PEGFA | Institute of Political Economy, Governance, Finance and Accountability

Greenwich Papers in Political Economy

What do we know about R&D spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power

Mehmet Ugur[a], Sefa Awaworyi Churchill[b] and Hoang M. Luong[a]

a: University of Greenwich; b: RMIT University

Abstract

As Mohnen (1996: 40) has indicated, research and development (R&D) externalities is a two-sided theoretical issue. Its 'dark' side concerns the under-investment problem caused by non-appropriability of R&D benefits. On the 'bright' side, R&D spillovers are a source of productivity gains. Both aspects have been invoked to justify public support for R&D investment directly and indirectly. To establish whether public support can be justified due to productivity gains from spillovers, we meta-analyse 983 productivity estimates for spillovers and 501 estimates for own-R&D from 60 empirical studies. Our findings indicate that the average spillover effect is: (i) positive but heterogenous and smaller than what is reported in most narrative reviews; (ii) usually smaller than that of own-R&D capital; (iii) too small to be practically significant when evidence with adequate statistical power is considered. Controlling for observable sources of heterogeneity and best-practice research, the meta-effect is insignificant in the full sample but significant and large among OECD firms/industries/countries. We discuss the implications of these findings for future research and public support for R&D investment.

Year: 2019

No: ???

Keywords: Knowledge externalities, R&D spillovers, productivity, public policy, meta-analysis

Acknowledgments:

This paper was presented at the Meta-Analysis of Economic Research Network (MAER-Net) 2017 Colloquium at Zeppelin University, Germany. We would like to thank the organising committee for reviewing the paper. We would also like to thank GPERC members Edna Solomon and Alex Guschanski for their insightful comments.

Corresponding authors: Mehmet Ugur, Department of International Business and Economics, University of Greenwich, Park Row, Old Royal Naval College, SE10 9LS, London, UK. Email: M.Ugur@greenwich.ac.uk

What do we know about R&D spillovers and productivity?
Meta-analysis evidence on heterogeneity and statistical power

Mehmet Ugur[a], Sefa Awaworyi Churchill[b] and Hoang M. Luong[a]

a: University of Greenwich;   b: RMIT University

Abstract

As Mohnen (1996: 40) has indicated, research and development (R&D) externalities is a two-sided theoretical issue. Its 'dark' side concerns the under-investment problem caused by non-appropriability of R&D benefits. On the 'bright' side, R&D spillovers are a source of productivity gains. Both aspects have been invoked to justify public support for R&D investment directly and indirectly. To establish whether public support can be justified due to productivity gains from spillovers, we meta-analyse 983 productivity estimates for spillovers and 501 estimates for own-R&D from 60 empirical studies. Our findings indicate that the average spillover effect is: (i) positive but heterogenous and smaller than what is reported in most narrative reviews; (ii) usually smaller than that of own-R&D capital; (iii) too small to be practically significant when evidence with adequate statistical power is considered. Controlling for observable sources of heterogeneity and best-practice research, the meta-effect is insignificant in the full sample but significant and large among OECD firms/industries/countries. We discuss the implications of these findings for future research and public support for R&D investment.

Keywords: Knowledge externalities, R&D spillovers, productivity, public policy, meta-analysis

JEL Codes: D24, O30, O32, O33, C1

1. Introduction

The effect of knowledge externalities on productivity has direct relevance for public policy and welfare. Griliches (1979; 1992) has contributed to the debate by introducing the notion of external knowledge capital, proxied by the level of external R&D capital stock. This approach ties in with macro-level endogenous growth models where investment in innovation is associated with increasing returns (Grossman and Helpman, 1991; Romer, 1990).

In an early review of the empirical research, Griliches (1992) acknowledges the risk of selection bias but concludes that the productivity effects or R&D spillovers are significant and usually larger than those of own R&D capital. Mohnen (1996) acknowledges that the rates of return on external R&D are estimated less precisely than elasticities, but he also affirms that returns on external R&D are larger than those of own R&D by 50%-100%. Similarly, Cincera and Van Pottelsberghe de La Potterie (2001) report that: (i) international spillovers contribute to productivity growth substantially; (ii) the productivity effects are larger in countries with a higher degree of openness to imports; and (iii) the spillover effects are often larger than those of domestic (own) R&D. Only a more recent review by Hall et al. (2010) reports that spillover and own-R&D effects are similar.

This paper aims to provide a comprehensive synthesis of the evidence base by drawing on latest developments in meta-regression analysis. It estimates the average productivity effects of own R&D and R&D spillovers in a unified framework, compares the effect-size estimates from the full sample with estimates based on adequately powered evidence, and distils evidence-based implications for future research and public policy. Other contributions of the paper include: (i) evaluation of the publication selection bias through alternative measures; (ii) using a hierarchical modelling framework that takes account of the nested nature of the evidence base; (iii) controlling for endogeneity that may result from correlation of the regressors with the random effects or idiosyncratic errors; (iv) reporting average effect-size estimates for different spillover types and comparing the latter with own R&D effects; (iii) using a frequentist model averaging method to specify the multivariate meta-regression model and obtain ‘best-practice’ meta-effect estimates in different contexts.

Our findings suggest that the productivity effect of R&D spillovers is smaller than what is reported in narrative reviews. Also, it is usually smaller and estimated with lower precision compared to the effect of own R&D. Furthermore, the effect is context specific: it is larger (smaller) in countries where absorptive capacity is higher (lower), depending on the history and level of investment in own R&D. We also find that more than two-thirds of the spillover effect estimates are based on evidence with low statistical power; and the average effect size based on adequately powered evidence is too small to be practically significant. Our findings do not invalidate the hypothesis that external R&D may be a source of productivity gains, but they indicate that the productivity effects of R&D spillovers is smaller than what is usually claimed; vary across countries, spillover measures and industries; and call for more innovation in measurement and estimation.

The paper is organised in six sections. In section 2, we present the theoretical/empirical framework that underpins the findings in the research field. In section 3, we present our search strategy and inclusion/exclusion criteria, comment on the distribution of the effect-size estimates and provide preliminary evidence on publication selection bias. Our meta-regression methodology is discussed in section 4 while Section 5 presents the empirical findings. Section 6 concludes by highlighting the implications our findings for policy and future research.

2. Spillovers and productivity: the theoretical/empirical framework

The effect-size estimates we analyse are extracted from studies that adopt the so-called *primal approach*, which draws on a Cobb-Douglas production function augmented with *own-R&D capital* and *external R&D capital* (Griliches, 1979). The augmented production function can be stated as follows:

$$Y_{it} = Ae^{\lambda t}C_{it}^{\alpha}L_{it}^{\beta}K_{it}^{\gamma}S_{it}^{\varphi}e^{u_{it}} \quad (1)$$

Here, Y is real output; C is physical capital stock net of depreciation; L is labour (number of employees or hours worked); K is own R&D capital stock net of depreciation; S is the spillovers pool (as specified below); λ is the rate of disembodied technological change; and A is a constant. Subscripts t and i denote time and cross-section units (firms, industries or countries/regions), respectively. Two standard assumptions underpinning (1) are constant returns to scale and continuous optimisation by the production unit.

The spillover pool (S) available to unit i is the weighted sum of the R&D capital stocks of other units (j) where $j \neq i$ and can be unscaled (1a) or scaled (1b).

$$S_{it} = \sum_{j=1}^{n} W_{ij} K_{jt} \quad (1a)$$

$$S_{it} = \sum_{j=1}^{n} a_i W_{ij} K_{jt} \quad (1b)$$

The weight W_{ij} (or W_{ijt} if the weight is calculated for each year rather than as an average for the analysis period) is a vector that captures either technological proximity or transaction intensity between i and j. In (1b), a_i is an additional weight that captures the spillover-recipient's openness to international imports or 'intermediate trade' with units in j. Several studies (e.g., Coe et al., 1997; Keller, 1998; Krammer, 2010) utilize the additional weight arguing that the productivity effects of spillovers depend not only on bilateral import or transaction shares but also on the beneficiary's openness to import or transaction with the 'rest of the world'. However, it must be noted that the spillover pools obtained from (1b) are smaller than that in (1a) by construction as both a_i and W_{ijt} are fractions. Finally, K_{jt} is the R&D capital of unit *j* in period *t*.

Taking natural logarithms and using lower-case letters to denote logged values, we obtain (2a) below. The log of technical progress ($Ae^{\lambda t}$) yields a unit-specific effect (η_i) and a time effect (λt). The coefficients of main interest are γ and φ: the *elasticities* of output with respect to own R&D and the spillover pool respectively.

$$y_{it} = \eta_i + \lambda t + \alpha c_{it} + \beta l_{it} + \gamma k_{it} + \varphi s_{it} + u_{it} \quad (2a)$$

Country-level studies that estimate the effects of international spillovers utilise a total factor productivity (TFP) specification as indicated in (2b) below. Coefficient estimates from (2a) and (2b) will be consistent if the assumptions of perfect competition and constant returns to scale are satisfied.

$$LogTFP_{it} = \eta_i + \lambda t + \gamma k_{it} + \varphi s_{it} + u_{it} \quad (2b)$$

In this study, we synthesize evidence from *primal-approach* studies only to ensure that we are pooling comparable evidence, derived from estimating a common production function with quantities as inputs. *Dual-approach* studies, which estimate a system of factor demand equations derived from a cost function representation of technology, are excluded because technology can be represented by different cost, profit or value functions. As such, estimates from the dual approach are less comparable not only with the primal-approach estimates but also among themselves (Hall et al., 2010). The focus on primal-approach evidence only delimits the extent of heterogeneity in the evidence base, but it does not eliminate it. Therefore, and in line with other meta-analysis studies, we extract information about observable sources of heterogeneity and use the latter to model the variation in the evidence base.[1] For this, we utilise a hierarchical modelling framework that allows for both intercept and slope heterogeneity in meta-regression models.

Our study shares a common ground with several meta-analysis studies on the productivity effects of foreign direct investment (FDI) spillovers (see, e.g., Demena and van Bergeijk, 2017; Gorg and Strobl, 2001; Iršová and Havránek, 2013; Mebratie and Bergeijk, 2013; Meyer and Sinani, 2009; and Wooster and Diebel, 2010). Like these studies, we differentiate between different spillover measures and verify whether their productivity effects differ. We also control for selection bias and assess the productivity effects after controlling for selection bias. Finally, we estimate multivariate meta-regression models to identify the observable sources of heterogeneity in the evidence base.

However, we also offer a number of extensions, including: (i) taking account of within-study dependence and nested nature of the meta-analysis data through hierarchical meta-regression models (HMRMs); (ii) addressing issues of endogeneity that may result from correlation between regressors and unobserved random effects or the regression error; (iii) investigating the level of statistical power in the evidence base and the implications of adequate power for effect-size estimates; and (iv) using a frequentist model averaging method to address the issue of model uncertainty in the context multivariate meta-regression models.

[1] See *Box 1* in the *Online Appendix* for discussion on the various sources of heterogeneity and the ways in which information is extracted to model them.

3. Inclusion/exclusion criteria and overview of the research field

To identify eligible studies, we began with work cited in the existing narrative reviews. This sample was augmented through electronic search in *Google Scholar* and in the *Science & Technology Management Bibliography (STMB)* [2] We used a pre-specified list of keywords, including: “R&D spillovers”, “knowledge externalities”, “R&D externalities”, “knowledge capital”, “technology diffusion”, “R&D and productivity”, and “spillovers and productivity”. The search period is from 1980 to 2016, starting one year after the publication of the seminal paper by Griliches (1979).[3]

Following the best-practice recommendations for meta-analysis of economics and business research in Stanley et al. (2013), we screened 2,324 potentially relevant studies using title and abstract information. Screening decisions identified a sample of 106 studies that we then evaluated by reading the full-text.[4] We excluded studies (e.g., Bersntein, 1989) that adopt the dual approach described above; and those that estimate translog or quadratic production functions (e.g., Aiello and Cardamone, 2008; Mairesse and Mulkay, 2008). The dual-approach studies are excluded because they draw on different specifications for factor-demand and cost functions as indicated above. They also use ex-post (as opposed to expected) output on the right-hand-side of their models, increasing the risk of biased estimates (Griliches, 1992: S40). Translog or quadratic model studies are excluded because their estimates of the spillover effects are non-linear. Finally, we also excluded studies that report starred coefficients without standard errors or t-values (e.g., Ang and Madsen, 2013; Coe and Helpman, 1995; Müller and Nettekoven, 1999).

At the end of the full-text evaluation, we obtained a sample of 76 studies that adopt the primal approach. Further evaluation indicated that some of these studies reported rate-of-return instead of elasticity estimates (e.g., Griliches and Lichtenberg, 1984; Hanel, 2000; Mansfield, 1980). We excluded such studies for two reasons. First, rate-of-return estimates are biased if the assumption of zero depreciation for R&D does not hold (Griliches, 1979). Secondly, rate-of-

[2] The STMB database contains references to more than 20,000 articles, books and conference proceedings on R&D management, the management of technological innovation & entrepreneurship, science & technology policy, and technology transfer. See, http://tomeclarke.ca/science.htm
[3] As indicated above, Griliches (1979) has provided a lasting theoretical and empirical framework for the knowledge capital model and identified the range of measurement and estimation issues that empirical researchers need to be aware of.
[4] Screening decisions were made by two researchers whilst a third researcher conducted random checks on the former’s decisions. Evaluation and the following inclusion/exclusion decisions were taken unanimously by three researchers.

return estimates for both own and external R&D are much less precise than the elasticity estimates; and the imprecision is more evident with respect to external R&D (Hall et al., 2010).

Our estimation sample consists of 60 primary studies that report 983 productivity-effect estimates for spillovers and 501 estimates for own R&D at the firm, industry or country/region levels.[5] We have extracted all reported effect-size estimates to ensure full use of existing information and avoid the risk of reviewer-induced selection bias. To control for observed sources of heterogeneity, we coded each estimate with respect to: (i) publication characteristics (publication type and date, journal quality, etc.); (ii) model specification (control for own R&D, time dummies, industry/country dummies, etc.); (iii) data and sample characteristics (unit of analysis, data origin, etc.); (iv) estimation methods (GMM, 2SLS, 3SLS, OLS, panel cointegration, FE, etc.); and (v) spillover types (knowledge, rent or mixed spillovers).[6]

Kernel densities of the effect-size estimates and associated t-values are presented in *Figure 1* for the spillover and own-R&D samples. Most effect-size estimates are positive, but their distribution has long tails to the right. Long right tails are also observed in the distribution of t-values.[7] Moreover, the t-values have the highest density around 2 – near the cut-off point associated with statistical significance at 5%. This may be a sign of publication selection, which reveals itself as a sudden increase in the frequency of effect sizes that just pass the 5% significance level.[8]

[5] In our sample, primary studies based on firm or industry data from developing countries are small in number (e.g., Parameswaran, 2009 and Raut, 1995 with Indian data and Wang and Chao, 2008 with Taiwanese data). This may be due to relative lack of data in developing countries. It is also the case that some studies based on developing country data had to be excluded due to non-compatibility with the primal approach. These include, but are not limited to: Chuang and Lin (1999) – with Taiwanese data, but the model is not specified in log-log form; Wei and Liu (2006) – with Chinese data, but it estimates the effects of intangible asset spillovers as opposed to R&D spillovers; and Johnson and Evenson (2000) – with African data, but the model is not specified in log-log form.

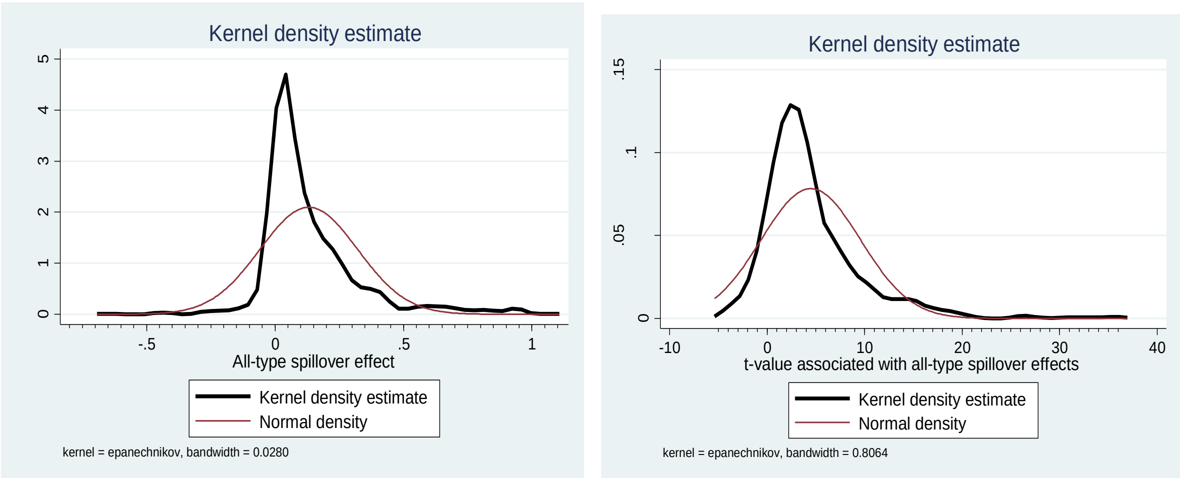
[6] The mixed spillovers category captures the spillover pool constructed with transaction intensity in knowledge-intensive goods/services (e.g., Branstetter, 2001; Griffith et al., 2006; Lee, 2005). As indicated in section 2 above, there is lack of agreement in the literature on whether such spillovers should be considered as knowledge or rent spillovers.

[7] The probabilities of the t-values by four cut-off points are given in *Table 2* below.

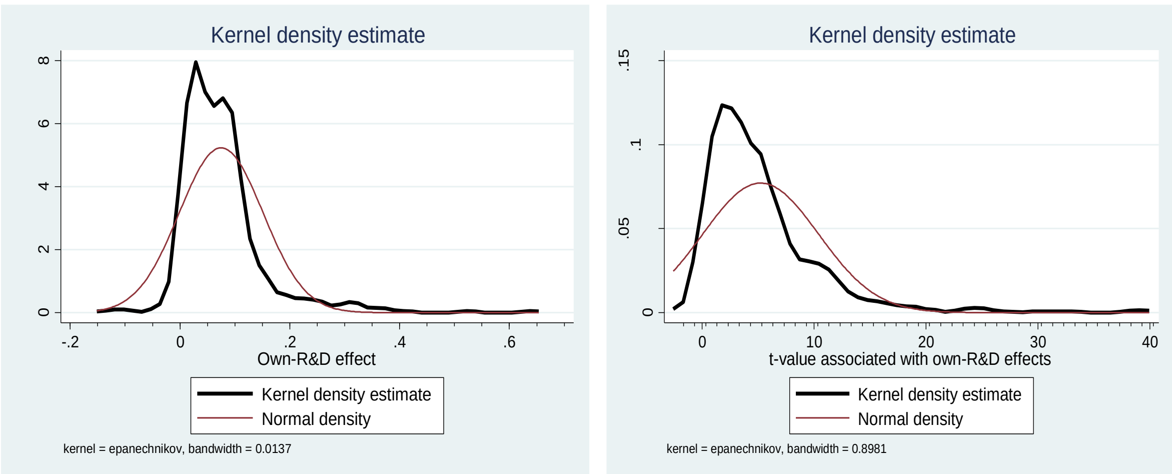
[8] This is similar to Andrews and Kasy (2019), who consider the jumps in the density of the reported estimates around the cut-off points for significance as potential indicators of selection bias.

Figure 1: Kernel densities of effect-size estimates and associated t-values

Panel A: Spillover effects and associated t-values



Panel B: Own-R&D effects and associated t-values



Further information is provided in *Tables A1* and *A2* in the *Online Appendix*. Most studies (97%) are journal articles, with the remaining 3% consisting of working papers. The median effect-size estimate and t-value, respectively, are 0.070 and 3.323 for spillovers and 0.061 and 4.261 for own-R&D. The summary measures indicate that spillover and own-R&D effects are similar, but the former tend to be estimated with lower precision compared to own R&D effects. The within-study median of the effect size is positive in a large majority of the primary studies, with the exception of four studies (Braconier and Sjöholm, 1998; Harhoff, 2000; Kwon, 2004; McVicar, 2002) where the median spillover effect is negative and two studies (Biatour et al., 2011; Braconier and Sjöholm, 1998) with a negative median productivity effect for own R&D.

It is also worth noting that most studies report multiple productivity estimates, with a range from 1 to 102 in the spillover sample and from 1 to 45 in the own-R&D sample. Whilst 22 studies utilise firm-level, 11 studies use industry-level data and 25 studies are based on country data. Of the remaining two, Acharya and Keller (2009) focuses on both countries and industries while Bronzini and Piselli (2009) focuses on regions (in Italy).

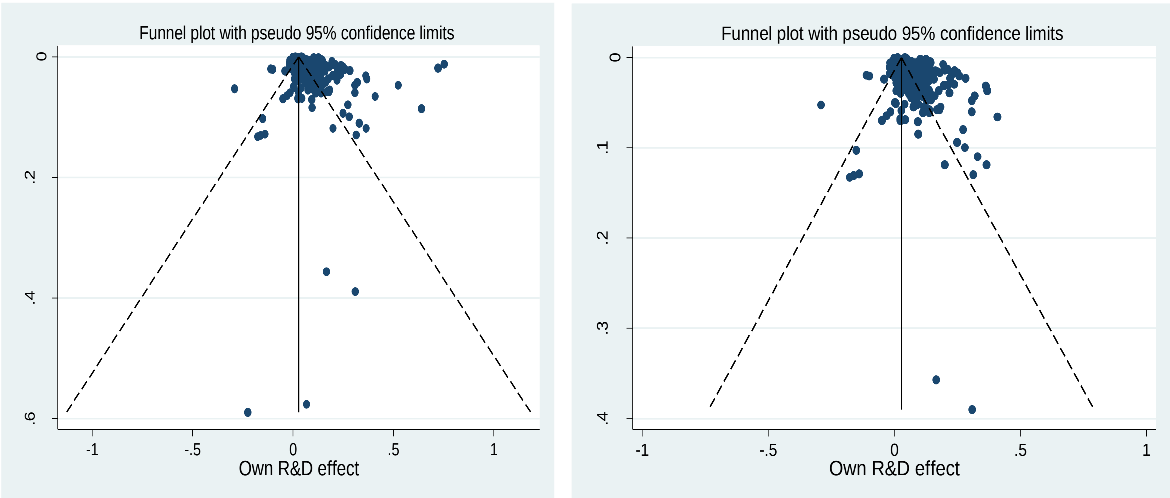
We probe the twin issues of heterogeneity and publication selection further through funnel plot asymmetry in *Figure 2*. The vertical line indicates the fixed-effect weighted mean (FEWM), estimated with weights equal to the reciprocal of the squared standard error. Asymmetric distribution around the FEWM is an indication of publication selection bias. Effect-size estimates beyond the 95% pseudo confidence interval limits (dashed lines) reflect the degree of residual heterogeneity that cannot be explained by sampling variations. In the left panel, outliers are excluded based on *dfbeta* routine in Stata. In the right panel, we implement a more severe restriction by excluding all effect-size estimates with a standard error of more than 0.5.

The funnel graphs indicate that the fixed-effects weighted means (FEWMs) of the effect sizes are similar for spillovers and own-R&D. Also, in both evidence pools, the distribution of the effect sizes around the FEWM is asymmetric, with evident concentration to the right. This is an indication of publication selection bias that needs to be verified using meta-regression and other methods. Third, the effect-size estimates for spillovers are associated with larger standard errors (ranging from 0 to 1.5) compared to standard errors associated with own-R&D effects (ranging from 0 to 0.6). Finally, the extent of residual heterogeneity that cannot be explained by sampling variation is very high in both samples (around 98%).

Figure 2: Funnel graphs for productivity effects of spillovers and own RD

Spillover pool

Own R&D pool



Notes: In the left panel, outlier restriction is based on dfbeta<=1; in the right panel, observation with standard error > 0.5 are excluded. Residual heterogeneity remains at 98.1% for spillovers and 97.9% for own R&D in both scenarios.

We conduct two further checks to verify if the funnel graph asymmetry is a significant indicator of selection bias. First, we follow Card and Krueger (1995) and Gorg and Strobl (2001) and regress the logarithm of the absolute t-value associated with reported estimates on the logarithm of the square root of the corresponding degrees of freedom (LSRDF). The assumption here is that estimates based on larger samples (i.e., estimates with higher degrees of freedom) should be more precise (i.e. should have smaller standard errors and higher t-ratios). The null hypothesis is that publication selection is absent if the coefficient on LSRDF is 1 – i.e., if a 1% increase in the degrees of freedom is associated with a 1% increase in the t-value of the estimate. Results presented in *Table 1* indicate clearly that the coefficients are much smaller than 1 and the null hypothesis of the Wald test is rejected strongly.

Table 1: Card and Krueger (1995) test for publication selection bias

Evidence pool	Coefficient (β)	SE	Wald test statistic H_0: β=1	p-value
All-type SPO	0.2183	0.0984	F(1,1035)= 63.07	0.0000
Own-R&D	0.1114	0.2030	F(1, 496) = 19.16	0.0001

Notes: The dependent variable is the logarithm of the absolute value of the t-ratio. Robust standard errors are clustered at the study level. Other tests for sub-samples consisting of knowledge, rent and mixed spillovers as well as spillovers and own-R&D sub-samples by unit of analysis (i.e., by firm, industry and country levels of data) are similar. The latter are not reported here to save space, but are available upon request.

We then exploit discontinuity (jumps) in the probability of publication around t-values that correspond to significance at 5% (Andrews and Kasy, 2019). This non-parametric test involves estimating probabilities of observing negative or insignificant productivity-effect estimates relative to the probability of observing estimates that are significant and positive in the latent population of studies. The cut-off points for the t-value and the resulting bands are specified in (3) below, where the t-values in the population are associated with probabilities $\beta_{p,1}$, $\beta_{p,2}$, $\beta_{p,3}$, $\beta_{p,4}$ depending on the band they are in.

$$P(t_{value}) \propto \begin{cases} \beta_{p,1} & t_{value} < -1.96 \ : band1 \\ \beta_{p,2} & -1.96 \leq t_{value} < 0 \ : band2 \\ \beta_{p,3} & 0 \leq t_{value} < 1.96 \ : band3 \\ \beta_{p,4} & t_{value} \geq 1.96 \ : band4 \end{cases} \quad (3)$$

Andrews and Kasy (2019) demonstrate that discontinuities (jumps) in the density of the t-values in the sample can be used for identifying discontinuities in their density in the latent population of studies. Therefore, the population probabilities in (3) can be estimated using sample data. We do this through a multinomial logit estimator, treating the probability of observing a t-value in band 4 as the excluded category. The results (not reported here to save space) indicate that the probability of observing t-values in bands 1 – 3 would be smaller than the probability of observing t-values in band 4. Using predictive margins, the relative probabilities for all bands are given in *Table 2*.[9]

Table 2: Relative publication probabilities of observed effect-size estimates

Publication probability	$t_{value} < -1.96$	$-1.96 \leq t_{value} < 0$	$0 \leq t_{value} < 1.96$	$t_{value} \geq 1.96$
All-type SPO	0.0412***	0.0825***	0.2109***	0.6654***
	(0.0096)	(0.0135)	(0.0265)	(0.0323)
Own R&D	0.0080	0.0141*	0.2309***	0.7470***
	(0.0062)	(0.0076)	(0.0434)	(0.0447)
Distribution of t-values				
All-type SPO	39	76	202	666
Own R&D	4	9	116	371

Note: Robust standard errors (in brackets) are clustered at the study level. *, *** indicates significance at 10% and 1%, respectively. Publication probabilities reflect the distribution of the t-values in the sample.

Predicted margins in the last column indicate that the probabilities of observing significant and positive effect-size estimates are about three-times the probabilities of observing positive but insignificant estimates (penultimate column). Findings in *Tables 1* and *2* provide quantitative measures of publication selection (funnel asymmetry) but the latter is estimated independently of the average effect size in the evidence base. In what follows, we discuss our preferred methodology for estimating the publication selection bias and obtaining average effect-size estimates at the same time.

4. Meta-regression methodology

Our methodology is informed by the publication selection model proposed by Egger et al. (1997) and the weighted least squares (WLS) model estimation discussed in Stanley (2005; 2008) and Stanley and Doucouliagos (2012; 2014; 2017) among others. We contribute to the existing effort by: (i) adopting a hierarchical modelling (HM) approach to the bivariate and multivariate meta-regression models; (ii) evaluating the statistical power in the evidence base and obtaining effect-size estimates based on adequately powered evidence; and (iii) utilising a weighted-average least squares (WALS) routine for addressing model uncertainty in the

[9] We estimated the multinomial logit model with a constant only and by allowing the t-value bands to be related to the degrees of freedom within the band. The predictive margins from both specifications are identical. Also, we have estimated relative probabilities for all spillover-type samples, with similar results. The latter are not presented here to save space, but they are available on request.

specification and estimation of the multivariate meta-regression model. These methodological considerations are discussed in sections 4.1 - 4.3 below.

4.1 Specification and estimation issues in the bivariate meta-regression model

In Egger et al. (1997), the estimates of a 'true' effect in a research field vary around the true effect with an idiosyncratic error (ξ_i) if publication selection does not exist. However, when publication selection exists, effect-size estimates in primary studies will be conditional on the standard error (SE_i) as indicated in (4a) below. The model in (4a) can be used to conduct tests for the null hypothesis of no publication bias by testing for $\alpha = 0$. Also, the estimate of β would be an unbiased estimator of the average productivity effect after controlling for selection.

$$effect\_size_i = \beta + \alpha SE_i + \xi_i \qquad (4a)$$

Yet, the bivariate model in (4a) raises *seven* estimation issues, three of which have been addressed in Stanley (2005; 2008) and Stanley and Doucouliagos (2012), among others. Of these, heteroskedasticity is addressed through a weighted least squares (WLS) estimator, using precision-squared ($1/SE_i^2$) as analytical weight. This is equivalent to dividing both sides of (4a) with the standard error (Stanley and Doucouliagos, 2012), leading to:

$$t_i = \alpha + \beta \left(1/_{SE_i} \right) + \omega_i \qquad (4b)$$[10]

Here t_i is the t-value associated with the effect-size estimate as reported in the primary study and ω_i is the error term in (4a) divided with the standard error. Testing for $\beta = 0$ is a precision-effect test (*PET*) for 'genuine' effect, whilst testing for $\alpha = 0$ is the funnel asymmetry test (*FAT*) for publication selection. However, *FAT* is known to have low power and as such it is advisable to compare the *FAT* results with other results. Therefore, we will rely on the *FAT* results only if they are consistent with the selection bias evidence we reported in *Tables 1* and *2*. On the other hand, rejection of the null hypothesis in *PET* indicates significant effect after controlling for publication selection.

---

[10] Note that the error term in (4b) is also divided by the standard error. In other words, $\omega_i = {}^{\xi_i}/_{SE_i}$. This is also the case in (5b) below.

The *second issue* relates to the observation in Andrews and Kasy (2019) that the interpretation of β as selection-corrected average affect across studies may be misguided. This is because the conditional expectation of the effect size across studies is not necessarily linear in the standard error (or precision). To avoid the risk of biased estimates, Andrews and Kasy (2019) propose a non-parametric method of correcting for publication selection.

We argue in favour of a parametric method for two reasons. First, a solution to non-linearity in the relationship between effect-size estimates and their standard errors has already been proposed in the context of parametric meta-analysis models. Simulation studies (Moreno et al., 2009, 2011; Stanley and Doucouliagos, 2014) clearly indicate that a non-linear (quadratic specification) is preferable if an effect size exists – i.e., if the PET rejects the null hypothesis. Then, the non-linear Egger model and its WLS equivalent are:

$$effect\_size_i = \gamma + \varphi SE_i^2 + \vartheta_i \quad (5a)$$

$$t_i = \gamma \left({}^{1}/_{SE_i} \right) + \varphi SE_i + \varepsilon_i \quad (5b)$$

Model (5b) is estimated without a constant term and is referred to as precision-effect test corrected for standard errors (*PEESE*)**.** The average effect size is γ, which is shown to have smaller bias and mean square error if the *PET* result indicates a significant effect. In addition, the performance of the *PEESE* specification remains better at different levels of selection bias (Moreno et al., 2009, 2011; Stanley and Doucouliagos, 2014).[11] Therefore, we address the issue raised by Andrew and Kasy (2019) by reporting *PEESE* estimates for the unbiased average productivity effects when the latter is significant in the *PET*.

The second argument in favour of parametric models is that they, unlike the non-parametric model proposed by Andrew and Kasy (2019), do not rely on a strong assumption that the publication probability is either known or can be estimated correctly as a function of the Z-statistic that indicates statistical significance. If the publication probability is unknown or estimated incorrectly from the sample, Andrews and Kasy (2019) clearly indicate that the median unbiased effect size they propose cannot be estimated correctly: only valid confidence intervals can be obtained.

[11] The better performance of the quadratic specification at different levels of bias is an important finding that both justifies and allays the concern in Andrews and Kasy (2019) that the linear Egger regression may yield biased effect-size estimates if selection bias exists.

Although we prefer a parametric modelling approach, we also argue that the latter may yield biased estimates if the hierarchical nature of the data is overlooked (*issue 3*). A well-known feature of the meta-analysis data is that primary studies report multiple effect-size estimates that may be correlated due to dependence on a given data source, estimator or time period or a combination thereof. Hence, the meta-regression model must take account of such dependencies by allowing for a more general covariance structure where effect-size estimates from the same study can be correlated. If within-study dependence (i.e., intra-study correlation) exists, the ordinary least squares (OLS) estimator would be inappropriate for two reasons: (i) the sample size is exaggerated due to treatment of all effect-size estimates as independent observations; and (ii) the risk of rejecting the null hypothesis erroneously (type-I error) is higher (Snijders and Bosker, 2012).

We address within-study dependence by adopting a hierarchical modelling (HM) framework that allows for intra-study correlation between effect-size estimates. This contrasts with the existing practice of relying on pooled OLS with clustered standard errors (Stanley and Doucouliagos, 2012). The latter corrects for standard errors but not for the effect-size estimate, which is estimated without taking account between-study heterogeneity. In the presence of unobserved heterogeneity, the HM estimator performs better than OLS (Katahira, 2016). Furthermore, they allow for flexibility in modelling unobserved heterogeneity, which can be modelled as random intercepts, random slopes or both. A two-level random-intercept and random-slope HM is stated below, with other specifications given in *Box 2* in the *Online Appendix*.[12]

$$t_{ij} = \alpha^{RIS2} + \beta^{RIS2}\left(1/_{SE_{ij}}\right) + h_{0j}^{RIS2} + h_{1j}^{RIS2}\left(1/_{SE_{ij}}\right) + u_{ij}^{RIS2} \qquad (6)$$

Here the superscript (^{RIS2}) indicates that this is a two-level HM with random intercepts and random slopes. Of the random-effect components, h_{0j}^{RIS2} is study-specific random intercepts that reflect between-study heterogeneity in terms of intercepts; and $h_{1j}^{RIS2}\left(1/_{SE_{ij}}\right)$ are the study-specific slopes that reflect between-study slope heterogeneity. Both random effects are

[12] See *Box 2* in the *Online Appendix*; and Ugur et al. (2016; 2018) for further discussion.

distributed normally with a zero mean and constant variance; and they are assumed orthogonal to the regressor (i.e., *1/SE*). Finally, the model error is u_{ij}^{RIS2} and is assumed to be i.i.d.

The choice between OLS and HM estimator and the choice between restricted (e.g., random-intercepts-only - RIO) and unrestricted (e.g., random-intercepts-and-slopes - RIOS) HMs is made through likelihood ratio (LR) tests. The null hypothesis in the LR tests is that the restricted model (OLS or restricted HM) is nested within the unrestricted HM. Rejection of the null hypothesis indicates that the unrestricted HM is more appropriate.

The *fourth issue* is endogeneity due to two sources. The first, which is a common problem in all estimators, is potential correlation between the regressors and the model's error term. To address this issue, we estimate the Egger regression (4a) with a two-stage least-squares (2SLS) estimator, using the inverse of the squared standard error ($1/SE^2$) as analytical weights. We also use the inverse of the sample size and its square as instruments for the standard error (*SE*). We conduct endogeneity and overidentification tests to verify if the *SE* is exogenous and the instruments are valid.[13]

The second type of endogeneity, specific to HMs, may arise due to correlation between the regressors (*1/SE* in 6) and study-level random-effect components (h_{0j}^{RIS2} or h_{1j}^{RIS2} in 6). Simulations in Ebbes et al. (2004) indicate that the hierarchical model would yield upward-biased effect-size estimates and downward-biased publication selection bias estimates if the regressors are correlated with the random-effect components. Both biases are larger if the first type of endogeneity also exists.

A well-known correction for the regressor's correlations with the random-effect components has been proposed by Mundlak (1978), who demonstrate that inclusion of within-group means of the regressor in the model ensures mean-independence between the regressors and the group-level random effects (see, also, Snijders and Bosker, 2012). The Mundlak correction, however, may yield biased estimates because it requires within-group means at the population level whereas only sample means are available (see, Grilli and Rampichini, 2011). Nevertheless, the potential estimation bias in the Mundlak correction does not invalidate the Mundlak test itself (Ebbes et al., 2004; Hanchane and Mostafa, 2012). Therefore, we follow the Mundlak approach

[13] Because we estimate the models with robust standard errors, the test reports Wooldridge's (1995) score test for endogeneity. Failure to reject the null hypothesis indicates that OLS should be used instead of 2SLS. Test results are reported in the last rows of the estimation tables below.

to test the null hypothesis that the regressors are uncorrelated with the study-level random effects. If the null hypothesis cannot be rejected, we estimate the HMs without Mundlak correction. Otherwise, we use the within-study mean of the precision as an additional regressor.

The *fifth issue* relates to multiple effect-size estimates reported by primary studies. To avoid domination by studies that report large number of estimates, we estimate (4b) and (5b) with frequency weights, specified as the inverse of the number of estimates reported in each study (see, Stanley, 2005; 2008; Stanley and Doucouliagos, 2012).

4.2 Statistical power and the weighted average of the adequately powered evidence

The *sixth issue* we address is whether the evidence base is adequately powered and what would the average effect size be when only adequately powered effect-size estimates are used in the meta-analysis. Technically, statistical power is the probability of rejecting the null hypothesis correctly – i.e., when the alternative is true. Stated differently, power is the probability of 'discovering' a significant effect when the latter actually exists (Ellis, 2010: 52). In our context, it provides important information about the probability of detecting a significant spillover (or own-R&D) effect when the latter is true. Statistical power (hence the probability of discovering a 'true' effect) is higher when the sample size is larger and the standard error of the productivity estimate is lower (i.e., the precision is higher).

As indicated in Ioannidis et al. (2017), adequate power in social-scientific research has been conventionally set at 80% or over. This corresponds to a probability of a Type II error that is not larger than four times the probability of the Type I error. With a 5% significance level, this 'power rule' implies the following relationship between the estimate of the 'true' effect (γ) and its standard error (SE) (Ioannidis et al., 2017: F239):

$$|\gamma| \,/\, SE_i \geq 2.80 \quad \text{or} \quad SE_i \leq |\gamma|/2.80. \qquad (7)$$

To identify the standard errors that satisfy the inequality in (6), we use the PEESE estimates of the average effect size (γ) from model (5b).[14] Then the percentage of the adequately-powered evidence in the sample is obtained by dividing the number of adequately-powered effect-size

[14] Ioannidis et al. (2017) use PEESE estimates obtained from a WLS estimator instead of HM estimator. When we replicate their approach, we obtain highly similar percentages of adequately-powered estimates and weighted averages of the adequately powered. The results, not reported here to save space, can be provided on request.

estimates with the total number of productivity effect estimates in the relevant sample. Finally, the weighted average of the adequately-powered (WAAP) is obtained as a weighted average of effect-size estimates that satisfy the power rule, using the precision-squared (*$1/SE^2$*) as analytical weights.

4.3 Addressing model uncertainty in the multivariate meta-regression model

The *seventh issue* arises in the context of the multivariate meta-regression model and concerns the choice of the moderating variables that capture the sources of observed heterogeneity in the evidence base. To address this issue, we estimate a hierarchical multivariate meta-regression model (HMRM) as specified in (8) below and discussed in *Box 2* in the *Online Appendix*. Observed heterogeneity is modelled through a set of binary (*Z*) variables that moderate the magnitude of the effect-size estimates and a set of continuous (*X*) variables that capture the sources of selection bias.

$$t_{ij} = \alpha^{HM} + \beta^{HM}(1/SE_{ij}) + \sum_m \gamma_m^{HM} Z_m(1/SE_{ij}) + \sum_k \gamma_k^{HM} X_k + h_{0j} + h_{1j}(1/SE_{ij}) + \varepsilon_{ij}^{HM} \quad (8)$$

Our choice of the moderating variables is informed by reporting guidelines for meta-analysis of economics research (Stanley et al., 2013). The continuous *X*-variables are not divided with standard error to verify whether higher levels of perceived quality and larger sample sizes are associated with less or more selection bias.[15] They include the H-index of the journal, the number of citations for the primary study, the number of observations on which the effect-size estimate is based, and the number of years in the data used. As a measure of perceived quality, we have chosen the journal H-index instead of other journal quality indicators in the light of review evidence on alternative measures. [16] We have also used the study-specific number of citations per year as an additional quality indicator.[17]

[15] Nevertheless, we have estimated the HMRM with continuous variables divided by the standard error. The results, not reported here, are highly similar and can be provided on request.
[16] We use the logarithm of the h-index for journals, as reported in the Scimago Journal & Country Rank (SJR) database (http://www.scimagojr.com/). Several reviews of the existing journal ranking schemes report that the alternative measures (e.g., eigenfactor, SJR score, source normalized impact per paper, etc.) are highly correlated (González-Pereira et al., 2010; Mingers and Yang, 2017; and Yuen, 2018). Furthermore, Mingers and Yang (2017) suggest that the H-index is one of the most efficient indicators for business and management journals, which include economics journals.
[17] This is in line with Philips (2016), who also investigates whether perceived paper quality is associated with selection bias. Number of citations are drawn from Google Scholar. Reviews of existing citation platforms indicate

In addition to the wide range of binary moderating variables utilised in meta-analysis, we have two binary variables informed by the characteristics of the research field: data midpoint in 1991 or before; and publication date after 2000. The binary variable for data mid-point in 1991 or before identifies effect-size estimates that would be more dependent on data collected before the OECD's publication of the Oslo Manual on collecting and interpreting technological innovation data. These studies account for 25% of the evidence base. The binary variable for studies published after 2000 identifies primary studies published after seminal contributions to the field by Griliches (1992), Coe et al. (1997) and Keller (1998). These studies account for 75% of the evidence base.[18]

Given the absence of theoretical guidance on the set of moderating variables to be included in (8), we address model uncertainty by a weighted-average least squares (WALS) routine that combines ideas from Bayesian and frequentist approaches to model averaging (De Luca and Magnus, 2011). We prefer the WALS routine to the Bayesian model averaging (BMA) approach in meta-regression (Havránek, 2015; Iršová and Havránek, 2013) because the former has a bounded risk profile, relies on transparent ignorance in the selection of the focus variable(s), and is less costly in terms of computation time. The focus variable in WALS is precision ($1/SE_i$), which reflects prior knowledge based on the Egger et al. (1997) model. All moderating variables are treated as auxiliary covariates, the relevance of which is determined by information from the data. The decision rule is to include an auxiliary covariate in the specific HMRM if its t-value in WALS is greater than one in absolute value (De Luca and Magnus, 2011).[19] Following this rule, we estimate the HMRM with 20 moderating variables out of 30 potential candidates.[20]

that Google Scholar provides a relatively more comprehensive account of citations compared to alternatives such as the Thomson ISI Web of Science and Scopus (Harzing, 2008; Meho and Yang, 2006; Nisonger, 2004).

[18] The coefficients on these and other binary variables described in *Table A5* in the *Online Appendix* indicate whether the effect-size estimates associated with the observed characteristic are larger or smaller than the effect-size estimates in the refence category.

[19] Please see *Box 3* in the *Online Appendix*, where we discuss the frequentist and Bayesian model averaging methods and indicate why the WALS routine is preferred.

[20] Results from the specific HMRM with 20 covariates are in *Table 6*, whereas results from the general HMRM with 30 covariates are reported in *Table A6* in the *Online Appendix*. The list of moderating variables (auxiliary covariates) that do not satisfy the WALS criteria for inclusion in the HMRM include the following: Cobb-Douglas production function specified with human capital; control for capital in the primal model; use of panel data for estimation as opposed to cross-section data; North American data as opposed to rest-of-the-world data; knowledge spillovers as opposed to rent or mixed spillovers; rent spillovers as opposed to knowledge or mixed spillovers; use of GMM estimator as opposed to other estimators; use of data at the country level as opposed to firm or industry level data; the logarithm of data years; and the logarithm of the h-index for the journal.

5. Meta-regression results

Table 3 presents the hierarchical model (HM) estimates of the average productivity effect by spillover types and own R&D; while *Table 4* presents the results by unit of analysis.[21] As indicated above, we have estimated the Egger et al. (1997) model with 2SLS for 11 evidence pools (samples) and tested for endogeneity and instrument validity before estimating the HM in (6). In all tests, we failed to reject the null hypotheses that the SE is exogenous.

Then, we checked if endogeneity exists due to correlation between the regressor (precision) and the study-level (level-2) random effects. For this, we conducted Mundlak (1978) tests and failed to reject the null hypothesis of no correlation at 5% in 9 out of 11 estimations. The null hypothesis is rejected in the case of rent spillover sample (*Table 3*) and own R&D sample at the firm level (*Table 4*). Here, we used the Mundlak correction by adding the within-study mean of the precision to the HM in (6) to ensure mean independence.

Finally, we have conducted likelihood ratio (LR) tests to verify if the HM specification we use is preferable to the constrained alternative, which can be OLS or a restricted HM. All LR tests clearly justify the HM specification we adopt, which provides for random slopes and: (i) random intercepts at the spillover type and study levels in the case of spillovers; and (ii) random intercepts at the study level only in the case of own R&D effects.

[21] Results in *Tables 3* and *4* are not based on frequency weights that account for multiple estimates reported in primary studies. However, estimations with frequency weights are in *Tables A3* and *A4* in the *Online Appendix*. The latter are largely similar to what is reported in *Tables 3* and *4*.

Table 3 – Productivity effect estimates by spillover (SPO) types and own R&D

Dependent variable: t-value	**Panel A - PET/FAT**					**Panel B - PEESE**		
	Knowledge SPOs	Mixed SPOs	Rent SPOs	All SPOs	Own R&D	Knowledge SPOs	All SPOs	Own R&D
Effect (β in PET/FAT, γ in PEESE)	0.048***	0.074	0.007†	0.038***	0.064***	0.069***	0.036***	0.073***
	(0.017)	(0.050)	(0.023)	(0.014)	(0.012)	(0.016)	(0.014)	(0.011)
Selection bias	2.065***	1.377	2.751***	2.195***	0.808***			
	(0.572)	(1.030)	(0.541)	(0.380)	(0.402)			
Standard error						-0.835	-2.736***	3.588
						(1.540)	(1.247)	(3.543)
Obs.	557	96	327	983	501	557	983	501
Studies	46	6	30	60	26	46	60	26
Log-likelihood (LL)	-1760.941	-306.995	-932.755	-3064.777	-1472.789	-1766.875	-3067.101	-1474.265
LL (comp. model)	-1853.435	-323.953	-1051.853	-3321.677	-1685.547	-1933.186	-3449.187	-1714.120
LR test: comp. mod. Preferred ($p > chi^2$)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mundlak test: no-corr. at study level ($p > chi^2$)	0.572	0.303	0.054	0.991	0.531	0.572	0.991	0.531
Exogeneity test ($p > chi^2$)	0.920	0.498	0.492	0.803	0.940	0.920	0.803	0.940

Notes: Random effects are modelled as random intercepts and random slopes. The null hypothesis in the likelihood ratio (LR) test is that the restricted model is preferable to the multi-level model chosen. The null hypothesis in Mundlak test is that the regressor (precision) is not correlated with study-level random effects. The null hypothesis in the exogeneity test is that the regressor (standard error) in the Egger model is exogenous. Robust standard errors are clustered at the study level. ***, **, * indicates significance at 1%, 5% and 10%, respectively. . † Estimated with Mundlak correction.

In *Panel A* of *Table 3*, the selection bias is severe in three out of four spillover samples but moderate in the own R&D sample.[22] The positive signs of the selection bias are consistent with the funnel plot asymmetry and with selection bias evidence reported in *Tables 1* and *2*. Given these congruent findings, we conclude publication selection exists and hence simple summary measures or vote-counting evidence relied upon in narrative reviews would be biased.

After controlling for selection bias, the average effect size in *Table 3* is positive and significant for knowledge spillovers, own R&D and all spillover types; but it is insignificant for mixed and rent spillovers. With PEESE correction in *Panel B*, the average productivity effects for knowledge spillovers and own R&D remain similar at 0.69 and 0.073, respectively. The corrected estimate for all spillover types is smaller (0.036). In the case of mixed spillovers, neither the effect-size nor the selection bias is significant. These results are in line with those obtained with frequency weights in *Table A3* in the *Online Appendix*.[23]

A similar pattern is evident in *Table 4*, where we report average effect-size estimates at the country, industry and firm levels. Here too selection bias is substantial (spillover samples at the country level) or severe (spillover samples at the firm and industry levels). In the case of own-R&D, it is moderate at the country level and substantial at the firm level. The PET estimates indicate that spillovers have a positive and significant productivity effect only at the country level; whereas own R&D has positive and significant effects at the country and firm levels. With PEESE correction, the average productivity effects of spillovers and own-R&D are highly similar, ranging from 0.050 - 0.060. The results obtained with frequency weighting (*Table A4* in the *Online Appendix*) are mostly in line with *Table 4*.[24]

[22] The selection bias is considered as severe if $|\alpha| > 2$; substantial if $1 \leq |\alpha| \leq 2$; and moderate if $|\alpha| < 1$ (Doucouliagos and Stanley, 2013).
[23] The difference between spillovers and own R&D effect is insignificant as the confidence intervals for both estimates overlap.
[24] The only difference is that the average spillover effect at the country level (0.135) is larger than the rest. However, spillovers at the country level should be treated with caution. Keller (1998) has demonstrated that the spillover effect on country-level productivity remains positive and significant even when random weights are used instead of weights based on import shares.

Table 4 - Productivity effect estimates by spillover (SPO) types, own R&D and unit of analysis

Dependent variable: t-value	**Panel A - PET/FAT**						**Panel B - PEESE**		
	All SPO Country	All SPO Industry	All SPO Firm	Own R&D Country	Own R&D Industry	Own R&D† Firm	All SPO Country	Own R&D Country	Own R&D Firm
Effect (β in PET/FAT, γ in PEESE)	0.058***	-0.019	0.040	0.056***	0.088	0.050***	0.058**	0.060***	0.057***
	(0.015)	(0.038)	(0.029)	(0.007)	(0.059)	(0.016)	(0.016)	(0.006)	(0.016)
Selection bias	1.641***	2.370***	2.637***	0.623*	1.412	1.748**			
	(0.590)	(0.908)	(0.538)	(0.345)	(1.601)	(0.733)			
Standard error							-8.955**	13.934*	-0.047
							(4.115)	(8.208)	(2.587)
Obs.	459	223	299	283	89	126	459	283	126
Studies	26	12	22	25	9	19	26	25	19
Log-likelihood (Hierarch. Model)	-1408.742	-678.214	-955.515	-705.504	-313.91	-265.875	-1408.060	-705.669	-268.303
Log-likelihood (Comp. Model)	-1541.737	-756.472	-1008.283	-817.222	-333.312	-342.706	-1592.711	-824.433	-378.428
LR Test chi^2	265.989	156.515	105.535	223.437	38.805	153.662	369.302	237.530	220.250
p-value	0	0	0	0	0	0	0	0	0
Mundlak test ($P>Chi^2$)	0.355	0.617	0.573	0.106	0.519	0.0005	0.355	0.106	0.0005
Exogeneity test ($p > chi^2$)	0.710	0.594	0.780	0.821	0.244	0.757	0.710	0.821	0.757

Notes: See Table 3. † Estimated with Mundlak correction.

The findings in *Tables 3* and *4* do not support the claims in large majority of the narrative reviews that the productivity effects of spillovers are larger than those of own R&D.[25] Moreover, non-significant effects in *Table 3* raise doubt about whether rent or mixed spillovers capture true knowledge externalities or just measurement errors. Our findings lend support to Griliches (1992: S36), who indicates that "true spillovers are ideas borrowed by research teams of industry *i* from the research results of industry *j*" - and it is not clear whether this kind of borrowing is related to transactions (trade or patent flows) between the parties involved.

A further qualification to the received wisdom is called for by results in *Tables 4*, where spillover effects are insignificant at the firm and industry levels. This may be due to dominance of the creative destruction and/or market-stealing effects of external R&D (Aghion et al., 2014; Aghion and Howitt, 1992; Bloom et al., 2013). A third qualification is called for by positive and significant average spillover effects at the country level. This is because it is difficult to separate the rent spillovers through the imports channel from productivity gains due to the disciplining effect of international competition and wider gains from trade.

In *Table 5*, we provide evidence on another issue that has remained underworked in meta-analyses of economics research: the extent to which the existing evidence is adequately powered; and what would the average effect be when only adequately-powered evidence is used for estimations.[26] In Panel A, the percentage of adequately-powered evidence is 41% for knowledge spillovers and 30% for all spillover types.[27] In contrast, 73% of the primary-study estimates of own-R&D effects are adequately-powered. When spillover and own R&D effects are pooled by the unit of analysis (Panel B), the percentage of adequately-powered estimates is relatively higher (55%) for spillover effects at the country level but this is still lower than those related to own-R&D at the country level (67%) or firm level (74%). Overall, the findings confirm our observation in section 3 that the spillover effects are estimated with lower precision compared to own-R&D effects.

[25] It must be reiterated that Hall et al. (2010) is the only review that does not report larger spillover effects compared to own-R&D effects.
[26] To our knowledge, Ioannidis et al., (2017) is the only exception, where statistical power is investigated *ex post* using evidence from 159 meta-analysis studies.
[27] We identify the primary-study estimates with a statistical power of 80% or more by assuming that the best estimates of the average productivity effects in the sample are unbiased estimates of the 'true' population effects. The best estimates are the PEESE estimates in *Panels B* of *Tables 1* and *2*.

Table 5 – Weighted average effects from adequately-powered (WAAP) evidence: By spillover (SPO) type, own R&D and unit of analysis

	Panel A: By spillover (SPO) type and own R&D			Panel B: By unit of analysis		
	Knowledge SPO	All SPO	Own R&D	All SPO Country	Own R&D Country	Own R&D Firm
WAAP Effect	0.009***	0.012***	0.029***	0.027***	0.060***	0.012***
	(0.002)	(0.002)	(0.001)	(0.004)	(0.003)	(0.002)
Obs.	229	293	365	253	198	93
R-sq.	0.088	0.095	0.793	0.183	0.702	0.216
Adequately powered (%)†	41	30	73	55	67	74

***, **, * indicates significance at 1%, 5% and 10%, respectively. † = (number of adequately-powered estimates / all observations in sample)*100

Low power does not invalidate the effect-size estimates from the full samples, but it does suggest that such estimates constitute a poor basis for evidence-based policy. Our findings indicate that the average productivity effects based on adequately-powered evidence are much smaller. The weighted average effect from adequately-powered (WAAP) evidence for knowledge spillovers and all spillover types are too small (0.009 and 0.012, respectively) to be practically significant. The WAAP result for spillover effects based on country-level data is 0.027 and indicates a moderate effect, but this is still half of the own-R&D effect (0.06).

Findings so far allow for two observations. First, the evidence does not support the claims that the productivity effects of external R&D are substantial and larger than own-R&D effects. Such claims do not take account of either publication selection bias or the high percentage of low-powered estimates in the literature or both. Indeed, both in the full sample and in evidence pools with adequate statistical power, the productivity effects of own R&D are usually larger than those of spillover effects.

Second, our findings also indicate that the productivity effect of spillovers are usually estimated with lower precision and power compared to own R&D. As such, they indicate that our knowledge about the productivity effect of own R&D is based on firmer evidence. They also lend support to the argument that investing in R&D is a necessary first step to develop absorptive capacity, defined as organizational routines and processes that enable firms (hence industries and countries) to emulate, transform and exploit external knowledge (Cohen and Levinthal, 1989). A more recent study (Aldieri et al., 2018) sheds further light on the importance of absorptive capacity, which depends on the firms' learning strategies and the extent to which the latter is congruent with the spillover type they face.[28]

In what follows, we report evidence on observed sources of heterogeneity in the evidence base and how the meta-effect varies by each moderating variable. One moderating factor is the measurement of the external knowledge stock. Griliches (1992) has already drawn attention to challenges in constructing a measure for the external knowledge stock, which may well exist but is essentially unobserved. A more recent work (Di Lorenzo and Almeida, 2017) provides new evidence on why the weights used to construct spillovers through a transmission channel (e.g., researcher mobility, imports, patent flows, etc.) may be too aggregate to take account of the variations in the quality of the transmission links between the source and beneficiary of the spillovers.[29] Similar considerations apply to other transmission links, where different weighting schemes may be necessary to reflect quality variations in the composition of imports, foreign direct investment or patent flows that researchers use to construct weighted spillover pools.

Other sources of heterogeneity relate to variations in publication type and sampling, modelling and estimation strategies of the primary studies. With respect to publication type, journal quality or journal articles as opposed to working papers may have systematic effects on reported estimates. It is also necessary to check whether any systematic differences exist between different samples (e.g., whether the sample consists of firms, industries or countries; or the firms/industries/countries have high or low R&D intensity). Third, we need to check if different weights used to construct the measures of external knowledge matter and whether there are effect-size differences when R&D investments or R&D capital stocks are used to construct the external knowledge pool. Finally, we verify if different estimation methods are associated with different estimates.

Results in *Table 6* are based on two-level HMs with random intercepts and slopes (RIS) in columns 1 and 2; and with random intercepts only (RIO) in columns 3 and 4. In each specification, the model is estimated without frequency weights (columns 1 and 3) and with frequency weights (columns 2 and 4). The frequency weight is equal to the reciprocal of the number of effect-size estimates reported in

[28] Aldieri et al. (2018) report that firms closer to the technology frontier develops absorptive capacities that enable them to benefit more from knowledge spillovers whereas those away from the technology frontier build absorptive capacities that enable them to benefit more from rent spillovers.

[29] According to Di Lorenzo and Almeida (2017), the mobility of high-performance inventors is lower than that of low-performance inventors; and the difference between mobility rates is higher the larger is the performance differential. This implies that the quality of the knowledge that spills over through researcher mobility may be lower than the level implied by a single weight that reflects the total number of mobile researchers.

each primary study, ensuring that each study has an equal weight of one.[30] Given the log-likelihood values and LR test results, the preferred results are from the HM with RIS and without frequency weights (column 1), followed by RIS results with frequency weights in column 2. RIO results without and with frequency weights in columns 3 and 4 are reported as sensitivity checks.

Because the difference in the log-likelihood values is small between RIS and RIO specifications, and given that sign and significance consistency is around 70% across 4 columns, we interpret the results in *Table 6* as follows: (i) strongly-consistent results if sign and significance consistency is observed across 4 columns; (ii) consistent results if sign and significance consistency is observed between column 1 and two other columns; (iii) moderately-consistent evidence if sign and significance consistency is observed between columns 1 and 2 only; and (iv) weakly-consistent evidence if consistency exists between one of columns 1 or 2 and one of columns 3 or 4.

Starting with moderating variables for publication characteristics (top part of *Table 6*), we observe consistent evidence that journal articles report relatively smaller spillover effects on productivity compared to working papers. This finding indicates that journals, as opposed to working papers, do not exploit reputation to accommodate highly selected evidence. In other words, our finding does not lend support to the "winner's curse" argument in Costa-Font et al. (2013) that journal articles tend to report more selected evidence. We have also found consistent evidence that studies published after 2000 tend to report relatively smaller spillover effects compared to previous studies. This finding indicates a competition-related attenuation effect, which arises from method development and/or exploitation of richer datasets following the pioneering contributions by Griliches (1992), Coe et al. (1997) and Keller (1998). These results are consistent with those from the general model reported in *Table A6* in the *Online Appendix*.

---

[30] Note that the frequency-weighted HMs are double-weighted versions of the original Egger et al. (1997) model. Both sides of the latter are first weighted by precision to address heteroskedasticity. Then, the transformed model is estimated with frequency weights to address the issue of undue influence by studies reporting a large number of effect-size estimates. Frequency weighting is recommended as a sensitivity check for undue influence from studies reporting large numbers of estimates (Stanley and Doucouliagos, 2012).

Table 6 - Multivariate meta-regression analysis (WALS-determined specific model)

Dependent Variable: *t*-Value	(1)	(2)	(3)	(4)
Precision	0.204**	0.068	0.255***	0.232**
	(0.100)	(0.144)	(0.045)	(0.107)
Publication characteristics				
Journal article	-0.190**	-0.126	-0.238***	-0.244***
	(0.083)	(0.106)	(0.036)	(0.084)
Publication date after 2000	-0.037	0.020	-0.099***	-0.114***
	(0.031)	(0.050)	(0.011)	(0.036)
Log of citations	-0.067	0.330	-0.220	0.817
	(0.284)	(0.264)	(0.464)	(1.001)
Model specification				
TFP - Dependent variable is total factor productivity	0.017	0.032	0.021**	-0.012
	(0.023)	(0.034)	(0.009)	(0.023)
SPO coefficients in model <=2	0.037***	0.022**	0.044***	0.058**
	(0.010)	(0.011)	(0.008)	(0.026)
Control for own R&D in model	-0.053***	-0.033*	-0.031***	-0.005
	(0.019)	(0.019)	(0.011)	(0.032)
Industry/country dummies in model	0.051*	0.099***	0.026***	0.049***
	(0.031)	(0.029)	(0.009)	(0.019)
Year dummies in model	0.020	0.064	0.028***	0.046***
	(0.029)	(0.041)	(0.007)	(0.017)
Data and sample characteristics				
Unit of analysis: industry	-0.245***	-0.266***	-0.237***	-0.253***
	(0.058)	(0.103)	(0.024)	(0.073)
High R&D-intensity firm, industry	-0.039	-0.061**	-0.035***	-0.049
	(0.024)	(0.031)	(0.013)	(0.032)
South Asian data	0.050	0.005	0.075***	0.075***
	(0.037)	(0.050)	(0.011)	(0.010)
OECD data	0.092*	0.038	0.135***	0.118***
	(0.051)	(0.077)	(0.023)	(0.041)
Data mid-point < =1991	-0.045	-0.014	-0.038***	-0.033*
	(0.032)	(0.060)	(0.009)	(0.020)
Log of number observations	0.249	0.312	0.375	0.433
	(0.242)	(0.219)	(0.359)	(0.545)
Spillover characteristics				
Based on asymmetric weights	0.024***	0.045	0.021***	0.037
	(0.006)	(0.036)	(0.006)	(0.025)
Unweighted	0.003	0.003	0.005	0.006
	(0.003)	(0.005)	(0.003)	(0.007)
Based on R&D investment	0.107*	-0.014	0.222***	-0.012
	(0.063)	(0.064)	(0.026)	(0.154)
Estimation method				
Estimation with differenced data	-0.005*	-0.005***	-0.005*	-0.004***
	(0.003)	(0.001)	(0.003)	(0.001)
Estimation takes account of panel cointegration	-0.008	-0.009***	-0.008	-0.014*
	(0.005)	(0.003)	(0.005)	(0.008)
Instrumental variable (IV) estimation	-0.010	0.008	-0.004	0.032
	(0.007)	(0.019)	(0.007)	(0.038)
Constant	1.084	-0.970	0.898	-3.478
	(2.178)	(1.986)	(3.397)	(3.792)
Observations	983	983	983	983
Studies	60	60	60	60
Log-likelihood (HM)	-3015.533	-174.850	-3052.399	-186.039
LR Test chi2	96.545	182.603	400.093	341.093
P> chi2	0.000	0.000	0.000	0.000
converged	Yes	Yes	Yes	Yes
Log-likelihood (restricted model)	-3121.980	NA†	-3121.980	NA†

*** p<0.01, ** p<0.05, * p<0.1. (1) HM with random intercepts and slopes (RIS), without frequency weights; (2) HM with RIS and frequency weights; (3) HM with random intercepts only (RIO), without frequency weights; (4) HM with RIO and frequency weights. † log-likelihood statistics for the comparative model is not reported when the HM is estimated with frequency weights.

With respect to model specification, we find no systematic difference between studies that use total factor productivity (TFP) and those that use output or value added as dependent variable. This finding indicates that coefficient estimates based on both versions of the primal approach are consistent in this research field. In contrast, there is strongly consistent evidence that studies that control only for two or less spillover types (as opposed to three or more spillover types) tend to report relatively larger spillover effects. There is also consistent evidence that studies that control for own R&D capital tend to report smaller effect-size estimates compared to those that do not.

Controlling for own R&D capital is an explicit requirement of the knowledge capital model, where own-R&D capital and spillovers are complements and should be both included in the model to be estimated (Griliches, 1979; 1992). Therefore, our finding suggests that larger spillover effects in primary studies that do not control for own R&D are likely to be upward-biased due to omitted variable bias. However, the theory is silent on whether different spillover types should be treated as complements. Therefore, all we can conclude is that controlling for at least three spillover types (knowledge, rent and mixed spillovers) may be good practice in future research as it allows for comparing productivity effects from different spillover types.

Another model specification issue is inclusion of industry/country and year dummies in the estimated models. We find strongly-consistent evidence that the spillover effects are larger when industry/country dummies are included in the empirical models. The evidence on inclusion of year dummies is only weakly consistent but points in the same direction. The case for including industry/country or year dummies is not clear-cut (Hall et al., 2010). On the one hand, such dummies can account for erroneous omission of industry/country or year characteristics. On the other hand, they may be a source of bias if productivity effects differ because of different technological opportunities in different industries or during different phases of business cycles.

Concerning moderating factors that reflect data and sample characteristics, we find strongly-consistent evidence that data at the industry level is associated with lower spillover effects compared to the reference category that consists of effect-size estimates based on firm or country data.[31] The smaller spillover effects at the industry level suggest that the creative destruction and/or market-stealing effects of external R&D are stronger between industries compared to between firms or countries. This finding

[31] This is also line with the PET-FAT-PEESE results discussed above.

indicates that it is necessary to pay more attention to the extent to which the productivity effects of the external R&D stock may be attenuated by the creative destruction and/or market-stealing effects of external R&D (Aghion et al., 2014; Schumpeter, 1942; Bloom et al., 2013). It also indicates that the creative destruction and market-stealing effects may unfold at different speeds at different levels of analysis.

We find weakly-consistent evidence that the spillover effects are relatively smaller when the data relate to firms/industries with high R&D intensity. This finding is in line with insights from Schumpeterian models of innovation, where productivity growth in firms/industries closer to the technology frontier depends more on breakthrough innovation rather than emulation of external knowledge (Aghion et al., 2014). However, it must be indicated that only 21% of the effect-size estimates in the primary studies control for R&D intensity.[32]

In contrast, we find consistent evidence supporting the absorptive capacity argument in that the productivity effects of spillovers are larger when the data relate to OECD firms/industries/countries (73% of the evidence base). This finding indicates that a longer history of innovation and R&D investment, coupled with relatively higher levels of R&D intensity over the analysis period, is an important factor that enables OECD firms/industries/countries to benefit more from R&D spillovers compared to the reference category of non-OECD firms/industries/countries.[33] This is in line with Cohen and Levinthal (1989) and Aldieri et al. (2018), who report that investment in R&D is a necessary condition for benefiting from external knowledge spillovers. It is also consistent with findings in Goñi and Maloney (2017), who find that returns on own R&D investment is lower in less developed countries and the latter's ability to benefit from external knowledge via transfers or spillovers depends on other absorptive capacity factors such as education, quality of the scientific infrastructure and the national innovation system, and the quality of entrepreneurship.

The last two findings indicate that countries that have higher R&D intensity secure larger productivity gains from spillovers, but the productivity gains diminish as the firm/industry becomes more R&D intensive and approaches the technology frontier. Therefore, in future research, it is necessary to investigate how and whether the productivity effects of spillovers are mediated through own R&D

[32] See summary statistics in *Table A5* in the *Online Appendix*.

[33] Over the data period in the empirical studies (1965 – 2008), of the non-OECD countries only Hong Kong, Singapore and Israel have registered R&D intensity or patent-per-capita rates comparable to the OECD average (OECD *Science, Technology and Industry Scoreboard*)

intensity and whether the mediation is subject to threshold effects – both at the country and the firm/industry levels.

Turning to spillover characteristics, there is weakly-consistent evidence that studies that adopt a scaled (asymmetric) spillover weight tend to report relatively larger estimates compared to those that use a symmetric weight. However, it is difficult to ascertain whether this represents a genuinely larger effect for two reasons. First, the evidence is only weakly-consistent under our decision criteria. Secondly, the larger effect may well be due to smaller magnitude of the spillover pool constructed with asymmetric weights.[34] The same argument applies to the weakly-consistent evidence that spillover pools constructed with R&D investment (instead of R&D capital stock) are associated with larger productivity effects. Finally, there is no systematic difference between weighted and unweighted spillovers. The findings suggest that there may be several candidates for measuring R&D (knowledge) externalities, which are essentially unobserved in the data. As indicated by Griliches (1992), however, knowledge spillovers based on technology proximity may be theoretically more relevant.

With respect to estimation methods, we find strongly-consistent evidence that effect-size estimates obtained from first-differenced data are smaller than those obtained from non-differenced data. We also find moderately-consistent evidence that estimators that take account of panel cointegration yield smaller estimates than others that do not control. Both findings are in line with econometric theory. First-differencing is known to produce an attenuation bias because mismeasurement errors in the level variables are exacerbated when they are time-differenced (Draca et al., 2007; Ugur et al., 2016). Also, in the presence of a cointegrating relationship between panels, effect-size estimates from estimators such as dynamic OLS or similar methods converge on the true effect values much faster compared to cases where the variables are assumed stationary (Stock, 1987).

With respect to two continuous variables (log citations and log number of observations) we find that they have no effect on reported productivity estimates, but they reduce the magnitude of the selection bias (the constant term) considerably. The selection bias is reduced from 'substantial' and 'severe' levels in the bivariate meta-regression results of *Tables 3* and *4* to insignificant in the multivariate meta-regression results of *Table 6*.

In what follows, we exploit the findings from the WALS-selected HMRM (*Table 6*) to obtain meta-effect estimates from 'best-practice' research that satisfies three criteria: (i) the study is published after

---

[34] Recall that the two weights (openness to import and bilateral import shares) used in the construction of the scaled (asymmetric) spillover pool are fractions. As indicated above, the spillover pool based on bilateral impost shares only (eq. 1a) would be larger than the spillover pool constructed with bilateral import shares and openness to imports.

2000; (ii) own R&D capital is controlled for in the primary-study model; (iii) primary-study estimates take account of panel cointegration or utilise instrumental variable (IV) estimators. We use these criteria to define 'best practice' because studies published after 2000 take account of the modelling and estimation contributions by pioneering studies such as Griliches (1992), Coe et al. (1997) and Keller (1998). Second, controlling for own R&D is theoretically necessary under the knowledge capital model that underpins the reported effect-size estimates (Griliches, 1979; 1992) and its TFP equivalent (Coe et al., 1997). Third, checking for panel cointegration is standard practice when the number of cross section units in panel data is small (see Coe et al., 1997; 2009; Keller, 1998; and Eberhardt et al., 2013). Finally, IV estimators take account of endogeneity that may result from measurement errors, model misspecification or simultaneity (see Guellec and Van Pottelsberghe de La Potterie, 2001; 2004 and Lehto, 2007).

We apply the best-practice scenario to full sample evidence and obtain *meta-effect1* in accordance with (8a) below. Then we condition on OECD firm/industry/country data only and obtain meta-effect2 (8b below). We use *meta-effect2* as a sensitivity check for two reasons. First, intra-OECD data for R&D is relatively more comparable due to harmonisation of innovation definitions and data collection rules after the adoption of the Oslo Manual in 1992. Secondly, most OECD countries have relatively higher levels of R&D intensity due to a longer history of investment in (and public support for) R&D. As such, OECD firms/industries/countries can be expected to have a higher degree of absorptive capacity. The meta-effect estimates from the WALS-selected HMRM are calculated as follows:

$$Meta\ effect1 = \frac{\partial tvalue}{\partial precision} + \frac{\partial tvalue}{\partial pbn2000} + \frac{\partial tvalue}{\partial ctrlownRD} + \frac{\partial tvalue}{\partial cointg} + \frac{\partial tvalue}{\partial IV} \tag{8a}$$

$$Meta\ effect2 = \frac{\partial tvalue}{\partial precision} + \frac{\partial tvalue}{\partial pbn2000} + \frac{\partial tvalue}{\partial ctrlownRD} + \frac{\partial tvalue}{\partial cointg} + \frac{\partial tvalue}{\partial IV} + \frac{\partial tvalue}{\partial OECD} \tag{8b}$$

The results are presented in *Table 7*. The best-practice meta-effect is insignificant when full-sample evidence is used. This is the case irrespective of whether the HMRM is estimated with or without frequency weights. When we restrict the sample to OECD data only, the meta-effect (0.189) is significant only when the model is estimated without frequency weights. This finding provides partial support to Griliches (1992), who reviewed the early work (mainly based on OECD data) and reported an effect-size interval ranging from 0.05-0.20. However, it must be noted that the OECD-specific meta-effect is estimated with a wide confidence interval, which ranges from 0.007 - 0.3710 and includes the average effect-size estimate (0.036) from the bivariate meta-regression reported in *Table 3*.

Table 7: Meta-effect based on best-practice: Full sample and OECD data only

	Meta-effect	Std. error	P>Z	95% c.i.
Meta-effect – using the WALS-selected model				
Meta-effect1a: Full sample, no freq. weights	0.096	0.093	0.298	-0.085, 0.278
Meta-effect1b: Full sample, with freq. weights	0.054	0.121	0.655	-0.183, 0.291
Meta-effect2a: OECD data, no freq. weights	0.189**	0.093	0.042	0.007, 0.371
Meta-effect2b: OECD data, with freq. weights	0.092	0.123	0.452	-0.148, 0.333

Notes: The meta-effect is obtained by taking the linear combination of the coefficients on the moderating variables, using the *lincom* command after estimating the WALS-selected model in *Stata.*

The meta-effect estimate based on OECD data echoes the HMRM result discussed above, which indicates that OECD firm/industry/country data is associated with systematically larger effect-size estimates compared to non-OECD data. Both findings indicate that larger spillover effects are observed in countries with relatively higher levels of R&D intensity. Therefore, we reiterate the importance of absorptive capacity as a necessary condition for securing larger productivity gains from external knowledge (Cohen and Levinthal, 1989; Aldieri et al., 2018; Goñi and Maloney, 2017).

6. Conclusions

Griliches (1992: S44) concludes that vote counting calculations may exaggerate the magnitude and the effect of R&D spillovers due to "upward selectivity bias" in the results he evaluates and because of various measurement issues he discusses in Griliches (1979). Therefore, he calls for further work that would shed better light on the productivity effects of R&D spillovers. In this study, we have responded to Griliches' call by using rigorous meta-analysis methods and a rich dataset that reflects the research effort after his call in 1992. Our findings suggest that the effect-size interval suggested by Griliches (0.05 – 0.20) is context-specific – i.e., it reflects findings from studies utilising mainly OECD data. When data from all countries are considered, the meta-effect based on 'best-practice' research is insignificant. This context specificity suggests that the productivity effect of spillovers is large in countries with a longer history of investment in own R&D, which provides absorptive capacity for securing productivity gains from external R&D spillovers.

Our study has also discovered additional evidence that calls for downward revision of the narrative review conclusions in the field. The first concerns the claim that the spillovers' productivity effect is larger than that of own R&D. Our findings indicate that such optimism may be misplaced. The effect of *own R&D* (0.073) is twice that of *all spillover types* (0.036); and similar to that of *knowledge spillovers* (0.069). The second relates to statistical power in the evidence base. Our findings indicate that more than two-thirds of the spillover effect estimates are based on evidence with low statistical power. Low statistical power does not invalidate the effect-size estimates but raises concerns about their reliability as a basis for evidence-based policy. To place the issue in empirical context, we have shown that the productivity effect of *all spillover types* is insignificant when evidence with adequate statistical power is considered. The effect of *knowledge spillovers* with adequate statistical power is statistically significant but too small (0.009) to be practically insignificant. Furthermore, the spillover effect with adequate statistical power is less than one-third of own-R&D effect (0.029).

With respect to future research, our findings allow for three recommendations. First, there is a case for taking account of the distance to the technology frontier when constructing the external R&D stock. Second-generation endogenous growth models with creative destruction (Aghion and Howitt, 1992; 2006) suggest that firms/industries/countries closer to the technology frontier are less likely to benefit from knowledge spillovers. This is confirmed to some extent in our findings concerning productivity effects among firms with high R&D intensity. Therefore, we call for two types of innovations in modelling the productivity effect of R&D spillovers: (i) using intercept and/or slope dummies that reflect high-R&D-intensity firms or industries; and/or (ii) augmenting the knowledge capital model with an interaction term between the external knowledge stock and distance (proximity) to the technology frontier. Whereas the former would allow for verifying whether the spillovers' productivity effects differ between high- and low-R&D-intensity firms/industries; the latter would allow for obtaining effect-size estimates corrected for distance to the technology frontier.

Secondly, we call for further attention to the lagged effects of both own-R&D and external-R&D capital, which are rarely discussed in the primary studies. However, both Griliches (1992) and Hall et al. (2010) have pointed out the importance of time-lags between R&D expenditures and innovation, between the latter and commercialization; and in the case of spillovers, between innovation and diffusion. Therefore, we recommend either auto-regressive distributed lags (ARDL) modelling or estimations with different lags as sensitivity checks. Such exercises are more feasible in spillover studies compared to own-R&D studies. This is because own-R&D capital tends to follow a random walk and this makes the lag structure more difficult to pin down (Hall et al., 2010).

Our third recommendation relates to explicit modelling of heterogeneity in the productivity effects of spillovers. One source of heterogeneity is the variations in the level of R&D intensity among firms/industries/countries in the sample. A second source of heterogeneity could be the variations in the quality of the external knowledge that diffuses through a given transmission channel (Di Lorenzo and Almeida, 2017). A third source is variations in absorptive capacity, which must be commensurate with the external knowledge pools faced (Aldieri et al., 2018); and varies with the level of own R&D investment (Cohen and Levinthal, 1989) and the level of domestic institutional/entrepreneurship quality (Goñi and Maloney, 2017). A fourth source could be unobserved factors such as management quality or governance/institutional norms.

One way for addressing such sources of heterogeneity would be to augment the knowledge capital model with interaction terms capturing the role of absorptive capacity, which can be measured as the deviation of the firm/industry/country from the sample average for R&D intensity. Another way would be to model heterogeneity as unobserved random effects through a hierarchical model (HM) framework - as is the case in Aiello and Ricotta (2016). HMs allows for nesting the individual observations within firms, industries, regions or countries; and for estimating spillover effects after controlling for between-industry or between-region variations modelled as random intercepts, random slopes or both. As discussed in the methodology, HMs can be estimated with Mundlak (1978) corrections to take account of endogeneity that may be due to correlations between unit-level covariates and random-effect components.

In studies based on industry or country data, the case for modelling heterogeneity is even stronger because the number of cross-section units is relatively small and this calls for panel time-series models instead of standard panel-data models (Eberhardt, 2012). The former allows for heterogeneous slope coefficients and can take account of cross-sectional dependence, which may be due to common unobservable factors. The benefits here are twofold. On the one hand, one can test if the productivity effect of own R&D is reduced when spillovers are modelled as unobservable common factors (Eberhardt et al., 2013). On the other hand, one can test whether explicit inclusion of the spillover types in the model wipes out the cross-sectional dependence.

From a public policy perspective, our findings indicate that the case for public support for R&D investment (input subsidies) may not be as strong as has been assumed so far. We have established empirically that the direct productivity effect of own R&D is comparable to and usually larger than that of spillovers. We have also demonstrated that the spillover effect is larger when firms/industries/countries possess a higher level of absorptive capacity, which is a positive function of own-R&D investment. Given these findings, the *R&D gap* - i.e., the gap between actual and socially optimal levels of R&D investment at the firm level - maybe smaller and more heterogeneous than what has been assumed so far. The R&D gap may be narrower because firms are aware of the need to invest in own R&D and build absorptive capacity as a basis for gains from knowledge externalities. Also, the R&D gap is likely to be heterogeneous, due to different firm characteristics (age, size, market share, etc.) and different industry characteristics (e.g., technology type, technology frontier and nature of competition). Therefore, we argue that direct or indirect support for R&D investment may be too blunt an instrument for securing additional R&D effort by supported firms. Instead, 'innovation prizes' for successful innovators as suggested by Akcigit et al. (2017), may enable funders to adjust the innovation prize to the quality of the innovation and the level of knowledge externalities it entails. Outcome-based support schemes also enable the funders to overcome information asymmetry and agency problems inherent in direct or indirect R&D support schemes.

References

Acharya, R.C., Keller, W., 2009. Technology transfer through imports. The Canadian Journal of Economics / Revue canadienne d'Economique 42 (4), 1411-1448.

Aghion, P., Akcigit, U., Howitt, P., 2014. What do we learn from Schumpeterian growth theory?, in: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. Elsevier, Amsterdam, pp. 515-563.

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. Econometrica 60 (2), 323-351.

Aghion, P., Howitt, P,. 2006. Joseph Schumpeter Lecture: Appropriate growth policy - A unifying framework. Journal of the European Economic Association 4 (2-3), 269-314.

Aiello, F., Cardamone, P., 2008. R&D spillovers and firms' performance in Italy. Empirical Economics 34 (1), 143-166.

Aiello, F., Ricotta, F., 2016. Firm heterogeneity in productivity across Europe: Evidence from multilevel models. Economics of Innovation and New Technology 25 (1), 57-89.

Aldieri, L., Sena V., Vinci C. P., 2018. Domestic R&D spillovers and absorptive capacity: Some evidence for US, Europe and Japan. International Journal of Production Economics 198 (April), 38-49.

Akcigit U, Hanley D, Stantcheva S., 2017. Optimal taxation and R&D policies. National Bureau of Economic Research Working Paper No. w22908.

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

Ang, J.B., Madsen, J.B., 2013. International R&D spillovers and productivity trends in the Asian miracle economies. Economic Inquiry 51 (2), 1523-1541.

Bernstein, J. I., 1989. The structure of Canadian inter-industry R & D spillovers, and the rates of return to R & D. *The Journal of Industrial Economics* 37(3), 315-328.Biatour, B., Dumont, M., Kegels, C., 2011. The determinants of industry-level total factor productivity in Belgium. Federal Planning Bureau Working Paper 7-11.

Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81 (4), 1347-1393.

Braconier, H., Sjöholm, F., 1998. National and international spillovers from R&D: Comparing a neoclassical and an endogenous growth approach. Weltwirtschaftliches Archiv 134 (4), 638-663.

Branstetter, L.G., 2001. Are knowledge spillovers international or intranational in scope?: Microeconometric evidence from the US and Japan. Journal of International Economics 53 (1), 53-79.

Bronzini, R., Piselli, P., 2009. Determinants of long-run regional productivity with geographical spillovers: the role of R&D, human capital and public infrastructure. Regional Science and Urban Economics 39 (2), 187-199.

Card, D., Krueger, A., 1995. Time-series minimum-wage studies: a meta-analysis. American economic review 85, 238-243.

Caves, D.W., Christensen, L.R., Diewert, W.E., 1982. The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50 (6), 1393-1414.

Chuang, Y., Lin, C., 1999. Foreign direct investment, R&D and spillover efficiency: Evidence from Taiwan's manufacturing firms. The Journal of Development Studies 35 (4), 117-137.

Cincera, M., Van Pottelsberghe de La Potterie, B., 2001. International R&D spillovers: a survey. Cahiers Economiques de Bruxelles 169 (1), 3-32.

Coe, D.T., Helpman, E., 1995. International R&D spillovers. European Economic Review 39 (5), 859-887.
Coe, D.T., Helpman, E., Hoffmaister, A.W., 1997. North-South R & D spillovers. The Economic Journal 107 (440), 134-149.
Coe, D.T., Helpman, E., Hoffmaister, A.W., 2009. International R&D spillovers and institutions. European Economic Review 53 (7), 723-741.
Cohen, W. M., Levinthal, D. A., 1989. Innovation and learning: the two faces of R & D. Economic Journal 99 (397), 569-596.
Costa-Font, J., McGuire, A., Stanley, T., 2013. Publication selection in health policy research: The winner's curse hypothesis. Health Policy 109 (1), 78-87.
D'Aspremont, C., Jacquemin, A., 1988. Cooperative and noncooperative R & D in duopoly with spillovers. The American Economic Review 78 (5), 1133-1137.
Dasgupta, P., Stiglitz, J., 1980. Uncertainty, industrial structure, and the speed of R&D. The Bell Journal of Economics 11 (1), 1-28.
Dasgupta, P., Stiglitz, J., 1988. Learning-by-doing, market structure and industrial and trade policies. Oxford Economic Papers 40 (2), 246-268.
De Luca, G., Magnus, J.R., 2011. Bayesian model averaging and weighted average least squares: equivariance, stability, and numerical issues. Stata Journal 11 (4), 518–544.
Demena, B. A., van Bergeijk, P. A., 2017. A meta-analysis of FDI and productivity spillovers in developing countries. Journal of Economic Surveys, 31 (2), 546-571.
Demidenko, E. 2004. Mixed Models: Theory and Applications. Hoboken, NJ: Wiley.
Di Lorenzo, F. Almeida P., 2017. The Role of relative performance in inter-firm mobility of inventors, Research Policy 46 (6), 1162-1174.
Doucouliagos, C., Stanley, T. D., 2013. Are all economic facts greatly exaggerated? Theory competition and selectivity. Journal of Economic Surveys, 27 (2), 316-339.
Draca, M., Sadun, R., Van Reenen, J., 2007. Productivity and ICT: A review of the evidence, in: Mansell, R. (Ed.), The Oxford Handbook of Information and Communication Technologies. Oxford University Press, Oxford and New York, pp. 100–147.
Ebbes, P., Böckenholt, U., Wedel, M., 2004. Regressor and random-effects dependencies in multilevel models. Statistica Neerlandica 58 (2), 161-178.
Eberhardt, M., 2012. Estimating panel time-series models with heterogeneous slopes. Stata Journal 12 (1), 61-71.
Eberhardt, M., Helmers, C., Strauss, H., 2013. Do spillovers matter when estimating private returns to R&D?. Review of Economics and Statistics 95 (2), 436-448.
Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315 (7109), 629-634.
Ellis, Paul 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.
Goñi, E., Maloney, W.F., 2017. Why don't poor countries do R&D? Varying rates of factor returns across the development process. European Economic Review, 94, 126-147.
González-Pereira, B., Guerrero-Bote, V. P., Moya-Anegón, F., 2010. A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics, 4 (3), 379-391.
Gorg, H., Strobl, E., 2001. Multinational companies and productivity spillovers: A meta-analysis. The Economic Journal 111 (475), 723-739.
Griffith, R., Harrison, R., Van Reenen, J., 2006. How special is the special relationship? Using the impact of US R&D spillovers on UK firms as a test of technology sourcing. American Economic Review 96 (5), 1859-1875.
Griliches, Z., 1979. Issues in assessing the contribution of research and development to productivity growth. The Bell Journal of Economics 10 (1), 92-116.
Griliches, Z., 1992. The search for R&D spillovers. Scandinavian Journal of Economics 94, S29-47.

Griliches, Z., Lichtenberg, F., 1984. Interindustry technology flows and productivity growth: A reexamination. The Review of Economics and Statistics 66 (2), 324-329.
Grilli, L., Rampichini, C. 2011. The role of sample cluster means in multilevel models: A view on endogeneity and measurement error issues. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 7 (4), 121-133.
Grossman, G.M., Helpman, E., 1991. Innovation and Growth in the Global Economy MIT Press, Cambridge, MA.
Guellec, D., Van Pottelsberghe de la Potterie, B., 2001. R&D and productivity growth: Panel data analysis of 16 OECD countries. OECD Economic Studies 33, 1-24.
Guellec, D., Van Pottelsberghe de la Potterie, B., 2004. From R&D to productivity growth: Do the institutional settings and the source of funds of R&D matter? Oxford Bulletin of Economics and Statistics 66 (3), 353-378.
Hall, B.H., Mairesse, J., Mohnen, P., 2010. Measuring the returns to R&D, in: Hall, B.H., Rosenberg, N. (Eds.), Handbook of the Economics of Innovation. Elsevier, New York, pp. 1033-1082.
Hanchane, S., Mostafa, T., 2012. Solving endogeneity problems in multilevel estimation: an example using education production functions. Journal of Applied Statistics 39 (5), 1101-1114.
Hanel, P., 2000. R&D, interindustry and international technology spillovers and the total factor productivity growth of manufacturing industries in Canada, 1974-1989. Economic Systems Research 12 (3), 345-361.
Harbord, R.M., Higgins, J., 2008. Meta-regression in Stata. The Stata Journal 8 (4), 493-519.
Harhoff, D., 2000. R&D spillovers, technological proximity, and productivity growth. Schmalenbach Business Review 52 (3), 238-260.
Harzing, A.-W., 2008. Google Scholar-a new data source for citation analysis. Ethics in Science and Environmental Politics 8 (1), 61-73.
Havránek, T., 2015. Measuring intertemporal substitution: The importance of method choices and selective reporting. Journal of the European Economic Association 13 (6), 1180-1204.
Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ: British Medical Journal 327 (7414), 557–560.
Huang, F. L. 2018. Multilevel modeling and ordinary least squares regression: How comparable are they?. The Journal of Experimental Education 86 (2), 265-281.
Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H., 2017. The power of bias in economics research. The Economic Journal 127 (605), F236-F265.
Iršová, Z., Havránek, T. 2013. Determinants of horizontal spillovers from FDI: Evidence from a large meta-analysis. World Development 42(February), 1-15.
Johnson, D.K., Evenson, R.E., 1999. R&D spillovers to agriculture: Measurement and application. Contemporary Economic Policy 17 (4), 432-456.
Katahira, K., 2016. How hierarchical models improve point estimates of model parameters at the individual level. Journal of Mathematical Psychology 73(August), 37-58.
Ke, S., Luger, M.I., 1996. Embodied technological progress, technology-related producer inputs, and regional factors in a firm-level model of growth. Regional Science and Urban Economics 26 (1), 23-50.
Keller, W., 1998. Are international R&D spillovers trade-related?: Analyzing spillovers among randomly matched trade partners. European Economic Review 42 (8), 1469-1481.
Krammer, S.M., 2010. International R&D spillovers in emerging markets: The impact of trade and foreign direct investment. The Journal of International Trade & Economic Development 19 (4), 591-623.
Kwon, H.U., 2004. Productivity growth and R&D spillovers in Japanese manufacturing industry. Hi-Stat Discussion Paper Series No.16.
Lehto, E., 2007. Regional impact of research and development on productivity. Regional Studies 41 (5), 623-638.

Magnus, J.R., Powell, O., Prüfer, P., 2010. A comparison of two model averaging techniques with an application to growth empirics. Journal of Econometrics 154 (2), 139-153.
Mairesse, J., Mulkay, B., 2008. An exploration of local R&D spillovers in France. NBER Working Paper 14552.
Mansfield, E., 1980. Basic research and productivity increase in manufacturing. The American Economic Review 70 (5), 863-873.
McCulloch, C. E., Searle, S. R., Neuhaus, J. M., 2008. Generalized, Linear, and Mixed Models. 2nd ed. Hoboken, NJ: Wiley.
McVicar, D., 2002. Spillovers and foreign direct investment in UK manufacturing. Applied Economics Letters 9 (5), 297-300.
Mebratie, A. D., van Bergeijk, P. A., 2013. Firm heterogeneity and development: A meta-analysis of FDI productivity spillovers. The Journal of International Trade & Economic Development 22 (1), 53-74.
Meho, L.I., Yang, K., 2006. A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. Journal of the American Society for Information Science and Technology 58 (13), 2105–2125.
Meyer, K. E., Sinani, E., 2009. When and where does foreign direct investment generate positive spillovers? A meta-analysis. Journal of International Business Studies 40 (7), 1075-1094.
Mingers, J., Yang, L., 2017. Evaluating journal quality: A review of journal citation indicators and ranking in business and management. European Journal of Operational Research, 257 (1), 323-337.
Mohnen, P., 1996. R&D externalities and productivity growth. OECD STI Review 17, 39-59.
Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R. , Peters, J. L., Cooper N. J., 2009. Assessment of regression based methods to adjust for publication bias through a comprehensive simulation study. BMC Medical Research Methodology 9 (2), 1-17.
Moreno, S. G., Sutton, A. J., Thompson, J. R., Ades, A. E., Abrams, K. R., Cooper, N. J., 2011. A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity. Statistics in Medicine 31 (14), 1407–1417.
Müller, W.G., Nettekoven, M., 1999. A panel data analysis: research and development spillover. Economics Letters 64 (1), 37-41.
Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46 (1), 69-85.
Nisonger, T.E., 2004. Citation autobiography: An investigation of ISI database coverage in determining author citedness. College & Research Libraries 65 (2), 152-163.
Parameswaran, M., 2009. International trade, R&D spillovers and productivity: Evidence from Indian manufacturing industry. Journal of Development Studies 45 (8), 1249-1266.
Philips, A. Q., 2016. Seeing the forest through the trees: A meta-analysis of political budget cycles. Public Choice, 168 (3-4), 313-341.
Raut, L.K., 1995. R & D spillover and productivity growth: Evidence from Indian private firms. Journal of Development Economics 48 (1), 1-23.
Romer, P.M., 1990. Endogenous technological change. Journal of Political Economy 98 (5, Part 2), S71-S102.
Schumpeter, J.A., 1942. *Capitalism, Socialism, and Democracy.* New York: Harper and Brothers. (Harper Colophon edition, 1976).
Shieh, Y.-Y., Fouladi, R.T., 2003. The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. Educational and Psychological Measurement 63 (6), 951-985.
Snijders, T., Bosker, R., 2012. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Analysis. Sage, London.
Stanley, T.D., 2005. Beyond publication bias. Journal of Economic Surveys 19 (3), 309-345.
Stanley, T.D., 2008. Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. Oxford Bulletin of Economics and Statistics 70 (1), 103-127.

Stanley, T.D., Doucouliagos, H., 2012. Meta-regression Analysis in Economics and Business. Routledge, London and New York.
Stanley, T.D., Doucouliagos, H., 2014. Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods 5 (1), 60-78.
Stanley, T.D., Doucouliagos, H., 2017. Neither fixed nor random: Weighted least squares meta-regression. Research Synthesis Methods 8 (1), 19-42.
Stanley, T. D., H. Doucouliagos, M. Giles, J.H. Heckemeyer, R. Johnston, , P. Laroche, Nelson, J. P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R.S., Rost, K. (2013). Meta-analysis of economics research reporting guidelines. Journal of Economic Surveys, 27(2), 390-394.
Stock, J.H., 1987. Asymptotic properties of least squares estimators of cointegrating vectors. Econometrica 55 (5), 1035-1056.
Ugur, M., Awaworyi Churchill, S., Solomon, E., 2018. Technological innovation and employment in derived labour demand models: A hierarchical meta-regression analysis. Journal of Economic Surveys 32 (1), 50-82.
Ugur, M., Trushin, E., Solomon, E., Guidi, F., 2016. R&D and productivity in OECD firms and industries: A hierarchical meta-regression analysis. Research Policy 45 (10), 2069-2086.
Wang, Y., Chao, C.W., 2008. "Spillovers" and productivity: The case of the Taiwanese high-tech firms. Contemporary Economic Policy 26 (2), 248-258.
Wei, Y., Liu, X., 2006. Productivity spillovers from R&D, exports and FDI in China's manufacturing sector. Journal of International Business Studies 37 (4), 544-557.
Wooldridge, J. M., 1995. Score diagnostics for linear models estimated by two stage least squares. In Advances in Econometrics and Quantitative Economics: Essays in honor of Professor C. R. Rao, eds. G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, 66–87. Cambridge, MA: Blackwell Publishers.
Wooster, R. B., Diebel, D. S., 2010. Productivity spillovers from foreign direct investment in developing countries: a meta-regression analysis. Review of Development Economics 14 (3), 640-655.
Yu, H., Jiang, S., Land, K.C., 2015. Multicollinearity in hierarchical linear models. Social Science Research 53, 118-136.
Yuen, J., 2018. Comparison of impact factor, eigenfactor metrics, and SCImago journal rank indicator and h-index for neurosurgical and spinal surgical journals. World Neurosurgery, 119(November), e328-e337.

Primary studies included in the meta-analys

Acharya, R.C., Keller, W., 2009. Technology transfer through imports. The Canadian Journal of Economics / Revue canadienne d'Economique 42 (4), 1411-1448.

Adams, J.D., Jaffe, A.B., 1996. Bounding the effects of R&D: An investigation using matched establishment-firm Data. The RAND Journal of Economics 27 (4), 700-721.

Aiello, F., Cardamone, P., 2005. R&D spillovers and productivity growth: evidence from Italian manufacturing microdata. Applied Economics Letters 12 (10), 625-631.

Aldieri, L., Cincera, M., 2009. Geographic and technological R&D spillovers within the triad: Micro evidence from US patents. The Journal of Technology Transfer 34 (2), 196-211.

Belitz, H., Mölders, F., 2016. International knowledge spillovers through high-tech imports and R&D of foreign-owned firms. The Journal of International Trade & Economic Development 25 (4), 590-613.

Biatour, B., Dumont, M., Kegels, C., 2011. The determinants of industry-level total factor productivity in Belgium. Federal Planning Bureau Working Paper 7-11.

Bitzer, J., Geishecker, I., 2006. What drives trade-related R&D spillovers? Decomposing knowledge-diffusing trade flows. Economics Letters 93 (1), 52-57.

Bitzer, J., Kerekes, M., 2008. Does foreign direct investment transfer technology across borders? New evidence. Economics Letters 100 (3), 355-358.

Bloch, C., 2013. R&D spillovers and productivity: an analysis of geographical and technological dimensions. Economics of Innovation and New Technology 22 (5), 447-460.

Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81 (4), 1347-1393.

Braconier, H., Ekholm, K., Knarvik, K.H.M., 2001. In search of FDI-transmitted R&D spillovers: A study based on Swedish data. Review of World Economics 137 (4), 644-665.

Braconier, H., Sjöholm, F., 1998. National and international spillovers from R&D: Comparing a neoclassical and an endogenous growth approach. Weltwirtschaftliches Archiv 134 (4), 638-663.

Branstetter, L.G., 2001. Are knowledge spillovers international or intranational in scope?: Microeconometric evidence from the US and Japan. Journal of International Economics 53 (1), 53-79.

Bronzini, R., Piselli, P., 2009. Determinants of long-run regional productivity with geographical spillovers: the role of R&D, human capital and public infrastructure. Regional Science and Urban Economics 39 (2), 187-199.

Cincera, M., 2005. Firms' productivity growth and R&D spillovers: an analysis of alternative technological proximity measures. Economics of Innovation and New Technology 14 (8), 657-682.

Coe, D.T., Helpman, E., Hoffmaister, A.W., 1997. North-South R & D spillovers. The Economic Journal 107 (440), 134-149.

Coe, D.T., Helpman, E., Hoffmaister, A.W., 2009. International R&D spillovers and institutions. European Economic Review 53 (7), 723-741.

del Barrio-Castro, T., López-Bazo, E., Serrano-Domingo, G., 2002. New evidence on international R&D spillovers, human capital and productivity in the OECD. Economics Letters 77 (1), 41-45.

Edmond, C., 2001. Some panel cointegration modelsof international R&D spillovers. Journal of Macroeconomics 23 (2), 241-260.

Engelbrecht, H.-J., 1997. International R&D spillovers, human capital and productivity in OECD economies: An empirical investigation. European Economic Review 41 (8), 1479-1488.

Frantzen, D., 2000. R&D, human capital and international technology spillovers: A cross-country analysis. The Scandinavian Journal of Economics 102 (1), 57-75.

Frantzen, D., 2002. Intersectoral and international R&D knowledge spillovers and total factor productivity. Scottish Journal of Political Economy 49 (3), 280-303.
Funk, M., 2001. Trade and international R&D spillovers among OECD countries. Southern Economic Journal 67 (3), 725-736.
Griffith, R., Harrison, R., Van Reenen, J., 2006. How special is the special relationship? Using the impact of US R&D spillovers on UK firms as a test of technology sourcing. American Economic Review 96 (5), 1859-1875.
Guellec, D., Van Pottelsberghe de la Potterie, B., 2001. R&D and productivity growth: Panel data analysis of 16 OECD countries. OECD Economic Studies 33, 1-24.
Guellec, D., Van Pottelsberghe de la Potterie, B., 2004. From R&D to productivity growth: Do the institutional settings and the source of funds of R&D matter? Oxford Bulletin of Economics and Statistics 66 (3), 353-378.
Gutierrez, L., Gutierrez, M.M., 2003. International R&D spillovers and productivity growth in the agricultural sector. A panel cointegration approach. European Review of Agricultural Economics 30 (3), 281-303.
Harhoff, D., 2000. R&D spillovers, technological proximity, and productivity growth. Schmalenbach Business Review 52 (3), 238-260.
Hejazi, W., Safarian, A.E., 1999. Trade, foreign direct investment, and R&D spillovers. Journal of International Business Studies 30 (3), 491-511.
Higon, D.A., 2007. The impact of R&D spillovers on UK manufacturing TFP: A dynamic panel approach. Research Policy 36 (7), 964-979.
Jacobs, B., Nahuis, R., Tang, P.J., 2002. Sectoral productivity growth and R&D spillovers in the Netherlands. De Economist 150 (2), 181-210.
Jaffe, A.B., 1988. Demand and supply influences in R & D intensity and productivity growth. The Review of Economics and Statistics 70 (3), 431-437.
Jaffe, A.B., 1989. Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. Research Policy 18 (2), 87-97.
Johnson, D.K., Evenson, R.E., 1999. R&D spillovers to agriculture: Measurement and application. Contemporary Economic Policy 17 (4), 432-456.
Kao, C., Chiang, M.H., Chen, B., 1999. International R&D spillovers: an application of estimation and inference in panel cointegration. Oxford Bulletin of Economics and Statistics 61 (S1), 691-709.
Ke, S., Luger, M.I., 1996. Embodied technological progress, technology-related producer inputs, and regional factors in a firm-level model of growth. Regional Science and Urban Economics 26 (1), 23-50.
Keller, W., 1998. Are international R&D spillovers trade-related?: Analyzing spillovers among randomly matched trade partners. European Economic Review 42 (8), 1469-1481.
Krammer, S.M., 2010. International R&D spillovers in emerging markets: The impact of trade and foreign direct investment. The Journal of International Trade & Economic Development 19 (4), 591-623.
Kwon, H.U., 2004. Productivity growth and R&D spillovers in Japanese manufacturing industry. Hi-Stat Discussion Paper Series No.16.
Lee, G., 2005. Direct versus indirect international R&D spillovers. Information Economics and Policy 17 (3), 334-348.
Lee, G., 2006. The effectiveness of international knowledge spillover channels. European Economic Review 50 (8), 2075-2088.
Lehto, E., 2007. Regional impact of research and development on productivity. Regional Studies 41 (5), 623-638.
Lichtenberg, F.R., Van Pottelsberghe de La Potterie, B., 1998. International R&D spillovers: a comment. European Economic Review 42 (8), 1483-1491.

López-Pueyo, C., Barcenilla-Visús, S., Sanaú, J., 2008. International R&D spillovers and manufacturing productivity: A panel data analysis. Structural Change and Economic Dynamics 19 (2), 152-172.
Los, B., Verspagen, B., 2000. R&D spillovers and productivity: evidence from US manufacturing microdata. Empirical Economics 25 (1), 127-148.
Lumenga-Neso, O., Olarreaga, M., Schiff, M., 2005. On `indirect' trade-related R&D spillovers. European Economic Review 49 (7), 1785-1798.
Lychagin, S., Pinkse, J., Slade, M.E., Reenen, J.V., 2016. Spillovers in space: does geography matter? The Journal of Industrial Economics 64 (2), 295-335.
McVicar, D., 2002. Spillovers and foreign direct investment in UK manufacturing. Applied Economics Letters 9 (5), 297-300.
Negassi, S., 2009. International R&D spillovers and economic performance of firms: an empirical study using random coefficient models. Applied Economics 41 (8), 947-976.
Orlando, M.J., 2004. Measuring spillovers from industrial R&D: On the importance of geographic and technological proximity. The RAND Journal of Economics 35 (4), 777-786.
Ornaghi, C., 2006. Spillovers in product and process innovation: Evidence from manufacturing firms. International Journal of Industrial Organization 24 (2), 349-380.
Parameswaran, M., 2009. International trade, R&D spillovers and productivity: Evidence from Indian manufacturing industry. Journal of Development Studies 45 (8), 1249-1266.
Park, J., 2004. International student flows and R&D spillovers. Economics Letters 82 (3), 315-320.
Park, W.G., 1995. International R&D spillovers and OECD economic growth. Economic Inquiry 33 (4), 571-591.
Raut, L.K., 1995. R & D spillover and productivity growth: Evidence from Indian private firms. Journal of Development Economics 48 (1), 1-23.
Van Pottelsberghe de La Potterie, B., Lichtenberg, F., 2001. Does foreign direct investment transfer technology across borders? Review of Economics and Statistics 83 (3), 490-497.
Verspagen, B., 1997. Estimating international technology spillovers using technology flow matrices. Review of World Economics 133 (2), 226-248.
Wang, Y., Chao, C.W., 2008. "Spillovers" and productivity: The case of the Taiwanese high-tech firms. Contemporary Economic Policy 26 (2), 248-258.
Xu, B., Wang, J., 1999. Capital goods trade and R&D spillovers in the OECD. The Canadian Journal of Economics / Revue canadienne d'Economique 32 (5), 1258-1274.
Zhu, L., Jeon, B.N., 2007. International R&D spillovers: Trade, FDI, and information technology as spillover channels. Review of International Economics 15 (5), 955-976.