the Dependent Variable

Since the distribution of firms' R&D expenditures is largely skewed, the primary literature mainly specifies the logarithm of R&D spending as the dependent variable, while limiting the sample to observations with strictly positive values for R&D expenditure (69 percent of the estimates). However, explicitly omitting observations with zero values can induce endogenous selection bias. Therefore, other studies use R&D intensity (R&D expenditures scaled by assets or sales) as the outcome variable (11 percent of the estimates). In contrast to former MRA, I code two binary moderator variables that capture the definition of the dependent variable: *Logarithm* and *R&D Intensity*. Both moderator variables *Logarithm* and *R&D Intensity* are equal to one if an estimate relies on the logarithm of R&D expenditures or the R&D intensity as the outcome measure, and zero otherwise.

3.3.3 Estimation Strategy

I further introduce three dummy variables indicating the underlying estimator: *OLS Estimator, Matching,* and *DiD.* I use *DiD* as the base category to reduce collinearity between the variables. Since eligible firms could self-select into the treatment group, I expect an upward or downward bias for simple OLS estimates. In my meta-sample, most studies correct for selection bias by using matching (50 percent of the estimates). Others apply the DiD identification strategy to explore changes over time within firms while controlling for unobserved but fixed heterogeneity between firms, such as managerial ability or attitude to risk (36 percent of the estimates). A causal interpretation of the resulting estimates, however, depends heavily on whether the underlying key conditions are met (e.g., selection-on-observables and common trends assumptions). 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 affect the group of user firms more strongly than non-user firms.

3.3.4 Firm Characteristics: Size and Industry

The seminal work by Schumpeter (1942) hypothesizes the dominance of larger firms in the technological process through greater market power, leading to better financial and human resources. Theory and empirical evidence suggest that SMEs are less likely to innovate, since they face liquidity constraints (Czarnitzki and Hottenrott, 2011) or are unaware of tax incentives (Corchuelo and Martínez-Ros, 2010). One could expect the stronger additionality of tax incentives for SMEs than for large firms, due to designs that are more generous toward SMEs and the provision of additional liquidity. Surprisingly, the empirical evidence on heterogeneous additionality effects by firm size is mixed (e.g., Koga, 2003; Baghana and Mohnen, 2009; Corchuelo and Martínez-Ros, 2010; Kobayashi, 2014). The MRA results by Castellacci and Lie

(2015) indicate a larger additionality effect of tax credits for SMEs in the underlying primary literature, whereas Gaillard-Ladinska et al. (2015) report contrasting results. Therefore, I introduce a moderator variable for the estimates of the effects based on a subsample of large firms (*Large Firms*).

According to Castellacci and Lie (2015), the impact of tax credits for R&D varies among sectors, possibly because firms across various industries differ greatly regarding their innovation strategies and technological performance. The authors' results suggest, on average, smaller additionality 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. Thus, I categorize whether effect estimates are related to a subsample of high-tech firms (*High-Tech Firms*) or a subsample of manufacturing firms (*Manufacturing Firms*).

3.3.5 Design of R&D Tax Policy

A large body of literature examines the effectiveness of R&D tax incentives using firm-level data. However, this literature rarely investigates whether firms' responses to tax incentives can vary with design aspects, and, if so, studies have shown conflicting results (Klassen et al., 2004; Straathof et al., 2014; Spengel et al., 2017). Therefore, I include various moderator variables to separate the incentive effects of different design aspects. The information regarding the design features are taken from the underlying studies.

I introduce three moderator variables relating to countries' tax regime types: *Volume-Based Scheme*, *Incremental Scheme*, and *Hybrid Scheme*. As seen in Table 2, 44 percent of the underlying estimates refer to countries with incremental schemes, 30 percent to volume-based regimes, and 26 percent to hybrid ones. It is reasonable to expect weaker additionality effects for incremental schemes, since these regimes are more complex than volume-based ones, resulting in higher administrative costs that can discourage firms from applying for such incentives if the (perceived) application costs exceed the expected tax benefits (Appelt et al., 2016). By contrast, incremental instruments promote R&D that would not have been conducted without tax relief (Lokshin and Mohnen, 2012). Whether hybrid schemes (i.e., combining volume-based and incremental elements) exhibit the incentives of both schemes is an empirical question.

3.4 MRA

In my main analysis, I rely on MRA to explain the heterogeneity among estimates. I apply the following fixed effects (FE) meta-regression model:

$$\hat{e}_{is} = \alpha + \beta \cdot x_{is} + \varepsilon_{is} \tag{1}$$

where the estimate $\hat{e}_{is}$ (PCC_{is} or $Additionality\ Effect_{is}$) of regression $i = 1, \dots, I$ referring to primary study $s = 1, \dots, S$ is the outcome variable, x_{is} is a vector of moderator variables (i.e., characteristics within and between studies that explain the variation across estimates; see Section 3.3 for descriptions of the moderator variables), α is the constant, and ε_{is} is the error term.

In MRA, the residuals are clearly heteroscedastic because the variance of the error term ε_{is} rises in estimates' standard error, which is a moderator. To correct for heteroscedasticity, equation (1) is weighted by the inverse variances of the estimates as analytical weights (Stanley, 2008). The analytical weights are the standard errors taken from the primary studies ($SEPCC_{is}$ or SE_{is}). Beyond correcting for heteroscedasticity, weighting by the inverse variances corrects for low-quality estimates, since imprecise ones are given less weight in the MRA (Stanley and Doucouliagos, 2012). I consider multiple estimates from each primary study in my meta-sample, which bears the risk of within-study dependency (i.e., autocorrelation). I allow for autocorrelation between estimates per primary study due to unobserved study-level heterogeneity and cluster standard errors at the study level.

Despite heterogeneity among estimates, I apply an FE model assuming a common effect among estimates, where variations are solely due to sampling errors. A crucial assumption in any random effects model is that the unobserved study-level effects and moderator variables are uncorrelated. Due to the presence of publication bias, however, the unobserved study-level effects could be dependent on the estimates' standard errors, which is a moderator variable in my MRA. Then, only an FE model with clustered standard errors at the study level is appropriate (Feld and Heckemeyer, 2011).

3.5 MetaForest

MRA is likely to suffer from low statistical power when the number of observations (192 in my case) is low in comparison to the number of moderators (e.g., 16). A simple way to reduce the list of moderators is through sequential *t*-testing (or general-to-specific modeling), by removing statistically nonsignificant predictors one by one (Stanley and Doucouliagous, 2012). In this paper, I use the random forest algorithm MetaForest, by Van Lissa (2020a) to determine the importance of each underlying moderator in terms of its predictive performance. I follow the meta-analyses of Bonapersona et al. (2019) and Curry et al. (2018) in behavioral science.

A random forest builds a number of decision trees on bootstrap samples (5,000 in my case) of the original dataset by separating the samples into subgroups with similar estimates, based on a random subset of the splitting variables. The algorithm averages the predictions of the decision trees and, thus, is more robust to data overfitting than MRA. With the splitting variables randomly selected from the full set of moderators, the trees will be uncorrelated and the mean value from the resulting trees becomes less variable and hence more credible (James et al., 2017). The application of a random forest to meta-analysis (MetaForest) applies a weighting scheme to the bootstrap sampling by selecting more precise estimates with a higher probability. Van Lissa (2020a) provides further explanations on the random forest algorithm MetaForest.

I use the R package *metaforest* (Van Lissa, 2020b). The optimal number of splitting moderators, the minimum number of cases remaining in the subgroups, and the type of weights (e.g., uniform, fixed, or random effects) are chosen through *k*-fold cross-validation. The cross-validation involves (1) randomly portioning the dataset into *k* equal subsets, *k* - 1 training sets, and a remaining validation set, (2) fitting the models on the *k* - 1 training sets for all possible tuning parameter combinations, (3) validating the models on the validation set by estimating the error rates (i.e., the root mean-squared prediction error), and (4) averaging the error rate estimations. The combination of tuning parameters with the lowest mean cross-validation error rate is chosen for the final model.

4. Results

I provide a variety of results to explain the heterogeneity in estimates through the methodological variations and design aspects of R&D tax incentives across studies. Sections 4.1 and 4.2 discuss my main MRA results, focusing on the most comprehensive sample where the PCC is the dependent variable and the explanatory variables are several characteristics within and between the underlying studies, potentially driving the heterogeneity in estimates. In Section 4.3, instead of MRA, I use the machine learning algorithm MetaForest to explain the heterogeneity among estimates. Since the PCC is more of a statistical measure, the MRA in Section 4.4 assesses the average true effect across subsamples, allowing for an economical interpretation of the additionality effects of R&D tax incentives.

4.1 Main MRA Results

4.1.1 Publication Bias and Average True Effect

Table 4 presents alternative models to test whether the estimates are robust to a potential publication bias. Models (1) and (2) do not correct for publication bias, whereas models (3) to (6) include the moderator variable *PCC Standard Error* or *PCC Squared Standard Error*, to control for publication bias, since, according to simulations, a linear publication bias correction is likely to underestimate the average true effect (Moreno et al., 2009; Stanley and Doucouliagous, 2012).

The coefficients on the variables *PCC Standard Error* and *PCC Squared Standard Error* detect the presence of publication bias (funnel asymmetry test), and the value of the constant reflects the average true effect after accounting for publication selection (precision effect test). The regression coefficients on both variables are positive throughout my models, but only significant in model (3), providing only weak or even no evidence of the presence of publication bias after controlling for heterogeneity. I observe, however, significantly positive coefficients on *Published Study* in the majority of models, consistent with the slightly larger PPCs reported in published studies than in discussion papers. Those more formal findings on the issue of publication bias only partly support my previous conclusion based on the funnel plot in Figure 2.

Since the constants are positive and significant at the 1 percent level, there exists, on average, a true relation between tax incentives and R&D expenditures. More precisely, depending on the respective model, the average true effect varies between 0.013 and 0.035, whereby it turns out that the magnitudes in models (5) and (6) with *PCC Squared Standard Error* are indeed similar to those in models (1) and (2) of Table 4 without *PCC (Squared) Standard Error*. The absolute values range far below 0.07, and, thus, the correlation between tax incentives and R&D

expenditures seems to be very small in the literature based on the direct approach (Doucouliagos, 2011).[2]

(insert Table 4 about here)

4.1.2 Heterogeneity in Estimates

I consider two variables relating to the definition of the dependent variable because the estimates are quite heterogeneous with respect to the underlying specification. The coefficients on the moderator *Logarithm* are negative and highly significant, showing that estimates are, on average, smaller when specifying the logarithm of R&D expenditures. The coefficients on *R&D Intensity* are negative as well, but not significant or only weakly so. Accordingly, the estimates considering firm size by scaling R&D expenditures by total sales or assets do not exhibit lower sensitivity than the other estimates. Therefore, restricting the sample to observations with strictly positive values on R&D expenditures could induce selection bias, and R&D intensity could serve as a better outcome variable.

The regression coefficients on the variables *OLS Estimator* and *Matching* are positive, but only statistically significant (at the 1 percent level) throughout all the models for the former. This is in line with my predictions above: studies with OLS rarely control for selection bias; the results show that simple OLS estimates seem to overstate the impact of tax incentives on R&D expenditures because of endogeneity.

I distinguish between firm size groups and industries by adding the moderators *Large Firms, High-Tech Firms,* and *Manufacturing Firms*. The coefficients on *Large Firms* are positive and weakly significant in most models. Large firms' R&D expenditures are a little more sensitive (by 0.005–0.007) to tax incentives, in comparison to those of the overall firm population. Thus, my conclusion differs from the findings of Castellacci and Lie (2015), who find larger additionality effects of tax incentives for SMEs. The disparity in results potentially arises from the differing sets of estimates and explanatory variables under investigation. Moreover, Castellacci and Lie (2015) report smaller additionality effects for high-tech firms, and my estimates confirm this finding, since the coefficients on *High-Tech Firms* are significantly negative in all of my models. The estimates regarding the variable *Manufacturing Firms* are statistically nonsignificant, showing no difference between the estimates that are based on a subsample of manufacturing firms and those relying on a general sample.

According to my results, different design aspects among countries seem to be an important source of variation across estimates. The regression coefficients on *Hybrid Scheme* and *Volume-Based Scheme* are both positive and mostly statistically significant, while those on *Incremental Scheme* are significantly negative across my models. The magnitudes of the coefficients on *Hybrid Scheme* are considerably larger than those of the coefficients on *Volume-Based Scheme*. In other words, hybrid schemes provide, on average, stronger incentives for firms compared to volume-based ones, whereas the additionality effects in countries with incremental measures are, on average, lower in the underlying primary literature. The result could indicate that hybrid schemes increase the level of R&D more heavily due to lesser complexity, while simultaneously minimizing the deadweight loss through forgone tax revenue. I further investigate the design aspects of tax incentives in Section 4.2.

4.1.3 Robustness Checks

I perform a set of robustness checks to examine the sensitivity of the main MRA results to changes in the meta-sample composition. My first robustness test in models (1) and (2) of Table 5 investigates whether the results hold when I exclude the PCCs of Ho (2006), who reports considerably more estimates than the average primary study. Due to the smaller meta-sample size (120 observations) and the resulting correlation between moderator variables, I omit *Logarithm* from my MRA specification in models (1) and (2). Despite this, the results on the covariates for the variables *PCC Squared Standard Error, High-Tech Firms, Large Firms, Incremental Scheme*, and *Volume-Based Scheme* change because these variables hinge on diminished variability after dropping the estimates of Ho (2006).

As a second robustness test, I drop six outliers referring to the studies by Hægeland and Møen (2007) and Holt et al. (2016) in models (3) and (4) of Table 5. Those PCCs are far to the left and right of the remainder of the funnel plot in Figure 2 and, thus, can be seen as outliers (Stanley and Doucouliagous, 2012). Since the meta-sample is rather small (19 primary studies), the results could be quite sensitive to outliers. The findings change little: while the coefficient on *Large Firms* loses significance, the coefficient on *Manufacturing Firms* gains significance in model (3). Overall, my robustness checks support the main conclusions derived from the MRA results in Table 4.

(insert Table 5 about here)

4.2 Extended MRA Results

The goal of the extensions in Table 6 is to explore additional useful moderator variables relating to the design aspects of R&D tax incentives. I do not consider these variables in my main MRA, because the inclusion of these new moderators greatly increases collinearity (with some variables having VIFs above 10), so that the signs and magnitudes of the coefficients should be interpreted carefully.

In models (1) and (2) of Table 6, I only consider the moderator variables concerning the design aspects of tax incentives, because these explanatory variables are correlated with several methodological choices. My prior results for *Hybrid Scheme*, *Volume-Based Scheme,* and *Incremental Scheme* remain qualitatively stable.

Tax incentives reduce either the taxable base (e.g., enhanced allowance) or the tax liability (e.g., tax credit). I code the variable *Enhanced Allowance* as either one, to indicate that the underlying estimate refers to a country offering an enhanced allowance, or zero, to denote a country implementing a tax credit. Since the variable *Enhanced Allowance* is strongly correlated with its corresponding type, I only include the moderator variable *Enhanced Allowance* in model (3). I find only a low correlation between the variable *Enhanced Allowance* and the PCCs. This result demonstrates that the estimates for enhanced allowances do not strongly differ from those for tax credits.

I add in models (4) and (5) the variable *Average Sample Year,* which is normalized between zero and one, while setting the oldest average sample year (1982) to zero and the latest one (2012) to one. A large majority of studies in my sample refer to early periods of investigation (e.g., 1990s). Many countries, however, have adapted several features of their tax regimes for R&D (e.g., the United States, Ireland, France, Australia, or Italy), increasing the total volume of funding in relation to their gross domestic product over the past decade (OECD, 2020). Therefore, the design of the underlying tax measures to promote R&D differs greatly from the current ones. In particular, the estimates for incremental schemes relate to earlier periods (1975–1999). This fact could explain the smaller additionality effects of incremental schemes. My estimates confirm that the incentive effect of hybrid schemes is much stronger than that from the two other schemes (volume based and incremental incentives). In model (5), however, the coefficient on *Average Sample Year* is significantly positive, showing that the estimates have increased over time. After the variable is added, the significance of *Incremental Scheme* disappears. Thus, those results imply that lower incentive effects of incremental schemes can be partly explained by the underlying investigation period. This is an important insight for the interpretation of prior empirical findings on R&D tax incentives.

(insert Table 6 about here)

4.3 MetaForest Results

I use 10-fold cross-validation to choose the optimal tuning parameters: random effects weights, two splitting variables, and a minimum of four cases per subgroup. I construct 5,000 trees, since the model converges with approximately 5,000 trees (see Figure A1 in the Appendix). Thus, I conduct a random effects MetaForest with 5,000 trees based on clustered bootstrap samples to account for within-study dependency.

Figure 3 illustrates the MetaForest results for the importance of the variables. The moderators are ranked according to 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 for a given moderator after random permutation, averaged over all trees. When a moderator has no predictive power, randomly permuting coincidentally strengthens the relation, and the variable importance score becomes negative.

Figure 3 shows that the variables on the design aspects of R&D tax incentives (e.g., *Incremental Scheme*, *Volume-Based Scheme*, and *Hybrid Scheme*) largely contribute to the overall prediction, while *Hybrid Scheme* is by far the most important moderator. In contrast, the moderators on industry classifications, *High-Tech Firms* and *Manufacturing Firms,* have zero predictive power, since the importance of the variable is negative.

(insert Figure 3 about here)

4.4 Average True Effect across Methodological Aspects

I split my total meta-sample according to the underlying estimation strategy (*OLS Estimator, Matching,* or *DiD*). To ensure consistency between estimates, I only consider estimates with the logarithm of R&D spending as the dependent variable. Now, to interpret my MRA results economically, the dependent variable is the initial additionality effect found in primary studies, and its corresponding *Squared Standard Error* is the explanatory variable. I estimate my MRA without clustering standard errors at the study level, because the number of study clusters per subsample is too small.

As seen in Table 7, the issue of publication bias appears to be of varying importance between the different groups of estimates; the significant coefficient on *Squared Standard Error* in models (1) suggests that the DiD estimates suffer more strongly from publication bias than the OLS and matching ones. Moreover, the strength of the average true relation between tax incentives and R&D expenditures varies depending on the studies' methodological aspects. The constant for the OLS estimates are relatively large (e.g., 0.486) in comparison to the two other constants, confirming my earlier conclusions that selection bias leads to higher additionality effects in the literature. The subset of estimates using DiD represents the largest group of papers (10 primary studies). These estimates are, on average, much lower in absolute terms (e.g., 0.04) than the OLS and matching estimates. The average true average treatment effect on the treated is positive and equals 0.426 in model (2), indicating that the recipients of tax incentives appear to have, on average, 40 percent higher R&D expenditures than non-recipients. These findings are comparable to prior findings by Gaillard-Ladinska et al. (2015), who report weighted means of 0.13 for DiD and 0.90 for matching.

Overall, the relation between tax incentives and R&D expenditures is significantly positive, but methodological variations across studies drive the value of the average true effect. Even though studies using DiD and matching 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; see Section 2.1).[3]

(insert Table 7 about here)

5. Concluding Remarks

R&D tax incentives are major policy tools for stimulating firms' private R&D expenditures. Despite the extensive literature on tax incentives for R&D, little is known about the effects of diverse design aspects of the underlying instruments. In this paper, I set up a comprehensive meta-sample containing 192 estimates from 19 primary studies exploiting the direct approach.

Previous MRA by Castellacci and Lie (2015) and Gaillard-Ladinska et al. (2015) focus on heterogeneous effects among industrial sectors and firm size classes. I contribute to the literature by investigating the impact heterogeneity of diverse R&D tax incentives among countries (i.e., incremental, hybrid, and volume-based schemes), using MRA. Unlike former studies, beyond MRA, I additionally use the random forest algorithm MetaForest, by Van Lissa (2020a), to rank moderators according to their predictive power.

For the underlying meta-database, my results reveal that firms' different responses by size and sector do not play an important role in explaining the variations across estimates. Instead, I find evidence of stronger input additionality effects in countries with hybrid schemes, followed by volume-based tax regimes, while countries with incremental tax measures appear to provide lower incentives for firms. In additional MRA, I find that the reported additionality effects increase over time. Therefore, since the estimates for incremental schemes relate to earlier periods (1975–1999), the latter effect seems to be driven by a time effect. MetaForest supports these conclusions: the design aspects of tax incentives for R&D seem to be more important moderators than firm characteristics, while the hybrid scheme is by far the most important. Overall, my results suggest that hybrid schemes combine the advantages of the other two incentives and increase the level of R&D more strongly due to less complexity and a lower deadweight loss.

The primary studies vary in terms of the underlying specifications and estimation strategies. Therefore, I convert the initial coefficients into PCCs to ensure the equality of the estimates. The average true estimates range far below 0.07 and are thus considered very small. Since the PCC is more of a statistical measure, I assess the average true effect across more comparable subsamples. To interpret the relation between tax incentives and firms' R&D spending economically, the dependent variable is the original additionality effect found in the primary studies. I discover only weak publication bias in the underlying strand of literature, and only in studies using DiD. The true average association between tax incentives and R&D expenditures seem to be positive, but varies considerably from 0.04 to 0.486 with respect to methodological choices. Studies using OLS rarely control for selection bias and, due to endogeneity, overstate the impact of tax incentives on R&D expenditures. Even though studies with DiD and matching try to address selection bias, I interpret the average true estimates as associations rather than a causal interpretation, since the underlying key assumptions (e.g., selection-on-observables and common trends assumption) often seem to not be met in the underlying literature.

Notes

1. The underlying studies were coded by the author. The meta-database is available from the author upon request. The reporting guidelines for meta-analysis in economics (Havránek et al., 2020) suggest the participation of more than one coder. I compared my coding with that of Castellacci and Lie (2015) and find strong agreement. Disparities mainly arise because I sample estimates referring to a published version of the study and do not consider estimates for specific subsamples.
2. According to Doucouliagos (2011), a PCC is considered small if its absolute value ranges between 0.07 and 0.17, medium if its absolute value ranges between 0.17 and 0.33 and large if its absolute value is larger than 0.33.

3. Guceri (2018) and Guceri and Liu (2019) use medium firms as the treatment group and large firms as the control group, and their estimates are thus likely to be plagued by a confounding bias, because eligible (or medium) firms could be more affected by capital market imperfections.

Figure 1: Histogram of the PPCs

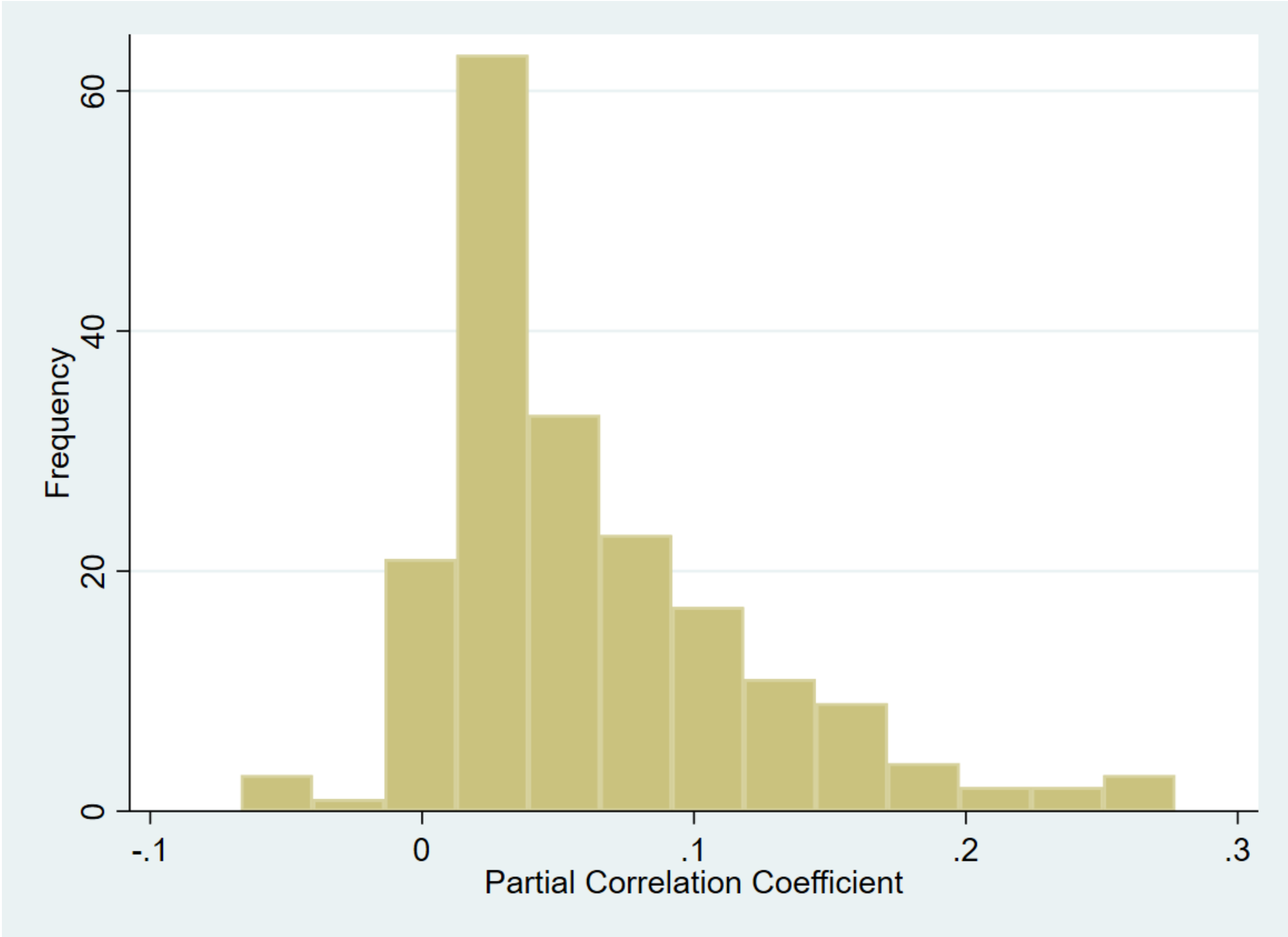


Figure 2: Funnel plot of the PCCs

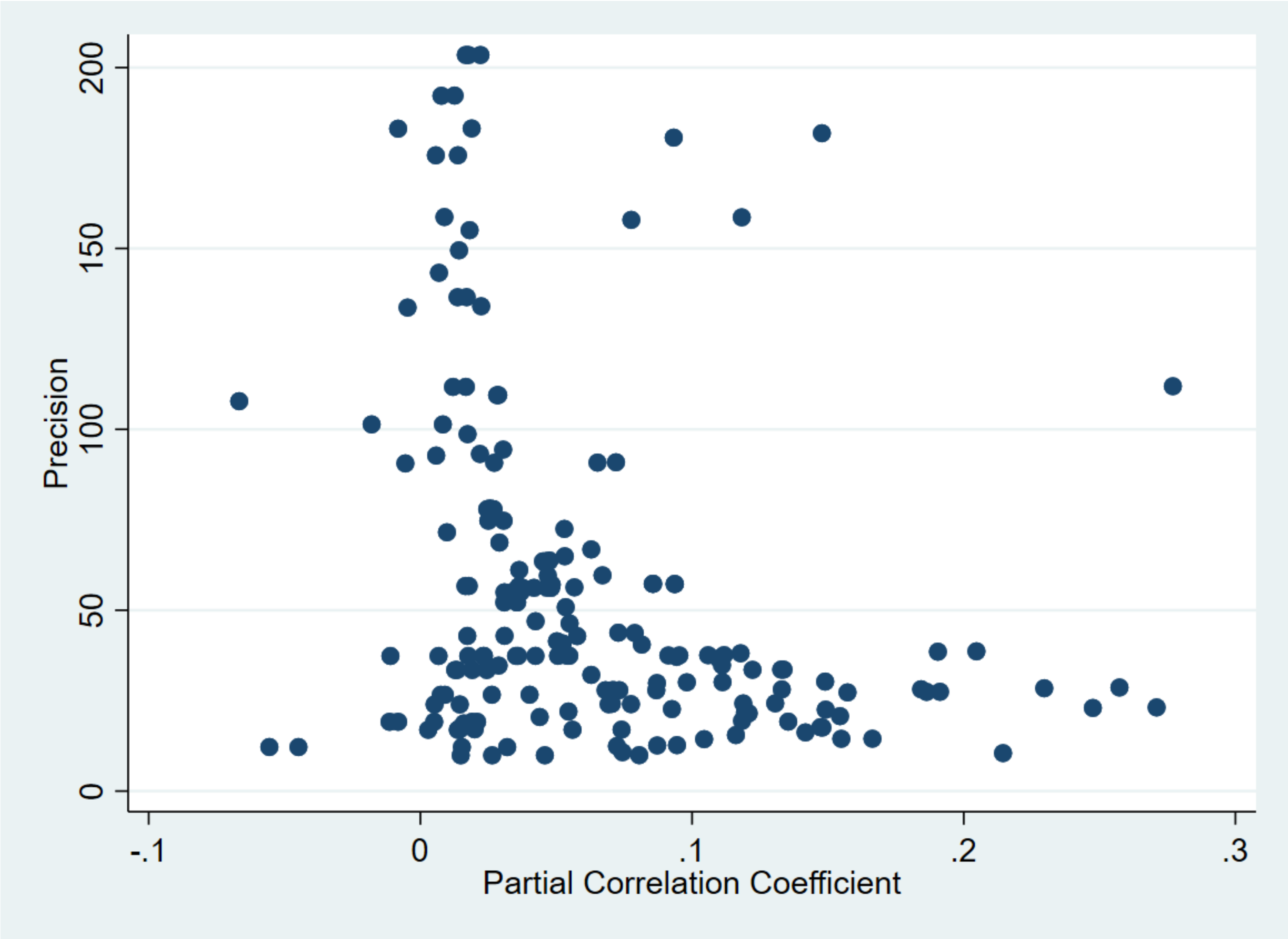
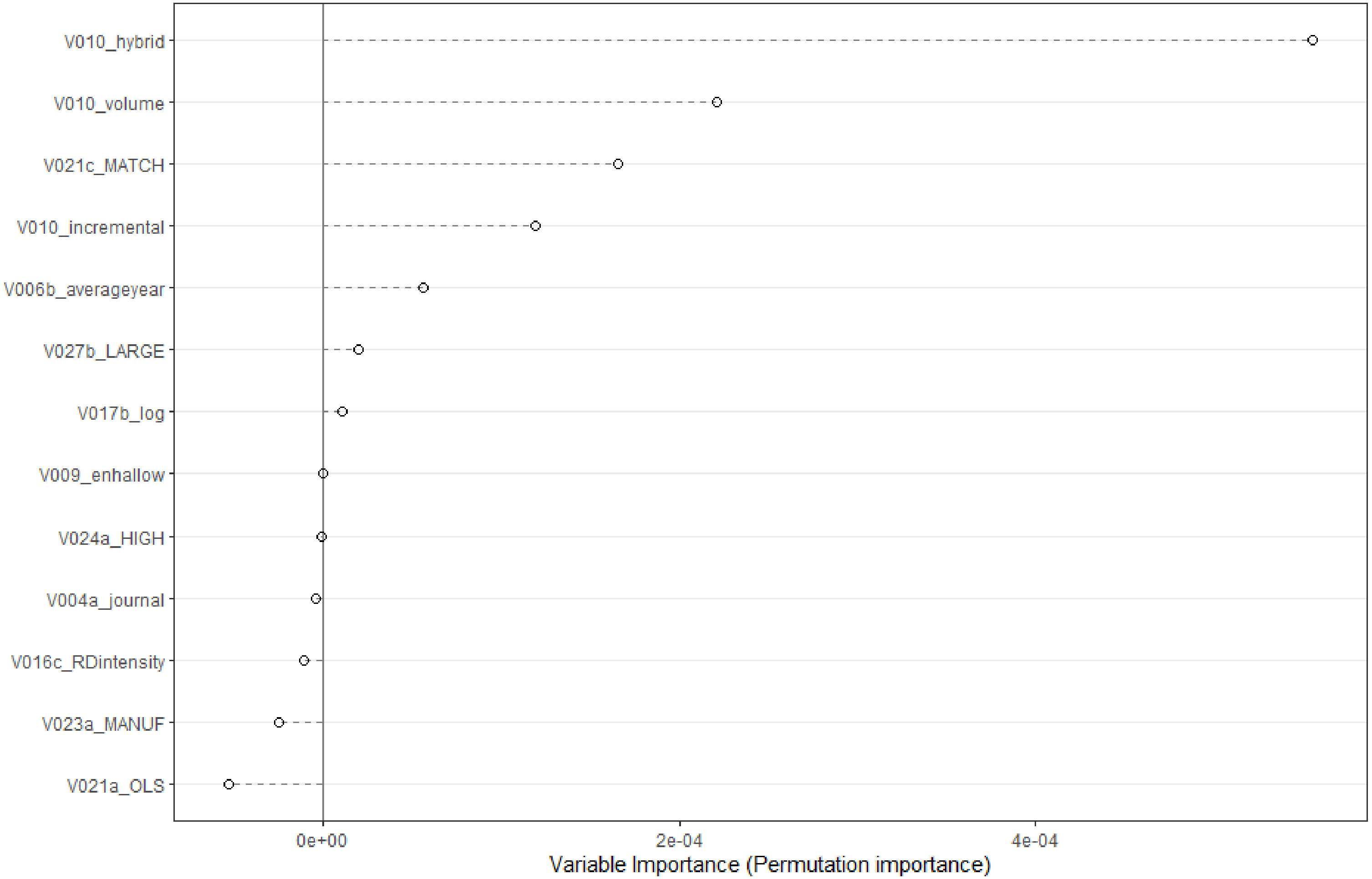


Figure 3: Variable importance plot



Notes: This figure illustrates the importance of the variables from MetaForest. The explanatory variables are ranked according to their predictive power, from the highest, at the top, to the lowest, at the bottom.

Table 1: Distribution of the additionality effects and PCCs

	Min	**1st Quantile**	**Median**	**3rd Quantile**	**Max**	**Mean**	**Std. Dev.**
Additionality Effects	-2.810	0.137	0.341	0.978	36.430	1.752	5.120
PCCs	-0.067	0.019	0.046	0.092	0.277	0.061	0.060

Table 2: Primary studies

#	Author and Publication Year	Country	Period	Instrument	Type	PCCs					Std. Dev.
						N	Mean	Median	Min	Max	
1	Acheson and Malone (2016)[n]	Ireland	2007–2014	tax credit	incremental	1	0.042	0.042	0.042	0.042	-
2	Agrawal et al (2017)[n]	Canada	2000–2007	tax credit	volume-based	3	0.019	0.018	0.017	0.022	0.003
3	Aristei et al. (2015)[n]	France Spain	2007–2009	tax credit	volume-based hybrid	2	0.086	0.086	0.063	0.110	0.033
4	Berger (1993)	USA	1997–1989	tax credit	incremental	2	0.057	0.057	0.047	0.067	0.014
5	Billings and Fried (1999)	USA	1994	tax credit	incremental	1	0.214	0.214	0.214	0.214	-
6	Billings et al. (2001)	USA	1992–1998	tax credit	incremental	2	0.024	0.024	0.017	0.031	0.010
7	Bozio et al. (2014)	France	2004–2010	tax credit	volume-based	6	0.035	0.030	0.018	0.053	0.014
8	Cantabane and Nascia (2014)[n]	Italy	2007–2009	tax credit	volume-based	2	0.063	0.063	0.052	0.073	0.014
9	Corchuelo and Martínez-Ros (2010)	Spain	1998–2002	tax credit	hybrid	18	0.101	0.105	0.016	0.154	0.039
10	Guceri (2015)[n]	UK	1998–2006	enhanced allowance	volume-based	10	0.031	0.035	0.017	0.037	0.008
11	Guceri (2018)[n]	UK	1999–2013	enhanced allowance	volume-based	10	0.036	0.036	0.025	0.048	0.011
12	Guceri and Liu (2019)[n]	UK	2003–2011	enhanced allowance	volume-based	6	0.046	0.047	0.038	0.057	0.006
13	Hægeland and Møen (2007)	Norway	1993–2005	tax credit	volume-based	16	0.077	0.075	-0.067	0.277	0.067
14	Ho (2006)	USA	1975–1999	tax credit	incremental	72	0.024	0.018	-0.056	0.112	0.030
15	Holt et al. (2016)[n]	Australia	2005–2011 2011–2012	enhanced allowance tax credit	hybrid volume-based	8	0.065	0.066	0.006	0.148	0.054
16	Huang and Yang (2009)	Taiwan	2001–2005	tax credit	hybrid	9	0.130	0.133	0.098	0.166	0.023
17	Kobayashi (2014)[n]	Japan	2009	tax credit	hybrid	15	0.180	0.186	0.070	0.271	0.057
18	Paff (2005)	USA	1994–1999	tax credit	incremental	6	0.084	0.072	0.068	0.133	0.025
19	Yang et al. (2012)	Taiwan	2001–2005	tax credit	hybrid	3	0.059	0.053	0.09	0.094	0.033
	Total Meta-Sample	-	1975–2014	-	-	192	0.061	0.046	-0.067	0.277	0.060

[n] Primary study was not subject to prior MRA by Castellacci and Lie (2015) or Gaillard-Ladinska et al. (2015).

Table 3: Summary statistics of moderator variables

Moderator Variables	Description	Summary Statistics		
		Estimations	Mean	Std. Dev.
Publication Bias				
PCC Standard Error	= PPCs' standard errors	-	0.031	0.022
PCC Squared Standard Error	= PPCs' squared standard errors	-	0.001	0.002
Squared Standard Error	= additionality effects' squared standard errors	-	1.577	6.275
Published Study	= 1 if the study is published in a journal, and 0 otherwise	65	0.339	0.474
Definition of Dependent Variable				
Logarithm	= 1 if the estimate relies on the logarithm of R&D expenditures as the dependent variable, and 0 otherwise	132	0.688	0.465
R&D Intensity	= 1 if the estimate relies on R&D intensity as the dependent variable, and 0 otherwise	23	0.120	0.326
Estimation Strategy				
OLS Estimator	= 1 if the estimate relies on an OLS estimator, and 0 otherwise	26	0.135	0.343
Matching	= 1 if the estimate relies on matching, and 0 otherwise	96	0.500	0.501
DiD	= 1 if the estimate relies on a DiD estimator, and 0 otherwise	70	0.365	0.483
Firm Characteristics: Size and Industry				
Large Firms	= 1 if the estimate relies on a subsample of large firms, and 0 otherwise	50	0.260	0.440
High-Tech Firms	= 1 if the estimate relies on a subsample of high-tech firms, and 0 otherwise	47	0.245	0.431
Manufacturing Firms	= 1 if the estimate relies on a subsample of manufacturing firms, and 0 otherwise	38	0.198	0.399
Design of R&D Tax Policy				
Incremental Scheme	= 1 if a country to which the respective estimate refers has adopted an incremental scheme, and 0 otherwise	84	0.438	0.497
Hybrid Scheme	= 1 if a country to which the respective estimate refers has adopted a hybrid scheme, and 0 otherwise	50	0.260	0.440
Volume-Based Scheme	= 1 if a country to which the respective estimate refers has adopted a volume-based scheme, and 0 otherwise	58	0.302	0.460
Enhanced Allowance	= 1 if a country to which the respective estimate refers has adopted an enhanced allowance, and 0 otherwise	30	0.156	0.364
Time Effect				

Average Sample Year	= average sample year of the respective estimate, normalized between 0 and 1	-	0.504	0.295

Table 4: Main results of MRA

Moderator Variables	Without PCC Standard Error (1)	(2)	PCC Standard Error (3)	(4)	PCC Squared Standard Error (5)	(6)
Constant	0.021***	0.034***	0.013***	0.033***	0.020***	0.035***
	(0.001)	(0.004)	(0.003)	(0.005)	(0.002)	(0.004)
Publication Bias						
PCC Standard Error			1.166***	0.214		
			(0.398)	(0.505)		
PCC Squared Standard Error					11.211	4.149
					(7.001)	(8.098)
Published Study	0.026**	0.027*	0.014	0.025*	0.023**	0.026*
	(0.011)	(0.014)	(0.008)	(0.014)	(0.009)	(0.013)
Definition of Dependent Variable						
Logarithm	-0.012***	-0.013***	-0.012***	-0.014***	-0.012***	-0.013***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
R&D Intensity	-0.034*	-0.039	-0.041**	-0.040	-0.041**	-0.041
	(0.018)	(0.028)	(0.014)	(0.028)	(0.016)	(0.028)
Estimation Strategy						
OLS Estimator	0.056***	0.077***	0.049***	0.076***	0.054***	0.077***
	(0.005)	(0.021)	(0.006)	(0.021)	(0.005)	(0.021)
Matching	0.014	0.048***	-0.001	0.046***	0.007	0.046***
	(0.010)	(0.010)	(0.005)	(0.008)	(0.009)	(0.008)
Firm Characteristics: Size and Industry						
Large Firms	0.006*	0.007*	0.001	0.006*	0.005*	0.007**
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
High-Tech Firms	-0.010***	-0.007**	-0.011***	-0.007**	-0.010***	-0.007**
	(0.002)	(0.003)	(0.001)	(0.003)	(0.002)	(0.003)
Manufacturing Firms	-0.006	0.007	-0.018	0.005	-0.012	0.005
	(0.019)	(0.023)	(0.015)	(0.023)	(0.017)	(0.023)
Design of R&D Tax Policy						
Incremental Scheme		-0.018***		-0.018***		-0.018***
		(0.006)		(0.006)		(0.006)

Hybrid Scheme	0.072***		0.083***		0.076***	
	(0.008)		(0.007)		(0.007)	
Volume-Based Scheme	0.008		0.010***		0.009**	
	(0.005)		(0.003)		(0.004)	
Number of Observations	192	192	192	192	192	192
Number of Primary Studies	19	19	19	19	19	19
Variance Inflation Factor (VIF)	1.73	1.53	1.91	1.64	1.76	1.56
Adj. R-squared	0.550	0.438	0.569	0.436	0.556	0.436

Notes: The PCC is the dependent variable. All moderator variables are coded as binary dummy variables, except for *PPC Standard Error* and *PCC Squared Standard Error*. I use an FE pooled weighted least squares (WLS) estimator. The analytical weights are *PPC Squared Standard Errors*. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively. Standard errors are in parentheses and clustered at the study level to control for autocorrelation (within-study dependency).

Table 5: Robustness tests

Moderator Variables	Without Ho (2006)		Without Outliers	
	(1)	(2)	(3)	(4)
Constant	0.011	0.017***	0.022***	0.031***
	(0.017)	(0.004)	(0.001)	(0.005)
Publication Bias				
PCC Squared Standard Error	24.784**	26.226**	9.160	11.299
	(9.693)	(10.523)	(6.459)	(10.070)
Published Study	0.018**	0.021*	0.020***	0.029*
	(0.008)	(0.010)	(0.006)	(0.014)
Definition of Dependent Variable				
Logarithm			-0.011***	-0.013***
			(0.002)	(0.002)
R&D Intensity	-0.038*	-0.040	-0.038**	-0.030
	(0.019)	(0.025)	(0.015)	(0.021)
Estimation Strategy				
OLS Estimator	0.053***	0.080***	0.047***	0.039***
	(0.004)	(0.022)	(0.003)	(0.006)
Matching	0.002	0.054***	0.005	0.023
	(0.006)	(0.009)	(0.012)	(0.019)
Firm Characteristics: Size and Industry				
Large Firms	0.002	0.003	0.003	0.008**
	(0.008)	(0.008)	(0.003)	(0.003)
High-Tech Firms	0.004	0.006	-0.011***	-0.008***
	(0.015)	(0.020)	(0.001)	(0.003)
Manufacturing Firms	-0.021	-0.017	-0.029*	0.022
	(0.017)	(0.018)	(0.016)	(0.021)
Design of R&D Tax Policy				
Incremental Scheme		-0.026		-0.013**
		(0.026)		(0.006)
Hybrid Scheme	0.075***		0.110***	
	(0.018)		(0.019)	

Volume-Based Scheme	0.007		0.007	
	(0.017)		(0.004)	
Number of Observations	120	120	186	186
Number of Primary Studies	18	18	19	19
VIF	1.76	1.42	1.76	1.55
Adj. R-squared	0.481	0.349	0.669	0.523

Notes: The PCC is the dependent variable. All moderator variables are coded as binary dummy variables, except for *PCC Squared Standard Error*. I use an FE pooled WLS estimator. The analytical weights are *PPC Squared Standard Errors*. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively. Standard errors are in parentheses and clustered at the study level to control for autocorrelation (within-study dependency).

Table 6: Design aspects of R&D tax incentives

Moderator Variables	Design Aspects			Time Effect	
	(1)	(2)	(3)	(4)	(5)
Constant	0.012***	0.049***	0.023**	0.024***	-0.012
	(0.002)	(0.011)	(0.008)	(0.004)	(0.019)
Publication Bias					
PCC Squared Standard Error	12.009**	19.704*	20.857*	11.179	5.278
	(5.647)	(9.745)	(10.217)	(7.016)	(8.030)
Published Study				0.024**	0.023*
				(0.010)	(0.013)
Definition of Dependent Variable					
Logarithm				-0.012***	-0.015***
				(0.001)	(0.003)
R&D Intensity				-0.043**	-0.035
				(0.017)	(0.027)
Estimation Strategy					
OLS Estimator				0.052***	0.081***
				(0.006)	(0.016)
Matching				0.007	0.043***
				(0.009)	(0.007)
Firm Characteristics: Size and Industry					
Large Firms				0.004	0.010**
				(0.003)	(0.004)
High-Tech Firms				-0.010***	-0.006
				(0.001)	(0.004)
Manufacturing Firms				-0.013	0.006
				(0.017)	(0.026)
Design of R&D Tax Policy					
Incremental Scheme		-0.039***			0.017
		(0.010)			(0.017)
Hybrid Scheme	0.098***			0.089***	
	(0.006)			(0.020)	

Volume-Based Scheme	0.019*			0.020	
	(0.010)			(0.014)	
Enhanced Allowance			0.050*		
			(0.027)		
Time Effect					
Average Sample Year				-0.018	0.060**
				(0.024)	(0.021)
Number of Observations	192	192	192	192	192
Number of Primary Studies	19	19	19	19	19
VIF	1.12	1.00	1.00	4.38	3.28
Adj. R-squared	0.443	0.172	0.194	0.555	0.447

Notes: The PCC is the dependent variable. All moderator variables are coded as binary dummy variables, except for *PCC Squared Standard Error* and *Average Sample Year*. I use an FE pooled WLS estimator. The analytical weights are *PPC Squared Standard Errors*. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively. Standard errors are in parentheses and clustered at the study level to control for autocorrelation (within-study dependency).

Table 7: Average true effects across methodological aspects

Definition of the Dependent Variable	*Logarithm*		
Estimation Strategy	*OLS Estimator*	*Matching*	*DiD*
Moderator Variables	(1)	(2)	(3)
Constant	0.486***	0.426***	0.040***
	(0.062)	(0.041)	(0.007)
Publication Bias			
Squared Standard Error	4.811	1.238	11.573***
	(4.622)	(1.076)	(1.863)
Number of Observations	21	54	57
Number of Primary Studies	3	5	10
VIF	1.00	1.00	1.00
Adj. R-squared	0.004	0.006	0.402

Notes: The dependent variable is the additionality effect. I use an FE pooled WLS estimator. The analytical weights are *Squared Standard Errors*. ***, **, and * indicate significance levels of 0.01, 0.05, and 0.1, respectively. Standard errors are in parentheses.

Appendix

Figure A1: Convergence plot

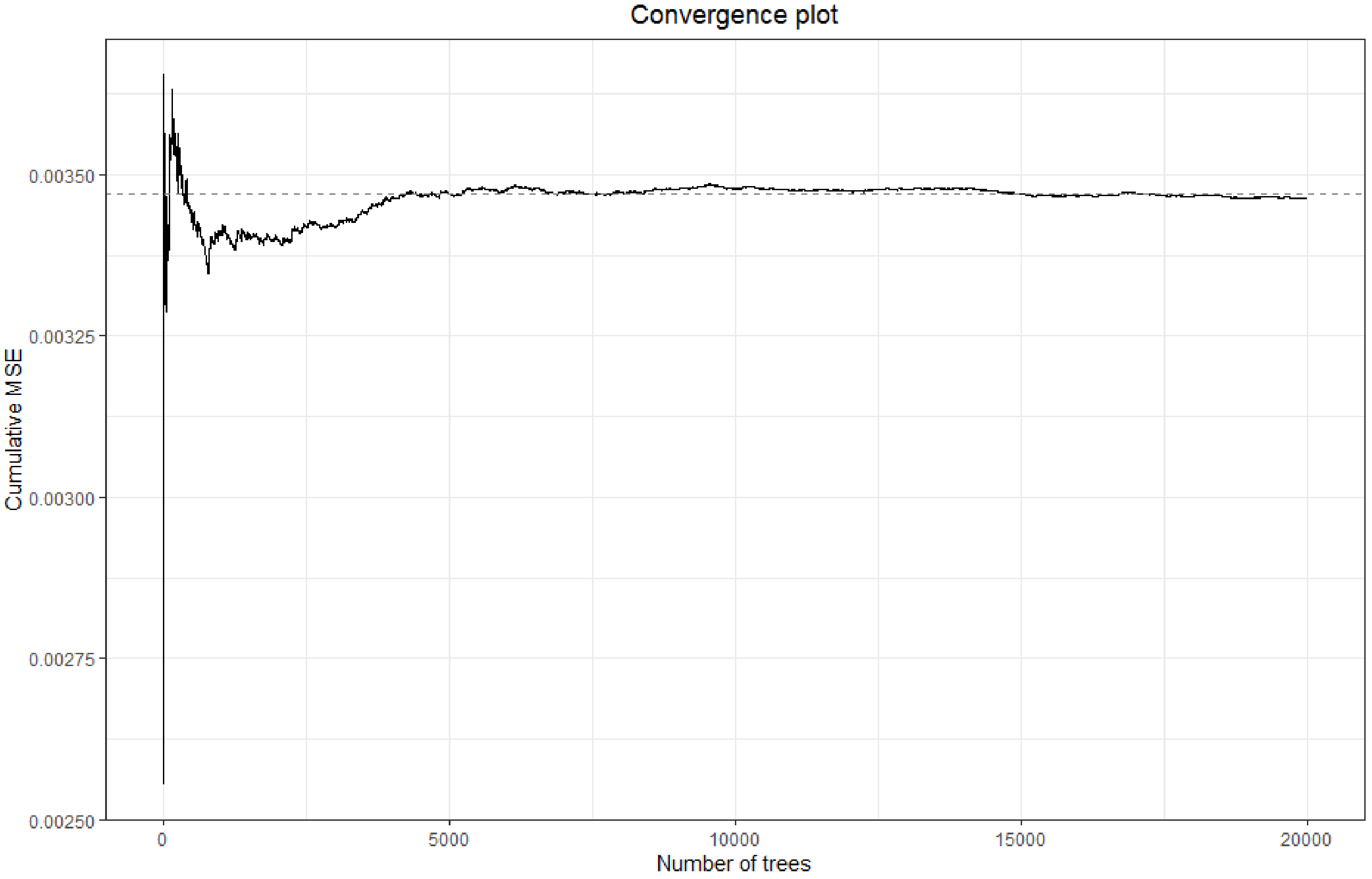


Table A1: Correlations between moderator variables

Moderator Variables	*Inverse Variance*	*Published Study*	*Logarithm*	*R&D Intensity*	*OLS Estimator*	*Matching*	*DiD*	*Large Firms*	*High-Tech Firms*	*Manufacturing Firms*	*Incremental Scheme*	*Hybrid Scheme*	*Volume-Based Scheme*	*Enhanced Allowance*	*Average Sample Year*
Inverse Variance	1.000														
Published Study	-0.093	1.000													
Logarithm	0.844	-0.045	1.000												
R&D Intensity	-0.080	0.190	-0.104	1.000											
OLS Estimator	0.258	-0.069	0.314	0.058	1.000										
Matching	0.264	-0.073	0.320	-0.037	-0.054	1.000									
DiD	0.858	-0.042	0.647	-0.094	-0.105	-0.100	1.000								
Large Firms	0.167	-0.108	0.053	-0.056	-0.057	-0.048	0.221	1.000							
High-Tech Firms	0.407	-0.119	0.095	-0.058	-0.063	-0.054	0.482	0.244	1.000						
Manufacturing Firms	-0.137	0.027	-0.106	0.085	0.013	-0.009	-0.148	-0.028	-0.081	1.000					
Incremental Scheme	0.558	-0.136	0.164	-0.021	-0.076	-0.070	0.655	0.351	0.736	-0.106	1.000				
Hybrid Scheme	0.365	-0.059	0.427	-0.027	0.587	0.582	-0.082	-0.043	-0.049	-0.004	-0.065	1.000			
Volume-Based Scheme	0.617	0.036	0.731	-0.073	0.055	0.062	0.612	-0.071	-0.097	-0.086	-0.127	-0.067	1.000		
Enhanced Allowance	0.354	0.184	0.428	-0.055	0.546	0.532	-0.056	-0.033	-0.070	-0.087	-0.091	0.939	-0.018	1.000	
Average Sample Year	0.833	-0.018	0.901	-0.090	0.350	0.401	0.588	-0.030	0.013	-0.101	0.023	0.525	0.786	0.527	1.000

Notes: This table shows the correlations between the moderator variables. The moderator variables are weighted by their inverse variances.

References

Agrawal, A., Rosell, C. and Simcoe, T.S. (2017) Tax credits and small firm R&D spending Boston University Questrim School of Business Research Paper.

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

Appelt, S., Bajgar, M., Criscuolo, C. and Galindo-Rueda, F. (2016) R&D tax incentives: Evidence on design, incidence and impacts. OECD Science, Technology and Industry Policy Paper No. 32.

Aristei, D., Sterlacchini, A. and Venturini, F. (2015) The effects of public supports in business RD: Firm-level evidence across EU countries. MPRA Paper No. 64611.

Arrow, K.J. (1962) Economic welfare and the allocation of resources for invention. In Nelson, R.R. (ed.) *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 609–625). Princeton: Princeton University Press.

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

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

Billings, B.A. and Fried, Y. (1999) The effects of taxes and organizational variables on research and development intensity. *R&D Management 29*(3): 289-301.

Billings, B.A., Glazunov, S.N. and Houston, M. (2001) The role of taxes in corporate research and development spending. *R&D Management 31*(44): 465-477.

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

Bozio, A., Irac, D. and Py, L. (2014) Impact of research tax credit on R&D and innovation: Evidence from 2008 French reform. BDF Working Paper Series No. 532.

Cantabane, 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.

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.

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.

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, Economics Series Working Paper No. 2011/5.

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

Feld, L.P. and Heckemeyer, J.H. (2011) FDI and taxation: A meta-study. *Journal of Economic Surveys 25*(2): 233-272.
Gaillard-Ladinska, E., Non, M. and Straathof, B. (2015) More R&D with tax incentives? A meta-analysis. CPB Discussion Paper No. 309.
Guceri, I. (2015) Tax incentives and R&D: An evaluation of the 2002 UK reform using micro data. Oxford University Centre for Business Taxation, Working Paper No. 15/11.
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): 266-291.
Hægeland, T., Møen, J. (2007) Input additionality in the Norwegian R&D tax credit scheme. Statistics Norway, Report No. 2007/47.
Hall, B.H. and Jorgenson, D.W. (1967) Tax policy and investment behavior. *American Economic Review 57*(3): 391-414.
Hall, B.H. 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.
Harris, R., Li, Q.C. and Trainor, M. (2009) Is a higher rate of R&D tax credit a panacea for low levels of R&D in disadvantaged regions? *Research Policy 38*(1): 192-205.
Havránek, T., Stanley, T.D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwaski, I., Reed, W.R., Rost, K. and Van Aert, R.C.M. (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. MPRA Paper No. 47937.
Holt, J., Skali, A. and Thomson, R. (2016) The additionality of R&D tax policy in Australia. Centre for Transformative Innovation, Working Paper No. 3/16.
Huang, C.-H. and Yang, C-H. (2009) Tax incentives and R&D activity: Firm-level evidence from Taiwan. Hitotsubashi University, Hi-Stat Discussion Paper No. 102.
James, G.,Witten, D., Hastie, T., Tibshirani, R. (2017) *An Introduction to Statistical Learning: With Applications in R.* New York: Springer.
Klassen, K.J., Pittman, J.A. and Reed, M.P. (2004) A cross-national comparison of R&D expenditure decisions: Tax incentives and financial constraints. *Contemporary Accounting Research 21*(3): 639-680.
Kobayashi, Y. (2014) Effect of R&D tax credits for SMEs in Japan: A microeconometric analysis focused on liquidity constraints. *Small Business Economy 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. *School of Public Policy Publications 7*(36): 1-45.
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.
Moreno, S.G., Sutton, A.J., Ades, A.E., Stanley, T.D., Abrams, K.R., Peters, J.L. and Cooper, N.J. (2009) Assessment of regression-based methods to adjust for publication bias through comprehensive simulation study. *BMC Medical Research Methodology 9*(1): 2.

Organisation for Economic Co-operation and Development (2018) OECD Review of National R&D Tax Incentives and Estimates of R&D Tax Subsidy Rates, 2017. Retrieved from http://www.oecd.org/sti/rd-tax-stats-design-subsidy.pdf (accessed August 17, 2020).

Organisation for Economic Co-operation and Development (2020) *Gross Domestic Spending on R&D (indicator)*. Retrieved from https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm (accessed April 17, 2020).

Paff, L.A. (2005) State-level R&D tax credits: A firm-level analysis. *Journal of Economic Analysis & Policy 5*(1): 17-41.

Pearl, J. (2009) *Understanding Propensity Scores. Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Romer, P.M. (1986) Increasing returns and long-run growth. *Journal of Political Economy 94*(5):1002-1037.

Schumpeter, J.A. (1942) *Capitalism, Socialism, Democracy*. New York: Harper and Brothers.

Solow, R.M. (1956) A contribution to the theory of economic growth. *Quarterly Journal of Economics 70*(1): 65-94.

Spengel, C., Rammer, C., Nicolay, K., Pfeiffer, O., Werner, A.-C., Olbert, M., Blandinières, F., Hud, M. and Peters, B. (2017) Steuerliche FuE-Förderung, Studie im Auftrag der Expertenkommission Forschung und Innovation. Studien zum deutschen Innovationssystem No. 15-2017.

Stanley, T. (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.* Abingdon: Routledge.

Straathof, B., Gaillard-Ladinska, E, Kox, H., Mocking, R., Goldberg, I., Jensen, C., Lindholm, P., Sobolewksi, M., Berg-Andersson, B., Kaseva, H., Rouvinen, P., Valkonen, T., Sellner, R., Schönpflug, K. and Paterson, I. (2014) A study on R&D tax incentives: Final report of CPB, Den Haag. Working Paper No. 52 - 2014.

Van Lissa, C.J. (2020a) Small sample meta-analyses: Exploring heterogeneity using MetaForest. In Miocevic, M., Van de Schoot, R. (ed.) *Small Sample Size Solutions* (pp. 186-202). Abingdon: Routledge.

Van Lissa, C.J. (2020b) MetaForest: Exploring heterogeneity in meta-analysis using random forests (version 0.1.3) (R-package). Retrieved from https://CRAN.R-project.org/package=metaforest (accessed August 17, 2020).

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.

Impressum:
Arbeitskreis Quantitative Steuerlehre, arqus, e.V.
Vorstand: Prof. Dr. Ralf Maiterth (Vorsitzender),
Prof. Dr. Kay Blaufus, Prof. Dr. Dr. Andreas Löffler
Sitz des Vereins: Berlin

Herausgeber: Kay Blaufus, Jochen Hundsdoerfer,
Martin Jacob, Dirk Kiesewetter, Rolf J. König,
Lutz Kruschwitz, Andreas Löffler, Ralf Maiterth,
Heiko Müller, Jens Müller, Rainer Niemann,
Deborah Schanz, Sebastian Schanz, Caren Sureth-Sloane, Corinna Treisch

Kontaktadresse:
Prof. Dr. Caren Sureth-Sloane, Universität Paderborn,
Fakultät für Wirtschaftswissenschaften,
Warburger Str. 100, 33098 Paderborn,
www.arqus.info, Email: info@arqus.info

ISSN 1861-8944