UNIVERSITY of GREENWICH
FOUNDATION FOR EUROPEAN PROGRESSIVE STUDIES
FONDATION EUROPÉENNE D'ÉTUDES PROGRESSISTES
GPERC

Greenwich papers in political economy

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

Mehmet Ugur (University of Greenwich)
Sefa Awaworyi Churchill (RMIT University)
Hoang Minh Luong (University of Greenwich)

Abstract

Endogenous growth theory and the knowledge capital model predict that research and development (R&D) investment is associated with increasing returns and positive externalities. These insights have informed public support for R&D investment directly and indirectly. We aim to establish where the balance of the evidence lies, the extent to which the evidence has adequate statistical power, and which factors may explain the variation in the empirical findings. Drawing on 983 spillovers and 501 own-R&D effect-size estimates from 60 empirical studies, we find that the average productivity effect of spillovers: (i) is smaller than what is reported in most narrative reviews; (ii) is even smaller when only adequately-powered evidence is considered; (iii) differs by spillover types; and (iv) is not larger than that of own-R&D. We also report that the percentage of adequately-powered evidence is low (30%-55%). We highlight the implications of these findings for future research and public policy design.

YEAR 2018

NO: GPERC67

GREENWICH POLITICAL ECONOMY RESEARCH CENTRE (GPERC)

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.

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

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]

Abstract

Endogenous growth theory and the knowledge capital model predict that research and development (R&D) investment is associated with increasing returns and positive externalities. These insights have informed public support for R&D investment directly and indirectly. We aim to establish where the balance of the evidence lies, the extent to which the evidence has adequate statistical power, and which factors may explain the variation in the empirical findings. Drawing on 983 spillovers and 501 own-R&D effect-size estimates from 60 empirical studies, we find that the average productivity effect of spillovers: (i) is smaller than what is reported in most narrative reviews; (ii) is even smaller when only adequately-powered evidence is considered; (iii) differs by spillover types; and (iv) is not larger than that of own-R&D. We also report that the percentage of adequately-powered evidence is low (30% - 55%). We highlight the implications of these findings for future research and public policy design.

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

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

a: University of Greenwich; b: RMIT University

1. Introduction

The effect of knowledge externalities on productivity has direct relevance for public policy and welfare. In early modelling efforts (e.g., Arrow, 1962; Chipman, 1970; Meade, 1952; Sheshinski, 1967; Simon, 1947), the externality is due to 'learning from others' and the latter is a positive function of the learner's capital stock. 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 modelling efforts, where investment in innovation is associated with increasing returns (Grossman and Helpman, 1991; Romer, 1990).

Of the narrative reviews that evaluate the empirical research, Griliches (1992) acknowledges the risk of publication selection bias but goes on to conclude that the elasticity estimates of the spillover effect are practically 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 elasticity estimates that are similar to those of own R&D.

The aim of this study is to contribute to existing knowledge and inform evidence-based policy by providing unbiased effect-size estimates for different spillover types, verifying the statistical power in the evidence base and accounting for sources of heterogeneity therein. Our findings are based on data from 60 primary studies that report 983 spillover and 501 own-R&D effects on productivity at the firm, industry or country/region levels.

Drawing on best practice in meta-regression analysis (Stanley et al., 2013), we address a number of research questions with relevance for public policy, business decision-making and future research: (i) What is the average productivity effect of spillovers after controlling for publication selection bias and choosing the appropriate estimation method? (ii) How does the average effect vary between spillover types (e.g., knowledge spillovers, rent spillovers, and mixed spillovers) and between cross-section units of the analysis (firms, industries, countries/regions)? (iii) How do spillover effects compare to those of own R&D? (iv) Does the evidence in this research field have adequate statistical power and what are the implications for average effect size when only adequately-powered evidence is used? (v) What moderating factors related to publication type, sample characteristics, estimation methods, etc. are significant in explaining the heterogeneity in the evidence base?

To address these questions, the paper is organised in six sections. In section 2, we present the theoretical/empirical framework that characterises the research field. Here, we introduce the primal approach adopted in the primary-studies; the different notions of external knowledge stock that yield knowledge, rent or mixed spillover effects; the different cross-section units for which spillover effects are estimated (firm, industry, country/region); the channels through which spillover effects unfold (technological proximity, patent citations/flows, intermediate inputs, imports, etc.); and the weights used for calculating the spillover pools. In section 3, we present our search strategy, inclusion/exclusion criteria and an overview of the evidence base.

Our meta-regression methodology is presented in section 4, where we first make the case for hierarchical models (HMs) that allow for flexibility in modelling unobserved heterogeneity and take account of nested nature of the data.[1] Then we verify the statistical power of the effect-size estimates in the literature and report the proportion of the adequately-powered evidence in the sample. Third, we compare the average effect-size estimates from the full sample with those obtained from adequately-powered evidence only. Finally, we go beyond the *ad hoc* manner in which the covariates (moderating variables) are selected in the multivariate meta-regression models (MVMRM). We propose a weighted average least squares (WALS) routine that provides information similar to a Bayesian model averaging (BMA) approach (De Luca and Magnus, 2011).

Section 5 presents the empirical findings on: (i) average effect-size estimates from the full sample, broken down by spillover types and cross-section units in the data; (ii) the proportion of adequately-powered estimates in the research field; (iii) average effect-size estimates based on adequately-powered evidence; and (iv) MVMRM estimates that provide information about observed sources of heterogeneity in the evidence base. Finally, in the conclusions, we distil the main findings and discuss their implications for public policy, business decision-making and future research.

---

[1] We consider the effect-size estimates as nested within primary studies or within clusters of different spillover types such as knowledge, rent or mixed spillovers. We model heterogeneity as random intercepts, random slopes or both – depending on likelihood ratio tests.

2. Spillovers and productivity: the theoretical/empirical framework

The effect-size estimates we analyse 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:

$$Y_{it} = A_i e^{\lambda t} C_{it}^{\alpha} L_{it}^{\beta} K_{it}^{\gamma} S_{it}^{\varphi} e^{u_{it}} \tag{1a}$$

Here, Y is real output; C is deflated physical capital stock; L is labour (number of employees or hours worked); K is own R&D capital stock; S is the spillovers pool (as specified below); λ is the rate of disembodied technological change; and A is a unit-specific constant. Subscripts t and i denote time and cross-section units (firms, industries or countries/regions), respectively. Two standard assumptions underpinning (1a) 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 stock in other units (j) where $j \neq i$; and can be unscaled (1b) or scaled (1c).

$$S_{it} = \sum_{j=1}^{n} W_{ij} K_{jt} \tag{1b}$$

$$S_{it} = \sum_{j=1}^{n} a_i W_{ij} K_{jt} \tag{1c}$$

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 (1c), a_i is an additional weight that captures the spillover-recipient's openness to international imports from or to 'intermediates trade' with units in j. Several studies (e.g., Coe et al., 1997; Keller, 1998; Krammer, 2010; Lee, 2005) 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'. 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 coefficient of main interest is φ - the *elasticity* of output with respect to external R&D. For comparison, however, we also extract the estimates of γ, the elasticity of output with respect to own R&D.

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

Some studies utilise total factor productivity (TFP) instead of output as the outcome variable. This is obtained either by estimating the model with conventional inputs (C and L) and obtaining TFP as the residual; or by constructing an input-based TFP index *a la* Malmqvist, Tornqvist or Caves *et al*. (1982). Then the model is:

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

Coefficient estimates from (2a) and (2b) will be consistent if the assumptions of perfect competition and constant returns to scale are satisfied.

Some studies estimate rates of return to own and external R&D investment by assuming rate-of-return equality between cross-section units. We exclude such studies (e.g., Griliches and Lichtenberg, 1984; Mansfield, 1980; Wolff and Nadiri, 1993) 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-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).

Focusing on the primal-approach and only elasticity estimates allow for pooling comparable evidence derived from a common model. Nevertheless, such demarcation does not eliminate the risk of heterogeneity among reported findings for several reasons. First, R&D spillover is a theoretical construct that is not observable in the data at hand. The common practice is to construct a stock of external R&D (knowledge) capital, *S*, which may capture different types of spillovers depending on the channels through which external knowledge is diffused and the weights (*W*) used (Griliches, 1992; Hall et al., 2010).

One channel is technological proximity between the sources and recipients of the spillover effects. In this case, the weight captures the recipient's proximity to its counterparts in the technology space. External R&D pools constructed with such weights are considered as sources of *knowledge spillovers* because the latter are not mediated through bilateral transactions between the sources and recipients of the knowledge externalities (Cincera and Van Pottelsberghe de La Potterie, 2001; Griliches, 1992; Hall et al., 2010; Mohnen, 1996; Verspagen, 1997). We have also described the reported estimates as *knowledge spillovers* if the spillover pool is constructed with equal weights (pure knowledge spillovers) or with weights that reflect geographical distance between units (spatial knowledge spillovers).

A large number of studies use bilateral import shares (see, Coe and Helpman, 1995) whereas some utilise intermediate input flows between industries (see, Biatour et al., 2011). In line with the literature, we classify the effect-size estimates based on such weights as *rent spillover effects*, which arise from the wedge between market prices paid by the buyers and the quality-adjusted true prices not observed in the data (Griliches, 1979; 1992).[2]

There is less clarity about how to characterise the external knowledge pools when the weights capture the intensity of bilateral transactions in knowledge-intensive goods/services (e.g., patent flows/citations; R&D collaborations, movements of R&D personnel, etc). Mohnen (1996) and Verspagen (1997) are in favour of classifying such pools as knowledge spillovers. However, Griliches (1992) and Hall et al., (2010) are less sanguine. They draw attention to the difficulty of classifying them as such because externalities mediated through bilateral transactions do not fit the theoretical concept of knowledge externalities – even though the transactions may involve knowledge-intensive goods/services. Therefore, we have coded for a third category of *mixed spillover effect* when the primary-study estimates are based on weights reflecting transaction intensity in knowledge-intensive goods/services (e.g., Branstetter, 2001; Griffith et al., 2006; Lee, 2005; Los and Verspagen, 2000).

Some primary studies (e.g., Keller, 1998; Krammer, 2010; Lee, 2005) argue that the level of rent and/or mixed spillovers would depend on two parameters: bilateral import/purchase shares and the spillover recipient's openness to 'trade'. The argument here is that the spillover recipient that is more open to 'trade' would derive a higher level of benefit from R&D externalities compared with another that is less open to trade. Hence a further source of heterogeneity is whether the weight is unscaled (i.e., reflects only bilateral import or purchase shares as in 1b above) or scaled (as in 1c above).

A third source of heterogeneity is due to different cross-section units in the panel datasets. Some studies use firm-level data while others use industry or country data. Hence we are faced with different levels of aggregation in the construction of the external knowledge stock and this may lead to different effect-size estimates. Also, the productivity effect of external knowledge may coexist with different countervailing effects that differ by the level of analysis. For example, at the industry level, knowledge externalities may co-exist with creative destruction effects (Aghion et al., 2014; Aghion and Howitt, 1992; Schumpeter, 1942) or

[2] Market prices are expected to be lower than quality-adjusted prices unless the innovative supplier has full monopoly power.

market-stealing effects (Bloom et al., 2013). Such adverse effects may take longer to unfold when the analysis is at the country level.

Finally, heterogeneity in the evidence base can result from between-study and within-study variations with respect to sample characteristics (e.g., high *versus* low R&D-intensity firms or industries), model specification (e.g., number of spillover types included in the model), estimation methods (standard OLS, dynamic OLS, panel cointegration, instrumental variable estimations, etc.), data period (relatively old or recent data), and publication types (journal articles, working papers, reports, etc.).

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

To identify the eligible studies, we began with studies cited in the existing narrative reviews mentioned above. This sample was augmented with potentially eligible studies identified through electronic search in *Google Scholar* and in the *Science & Technology Management Bibliography (STMB).*[3] The search period is set from 1980 to 2016, using keywords such as R&D spillovers, knowledge spillovers, productivity, R&D externalities, knowledge capital, etc. Following the best-practice recommendations for meta-analysis of economics and business research in Stanley *et al.* (2013), we have screened 2,324 potentially-relevant studies on the basis of title and abstract information. Screening decisions identified a sample of 106 studies that we then evaluated on the basis of full-text information.[4] We included studies that adopt the primal approach as indicated above. We excluded studies based on translog production functions (e.g., Aiello and Cardamone, 2008; Mairesse and Mulkay, 2008), those that adopt a dual approach (e.g., Bernstein and Nadiri, 1988) and those that estimate the effect of own R&D capital only.

Studies that adopt a translog model report coefficient estimates for both linear and non-linear terms for the spillovers variable. We excluded such studies because their estimates of the spillover effects are conditional on the level of the spillover stock and on the interaction of the

[3] 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, technology transfer. See, http://tomeclarke.ca/science.htm

[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.

latter with other inputs. Studies that adopt the so-called *dual approach* (e.g., Bernstein and Nadiri, 1988; Bernstein and Yan, 1997) are also excluded because they draw on different specifications for factor-demand and cost functions and thus yield more heterogeneous estimates. Another reason is that the dual-approach studies use ex-post (as opposed to expected) output on the right-hand-side of their models, and hence are suspected of reporting upward-biased estimates of both spillover and own-R&D effects (Griliches, 1992: S40). 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).

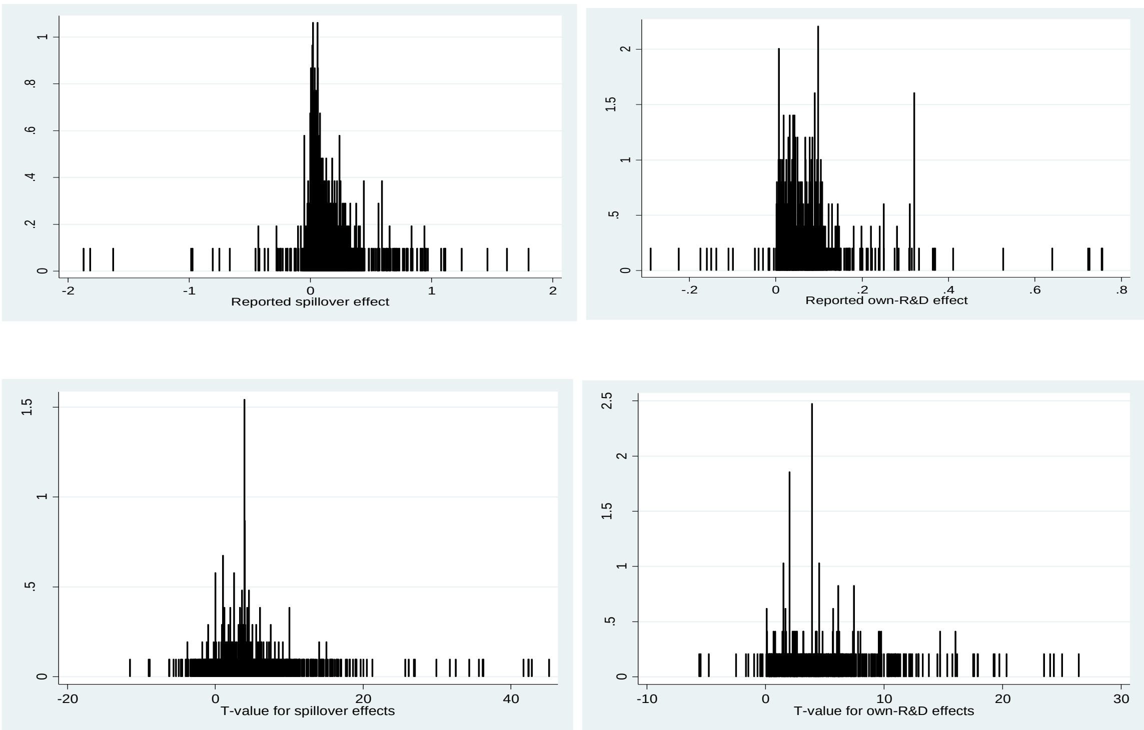
At the end of the full-text evaluation, we obtained a sample of 76 studies that adopt the primal approach. Further evaluation indicated that some of these studies reported rate-of-return estimates instead of elasticity estimates (e.g., Griliches and Lichtenberg, 1984; Hanel, 2000; Mansfield, 1980). Excluding these studies for reasons indicated in section 1, we have obtained a sample of 60 studies on which this meta-analysis is based.

We have extracted all reported effect-size estimates to ensure full use of existing information and avoid the risk of reviewer-induced selection bias. Data extraction led to a sample of 983 elasticity estimates for the effects of spillovers on productivity and 501 elasticity estimates for the effect of own-RD on productivity at the firm, industry and country/region levels. We coded each estimate to capture the observed sources of heterogeneity, which include: (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) spillover types; and (v) estimation methods (GMM, 2SLS, 3SLS, OLS, panel cointegration, FE, etc.).

Tables A1 and **A2** in the ***Appendix*** provide an overview of the primary-study characteristics. Most of the included studies (97%) are journal articles while the remaining 3% are working papers. All studies report one or more coefficient estimates for knowledge, rent, or mixed spillovers) but only 53 studies report estimates for own-R&D effects. It is also worth noting that 22 studies utilise firm-level data whilst 11 focus on industries and 25 focus on countries. One study (Acharya and Keller, 2009) focuses on both countries and industries while another (Bronzini and Piselli, 2009) focuses on Italian regions.

The median effect-size estimate and t-value, respectively, are 0.070 and 3.323 for the spillovers sample and 0.061 and 4.050 for the own-R&D sample.[5] The within-study median of the effect size is positive in most spillover studies, with the exception of four studies (Braconier and Sjöholm, 1998; Harhoff, 2000; Kwon, 2004; McVicar, 2002). Similarly, the within-study median of the effect size for own-R&D is also positive, with the exception of two studies (Biatour et al., 2011; Braconier and Sjöholm, 1998). The distribution of both effect sizes and associated t-values can be seen in **Figure 1**.

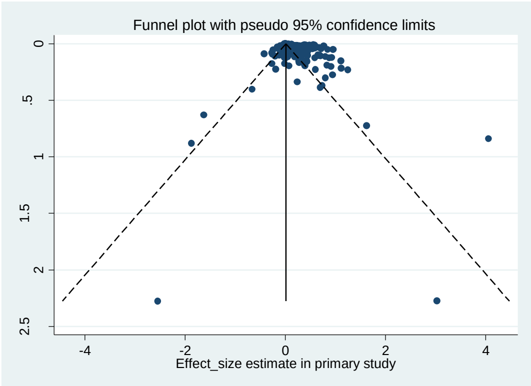
Figure 1: Histograms of effect-size estimates and associated t-values in primary studies



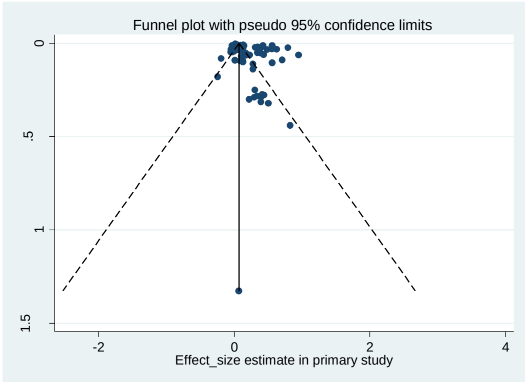
The descriptive information in Tables A1 and A2 and Figure 1 indicate that the majority of the effect-size estimates are positive; and the majority of the associated t-values are greater than 2. However, this summary information conceals a high degree of heterogeneity and may be affected by publication selection bias. Funnel graphs in **Figure 2a** and **2b** provide visual evidence on heterogeneity and selection bias, by spillover type and unit of analysis.

[5] The number of effect-size estimates reported by each study varies significantly, ranging from 1 to 102 in the case of spillover effects and from 1 to 45 in the case of own-R&D effects.

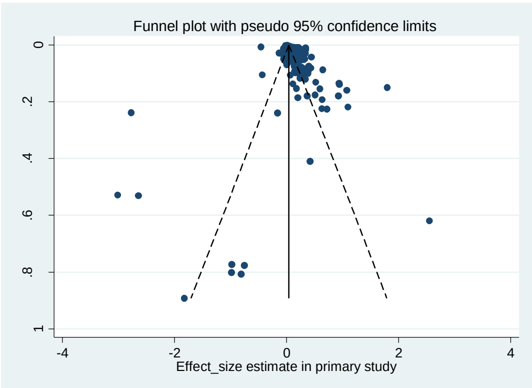
Figure 2a: Funnel graphs for productivity effects of spillover types and own RD



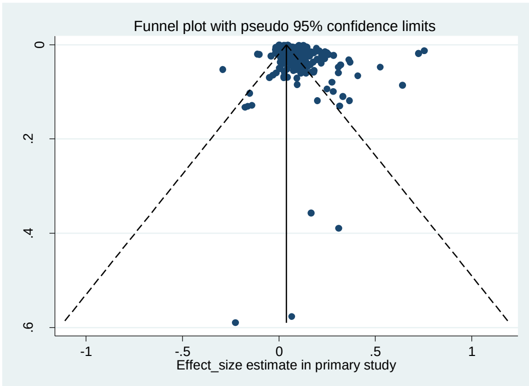
1. Knowledge spillover effects
Heterogeneity: 97.82%



2. Mixed spillover effects
Heterogeneity: 98.08%



3. Rent spillover effects
Heterogeneity: 97.37%

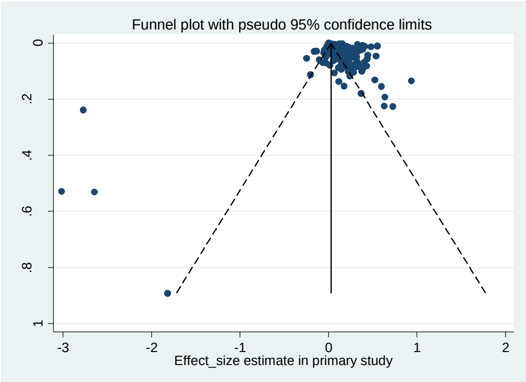


4. Own R&D effects
Heterogeneity: 97.87%

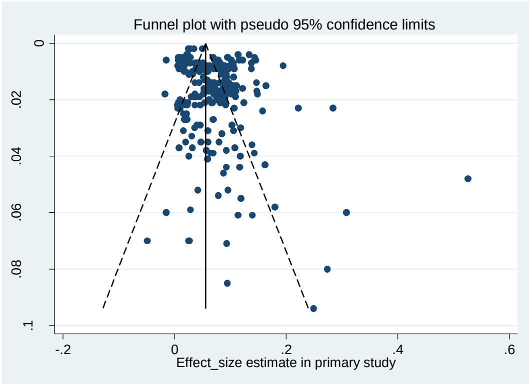
Funnel graphs in **Figure 2a** depict the distribution of the effect-size estimates for three spillover types (knowledge, mixed and rent spillovers) and own R&D. The vertical line indicates the fixed-effect weighted mean (FEWM), estimated with weights equal to the reciprocal of the squared standard error. The dotted lines represent the lower and upper limits of a 95% pseudo confidence interval that demarcates the range of heterogeneity due to sampling variations. Hence, effect-size estimates beyond the 95% confidence interval reflect *residual heterogeneity* that cannot be explained by sampling variations. The proportion of residual heterogeneity to sampling-related variation is obtained from a meta-regression model proposed by Harbord and

Higgins (2008) and reported under each graph.[6] Finally, the skewness in the distribution of the effect-size estimates around the FEWM provides indication of publication selection bias.

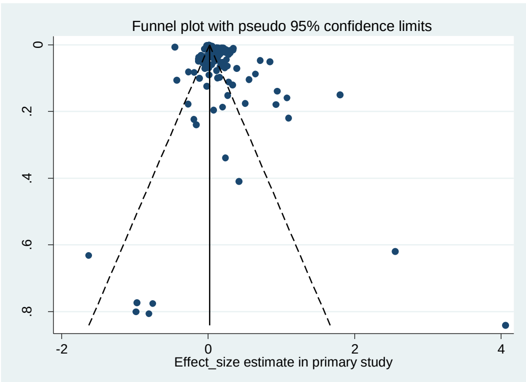
Figure 2b: Funnel graphs for units of analysis - spillover and own R&D effects



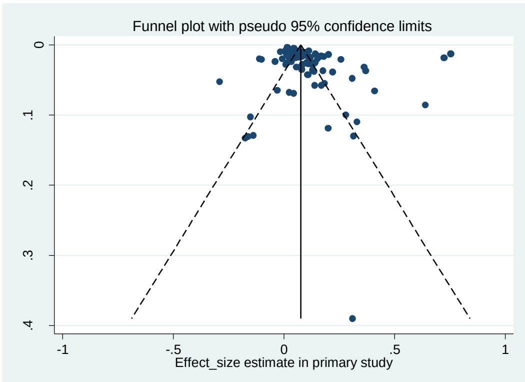
1. Spillover effects – country level
Heterogeneity: 97.93%



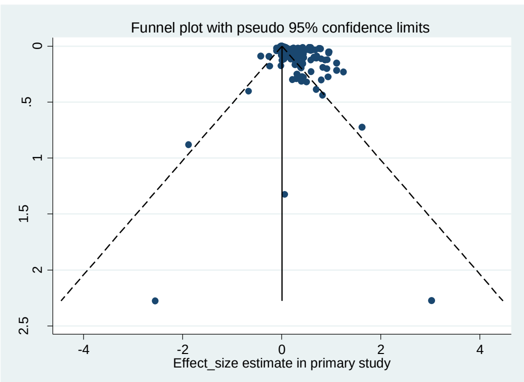
2. Own R&D effects – country level
Heterogeneity: 94.79%



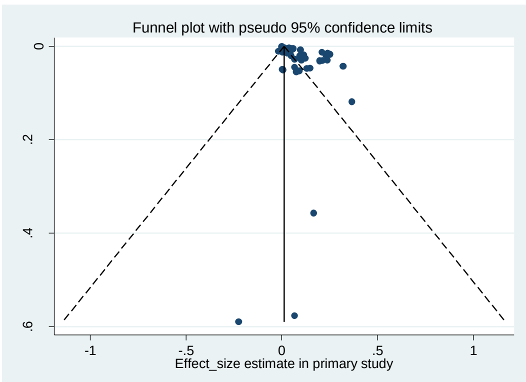
3. Spillover effects – industry level
Heterogeneity: 98.16%



4. Own R&D effects – industry level
Heterogeneity: 99.03%



5. Spillover effects – firm level
Heterogeneity: 98.03%



6. Own R&D effects – firm level
Heterogeneity: 92.72%

[6] There is no agreed-upon threshold for heterogeneity that is too high to yield a reliable confidence interval for the average effect-size, but Higgins et al. (2003) suggest that heterogeneity is low when I^2 is between 25%–50%, moderate for 50%–75%, and high for ≥75%.

Hence, the funnel graphs indicate that the proportion of residual heterogeneity that cannot be explained by sampling variation is high – around 97%-98%.[7] Also, the FEWM is usually smaller for the spillover effects compared to own R&D effects. Third, a larger proportion of the effect-size estimates reported in the primary studies are to the right of the vertical line that represents the FEWM. This is an indication of publication selection bias, the risk of which is acknowledged only in the narrative review by Griliches (1992).

These patterns are similar when we cluster the evidence by the cross-section units (see **Figure 2b**). In these clusters too, residual heterogeneity is high, the FEWMs of spillover effects is smaller or similar to the FEWMs of own R&D effects, and there are indications of selection bias. In the next section, we conduct formal tests to verify if selection bias exists, estimate the average effect size after taking account of selection bias and discuss the routines through which unobserved heterogeneity can be modelled.

4. Meta-regression methodology

The meta-regression model we utilise is based on the selection model proposed by Egger et al. (1997), who postulate that researchers search across model specifications, econometric techniques and data measures to find sufficiently large (hence statistically-significant) effect-size estimates. Assuming that the 'true' average effect is β, effect sizes reported in primary studies will vary around the true effect as follows:

$$effect\_size_i = \beta + \alpha SE_i + \xi_i \qquad (3a)$$

In model 3a, the effect size in primary studies is subject not only to an idiosyncratic error (ξ_i) but it is also conditional on the selection bias (α). The model in (3a) raises several estimation issues, some of which have been addressed in Stanley (2005; 2008) and Stanley and Doucouliagos (2012) among others. For example, heteroskedasticity is addressed through a weighted least squares (WLS) estimator where precision-squared ($1/SE_i^2$) is used as analytical weight. This is equivalent to dividing both sides of (3a) with the standard error (multiplying both sides with precision) (Stanley and Doucouliagos, 2012), leading to:

[7] There is no agreed-upon rule for when heterogeneity is too high for obtaining a reliable confidence interval for the average effect-size, but Higgins *et al.* (2003) suggest that heterogeneity is low when I^2 is between 25%–50%, moderate for 50%–75%, and high for ≥75%.

$$t_i = \alpha + \beta \left({}^{1}/_{SE_i} \right) + \omega_i \tag{3b}$$

Here t_i is the t-value associated with the effect-size estimate as reported in the primary study and v_i is the error term in (1) divided with the standard error. Testing for $\beta = 0$ is referred to as precision-effect test (**PET**), whilst testing for $\alpha = 0$ is the funnel asymmetry test (**FAT**). Rejection of the null hypothesis in PET indicates 'genuine' effect after controlling for publication selection. On the other hand, rejecting the null in FAT indicates the presence of selection bias, which can be interpreted as over-reporting of findings with larger standard errors (low precision).

The second issue is whether the relationship between primary-study estimates and their standard errors is linear or non-linear. Stanley and Doucouliagos (2014) provide evidence that a quadratic specification is preferred if the PET rejects the null hypothesis. Then, the non-linear Egger model and its WLS equivalent are:

$$effect\_size_i = \gamma + \varphi SE_i^2 + \vartheta_i \tag{4a}$$

$$t_i = \gamma \left({}^{1}/_{SE_i} \right) + \varphi SE_i + \varepsilon_i \tag{4b}$$

Model (4b) is estimated without a constant term and is referred to as precision-effect test corrected for standard errors (**PEESE).** The average effect-size estimate is γ. The latter is shown to have smaller bias and mean square error (Stanley and Doucouliagos, 2014).

The third issue relates to multiple effect-size estimates reported by primary studies. Because studies that report large number of estimates may dominate the informational content of the evidence base, we estimate (3b and 4b) with frequency weights, which are equal to the reciprocal of the number of estimates reported in each study.

One of the remaining four issues is within-study dependence, which may arise because the multiple effect-size estimates reported in each primary study are usually based on the same data set (or part thereof) and as such they are not necessarily random realizations of the true effect-size estimate. In the presence of within-study dependence, WLS would yield biased estimates for two reasons: (i) the sample size is exaggerated as the effect-size estimates within a study are treated as independent; and (ii) there will be a higher risk of committing type-I error (i.e., rejecting the null hypothesis when the latter is true) irrespective of the sample size (Snijders and Bosker, 2012).

The fifth issue is heterogeneity, which is a common feature of the empirical findings in economics (Stanley and Doucouliagos, 2017) and medical research (Turner et al., 2012). To address heterogeneity, we propose a hierarchical model (HM) framework that allows for nesting the effect-size estimates (Level 1 observations) within primary studies (Level 2 groups) that may, in turn, be nested within a spillover type (Level 3 cluster).

Stanley and Doucouliagos (2015; 2017) make a strong case for addressing heterogeneity through a weighted least-squares (WLS) estimator, as opposed to random-effects (RE) or fixed-effects (FE) estimators. In the WLS, between-study heterogeneity is modelled as a multiple of the variance of the individual effect size, whereas it is an additive term in the RE variance structure. In the FE, only within-study heterogeneity matters and therefore between-study heterogeneity is assumed as zero. Because simulations in Stanley and Doucouliagos (2015; 2017) exclude the HM estimator, we do not have evidence on how the latter would have performed in comparison to WLS, RE, and FE estimators. However, HMs allow for testing different assumptions about the variance structure through a likelihood ratio (LR) test. Hence, we choose a HM specification only if the LR tests indicate that a particular HM specification is preferable to WLS (the best-performer in Stanley and Doucouliagos, 2015; 2017) and other (more restricted) HM specifications.

The suite of HMs we test allow for different sources of heterogeneity, including: (i) *between-cluster* (Level 3) variation modelled as cluster-specific intercepts; (ii) *between-study* (Level 2) variation modelled as study-specific intercepts; (iii) *within-study* variation modelled as random slopes at Level 2; and (iv) any combination of the above. Formal HMs that correspond to such specifications are stated in **Box 1** in the ***Appendix***.[8]

The bivariate HM framework can be extended easily to multivariate meta-regression models (MVMRMs), which allow for modelling observed sources of heterogeneity indicated in section 3 above. These are binary variables with a value of 1 if a particular source is observed and 0 otherwise. Augmented with these additional covariates, the WLS and HM versions of the multivariate meta-regression are given in 7a and 7b in **Box 1** in the ***Appendix***.

The ~~fifth~~ sixth issue 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 estimation. As indicated in Ioannidis et al. (2017), adequate power in social-scientific research

[8] See also Ugur et al. (2016; 2018) for further discussion.

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 (usually, 0.05). 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*): $|\gamma| \,/\, SE_i \geq 2.80$ or $SE_i \leq |\gamma|/2.80$ (Ioannidis *et al.*, 2017: F239).

First, we use this condition to report the percentage of the effect-size estimates with adequate power. We do this for all evidence pools where the PET/FAT/PEESE tests indicate significant effect. To obtain the weighted average of the adequately-powered (WAAP) evidence, we use the average effect-size estimates form the HM specification of the PEESE model (model 4b above).[9] The WAAP evidence satisfies the 80% power rule – i.e., it consists of primary-study estimates with associated standard errors (*SE*) that satisfy the inequality that $SE_i \leq |\gamma|/2.80$.

The seventh issue relates to the choice of the moderating variables in the MVMRM. The challenge here is the absence of a theory on which moderating factors should be included in the MVMRM. To address this issue, we follow a model averaging routine that allows for identifying the relevant (robust) covariates. Tomas Havránek and his co-authors have ~~already~~ suggested a Bayesian model-averaging (BMA) method for this purpose (see, e.g., Havránek, 2015; Iršová and Havránek, 2013). In this study, we utilise a weighted-average least squares (WALS) routine that provides similar results to the BMA but at a much lower time cost (De Luca and Magnus, 2011; Magnus et al., 2010).

Both the WALS and BMA compute a weighted average of the conditional estimates taking into account all possible combinations or subsets of the explanatory variables and the regressions parameters (De Luca and Magnus, 2011). The BMA calculates posterior inclusion probabilities (*PIPs*) that indicate which variables are relevant for inclusion in a ‘true’ model. The common rule is to include the covariates with PIPs > 0.5. WALS does not report *PIPs*, but the *t-values* in WALS provide an indication about which variables would have *PIP* $\geq$ *0.5* in the BMA. As demonstrated in Masanjala and Papageorgiou (2008), a *t-ratio* of 1 in absolute value in the

[9] Because Ioannidist *et al.* (2017) use the average effect-size estimates from WLS specification of the PEESE model, we replicate their approach. The results, not reported here but can be provided on request, are similar to what report in Table 3 below.

WALS corresponds to a *PIP* of 0.5 in BMA. Therefore, variables with absolute *t-values* of 1 or over in the WALs routine are eligible for inclusion in the multivariate meta-regression.[10]

5. Meta-regression results

Table 1 presents the estimates of the average effect size by spillover types and own R&D; while **Table 2** presents the results by unit of analysis. In both Tables, **Panel A** presents the PET/FAT results while **Panel B** presents the PEESE results.[11] In Panel A of Table 1, the selection bias is substantial for spillover effects (columns A-1 to A-3) but moderate for own R&D effects (A-4). The bias is positive and consistent with the funnel plot asymmetry in Figures 1 and 2 above. These results indicate that researchers are predisposed to report empirical results that are statistically significant in the 'right' direction – i.e., in the direction of detecting positive spillover effects. Given this finding, the simple summary measures relied upon in narrative reviews would be biased and lead to incorrect conclusions.

[10] This approach is more tractable and statistically-coherent compared to the general-to-specific modelling routine where insignificant covariates are excluded one by one until all remaining covariates are significant.
[11] We do not use frequency weights in estimations in Tables 1 and 2. Estimations with frequency weights in **Tables A4** and **A5** in the ***Appendix***; and are largely similar to what is reported in Tables 1 and 2.

Table 1 – Effects size estimates by spillover types

Dependent variable: t-value	**Panel A - PET/FAT**					**Panel B - PEESE**		
	(A-1)	(A-2)	(A-3)	(A-4)	(A-5)	(B-1)	(B-4)	(B-5)
Effect (β in PET/FAT, γ in PEESE)	0.048***	0.074	0.007	0.064***	0.038***	0.069***	0.073***	0.036***
	(0.017)	(0.050)	(0.023)	(0.012)	(0.014)	(0.016)	(0.011)	(0.014)
Selection bias	2.065***	1.377	2.751***	0.808***	2.195***			
	(0.572)	(1.030)	(0.541)	(0.402)	(0.380)			
Standard error						-0.835	3.588	-2.736***
						(1.540)	(3.543)	(1.247)
Obs.	557	96	327	501	983	557	501	983
Studies	46	6	30	26	60	46	26	60
Log-likelihood (LL)	-1760.941	-306.995	-932.755	-1472.789	-3064.777	-1766.875	-1474.265	-3067.101
LL (comp. model)	-1853.435	-323.953	-1051.853	-1685.547	-3321.677	-1933.186	-1714.120	-3449.187
LR chi^2	184.987	33.915	238.196	425.516	513.8	332.623	479.709	764.173
P > LRc^2	0	0	0	0	0	0	0	0

***, **, * indicates significance at 1%, 5% and 10%. LR chi^2 is based on likelihood ratio test where the null hypothesis is that the WLS model is nested within the HM. Column (A-1 and B-1) is knowledge spillovers; (A-2) is mixed spillovers; (A-3) is rent spillovers; (A-4 and B-4) is own R&D; and (A-5 and B-5) is all spillovers types.

After controlling for selection bias, the average effect size in **Table 1** is positive and significant for knowledge spillovers (column A-1), own R&D (A-4) and all spillover types (A-5). In contrast, the average productivity effect of mixed spillovers (A-2) and rent spillovers (A-3) are insignificant. With PEESE correction (columns B-1, B-4 and B-5), the average effects for knowledge spillovers and own R&D remain similar at 0.69 and 0.073, respectively. The corrected estimate for all spillover types (B-5) is smaller (0.036). These results are in line with those obtained with frequency weights in **Table A4** in the ***Appendix***, where the average productivity effects of rent and mixed spillovers remain insignificant and the difference between the PEESE-based average effects for knowledge spillovers and own R&D is small.[12]

A similar pattern is evident in **Table 2**, where we report average effect-size estimates by the type of the cross-section units in the primary studies – i.e., by the beneficiary of the spillover effects. Here, spillovers have a positive and significant effect (0.058) only at the country level (columns A-1 and B-1). However, own R&D has positive and significant effects at the country and firm levels (A-4 and A-6). With PEESE correction, the own-R&D effects are 0.060 at the country and 0.057 at the firm level. The results obtained with frequency weighting (**Table A5** in the ***Appendix***) are mostly in line with those in **Table 2**. The only difference is that the average spillover effect at the country level (0.135) is larger than the rest. However, we need to treat spillovers at the country level with caution because Keller (1998) has demonstrated that the weights based on bilateral import shares, generally used in the construction of the country-level spillover pools, are open to criticism.[13]

[12] The difference between spillovers and own R&D effect is insignificant as the confidence intervals for both estimates overlap.
[13] Keller (1998) has demonstrated that the spillover effect on country-level productivity remains positive and significant when random weights are used instead of weights based on import shares.

Table 2 - Effects size estimates by unit of analysis

Dependent variable: t-value	**Panel A - PET/FAT**						**Panel B - PEESE**		
	(A-1)	(A-2)	(A-3)	(A-4)	(A-5)	(A-6)	(B-1)	(B-4)	(B-6)
Effect (β in PET/FAT, γ in PEESE)	0.058***	-0.019	0.040	0.056***	0.088	0.049***	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 (LL)	-1408.742	-678.214	-955.515	-705.504	-313.91	-265.875	-1408.060	-705.669	-268.303
LL (comp. model)	-1541.737	-756.472	-1008.283	-817.222	-333.312	-342.706	-1592.711	-824.433	-378.428
LR chi^2	265.989	156.515	105.535	223.437	38.805	153.662	369.302	237.530	220.250
P > LRc^2	0	0	0	0	0	0	0	0	0

***, **, * indicates significance at 1%, 5% and 10%. LR chi^2 is based on likelihood ratio test where the null hypothesis is that the linear model is nested within the multi-level model. (A-1 and B-1) is all spillovers – country-level; (A-2) is all spillovers – industry-level; (A-3) is all spillovers – firm-level; (A-4 and B-4) is own R&D – country-level; (A-5) is own R&D – industry-level; and (A-6 and B-6) is own R&D – firm-level.

The findings in **Tables 1** and **2** 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.[14] Also, our findings lend support to Griliches (1992: S36), who expresses doubt about whether rent or mixed spillovers capture true knowledge externalities or just measurement errors. The latter may be the case because "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 between the parties involved.

Results in **Tables 2** also indicate that the spillover effects are insignificant when the analysis is at the industry level. This is in line with the Schumpeterian models of innovation and growth, where the external R&D stock can be a source of positive productivity effect due to spillovers or a source of adverse effects due to creative destruction that reduces innovation rents. The latter arises if the existing knowledge and technology in an industry becomes obsolete at a faster speed when other industries increase their investment in R&D (Aghion et al., 2014; Aghion and Howitt, 1992).

In what follows, we probe another issue that has not been addressed so far 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.[15] The proportion of adequately-powered primary-study estimates and the weighted average of adequately powered (WAAP) estimates are in **Table 3**, obtained in accordance with the method explained in section 4.[16] We take the average effects in **Panel B** of **Tables 1** and **2** (the PEESE estimates from the HM estimator) as the 'true' population effect and identify the primary-study estimates with a statistical power of 80% or more on that basis.

[14] It must be reiterated that Hall et al. (2010) is the only review that does not conclude in favour of larger spillover effects compared to own-R&D effects.
[15] Ioannidis *et al.*, (2017) is the only exception, where statistical power is investigated *ex post* using evidence from 159 meta-analysis studies.
[16] Here we report the results only for the evidence pools in which PET/FAT and PEESE estimations yield statistically-significant average effects. These are the pools in **Panel B** of **Tables 1** and **2** above.

Table 3 – Weighted average effects from adequately-powered (WAAP) evidence: By spillover type and unit of analysis

	Panel A: Spillover type and own R&D			Panel B: Unit of analysis		
	(A-1)	(A-2)	(A-3)	(B-1)	(B-2)	(B-3)
WAAP Effect	0.009*** (0.002)	0.029*** (0.001)	0.012*** (0.002)	0.027*** (0.004)	0.060*** (0.003)	0.012*** (0.002)
Obs.	229	365	293	253	198	93
R-sq.	0.088	0.793	0.095	0.183	0.702	0.216
% of adequately powered†	41	73	30	55	67	74

***, **, * indicates significance at 1%, 5% and 10%. In panel A~~ and B~~, (A-1) is knowledge spillovers; (A-2) is own R&D; (A-3) is all spillover types. In Panel B, (B-1) is all spillovers – country; (B-2) is own R&D – country; and (B-3) is own R&D – firm. † = (number of adequately-powered estimates / all observations in sample)*100

As can be seen in the last row of **Table 3**, the percentage of adequately-powered evidence is 41% for knowledge spillovers and 30% for all spillover types. If the average effects in Tables 1 and 2 are unbiased estimates of the true population effects, our results indicate that 59% of the evidence base is low-powered in the knowledge spillover pool and 70% is low-powered in the all-spillover pool. The percentage of adequately-powered estimates is 55% when the spillover effect is at the country level. However, this is still lower than those related to own-R&D at the country level (67%) or firm level (74%).

These findings indicate that the level of selection in favour of low-powered estimates is more prevalent in spillover pools compared to own-R&D pools. The low-power findings we report for this research field are consistent with those reported by Ioannidis et al. (2017), who also detect low-power and exaggerated effect-size estimates in 159 research areas in economics.

Low power does not invalidate the estimates reported in primary studies, but it does suggest that average effect-size estimates based on low-powered evidence may be inflated. Indeed, this can be seen in the main results in **Table 3**, where the weighted averages of the adequately-powered (WAAP) effect sizes are much smaller than those based on effect sizes in the full sample (Panel B of Tables 1 and 2 above). Although statistically significant, the WAAP of the effect-size estimates for knowledge spillovers (A-1) and all spillover types (A-3) is very small (0.009 and 0.012, respectively). The WAAP result for spillovers at the country level (0.027 in column B-1) indicates a moderate effect, but this is still half of own-R&D effect (0.06 in column B-2).

Findings so far allow for two interim conclusions. First, they indicate that the claims that external knowledge externalities are substantial and larger than own-R&D effects are based on low-powered evidence that reflects a higher degree of selection bias. Secondly, both in the full sample and in evidence pools with adequate statistical power, the productivity effects of spillovers is usually smaller than that of own R&D. This finding is in line with Cohen and Levinthal (1989), who demonstrate that firms need to invest in R&D and develop absorptive capacity before they can reap the benefits of knowledge spillovers. The finding also suggests that the gap between the actual and optimal levels of R&D investment may be narrower than what is usually assumed in the public policy debate on the need for subsidizing R&D investment. Put differently, the additionality effects of the public support for privately-funded R&D investment may be smaller than what is implied by the public good characterisation of the R&D investment in Arrow (1962) and the following neoclassical literature.

Given prior considerations regarding heterogeneity, in what follows we investigate how the observable sources of heterogeneity affect the effect-size estimates in primary studies. First, we consider whether journal quality[17] or journal articles as opposed to working papers have systematic effects on reported estimates. Second, we consider differences in model specification mentioned above. Third, we examine if data and sample characteristics matter. Here, we verify if data types, type of the cross-section unit (firm, country or industry), level of R&D intensity and country or region of origin of the data influence the primary-study estimates. Fourth, we examine if different weights used to construct the measures of external knowledge matter and whether there are effect-size differences when R&D investment is used instead of R&D capital stocks as envisaged in the theoretical model. Lastly, we verify if different estimation methods are associated with different estimates.

We fit the MVMRM (model 8a or 8b in **Box 1** in the ***Appendix***) with covariates that capture these sources of heterogeneity. The covariates take the value of 1 if the primary-study estimate is associated with the controlled characteristic and zero otherwise.[18] All covariates are

[17] We adopt the Scimago Journal & Country Rank (SJR) ranking system (http://www.scimagojr.com/), which is arguably the most comprehensive journal ranking system which cuts across all fields, to capture journal quality. Specifically, we collect information on the H-index of each journal and categorize journals with an H-index above 100 as high quality journals.

[18] Descriptions of and summary statistics for the binary moderating variable are presented in **Table A3** in the ***Appendix***.

interacted with precision $\left(1/_{SE_i}\right)$ to capture their influence on the effect-size estimates rather than their effects on selection bias. Hence, a significant positive (negative) coefficient on a covariate indicates that primary-study estimates characterized by the included dummy variable are larger (smaller) than those associated with the excluded categories.

We have chosen the moderating variables from a large list generated from the coding procedure described in section 3. The choice is made through a model averaging routine (a weighted average least squares routine -WALS) described in the methodology (De Luca and Magnus, 2011; Magnus et al., 2010). The decision rule is to include the covariate in the MVMRM if its t-value in WALS estimation is greater than one in absolute value (De Luca and Magnus, 2011).

It must be noted here that the WALS routine (or the BMA routine) does not preclude the risk of multicollinearity among the covariates in a multiple regression. In classical linear regression, high levels of multicollinearity could lead to well-known statistical problems such as unstable coefficients, large standard errors, and sign reversals in the estimated parameters. However, simulation results have shown that HMs have better tolerance to multicollinearity, particularly when the latter is at Level 1 (i.e., between effect-size estimates nested within studies). It has been reported HMs yield unbiased coefficient estimates, particularly for the 'fixed-effects' coefficients – i.e., for the coefficients on the covariates as opposed to random-effect parameters (see, Shieh and Fouladi, 2003; Yu et al., 2015). Therefore, we are of the view that the coefficient estimates we discuss below are reliable even in the presence of multicollinearity.[19]

In **Table 4**, we report four estimation results based on: a random intercepts and slopes HM without frequency weights (column 1); the same specification with frequency weights (column 2); a random-intercept-only HM without frequency weights (column 3); and a random-intercept-only HM with frequency weights (column 4). Based on log-likelihood values, our preferred estimates are in columns 1 and 2. Because the log-likelihood values differ only marginally between the specifications, we also report results from two random-intercepts-only HM specifications in columns 3 and 4. As can be seen in the bottom part of Table 4, both LR tests and log-likelihood values favour the HM specifications against WLS.

[19] Apart from other advantages of the HMs discussed above, better tolerance to high multicollinearity is another feature that makes them preferable to WLS. True, meta-analysts using WLS do try to address multicollinearity through a general-to-specific modelling routine. The latter, however, is more likely to lead to model misspecification as the insignificant covariates dropped in the routine may well be relevant to the MVRM at hand.

We categorize conclusions from **Table 4** as supported by: (i) highly consistent evidence if the coefficient on the moderating variable reflects sign and significance consistency across all columns; (ii) moderately-consistent evidence if sign and significance consistency is observed in one of the preferred specifications and two of the secondary specifications; and (iii) weakly-consistent if the coefficient is significant in one of the preferred specifications OR in two of the secondary specifications.

We start with moderating factors that capture publication characteristics (top part of Table 4). Here, we find moderately-consistent evidence that journal articles report relatively smaller spillover effects on productivity compared to working papers. Also, there is weakly-consistent evidence that studies published in higher ranked journals (journals with an *h-index* of over 100) report relatively smaller spillover effects. These findings indicate that journals in general, and high-ranked journals in particular, do not seem to be affected by the "winner's curse" – i.e., by the tendency to exploit reputation and accommodate highly-selected evidence (Costa-Font et al., 2013). We have also found weakly-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 in the research field.

With respect to model specification, we find no systematic difference between studies that use total factor productivity (TFP) as the dependent variable instead of output or net sales. This finding indicates that coefficient estimates based on the output and TFP versions of the primal approach are consistent. In contrast, there is highly-consistent evidence that studies that control only for one or two spillover types instead of three tend to report relatively larger spillover effects. There is also moderately-consistent evidence that studies that control for own R&D in their models would report relatively smaller estimates than those that do not.

Table 4 - Multivariate meta-regression analysis: observed sources of heterogeneity

Dependent Variable: *t*-Value	(1)	(2)	(3)	(4)
Publication characteristics				
Precision	0.209**	0.089	0.247***	0.194*
	(0.100)	(0.123)	(0.046)	(0.113)
Journal article	-0.189**	-0.116	-0.239***	-0.236***
	(0.082)	(0.088)	(0.036)	(0.079)
Journal quality	-0.033	-0.077**	-0.014	-0.021
	(0.031)	(0.035)	(0.012)	(0.033)
Publication date after 2000	-0.043	-0.008	-0.103***	-0.120***
	(0.031)	(0.039)	(0.012)	(0.033)
Model specification				
TFP – Estimation based on total factor	-0.005	-0.055	0.007	-0.036
productivity	(0.025)	(0.040)	(0.012)	(0.031)
SPO coefficients in model <=2	0.039***	0.023**	0.043***	0.054**
	(0.010)	(0.012)	(0.008)	(0.024)
Control for own R&D in model	-0.058***	-0.025	-0.036***	-0.008
	(0.019)	(0.018)	(0.011)	(0.033)
Industry/country dummies in model	0.040	0.101***	0.026***	0.054**
	(0.031)	(0.024)	(0.010)	(0.023)
Year dummies in model	0.036	0.087**	0.031***	0.056***
	(0.029)	(0.038)	(0.007)	(0.020)
Data and sample characteristics				
Unit of analysis: country	0.073**	0.175***	0.032**	0.058
	(0.032)	(0.061)	(0.014)	(0.050)
Unit of analysis: industry	-0.204***	-0.138	-0.224***	-0.222***
	(0.061)	(0.086)	(0.026)	(0.075)
High R&D-intensity firm, industry	-0.031	-0.053*	-0.029**	-0.048
	(0.025)	(0.027)	(0.014)	(0.031)
North American (US&Canada) data	0.127**	0.119*	0.153***	0.169**
	(0.053)	(0.066)	(0.024)	(0.071)
OECD data	0.048	-0.006	0.094***	0.121**
	(0.037)	(0.053)	(0.014)	(0.052)
Data mid-point < 1991	-0.066**	-0.085*	-0.048***	-0.059
	(0.033)	(0.049)	(0.012)	(0.037)
Spillover characteristics				
Based on asymmetric weights	0.024***	0.044	0.021***	0.041
	(0.006)	(0.036)	(0.006)	(0.026)
Unweighted	0.003	0.003	0.005*	0.006
	(0.003)	(0.005)	(0.003)	(0.006)
Based on R&D investment as	0.075	0.007	0.193***	-0.010
opposed to R&D capital	(0.064)	(0.052)	(0.029)	(0.115)
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	-0.008	-0.009***	-0.010*	-0.014*
cointegration	(0.005)	(0.002)	(0.005)	(0.008)
Instrumental variable (IV)	-0.011	0.007	-0.003	0.030

estimation				
	(0.007)	(0.020)	(0.007)	(0.035)
Constant	2.426***	2.577***	2.416***	3.275***
	(0.398)	(0.372)	(0.682)	(1.208)
Observations	983	983	983	983
Studies	60	60	60	60
Log-likelihood (HM)	-3013.228	-174.221	-3050.514	-185.478
LR Test chi2	101.770	1284.973	405.049	303.962
P> chi2	0.000	0.000	0.000	0.000
converged	Yes	Yes	Yes	Yes
Log-likelihood (least-sqaures)	-3128.805	Not available†	-3128.805	Not available†

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

The theory is silent on whether different types of spillovers should enter the empirical model as complements (in which case all spillover types should be in the model) or as substitutes (in which case they should be summed up). Hence primary studies are justified in making different choices. However, those that overlook complementarity when the latter is true run the risk of omitted variable bias. Therefore we call for sensitivity checks reflecting the two assumptions of complementarity and substitutions between different spillover types.

In the case of own R&D, however, the knowledge capital model is explicit: own-R&D capital should be treated as complement to external R&D capital and both should be included in empirical models (Griliches, 1979; 1992). A number of studies in this research field do not follow this recommendation (e.g., Jaffe, 1988; Ke and Luger, 1996). Our finding suggests that larger effect-size estimates from such studies may be due to omitted variable bias.

The final moderating factor in the 'model specification' category concerns controlling for industry/country or year dummies. We find that the spillover effects are larger when primary-study authors control for industry/country or year dummies. As indicated in Hall et al. (2010), the case for including industry/country or year dummies in panel data models is not clear-cut. On the one hand, such dummies can allow for taking account of erroneous omission of industry/country or year characteristics. On the other hand, however, the dummies may be a source of bias if productivity effects differ because of different technological opportunities in different industries or during different phases of the business cycle.

Concerning moderating factors that reflect data and sample characteristics, we find moderately-consistent evidence that data at the country level is associated with larger spillover effects compared to data at the industry or firm levels. As indicated above, the weights used for constructing country-level spillover pools are usually based on bilateral import shares, about which Keller (1998) has raised a serious concern. Using randomly-created trade patterns (hence random bilateral import shares), he demonstrates that the latter also give rise to international R&D spillover estimates that are even larger in magnitude. Therefore, and in line with Keller (1998), we call for caution against claims that bilateral trade patterns are highly important in driving international R&D spillovers.[20]

In contrast, the productivity effects tend to be smaller when the primary studies utilise industry-level data. Smaller or insignificant spillover effects at the industry level may be due to the adverse effect of creative destruction (Aghion et al., 2014; Schumpeter, 1942), which is more evident at the industry level.

We find moderately-consistent evidence that the spillover effects are relatively smaller when the data relate to firms with high R&D intensity. This finding is also in line with insights from Schumpterian models of innovation, where the productivity of firms/industries closer to the technology frontier is driven by own investment in R&D rather than emulation (Aghion et al., 2014). However, at the country-level, we find that data related to North American or OECD countries is associated with larger spillover effects. Given that these countries are known to have higher R&D intensities than the rest, this finding indicates that countries must invest in R&D to reap the benefits of external R&D (knowledge) stock. Combining both sets of findings, we conclude that industries and/or countries need to keep investing in own R&D either to reduce the creative destruction effects of the R&D investments by others or to increase the scope for benefiting from the spillover effects of the external R&D stocks they face.

Turning to spillover characteristics, there is moderately-consistent evidence that studies that adopt a scaled spillover weight tend to report relatively larger estimates compared to those that use symmetric weight. This finding lends support to Coe and Helpman (1995), who first makes the case for weighted spillovers scaled by openness to import (import/GDP ratio). However, it must be noted that the relatively larger productivity effect of the scaled spillover measure may

[20] This is particularly the case when the weights are based on bilateral import shares only. Unlike Keller (1998), however, we think that inter-country spillovers through bilateral imports would still remain a vulnerable construct even when they are scaled with openness to trade as an additional weight.

be a measurement construct: the scaled spillover measure is just a fraction of the weighted sum of the external knowledge stock, where the fraction is equal to the imports/GDP ratio. Other findings in Table 4 indicate that there is no systematic difference; (i) between weighted and unweighted spillovers and (ii) between spillover pools constructed with R&D capital srock and those based on current R&D investment only. Given these findings, we argue that there may be several candidates for measuring R&D (knowledge) externalities, which are essentially unobserved in the data. However, some measures (e.g., those based on technology proximity) may be better in terms of their theoretical underpinnings (Griliches, 1992).

With regards to estimation methods, we find highly-consistent evidence that effect-size estimates obtained from estimators that take account of panel cointegration and those based on time-differenced data are smaller than their respective reference categories. Both findings are in line with what econometric theory suggests. 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).

6. Conclusions

Griliches (1992) reports that the elasticity of output with respect to external R&D capital is between "about a half and double of the elasticity of output with respect to private R&D". Given that the own-R&D elasticity from the literature he reviews is 0.1, he suggests that the range for elasticity estimates of external R&D is between 0.05 and 0.20. Based on this scenario, Griliches concludes that about half of the productivity growth can be explained by returns on R&D; and "most of the explanatory effect" is due to the "spillover component." As a distinguished pioneer of this research field, however, Griliches is aware of the risk of "upward selectivity bias in the results" he relies upon. Therefore, he calls for further work to discover the actual magnitude of the spillover effects.

In this study, we have responded to Griliches' call using rigorous meta-analysis methods and a rich dataset. We find that, after controlling for selection bias, the average productivity effect of *rent* or *mixed* spillovers isinsignificant. On the other hand, the effect of *knowledge* spillovers is small (0.069) and similar to the productivity effect of own-RD (0.073). To put in context, the magnitude of the spillovers' productivity effect is nearer the minimum of the interval in Griliches (1992). When the average effect is based on adequately-powered (WAAP) estimates, however, the productivity effect of knowledge spillovers is much smaller that own R&D: the ratio is 1:3 in the case spillovers at all (firm, industry and country) levels and 1:2 when the evidence is at the country level only.

Our findings do not necessarily indicate that R&D spillovers do not exist. However, they do suggest that the existing evidence suffers from low power and own-R&D effects are perhaps stronger determinants of productivity compared to external R&D. Hence the case for direct or indirect public support aimed at inducing firms to undertake the optimal level of business R&D investment should be qualified along two dimensions. On the one hand, policy-makers should acknowledge that the 'true' productivity effects of spillovers are likely to be smaller than what is suggested by summary measures or vote counting. On the other hand, they should take account of excess heterogeneity in the evidence base and of the concomitant need for targeting firms/industries where the spillover effects are higher.

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. Both neoclassical and Schumpeterian theories of innovation suggest that firms/industries/countries closer to (further away from) the technology frontier are less (more) likely to benefit from external knowledge. This is confirmed to some extent in our findings indicating that the productivity effect of spillovers is relatively smaller when the firm or industry is characterised by high R&D intensity. Therefore, we call for scaling the stock of the external knowledge by an additional weight that reflects the distance to technology frontier. The procedure for this is simple and already applied to bilateral import shares, which are scaled by openness to trade (see equation 1c above).

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. Yet the perpetual inventory method used for obtaining the R&D stock presumes that the latest addition to the stock becomes productive immediately. However Griliches (1992) and Hall *et al.* (2010) question this assumption because of the time-lags between R&D expenditure and innovation, between innovation and commercialization, and in the case of spillovers, between innovation and diffusion. Therefore, we recommend either inclusion of distributed lags for own and external

R&D capital in a given empirical model or experimentations with different lags as sensitivity checks.

Our third recommendation relates to explicit modelling of heterogeneity in the productivity effects of spillovers. This is in line with the “new growth” literature where the homogenous technology (i.e., homogeneous slope coefficients) assumption of the Cobb-Douglas production function is questioned (Azariadis and Drazen, 1990; Banerjee and Newman, 1993; Eberhardt and Teal, 2013). In studies based on firm-level data with shorter time spans, it is difficult to rely on panel time-series estimators such as mean group or common correlated effect mean group estimators that allow for heterogeneous technology and heterogeneous unobservable effects. However even with firm-level data, it is possible to control for observed sources of heterogeneity through interaction dummies, which provide information on how the spillovers' productivity effect differs by firm size, R&D intensity, competition in own industry, or whether the firm is incumbent or a new entrant. Evidence from such models can inform differentiated public policy designs that target public support to firms associated with characteristics conducive to higher levels of externalities.

Another way in which heterogeneity can be modelled more explicitly in firm-level studies is to adopt a hierarchical modelling (HM) approach (Aiello and Ricotta, 2016). HMs allows for nesting the firms within industries and/or regions and estimating the spillover effects after controlling for between-industry or between-region variations modelled as random intercepts, random slopes or random intercepts and slopes. There are remedies for endogeneity that may be due to correlations between the firm-level covariates and industry- or region-specific random-effect components (Hanchane and Mostafa, 2012).

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 allow for heterogeneous slope coefficients and can take account of cross-sectional dependence, which may be due to common unobservable factors but is assumed away in the standard panel-data models (Eberhardt *et al.,* 2013). For these reasons, we call for explicit modelling of heterogeneity with respect to technology as well as unobserved common factors.

Our findings also offer some insights into business decision making about R&D investment and the latter's implications for public policy design. On the one hand, the gap between the actual and optimal levels of R&D investment can be expected to be small even in the absence of public support. This is due to the need for own R&D investment as a basis for benefiting from the external knowledge stock. Secondly, public support for R&D would induce heterogeneous firm efforts towards closing the gap between actual and optimal R&D investment. Our findings suggest that the additionality effect of the public support would be small if the firm is R&D-intensive and the level of creative destruction in its industry is high.

References - general

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.

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.

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.

Arrow, K.J., 1962. The economic implications of learning by doing. The Review of Economic Studies 29 (3), 155-173.

Azariadis, C., Drazen, A., 1990. Threshold externalities in economic development. The Quarterly Journal of Economics 105 (2), 501-526.

Banerjee, A.V., Newman, A.F., 1993. Occupational choice and the process of development. Journal of Political Economy 101 (2), 274-298.

Bernstein, J.I., Nadiri, M.I., 1988. Interindustry R&D spillovers, rates of return, and production in high-tech industries. The American Economic Review 78 (2), 429-434.

Bernstein, J.I., Yan, X., 1997. International R&D spillovers between Canadian and Japanese industries. The Canadian Journal of Economics / Revue canadienne d'Economique 30 (2), 276-294.

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.

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.

Chipman, J.S., 1970. External economies of scale and competitive equilibrium. The Quarterly Journal of Economics 84 (3), 347-385.

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.

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.
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.
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.
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.
Grossman, G.M., Helpman, E., 1991. Innovation and Growth in the Global Economy MIT Press, Cambridge, MA.
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.
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.
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, 1-15.

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.
Los, B., Verspagen, B., 2000. R&D spillovers and productivity: evidence from US manufacturing microdata. Empirical Economics 25 (1), 127-148.
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.
Masanjala, W.H., Papageorgiou, C., 2008. Rough and lonely road to prosperity: a reexamination of the sources of growth in Africa using Bayesian model averaging. Journal of Applied Econometrics 23 (5), 671-682.
McVicar, D., 2002. Spillovers and foreign direct investment in UK manufacturing. Applied Economics Letters 9 (5), 297-300.
Meade, J.E., 1952. External economies and diseconomies in a competitive situation. The Economic Journal 62 (245), 54-67.
Mohnen, P., 1996. R&D externalities and productivity growth. OECD STI Review 17, 39-59.
Müller, W.G., Nettekoven, M., 1999. A panel data analysis: research and development spillover. Economics Letters 64 (1), 37-41.
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).
Sheshinski, E., 1967. Optimal accumulation with learning by doing, in: Shell, K. (Ed.), Essays on the theory of optimal economic growth. MIT Press, Cambridge, MA, pp. 31-52.
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.
Simon, H., 1947. Some models for the study of the economic effects of technological change. Cowles Commission Discussion Paper DPP213.
Snijders, T., Bosker, R., 2012. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Analysis. Sage, London.
Stanley, T., Doucouliagos, H., 2015. Neither fixed nor random: weighted least squares meta-analysis. Statistics in Medicine 34 (13), 2116-2127.
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., Doucouliagos, H., Giles, M., Heckemeyer, J.H., Johnston, R.J., Laroche, P., Nelson, J.P., Paldam, M., Poot, J., Pugh, G., 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.
Turner, R.M., Davey, J., Clarke, M.J., Thompson, S.G., Higgins, J.P., 2012. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. International Journal of Epidemiology 41 (3), 818-827.
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.
Verspagen, B., 1997. Estimating international technology spillovers using technology flow matrices. Review of World Economics 133 (2), 226-248.
Wolff, E.N., Nadiri, M.I., 1993. Spillover effects, linkage structure, and research and development. Structural Change and Economic Dynamics 4 (2), 315-331.
Yu, H., Jiang, S., Land, K.C., 2015. Multicollinearity in hierarchical linear models. Social Science Research 53, 118-136.

References – Included primary studies

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.

Appendix

Table A1 - Spillovers and Productivity: Overview of the Evidence Base

Study	Pub type	Data type	Estimation method	Unit of analysis	Unit count	Country	Data start	Data end	Mean effect size	Median effect size	median t-value	No. of est.	Spillover type	Spillover weight
Acharya and Keller (2009)	Jnl	Panel	OLS	Country/Industry	17	Mixed	1973	2002	0.151	0.129	5.083	102	Knowledge	Unweighted
Adam and Jaffe (1996)	Jnl	Panel	Non Linear LS	Firm	80	US	1974	1988	0.175	0.215	9.150	6	Knowledge	Tech proximity
Aiello and Cardarmone (2005)	Jnl	Panel	GLS/GMM	Firm	1017	Italy	1995	2000	0.011	0.012	3.995	4	Knowledge	Tech proximity
Aldieri and Cincera (2009)	Jnl	Panel	GMM	Firm	808	US	1988	1997	0.500	0.500	17.777	4	Knowledge	Tech proximity/Distance
Belitz and Molders (2016)	Jnl	Panel	Cointegration/FE	Country	77	Mixed	1990	2008	0.034	0.036	3.500	19	Knowledge/Rent	FDI/Import shares shares
Biatour et al (2011)	WP	Panel	Dynamic	Industry	21	Belgium	1987	2007	0.089	0.090	1.630	27	Mixed/Rent	Patent/Input-Output Flows
Bitzer and Geishecker (2006)	Jnl	Panel	GLS	Industry_Country	170	Mixed	1973	2000	0.009	0.017	4.375	8	Rent	Import shares shares
Bitzer and Kerekes (2008)	Jnl	Panel	GLS	Industry_country	10	Mixed	1973	2000	0.019	0.015	7.103	18	Rent/Knowledge	Import shares shares/FDI
Bloch (2013)	Jnl	Panel	FE	Firm	n.a.	Denmark	1997	2005	0.052	0.062	2.952	5	Knowledge	Tech proximity
Bloom et al. (2013)	Jnl	Panel	OLS/2SLS	Firm	n.a.	US	1981	2001	0.165	0.191	4.125	5	Knowledge	Tech proximity
Braconier and Sjoholm (1998)	Jnl	Panel	OLS	Industry	49	Mixed	1979	1991	-0.015	-0.014	1.260	7	Knowledge	Unweighted
Braconier et al. (2001)	Jnl	Panel	OLS/FE/RE	Firm	66	Sweden	1978	1994	0.009	0.017	0.850	19	Knowledge	Tech proximity/FDI
Branstetter (2001)	Jnl	Panel	Diff	Firm	209	Japan/US	1985	1989	0.371	0.449	1.412	4	Mixed	Patent
Bronzini and Piselli (2009)	Jnl	Panel	Cointegration	Region	19	Italy	1985	2001	0.405	0.405	13.824	2	Knowledge	Distance
Cincera (2005)	Jnl	Panel	Diff/GMM/FE	Firm	625	Mixed	1988	1994	0.573	0.590	3.582	9	Knowledge	Tech proximity
Coe et al (1997)	Jnl	Panel	Diff	Country	77	Mixed	1971	1990	-0.010	0.061	2.813	15	Rent/Knowledge	Import shares /Unweighted
Coe et al (2009)	Jnl	Panel	FE	Country	24	Mixed	1971	2004	0.079	0.049	3.300	24	Knowledge/Rent	Unweighted/Import shares
del Barrio-Castro et al (2002)	Jnl	Panel	Cointegration	Country	21	Mixed	1966	1995	0.117	0.117	2.485	2	Rent	Import shares shares
Edmond (2001)	Jnl	Panel	FE	Country	21	Mixed	1971	1990	0.222	0.200	8.429	7	Rent/Knowledge	Import shares/Unweighted
Engelbrecht (1997)	Jnl	Panel	GLS/OLS	Country	21	Mixed	1971	1985	0.280	0.249	4.304	11	Rent	Import shares
Frantzen (2000)	Jnl	Panel	Cointegration/OLS	Country	21	Mixed	1991	1980	0.271	0.219	3.914	10	Rent	Import shares
Frantzen (2002)	Jnl	Panel	Cointegration/OLS	Industry_Country	308	Mixed	1972	1994	0.169	0.164	9.765	42	Rent	Import shares
Funk (2001)	Jnl	Panel	Cointegration/OLS	Country	22	Mixed	1971	1990	0.112	0.059	2.530	15	Rent	Import shares
Griffith et al (2006)	Jnl	Panel	GMM/OLS	Firm	188	UK	1990	2000	0.174	0.154	1.582	46	Mixed/Knowledge	Patent/Unweighted
Guellec and Van Pottelsberghe (2001)	Jnl	Panel	3SLS/GLS	Country	16	Mixed	1980	1998	0.094	0.092	7.885	6	Knowledge	Tech proximity
Guellec and Van Pottelsberghe (2004)	Jnl	Panel	3SLS	Country	16	Mixed	1980	1998	0.398	0.398	39.015	2	Knowledge	Tech proximity
Gutierrez and Gutierrez (2003)	Jnl	Panel	Coint./FM/DOLS	Country	47	Mixed	1970	1992	0.517	0.531	3.068	4	Rent	Import shares
Harhoff (2000)	Jnl	Survey	OLS	Firm	439	Germany	1977	1989	-0.016	-0.013	-0.412	4	Knowledge	Tech proximity
Hejazi and Safarian (1999)	Jnl	Panel	OLS	Country	20	Mixed	1971	1990	0.081	0.089	3.615	8	Rent/Knowledge	Import shares/FDI
Higon (2007)	Jnl	Panel	Coint/MG/PMG	Industry	8	UK	1970	1997	0.598	0.215	0.578	8	Rent	Import shares

Jacobs et al (2002)	Jnl	Panel	FE	Industry	11	Netherlands	1973	1992	0.962	0.926	5.144	5	Rent	Import shares/Input-Output Flows
Jaffe (1988)	Jnl	CrS	OLS	Firm	391	US	1972	1977	0.082	0.094	2.102	4	Knowledge	Tech proximity
Jaffe (1989)	Jnl	Panel	OLS	Firm	432	US	1973	1979	0.128	0.128	3.368	1	Knowledge	Tech proximity
Johnson and Evenson (1999)	Jnl	Panel	OLS	Country	6	Mixed	1973	1987	0.143	0.143	2.875	2	Knowledge	Unweighted
Kao et al. (1999)	Jnl	Panel	Coint/Dyn/OLS	Country	22	Mixed	1971	1990	0.167	0.114	3.269	16	Rent	Import shares
Ke and Luger (1996)	Jnl	CrS	OLS	Firm	210	US	1991	1991	0.094	0.093	1.529	6	Knowledge	Unweighted
Keller (1998)	Jnl	Panel	OLS	Country	22	Mixed	1971	1990	0.186	0.156	11.167	9	Rent	Import shares
Krammer (2010)	Jnl	Panel	Cointegration/FE	Country	47	Mixed	1990	2006	0.065	0.029	4.322	42	Rent/Knowledge	Import shares/FDI
Kwon (2004)	WP	Panel	GLS	Industry	34	Japan	1970	1998	-0.440	-0.378	-0.595	8	Knowledge/Rent	Tech proximity/Input-Output Flows
Lee (2005)	Jnl	Panel	Coint/OLS/Dyn/FE	Country	17	Mixed	1971	2000	0.031	0.032	2.569	40	Mixed/Knowledge/Rent	Patent/Tech Proximity/Import shares
Lee (2006)	Jnl	Panel	Cointegration/OLS	Country	16	Mixed	1981	2000	0.049	0.031	2.302	27	Rent/Knowledge	Import shares/FDI
Lehto (2007)	Jnl	Panel	2SLS/OLS/GLS	Firm	2171	Finland	1987	1998	-0.013	0.014	3.400	39	Knowledge	Unweighted
Lichtenberg and Van Pottelsberghe (1998)	Jnl	Panel	FE	Country	22	Mixed	1971	1990	-0.514	0.058	2.926	14	Rent	Import shares
Lopez-Pueyo et al (2008)	Jnl	Panel	Cointegration/OLS	Industry_country	10	Mixed	1979	2000	0.125	0.123	3.962	26	Knowledge	Tech proximity
Los and Verspagen (2000)	Jnl	Panel	Unspecified/FE/RE	Firm	680	US	1977	1991	0.389	0.393	8.485	48	Knowledge/Mixed	Tech proximity/Unweighted/Patent
Lumenga-Neso et al (2005)	Jnl	Panel	OLS	Country	22	Mixed	1971	1990	0.156	0.200	3.774	9	Rent/Knowledge	Import shares/Unweighted
Lychagin et al (2016)	Jnl	Panel	OLS/GMM	Firm	1383	US	1970	2000	0.462	0.654	3.250	29	Knowledge	Tech proximity/Distance
Mcvicar (2002)	Jnl	Panel	Non Linear LS	Industry	7	UK	1973	1992	0.464	-0.288	-1.889	4	Rent/Knowledge	Import shares/FDI/Unweighted
Negassi (2009)	Jnl	Panel	3SLS	Firm	2763	France	1990	1996	0.093	0.094	1.806	6	Knowledge/Rent	Tech proximity/Import shares
Orlando (2004)	Jnl	Panel/CrS	Diff/FE/OLS	Firm	515	US	1972	1995	0.006	0.003	0.875	24	Knowledge	Unweighted/Distance
Ornaghi (2006)	Jnl	Panel	GMM	Firm	3151	Spain	1991	1999	0.033	0.019	2.035	12	Knowledge	Size proximity
Parameswaran (2009)	Jnl	Panel	Unspecified	Firm	2100	India	1992	2001	0.057	0.044	3.192	3	Knowledge	Tech proximity
Park (1995)	Jnl	Panel	RE/OLS/FE	Country	10	Mixed	1970	1987	0.167	0.172	2.679	16	Knowledge	Tech proximity
Park (2004)	Jnl	Panel	OLS	Country	22	Mixed	1971	1990	0.079	0.072	5.224	14	Rent/Knowledge	Import shares/Student Flows
Raut (1995)	Jnl	Panel	RE/FE/OLS/3SLS	Firm	192	India	1975	1986	0.120	0.095	3.340	19	Knowledge	Unweighted
Van Pottelsberghe and Lichtenberg (2001)	Jnl	Panel	Diff/FE	Country	13	Mixed	1971	1990	0.065	0.053	5.308	19	Rent/Knowledge	Import shares/FDI
Verspagen (1997)	Jnl	Panel	FE/RE	Industry_country	22	Mixed	1974	1992	0.073	0.061	2.905	48	Knowledge/Mixed	Tech proximity/Patent
Wang and Chao 2008	Jnl	Panel	OLS	Firm	72	Taiwan	1994	2000	0.195	0.195	1.710	2	Knowledge	Unweighted
Xu and Wang (1999)	Jnl	Panel	OLS	Country	21	Mixed	1983	1990	0.063	0.062	2.241	30	Knowledge/Rent	Distance/Import shares/Export
Zhu and Jeon (2007)	Jnl	Panel	Cointegration/OLS	Country	22	Mixed	1981	1998	0.072	0.064	5.525	18	Knowledge/Rent	FDI/Import shares/Total Trade
TOTAL									**0.126**	**0.070**	**3.323**	**983**		

Notes Jnl is Journal article, WP is working paper, CrS is Cross Section

Table A2 - Own R&D and Productivity: Overview of the Evidence Base

Study	Pub type	Data type	Estimation method	Unit of analysis	Unit count	Country	Data start	Data end	Mean effect size	Median effect size	median t-value	No. of est.
Acharya and Keller (2009)	Jnl	Panel	OLS	Country	17	Mixed	1973	2002	0.112	0.108	7.286	45
Adam and Jaffe (1996)	Jnl	Panel	Non Linear LS	Firm	80	US	1974	1988	0.043	0.040	8.750	6
Aiello and Cardarmone (2005)	Jnl	Panel	GLS/GMM	Firm	1017	Italy	1995	2000	0.070	0.068	5.410	4
Aldieri and Cincera (2009)	Jnl	Panel	GMM	Firm	808	US	1988	1997	0.230	0.230	14.706	3
Belitz and Molders (2016)	Jnl	Panel	Cointegration/FE	Country	77	Mixed	1990	2008	0.008	0.009	1.550	14
Biatour et al. (2011)	WP	Panel	Dynamic	Industry	21	Belgium	1987	2007	0.076	-0.030	-0.460	11
Bitzer and Geishecker (2006)	Jnl	Panel	GLS	Industry_Country	170	Mixed	1973	2000	0.043	0.043	3.720	5
Bloch (2013)	Jnl	Panel	FE	Firm	n.a.	Denmark	1997	2005	0.212	0.206	6.629	4
Bloom et al. (2013)	Jnl	Panel	OLS/2SLS	Firm	n.a.	US	1981	2000	0.046	0.043	6.143	5
Braconier and Sjoholm (1998)	Jnl	Panel	OLS	Industry	49	Mixed	1979	1991	-0.071	-0.062	1.820	6
Braconier et al. (2001)	Jnl	Panel	RE/FE/OLS	Firm	66	Sweden	1978	1994	0.038	0.043	4.150	8
Branstetter (2001)	Jnl	Panel	Diff	Firm	209	Japan/US	1985	1989	0.187	0.187	1.607	2
Bronzini and Piselli (2009)	Jnl	Panel	Cointegration	Region	19	Italy	1985	2001	0.029	0.029	2850	2
Cincera (2005)	Jnl	Panel	FE/diff/GMM	Firm	625	Mixed	1988	1994	0.250	0.245	11.074	6
Coe et al (2009)	Jnl	Panel	FE	Country	24	Mixed	1971	2004	0.098	0.096	8.685	22
del Barrio-Castro et al (2002)	Jnl	Panel	Cointergration	Country	21	Mixed	1966	1995	0.043	0.043	1.298	2
Edmond (2001)	Jnl	Panel	FE	Country	21	Mixed	1971	1990	0.060	0.064	7.789	8
Engelbrecht (1997)	Jnl	Panel	GLS/OLS	Country	21	Mixed	1971	1985	0.090	0.090	6.700	11
Frantzen (2000)	Jnl	Panel	Cointegration/OLS	Country	21	Mixed	1991	1980	0.079	0.091	3.176	10
Funk (2001)	Jnl	Panel	Cointegration/OLS	Country	22	Mixed	1971	1990	0.077	0.075	5.080	12
Griffith et al (2006)	Jnl	Panel	GMM/OLS	Firm	188	UK	1990	2000	0.022	0.024	2.039	12
Guellec and Van Pottelsberghe (2001)	Jnl	Panel	GLS/3SLS	Country	16	Mixed	1980	1998	0.024	0.024	5.400	6
Guellec and Van Pottelsberghe (2004)	Jnl	Panel	3SLS	Country	16	Mixed	1980	1998	0.116	0.116	61.190	2
Gutierrez and Gutierrez (2003)	Jnl	Panel	Coint./FM/DOLS	Country	47	Mixed	1970	1992	0.236	0.062	3.046	4
Harhoff (2000)	Jnl	Survey	OLS	Firm	439	Germany	1977	1989	0.068	0.068	2.429	4
Hejazi and Safarian (1999)	Jnl	Panel	OLS	Country	20	Mixed	1971	1990	0.097	0.096	8.665	6
Higon (2007)	Jnl	Panel	Coint/MG/PMG	Industry	8	UK	1970	1997	0.309	0.313	2.617	4
Jacobs et al. (2002)	Jnl	Panel	FE	Industry	11	Netherland	1973	1992	0.315	0.336	8.205	4
Jaffe (1989)	Jnl	Panel	OLS	Firm	432	US	1973	1979	0.031	0.031	2.583	1
Johnson and Evenson (1999)	Jnl	Panel	OLS	Country	6	Mixed	1973	1987	0.052	0.052	4.410	1
Kao et al. (1999)	Jnl	Panel	Coint/Dyn/OLS	Country	22	Mixed	1971	1990	0.089	0.091	4.688	16
Keller (1998)	Jnl	Panel	OLS	Country	22	Mixed	1971	1990	0.054	0.047	10.778	9

Krammer (2010)	Jnl	Panel	Cointegration/FE	Country	47	Mixed	1990	2006	0.071	0.063	4.286	21
Lee (2005)	Jnl	Panel	Coint/Dyn/OLS/FE	Country	17	Mixed	1971	2000	0.040	0.026	3.922	20
Lee (2006)	Jnl	Panel	Cointegration/OLS	Country	16	Mixed	1981	2000	0.039	0.033	0.848	10
Lehto (2007)	Jnl	Panel	2SLS/OLS/GLS	Firm	2171	Finland	1987	1998	0.028	0.031	5.200	15
Lichtenberg and Van Pottelsberghe (1998)	Jnl	Panel	FE	Country	22	Mixed	1971	1990	0.077	0.082	9.556	10
Lopez-Pueyo et al. (2008)	Jnl	Panel	Cointegration/OLS	Industry_Country	10	Mixed	1979	2000	0.335	0.158	10.489	12
Lumenga-Neso et al. (2005)	Jnl	Panel	OLS	Country	22	Mixed	1971	1990	0.042	0.023	0.852	9
Lychagin et al. (2016)	Jnl	Panel	OLS/GMM	Firm	1383	US	1970	2000	0.019	0.006	0.720	15
Mcvicar (2002)	Jnl	Panel	Non Linear LS	Industry	7	UK	1973	1992	0.032	0.032	2.286	1
Negassi (2009)	Jnl	Panel	3SLS	Firm	2763	France	1990	1996	0.157	0.157	1.809	2
Orlando (2004)	Jnl	Panel	Diff/FE/OLS	Firm	515	US	1972	1995	-0.005	0.039	3.194	6
Ornaghi (2006)	Jnl	Panel	GMM	Firm	3151	Spain	1991	1999	0.093	0.098	4.261	9
Parameswaran (2009)	Jnl	Panel	Unspecified	Firm	2100	India	1992	2001	0.001	0.002	2.000	3
Park (1995)	Jnl	Panel	RE/OLS/FE	Country	10	Mixed	1970	1987	0.096	0.091	2.014	16
Park (2004)	Jnl	Panel	OLS	Country	22	Mixed	1971	1990	0.049	0.057	5.750	8
Raut (1995)	Jnl	Panel	RE/FE/OLS/3SLS	Firm	192	India	1975	1986	0.008	0.008	1.490	19
Van Pottelsberghe and Lichtenberg (2001)	Jnl	Panel	Diff/FE	Country	13	Mixed	1971	1990	0.052	0.048	3.575	14
Verspagen (1997)	Jnl	Panel	FE/RE	Industry_Country	22	Mixed	1974	1992	0.076	0.076	3.665	24
Wang and Chao (2008)	Jnl	Panel	OLS	Firm	72	Taiwan	1994	2000	0.118	0.118	5.345	2
Xu and Wang (1999)	Jnl	Panel	OLS	Country	21	Mixed	1983	1990	0.051	0.029	1.493	20
Zhu and Jeon (2007)	Jnl	Panel	Cointegration/OLS	Country	22	Mixed	1981	1998	0.060	0.061	7.720	12
TOTAL									**0.078**	**0.061**	**4.050**	**503**

Notes Jnl is Journal article, WP is working paper, CrS is Cross Section

Box 1: Hierarchical specifications for bivariate and multi-variate meta-regression models

The underlying model for the average effect size is that of Egger et al. (1997):

$$effect\_size_i = \beta + \alpha SE_i + \xi_i$$

Using the inverse of the squared standard error as weights to address heteroskedasticity, the precision-effect and funnel-asymmetry tests (PET/FAT) are based on:

$$t_i = \alpha + \beta \left({}^{1}\!/_{SE_i} \right) + \omega_i$$

Bivariate HMs for PET/FAT can be stated as follows:

$$t_{ij} = \alpha^{RI2} + \beta^{RI2} \left({}^{1}\!/_{SE_{ij}} \right) + h_{0j}^{RI2} + u_{ij}^{RI2} \qquad (5a)$$
$$t_{ij} = \alpha^{RIS2} + \beta^{RIS2} \left({}^{1}\!/_{SE_{ij}} \right) + h_{0j}^{RIS2} + h_{1j}^{RIS2} \left({}^{1}\!/_{SE_{ij}} \right) + u_{ij}^{RIS2} \qquad (5b)$$
$$t_{ijk} = \alpha^{RI3} + \beta^{RI3} \left({}^{1}\!/_{SE_{ijk}} \right) + h_{0j}^{RI3} + h_{0k}^{RI3} + u_{ijk}^{RI3} \qquad (5c)$$
$$t_{ijk} = \alpha^{RIS3} + \beta^{RIS3} \left({}^{1}\!/_{SE_{ijk}} \right) + h_{0j}^{RIS3} + h_{0k}^{RIS3} + h_{1j}^{RIS3} \left({}^{1}\!/_{SE_{ij}} \right) + u_{ijk}^{RIS3} \qquad (5d)$$

Here, subscript *i* denotes effect-size estimate, *j* denotes primary-study (level-2 cluster), and *k* denotes spillover type (level-3 cluster). The random-effect components ($h_{..}^{..}$) with subscript *0* denote study- or cluster-specific intercepts whereas those with subscript *1* denote study- or cluster-specific slopes. Finally, of the superscripts, *RI2* indicates two-level HM with random intercepts only; *RIS2* indicates two-level HM with random-intercepts and slopes; *RI3* indicates three-level HM with random intercepts only; and *RIS3* indicates a three-level HM with random intercepts and slopes. Hence, (5a) is a two-level HM with random intercepts; (5b) is a two-level HM with random intercepts and slopes; (5c) is three-level HM with random intercepts; and (5d) is a three-level HM with random intercepts at the study and cluster levels and random slopes at the study level.

If any of the HMs in 5a-5b is preferred against a weighted least squares (WLS) specification and if the average effect-size estimate is statistically significant, the PEESE versions of the HM specifications can be stated as follows:

$$t_{ij} = \gamma^{RI2} \left({}^{1}\!/_{SE_i} \right) + \varphi^{RI2} SE_i + h_{0j}^{RI2} + u_{ij}^{RI2} \qquad (6a)$$
$$t_{ij} = \gamma^{RIS2} \left({}^{1}\!/_{SE_i} \right) + \varphi^{RIS2} SE_i + h_{0j}^{RIS2} + h_{1j}^{RIS2} \left({}^{1}\!/_{SE_{ij}} \right) + u_{ij}^{RIS2} \qquad (6b)$$
$$t_{ijk} = \gamma^{RI3} \left({}^{1}\!/_{SE_i} \right) + \varphi^{RI3} SE_i + h_{0j}^{RI3} + h_{0k}^{RI3} + u_{ijk}^{RI3} \qquad (6c)$$

$$t_{ijk} = \gamma^{RIS3}\left(1/_{SE_i}\right) + \varphi^{RIS3}SE_i + h_{0j}^{RIS3} + h_{0k}^{RIS3} + h_{1j}^{RIS3}\left(1/_{SE_{ij}}\right) + u_{ijk}^{RIS3} \qquad (6d)$$

We rely on likelihood ratio (LR) tests to choose between HMs and WLS or between different HM specifications. The test compares different assumptions about the variance of the reported effect-size estimates, which is assumed to be distributed around the 'true' effect (γ) with a variance of θ_i.

$$effect\_size_i \sim N(\gamma, \theta_i) \qquad (7)$$

The WLS assume that individual variances are just a multiple of the idiosyncratic error variance ($\emptyset\sigma_i^2$). In contrast, the hierarchical models assume an additive variance structure in which the random-effect variances (τ^2) correspond to different assumptions about between-study heterogeneity. These assumptions can be stated as follows:

$\theta_i^{WLS} = \emptyset\sigma_i^2$ WLS
$\theta_i^{RI2} = \sigma_i^2 + \tau_{01}^2$ Two-level HM with random intercepts at the study level
$\theta_i^{RIS2} = \sigma_i^2 + \tau_{01}^2 + \tau_{11}^2$ Two-level HM with random intercepts and random slopes
$\theta_i^{RI3} = \sigma_i^2 + \tau_{01}^2 + \tau_{02}^2$ Three-level HM with random intercepts at study and spillover cluster levels
$\theta_i^{RIS3} = \sigma_i^2 + \tau_{01}^2 + \tau_{02}^2 + \tau_{11}^2$ Three-level HM with random intercepts and random slopes

The null hypothesis in the LR tests is that the restricted model (the model with one or several random-effect variances restricted to zero) is nested within the unrestricted model. A rejection of the null hypothesis indicates that the unrestricted model (the HM with more complex heterogeneity structure) fits the data better than the restricted model, which can be a WLS model with no additive term for heterogeneity or a HM with a relatively simpler heterogeneity structure.

One drawback of the HMs is that they assume normality of the model residuals and this is more explicit compared to WLS. However, violation of the normality assumption affects the confidence intervals but not the coefficient estimates. Therefore, we are of the view that HMs are capable of addressing a wide range of estimation issues with little or no cost in terms of consistency (Demidenko, 2004; McCulloch et al., 2008; Snijder and Bosker, 2012).

The HM framework for the bivariate meta-regression outlined above applies directly to a multivariate meta-regression context that allows for modelling the observed sources of heterogeneity. Augmented with covariates that capture observed sources of heterogeneity, the WLS and HM versions of the multivariate meta-regression can be stated as follows:

$$t_{ij} = \alpha^{WLS} + \beta^{WLS}(1/SE_{ij}) + \sum_m \gamma_m^{WLS} Z_m(1/SE_{ij}) + \varepsilon_{ij}^{WLS} \quad (8a)$$

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

Here Z is a vector of m binary variables that control for moderating factors as indicated above. All moderating variables are divided with the standard error of the primary-study estimates to capture their effects on the average effect (as opposed to their effects on the selection bias). Equation 7a is a WLS model and (7b) is a two-level HM with random intercepts and slopes at the study level. The choice between restricted and unrestricted models is based on LR tests indicated above.

Table A3. Summary Statistic for Moderating Variables

Moderating variables	Obs.	Mean	Std Dev	Min	Max
Effect indicators					
Effect size	983	0.126	0.352	-3.014	4.060
Standard error of effect size	983	0.055	0.144	0.001	2.276
t-value	983	4.712	7.109	-64.857	65.8
Precisions	983	86.169	146.319	0.439	999
Publication characteristics					
Journal article	983	0.964	0.185	0	1
Publication date after 2000	983	0.721	0.449	0	1
Journal quality	983	0.227	0.419	0	1
Model specification in primary study					
TFP - Dependent variable is total factor productivity	983	0.586	0.493	0	1
SPO coefficients in model <=2	983	0.687	0.464	0	1
Firm, industry, country dummy in model	983	0.774	0.418	0	1
Control for own R&D in model	983	0.846	0.361	0	1
Time dummy in model	983	0.586	0.493	0	1
Data and sample characteristics					
Data mid-point < 1991	983	0.783	0.412	0	1
Unit of analysis: country	983	0.467	0.499	0	1
Unit of analysis: industry	983	0.060	0.238	0	1
High R&D-intensity firm, industry Versus low or mixed R&D intensity	983	0.208	0.406	0	1
North American (US and Canada) data	983	0.131	0.338	0	1
OECD data	983	0.729	0.444	0	1
Spillover characteristics					
Asymmetric – weights are scaled by openness	983	0.195	0.397	0	1
Unweighted – spillover pool is constructed without weights	983	0.251	0.434	0	1
R&D flows	983	0.109	0.313	0	1
Estimation method					
Estimation is based on panel cointegration	983	0.172	0.378	0	1
IV – instrumental variable estimation	983	0.033	0.178	0	1
Differenced – within estimation	983	0.044	0.205	0	1

Table A4 – Effects size estimates by spillover types (frequency-weighted)

Dependent variable: t-value	**Panel A - PET/FAT**					**Panel B - PEESE**	
	(A1)	(A2)	(A3)	(A4)	(A5)	(B1)	(B4)
Effect (β in PET/FAT, γ in PEESE)	0.064**	0.052	-0.097	0.058***	0.043	0.097***	0.073***
	(0.028)	(0.037)	(0.099)	(0.012)	(0.033)	(0.026)	(0.011)
Selection bias	1.900***	1.259***	2.932***	1.231***	2.285***		
	(0.473)	(0.418)	(0.765)	(0.346)	(0.339)		
Standard error						1.424	3.203**
						(1.479)	(1.587)
Obs.	557	96	327	501	983	557	501
Studies	46	6	30	26	60	46	26
Log-likelihood (LL)	-1760.941	-306.995	-932.755	-1472.789	-3064.777	-1766.875	-1474.265
LL (comp. model)	-1853.435	-323.953	-1051.853	-1685.547	-3321.677	-1933.186	-1714.120
LR chi^2	184.987	33.915	238.196	425.516	513.8	332.623	479.709
P > LRc^2	0	0	0	0	0	0	0
Intra-class correlation	0.103	9.30e-18	0.252	1.51e-13	0.063	0.170	3.32e-14

***, **, * indicates significance at 1%, 5% and 10%. LR chi^2 is based on likelihood ratio test where the null hypothesis is that the linear model is nested within the multi-level model. (1) is knowledge spillovers; (2) is mixed spillovers; (3) is rent spillovers; (4) is own R&D; (5) is all spillovers types

Table A5 - Effects size estimates by unit of analysis (frequency-weighted)

Dependent variable: t-value	**Panel A - PET/FAT**						**Panel B - PEESE**				
	(A1)	(A2)	(A3)	(A4)	(A5)	(A6)	(B1)	(B2)	(B3)	(B4)	(B6)
Effect (β in PET/FAT, γ in PEESE)	0.102**	-0.106*	0.030***	0.049***	0.048	0.059***	0.135***	-0.078	0.069***	0.055***	0.073***
	(0.050)	(0.062)	(0.010)	(0.008)	(0.041)	(0.020)	(0.027)	(0.060)	(0.003)	(0.006)	(0.019)
Selection bias	1.822**	2.513***	2.501***	0.898**	2.059**	1.263***					
	(0.825)	(0.648)	(0.480)	(0.441)	(1.011)	(0.409)					
Standard error							-0.004	2.516***	-0.226	18.257**	0.655
							(0.931)	(0.619)	(0.911)	(7.584)	(1.002)
Obs.	459	223	299	283	89	126	459	223	299	283	126
Studies	26	12	22	25	9	19	26	17	22	25	19
Log-likelihood (LL)	-1408.742	-678.214	-955.515	-705.504	-313.91	-265.875	-1408.742	-678.214	-955.515	-705.504	-265.875
LL (comp. model)	-1541.737	-756.472	-1008.283	-817.222	-333.312	-342.706	-1541.737	-756.472	-1008.283	-817.222	-342.706
LR chi^2	265.989	156.515	105.535	223.437	38.805	153.662	265.989	156.515	105.535	223.437	153.662
P > LRc^2	0	0	0	0	0	0	0	0	0	0	0
Intra-class correlation	0.037	0.242	0.019	2.55e-16	5.00e-16	0.396	0.100	0.403	0.248	1.13e-15	0.617

***, **, * indicates significance at 1%, 5% and 10%. LR chi^2 is based on likelihood ratio test where the null hypothesis is that the linear model is nested within the multi-level model. (1) is spillovers - country; (2) is spillovers - industry; (3) is spillovers - firm; (4) is own R&D - country; (5) is own R&D – industry; and (6) is own R&D – firm. (1) to (3) is inter-unit only while (4) to (6) is intra-unit only.