Title page

R&D and productivity in OECD firms and industries: A hierarchical meta-regression analysis

Mehmet Ugur[a], Eshref Trushin[b] Edna Solomon[a] and Francesco Guidi[a]

Abstract

The relationship between R&D investment and firm/industry productivity has been investigated widely following seminal contributions by Zvi Griliches and others from late 1970s onwards. We aim to provide a systematic synthesis of the evidence, using 1,253 estimates from 65 primary studies that adopt the so-called primal approach. In line with prior reviews, we report that the average elasticity and rate-of-return estimates are positive. In contrast to prior reviews, however, we report that: (i) the estimates are smaller and more heterogeneous than what has been reported before; (ii) residual heterogeneity remains high among firm-level estimates even after controlling for moderating factors; (iii) firm-level rates of return and within-industry social returns to R&D are small and do not differ significantly despite theoretical predictions of higher social returns; and (iv) the informational content of both elasticity and rate-of-return estimates needs to be interpreted cautiously. We conclude by highlighting the implications of these findings for future research and evidence-based policy.

Key words: R&D, knowledge capital, productivity, meta-analysis
JEL Classification: D24, O30, O32, C49, C80

This paper is part of an ESRC Project titled *Evaluation of Research and Development (R&D) Expenditures, Firm Survival, Firm Growth and Employment: UK Evidence in the OECD Context.* Reference no ES/K004824/1. We thank the funders for their support, subject to usual disclaimer: the views expressed here are those of the authors only, who are responsible for any errors or omissions.

[a] *University of Greenwich Business School;* [b] *Durham University Business School*

Corresponding author: Mehmet Ugur, Professor of Economics and Institutions, University of Greenwich Business School, Park Row, London SE10 9LS.
Email: M.Ugur@gre.ac.uk

R&D and productivity in OECD firms and industries: A hierarchical meta-regression analysis

Mehmet Ugur[a], Eshref Trushin[b] Edna Solomon[a] and Francesco Guidi[a]

Abstract

The relationship between R&D investment and firm/industry productivity has been investigated widely following seminal contributions by Zvi Griliches and others from late 1970s onwards. We aim to provide a systematic synthesis of the evidence, using 1,253 estimates from 65 primary studies that adopt the so-called primal approach. In line with prior reviews, we report that the average elasticity and rate-of-return estimates are positive. In contrast to prior reviews, however, we report that: (i) the estimates are smaller and more heterogeneous than what has been reported before; (ii) residual heterogeneity remains high among firm-level estimates even after controlling for moderating factors; (iii) firm-level rates of return and within-industry social returns to R&D are small and do not differ significantly despite theoretical predictions of higher social returns; and (iv) the informational content of both elasticity and rate-of-return estimates needs to be interpreted cautiously. We conclude by highlighting the implications of these findings for future research and evidence-based policy.

Key words: R&D, knowledge capital, productivity, meta-analysis
JEL Classification: D24, O30, O32, C49, C80

[a] *University of Greenwich Business School;* [b] *Durham University Business School*

Corresponding author: Mehmet Ugur, Professor of Economics and Institutions, University of Greenwich Business School, Park Row, London SE10 9LS.
Email: M.Ugur@gre.ac.uk

R&D and productivity in OECD firms and industries: A hierarchical meta-regression analysis

1. Introduction

Productivity effects of research and development (R&D) investment has been a subject of major interest for researchers and policy makers. The pioneering work is that of Minasian (1969) and Griliches (1973) on elasticities of R&D capital; and Terleckyj (1974) on rates of return to R&D. Mairesse and Sassenou (1991), Mairesse and Mohnen (1994) and Hall (1996) provide early reviews of the empirical literature that flourished after Griliches (1979) had articulated a lasting framework for the analytical/empirical issues in the research field. More recently, Hall et al. (2010) provide an authoritative assessment of the findings and how the latter relate to variations in identification, estimation and level of analysis. Finally, Wieser (2005) and Møen and Thorsen (2015) meta-analyse the sources of variation in the evidence base and the extent of publication selection bias, respectively.

We have identified a number of issues that justify a novel review. First, existing reviews tend to rely on selected estimates rather than all available information. For example Wieser (2005) covers firm-level studies only and its sample consists of 22 and 16 studies that report 102 elasticity and 52 rate-of-return estimates respectively. Furthermore, it uses means and medians of the statistically-significant estimates to depict the balance of the evidence. Hall et al. (2010) covers both firm- and industry-level studies and is much more comprehensive in terms of primary studies reviewed. However, this review too reports only a single summary measure or a range for each study without spelling out how the measure is selected. Finally, Møen and Thorsen (2015) utilize 94 elasticity and rate-of-return estimates based on median of the estimates reported in 41 primary studies.

The evidence analysed in these reviews is clearly truncated because the number of primary studies and estimates we have identified is much larger and distributed as follows: 33 primary studies reporting 773 elasticity estimates at the firm level; 21 studies reporting 192 rate-of-return estimates at the firm level; 9 studies reporting elasticity estimates at the industry level; and 12 studies reporting 153 rate-of-return estimates at the industry level. A truncated sample not only constitutes an inefficient use of the existing information but also exacerbates the risk of selection bias discussed next.

The risk of selection bias arises when studies that fail to reject the null hypothesis are less likely to be published than those that do produce a statistically significant result. This is known as the file drawer problem in meta-analysis (Card and Krueger, 1995; Sterling et al., 1995; Stanley, 2008; Stanley and Doucouliagos, 2012). Of the existing reviews, only

Møen and Thorsen (2015) addresses the risk of selection bias through precision-effect tests (PET) and trim-and-fill methods. However, the selected nature of the estimates in this review and others constitutes an additional source of selection bias. In this review, we include all available information to avoid the reviewer-induced selection bias and we address the file-drawer problem systematically for both elasticity and rate-of –return estimates at the firm and industry levels.

The third issue is that none of the reviews addresses the twin problems of data dependence and heterogeneity. Data dependence arises when primary studies using a particular dataset report multiple estimates or when different studies use overlapping segments of the survey data compiled by the same national statistical (Doucouliagos and Laroche, 2009; Stanley and Doucouliagos, 2012). In such cases, the primary-study estimates are not based on random realisations of the data generating process and therefore the meta-analyst needs to take account of both within- and between-study dependence through hierarchical model estimations. Furthermore, heterogeneity in the evidence base must be quantified and its implications for the generalizability of the summary measures or meta-analysis estimates must be discussed (Sterne and Harbord, 2004; Harbord and Higgins, 2008). The existing reviews acknowledge the heterogeneous nature of the primary-study estimates but they neither provide a quantitative measure of the heterogeneity that cannot be explained by sampling/study characteristics nor do they caution about the extent to which the summary measures they report can be generalised.

Finally, and despite repeated cautions by leading contributors, the existing reviews do not address the question as to whether the elasticity and rate-of-return estimates measure what they are actually meant to measure – i.e., true productivity effects of R&D investment. We argue that the informational content of the primary-study estimates may be constrained for three reasons. First, the 'true productivity' effect at the firm level may differ from the 'revenue productivity' effect if the firm-specific output prices differ from the industry-level deflator used to deflate the firm's output. If the firm-specific price is higher (lower) than the average industry deflator, the 'real' value of its output will be biased upward (downward). To the extent that the wedge between firm and industry prices may reflect differences in the firms' market power, the latter's 'revenue elasticity' can diverge from the true 'productivity elasticity'. The two are the same only in the case of infinite price elasticity of demand, i.e. when all firms operate in perfectly competitive markets (Griliches and Mairesse, 1995; Mairesse and Jaumandreu, 2005; Foster et al, 2008 and Hall, 2011).

The second limitation is due to the contemporaneous nature of the rate-of-return estimates, which measure the effect of R&D intensity in year *t* on output or TFP growth in year *t*. However, R&D projects may take a long time to complete and even completed projects may affect productivity with a lag as firms convert the R&D knowledge into new products and services. Therefore, evidence from Añón Higón (2007) and Doraszelski and Jaumandreu (2013) indicates that the contemporaneous rates of return may be biased downward. The downward bias may be exacerbated by the fact that the R&D intensity in rate-of-return models is based on gross rather than net R&D after depreciations. That is

why Griliches and Mairesse (1991a) caution that the rate-of-return estimate could be considered only as a 'distant reflection' of the true rate-of-return concept.

The third limitation is due to the strict assumptions required to compare the social rates of return estimated at the industry level with private returns to R&D at the firm level. The industry-level estimates of social return are valid only if returns to scale are constant and all firms within an industry face a common factor price (Griliches, 1992). When these assumptions hold, the within-industry social returns are expected to be higher than private returns as the former capture both private returns to the firms in the industry and the effect of knowledge spillovers from the R&D capital stock in the industry (Griliches, 1979; 1992). When these assumptions do not hold, however, industry-level estimates may capture both spillover-effects and shifts in aggregate industry productivity caused by different combinations of firms with different firm-specific factor prices.

Having addressed the issues above, we find that the average elasticity and rate-of-return estimates are positive but smaller than most summary measures reported in prior reviews. We argue that this is because existing summary measures suffer from what we describe as double selection: *publication selection* that arises when primary-study authors search for samples, estimation methods or model specifications that yield statistically-significant estimates; and *sample selection* that arises when reviewers rely on 'representative' or 'preferred' estimates rather than all available information.

Secondly, we report that the elasticity and rate-of-return estimates are highly heterogeneous. Heterogeneity does not invalidate the synthesized findings but limits the extent to which they can be generalised. This is particularly the case with respect to firm-level private returns, where moderating factors explain only part of the heterogeneity that cannot be explained by sampling differences.

Our third finding indicates that the productivity effect of R&D at the industry level *does not* differ from that at the firm level - i.e., *private returns* to R&D are about the same as *within-industry social returns*. This is in contrast to theoretical predictions and may be due to data quality issues or absence of support for the assumptions of the theoretical model or both.

The fourth finding we report indicates that that the gross private rate of return at the firm level (14%) is less than the depreciation rate for R&D capital (15%) usually assumed in the primary studies. This anomaly clearly suggests that the existing estimates suffer from a serious downward bias as suspected by Griliches and Mairesse (1991a).

In the light of these findings, we argue that the informational content of the existing estimates is constrained by data quality problems and the limited extent to which the latter can be addressed satisfactorily by the econometrician. Therefore, we suggest that future research should utilise finely-grained industry or product-line data with long time

horizons to: (i) identify the lag structure of the R&D capital and estimate both short and long-run R&D productivity effects through autoregressive distributed lag (ARDL) estimations (Añón Higón, 2007) or Markov chain processes that capture the impact of R&D on the evolution of productivity (Doraszelski and Jaumandreu, 2013); (ii) take account of the interactions between R&D investments and market power with a view to distinguish between 'revenue' and 'true' productivity effects (Hall, 2011; Mairesse and Jaumandreu, 2005); (iii) disentangle private from social returns to R&D by taking account of cross-sectional dependence (Eberhardt et al., 2013) and separate technology spillovers from product-market rivalry (Bloom et al., 2013) or from creative destruction (Aghion et al., 2014).

The rest of the paper is organised as follows. Section 2 discusses the analytical and empirical dimensions of the research field and their implications for the heterogeneity of the evidence base. In section 3, we report the systematic review strategy we adopted to identify eligible studies and the meta-analysis methodology we followed to take account of selection bias, heterogeneity, and data dependence in the evidence base. Section 4 report both bivariate and multi-variate hierarchical regression results for four sets of evidence. Two of the evidence pools are related to private returns to R&D measured as elasticities and rates of return at the firm level; and the other two are related to within-industry social returns measured as elasticity and rates of return estimates at the industry level. We conclude in section 5 by providing a systematic summary of our findings and their implications for future research.

2. R&D and productivity: Analytical and empirical dimensions of the research field

Primary studies on R&D and productivity usually draw on a Cobb-Douglas production function, augmented with R&D (knowledge) capital. Estimates from these studies constitute the evidence base from the *primal approach* as opposed to the *dual approach* based on cost or profit functions.[1] Assuming perfect competition in factor markets and separability of the conventional inputs (capital and labour) from knowledge capital, the production function can be stated as follows:

$$Q_{it} = Y_{it}/P_{jit} = Ae^{\lambda t}C_{it}^{\alpha}L_{it}^{\beta}K_{it}^{\gamma}e^{u_{it}} \quad (1)$$

Here, Q_{it} is real output of firm or industry *i* at time *t*. Real output is nominal output (Y_{it}) deflated with industry price deflators (P_{jit}). C_{it} is deflated physical capital stock; K_{it} is deflated R&D capital; L_{it} is labour (number of employees or hours worked); and $Ae^{\lambda t}$ is technological progress with a rate of disembodied technological change λ. Taking natural logarithms and using lower-case letters, the empirical model can be written as:

[1] This review excludes the dual-approach studies as the latter are small in number and their model specifications are more varied than the primal-approach studies.

$$q_{it} = y_{it} - p_{sit} = \alpha c_{it} + \beta l_{it} + \gamma k_{it} + \eta_i + \lambda_t + u_{it} \quad (2)$$

The logarithm of technical progress yields a firm- or industry-specific effect (η_i) and a time effect (λ_t). Following Mairesse and Griliches (1988), the empirical work adopts various assumptions about the intercept (η_i) and the slope coefficient of interest (γ). Some studies assume that both the intercept and the slope coefficient are constant across firms/industries and hence use pooled OLS for estimation. Some others assume random intercept drawn from the same distribution and constant slopes. Then the parameters are estimated either with a random-effect estimator where the intercept is the expected value of the idiosyncratic intercept coefficients; or with a between estimator that consists of a cross-sectional (total) OLS with data averaged over time for each cross-sectional unit. Elasticity estimates from OLS, random-effect or between estimators are referred to as *elasticity estimates in the level dimension*.

Some studies assume firm-specific intercepts and a common slope parameter. To eliminate the firm-specific fixed effects, model (2) can be time-differenced as follows[2]:

$$\Delta q_{it} = \alpha \Delta c_{it} + \beta \Delta l_{it} + \gamma \Delta k_{it} + \Delta \lambda_t + \Delta u_{it} \quad (3)$$

Time-differencing eliminates the fixed-effect and the time effect is now a growth-rate effect. Estimations based on differencing between two subsequent years are referred to as *first-differenced* and those based on longer periods are usually referred to as *long-differenced*. Alternatively, the fixed-effect can be eliminated by estimating model (2) with a *within* estimator, where all terms in the model are expressed as deviations from the mean over the time period. The 'within' model corresponds to growth rates of the covariates in (2), given that the latter are in logarithms (Cincera, 1998). Hence, productivity estimates from time-differenced or within estimators are referred to as *elasticity estimates in the temporal dimension*.

Estimates from the *level* and *temporal* dimensions will be consistent if model (2) is specified correctly and the covariates are not subject to mismeasurement. Assuming correct model specification, time-differencing exacerbates any measurement errors and usually leads to downward bias in estimated coefficients of interest (Mairesse and Griliches, 1988; Hall et al., 2010). Hence, elasticity estimates in the temporal dimension are usually expected to be smaller than those in the level dimension.

In (2) and (3), the elasticities (γ) are assumed constant across firms or industries. However, firms may operate with different factor shares (hence elasticities), depending on the competitive equilibria they are faced with (Hall et al., 2010). In this case, it is more appropriate to assume *rate-of-return* rather than *elasticity* equalisation.

[2] In (3), $\Delta q_{it} = q_{it} - q_{it-1}$; $\Delta c_{it} = c_{it} - c_{it-1}$; $\Delta l_{it} = l_{it} - l_{it-1}$ and $\Delta k_{it} = k_{it} - k_{it-1}$.

For rate-of-return estimations, the change in R&D capital stock (Δk_{it}) is transformed into R&D intensity using the definition of the elasticity coefficient below.

$$\gamma = (\partial Q_{it}/\partial K_i)\,(K_{it}/Q_{it}) = \rho(K_{it}/Q_{it}) \quad (4a)$$

Here, $\rho = \partial Q_{it}/\partial K_i$ is the marginal product of R&D capital. If the depreciation rate (δ) between two years is close to zero, the rate of return to R&D investment can be estimated directly using (4b).[3]

$$\Delta q_{it} = \Delta\lambda_t + \alpha\Delta c_{it} + \beta\Delta l_{it} + \rho\frac{R_{it}}{Q_{it}} + \Delta u_{it} \quad (4b)$$

Some studies use a total factor productivity (TFP) version of (4b) by subtracting conventional inputs (capital and labour) from both sides, yielding[4]:

$$\Delta TFP_{it} = \Delta\lambda_t + \rho\frac{R_{it}}{Q_{it}} + \Delta u_{it} \quad (4c)$$

Rates of return in (4b) or (4c) are gross returns as they are based on the assumption that the depreciation rate for R&D capital is zero. They measure the gross return on $1 worth of investment in R&D.

If they are estimated with *firm-level data*, models (2), (3), (4a) and (4b) yield estimates of *private returns* to R&D. The latter may capture transfers when the innovating firm increases its revenue at the expense of its competitors. However, this transfer-inclusive private returns are different than *social returns*, which can be estimated only at the *industry level* (Griliches, 1979: 25).[5]

[3] Model (4b) is derived by substituting 4a in 3. Then, $\Delta q_{it} = \Delta\lambda t + \alpha\Delta c_{it} + \beta\Delta l_{it} + \rho(K_{it}/Q_{it})\,\Delta k_{it} + \Delta u_{it}$. The term for knowledge capital simplifies as follows: $\rho(K_{it}/Q_{it})\,\Delta k_{it} = \rho(K_{it}/Q_{it})\,(\Delta K_{it}/K_{it}) = \rho(\Delta K_{it}/Q_{it}) = \rho(K_{it} - K_{it-1})/Q_{it} = \rho\frac{(1-\delta)K_{it-1}+R_{it}-K_{it-1}}{Q_{it}} = \rho\frac{R_{it}-\delta K_{it-1}}{Q_{it}} \cong \rho\frac{R_{it}}{Q_{it}}$ if rate of depreciation (δ) is close to zero.

[4] The use of TFP as dependent variable is rare with estimations based on firm-level data. This is because TFP requires the imposition of a priori restrictions, the most important of which is that estimated elasticities are equal to factor shares observed in the data. As indicated in Griliches (1979), it is more plausible to impose such restrictions at the industry rather than firm level. Hence, we meta-analyse the estimates based on TFP at the industry level only.

[5] Within-industry social returns can be derived from the Cobb-Douglas production function at the firm level. Let Q_i be firm-specific output; X_i a vector of conventional inputs; K_i firm-specific R&D capital; and K_a R&D knowledge pool in the industry. The firm-specific output is given by: $Q_i = BX_i^{1-\gamma}K_i^{\gamma}K_a^{\mu}$. Assuming constant returns to scale and same relative factor prices for all firms, the output at the industry level is the sum of all firm-specific outputs. Hence: $\sum_i Q_i = B(\sum_i K_i / \sum_i X_i)^{\gamma}K^{\mu}\sum_i X_i$. Given that $\sum_i K_i = K_a$ and $\sum_i X_i = X_a$, the output at the industry level is: $\sum_i Q_i = BX_a^{1-\gamma}K_a^{\gamma+\mu}$. The return to R&D capital at the industry level ($\gamma+\mu$) is greater than the return at the firm level (γ) if knowledge spillovers (μ) are positive.

Within-industry social returns are only one of the variants estimated in the wider literature. On the one hand, some studies augment the Cobb-Douglas production function with inter-industry, inter-region or inter-country spillovers to estimates social returns at the firm, industry or country levels (see Griliches, 1992 on the underlying model). On the other hand, the spillover pool can be constructed without weights (e.g., Bernstein and Nadiri, 1989); by using a weight matrix that capture technological proximity between firms or industries (Jaffe, 1986; Griliches, 1992); or by using geographical distance as the weight matrix (see Keller, 2004 for a review). More recently, Bloom et al. (2013) have proposed a Mahalanobis distance measure based on the co-location of patenting technology classes, which can also be extended to construct a weight matrix that captures product-market rivalry. The research in this extant literature warrants a systematic review in its own right, but this is beyond the scope of our study due to space limitations.

The analytical and empirical framework summarized above is fairly tractable and allows for pooling the existing estimates for meta-analysis. However this quality should not conceal the potential for high levels of heterogeneity due to measurement, identification, sampling and estimation issues discussed widely by leading contributors to and reviewers of the research field (Griliches, 1979; Griliches and Mairesse, 1995, Hall et al., 2010). Table 1 summarizes the potential sources of heterogeneity under five headings: publication type; measurement of inputs and output; model specification, sample characteristics; and estimation methods. Under each heading, we indicate the moderating (dummy) variables we use to capture the sources of heterogeneity and the reference category(ies) against which they are defined. We also indicate the expected effects of the sources of heterogeneity on the primary-study estimates we meta-analyse.

Table 1: Sources of variation in the evidence on R&D and productivity

Sources of variation in the evidence base	Controlled category	Reference category	Expected effect on: Elasticities	Rates of return
A. Publication type				
Journal article, working paper, report, thesis	Journal article	Working paper, report, thesis	+ / -	+ / -
B. Model specification				
Control for spillovers	Yes	No	-	-
Control for capacity utilisation	Yes	No	+ / -	+ / -
Industry/sector-dummies in estimated models	Yes	No	(-) in level dimension; (+/-) otherwise	+ / -
Time dummies in estimated models	Yes	No	+ / -	+ / -
Variable returns to scale	Yes	No	-	-
Data corrected for double counting	Yes	No	+	+
C. Input and Output measurement				
R&D capital constructed with perpetual inventory method	Yes	No	+	n.a.
Output is measured as output, sales or value added	Value added	Output and sales	+	+
D. Sample characteristics				
Mid-point of data period is 1980 or after	Yes	No	+ / -	+ / -
Firm size	Small firms	Mixed-size firms	-	-
French, German, UK and US firm or industry data	Country(i) data	Data from other OECD countries and Country(j≠i)	+ / -	+ / -
Data relates to R&D-intensive firms or industries	Yes	No	+	+
Data relates to publicly-funded R&D	Yes	No	-	-
E. Estimation method				
Pooled/Total OLS; Common factor frame estimators; Time-differenced estimators; GMM; Instrumental variable (IV) estimators; within estimators	ALL – except Pooled/Total OLS and Long-differenced	Pooled/Total OLS in levels; Long-diff. in temporal dimension	First-differenced (+/-); GMM,IV (-)	First-differenced (+/-); GMM and IV (-)

Notes: n.a = not applicable; +/- indicates that reported estimates associated with the moderating factor we control for are larger/smaller than those associated with the reference category.

The expected effects in the last two columns are informed by the theoretical and empirical framework discussed in Griliches (1979) and Griliches and Mairesse (1988; 1995). Those related to differences in measurement and sample characteristics are informed by empirical patterns reported in Griliches (1979), Hall et al. (2010) and prior reviews. Finally, the expected effect of publication types is informed by meta-analysis studies, which report that selection may be related to publication types (Card and Krueger, 1995; Sterling et al., 1995; Stanley, 2008; Costa-Font et al., 2013).

3. **Meta-analysis: protocol and method**

We follow best-practice guidelines for meta-analysis recommended in Stanley et al. (2013). We searched in 9 databases, using 13 search terms in the *Title* and 20 search terms in the *Abstract* fields. We also used the snowballing approach and identified 32 studies through backward citations. We included 65 studies that adopt the so-called primal approach, which involves estimating a Cobb-Douglas production function augmented with R&D capital or R&D intensity. The included studies report R&D elasticities when the independent variable is the logarithm of R&D capital and rates of return to R&D when the independent variable is R&D intensity defined as the ratio of R&D to output. Finally, the included studies report *private returns* to R&D when the estimation is based on firm-level data or *within-industry social returns* when industry-level data is used and the intra-industry R&D capital is aggregated with equal weights (Griliches, 1979; 1992).

Our literature search strategy and study inclusion/exclusion criteria are specified in a protocol[6], which also contains information on the decisions we made at the study selection and critical evaluation stages. We have set 1980 as the initial and 2013 as the final year of publication. The initial year[7] has been chosen to capture the empirical studies published after the publication of the seminal paper by Griliches (1979), who articulated a lasting framework for the theoretical and empirical dimensions of the research field. The final year corresponded to the start of our research project. We have excluded studies that adopt the dual approach on the grounds that the latter are based on cost or profit functions, small in number and their model specification is more varied than those adopting the primal approach (Hall et al., 2010). We have also excluded studies that follow Crépon et al. (1998), where R&D is an input in the innovation production rather than in the output production function.

Finally, we have also excluded studies that estimate social returns to R&D at sector, region or country levels. This is because social returns in such studies vary in nature and measurement. For example, the social returns may capture rent spillovers when the weight matrix is based on input/output tables. On the other hand, they may capture pure public good spillovers when R&D capital is aggregated with equal weights or asymmetric technology spillovers when the weight matrix is based on technology proximity (Griliches, 1992; Hall et al., 2010). Such differences make the within-industry social returns we meta-analyse here non-comparable with social returns associated with external spillover pools at the industry, region or country levels.

[6] The protocol will be made available via live link after completion of the anonymous review process. It also contains examples of and reasons for excluded studies.

[7] We note here that the most comprehensive review by Hall et al. (2010) also adopts 1980 as the initial year of publication. Although this review refers to the pioneering studies published before 1980, the sample used for providing summary measures of the elasticity and rate-of-return estimates consists of studies published in 1980 and thereafter.

Data extraction yielded 1257 estimates, of which we excluded four outliers with undue influence.[8] Hence, the meta-analysis is based on 1,253 estimates, of which 773 are elasticities at the firm level; 135 are elasticity estimates at the industry level; 192 are rate-of-return estimates at the firm level; and 153 are rate-of-return estimates at the industry level. We take account of all available information by including all elasticity and rate-of-return estimates reported in the primary studies.

The meta-analysis methodology draws on Stanley (2005, 2008), Doucouliagos and Stanley (2013), and Stanley and Doucouliagos (2012). The underpinning theoretical framework is that of Egger et al. (1997), who postulate that researchers with small samples would search intensely across model specifications, econometric techniques and data measures to find sufficiently large (hence statistically-significant) effect-size estimates. This simple theoretical framework implies that reported estimates are correlated with their standard errors. Denoting the effect size with e_i and the standard error with SE_i, and assuming that the error term (u_i) is independently and identically distributed (i.i.d.), the selection process can be modelled as follows:

$$e_i = \beta + \alpha SE_i + u_i \quad (5)$$

However, the selection model in (5) poses three major estimation issues. First, the model is heteroskedastic because effect-size estimates have widely-different standard errors. To address this issue, we estimate a weighted least squares (WLS) version (6), where precision ($1/SE_i$) is used as weight (Stanley, 2008; Stanley and Doucouliagos, 2012):

$$t_i = \beta \left(1/SE_i\right) + \alpha + v_i \quad (6)$$

Here t_i is the t-value associated with the reported estimate and the error term v_i is the error term in (5) weighted by precision. If the Gauss-Markov conditions are satisfied, OLS estimation of (6) yields minimum-variance linear unbiased estimates. Testing for $\alpha = 0$ is a test for publication selection bias or funnel asymmetry test (FAT), whereas testing for $\beta = 0$ is a 'genuine effect' or precision-effect test (PET) after controlling for selection bias. The selection bias is considered as *substantial* if $|\alpha| \geq 1$ or *severe* if $|\alpha| \geq 2$ (Doucouliagos and Stanley, 2009; 2012). Testing for selection bias is justified given the evidence about its prevalence in both social-scientific and medical research (Card and Krueger, 1995; Dickersin and Min, 1993; Ioannidis, 2005; and Simmons et al., 2011).[9]

[8] Observations with undue influence are identified through the DFBETA influence statistics. This procedure first estimates the difference between the regression coefficients by excluding an observation and then scales this difference by the estimated standard error of the coefficient. An influential observation is identified if the difference exceed one standard error, i.e., if |DFBETA|>1 (Bollen and Jackman, 1990).

[9] There is a mistaken presumption that the model proposed by Egger et al. (1997) makes the detection of publication selection bias almost inevitable because of the positive association between effect-size estimates and their standard errors or because of the negative association between effect-size estimates and their precision. On the contrary, simulation results in Stanley (2008) indicate that the selection bias test

The second issue is about which estimator is better-suited for the data at hand. Most meta-analysis studies tend to estimate (6) with ordinary least squares (OLS). This is the case in the two meta-analysis studies on R&D productivity (Wieser, 2005; Møen and Thorsen, 2015); and in recent meta-analysis studies in related fields (e.g., Castellacci and Lie, 2015). However, OLS estimates from (6) would be biased if the primary-study estimates were affected by *data dependence*. The latter arises when primary studies using a particular dataset report multiple estimates or when different studies use overlapping segments of the survey data compiled by the same national statistical agency (Doucouliagos and Laroche, 2009). Data dependence is clearly an issue here as several studies make use of the same dataset several times – albeit at different time periods. (See Tables A1.1 – A1.4 in the Appendix).

Data dependence can be addressed by: (i) obtaining bootstrapped standard errors; (ii) conducting clustered data analysis; and (iii) using hierarchical models (Doucouliagos and Laroche, 2009). The first two approaches only correct the standard errors for within-study dependence. Hierarchical models (HM), however, allow for robust standard errors clustered on studies and take account of both within-study and between-study dependence explicitly. An added feature is that HMs allow for a range of likelihood ratio (LR) tests to choose between least-square and HM estimators and between the latter themselves with respect to how dependence should be modelled. Therefore, we estimate model (6) as a HM – provided that the choice is justified by LR tests.

Data dependence is modelled by allowing for random variation between study-specific estimates, which may be due to study-specific intercepts and/or study-specific slopes (Demidenko, 2004; McCulloch et al., 2008). Stated differently, productivity estimates reported by primary studies are nested within each study; and the estimates are modelled to differ between studies either because they share a common intercept (a fixed component) and/or a common slope within each primary study.

The random-intercept-only and random-intercept-and-slope versions of the HM are stated in (7) and (8), respectively.

$$t_{ij} = \alpha + \beta\left(1/SE_{ij}\right) + v_{0j} + \varepsilon_{ij} \quad (7)$$

$$t_{ij} = \alpha + \beta\left(1/SE_{ij}\right) + v_{0j} + v_{1j}\left(1/SE_{ij}\right) + u_{ij} \quad (8)$$

Here, subscripts *j* and *i* refer to primary studies and effect-size estimates respectively; and ε_{ij} and u_{ij} are normally distributed error terms with zero mean and fixed variance. The random effects (v_{0j}) and (v_{1j}) are not estimated directly, but their variances are. Finally, parameters α and β are as defined above and estimated with maximum likelihood (ML).

based on Egger et al (1997) has low power - i.e., it tends to fail detecting publication selection when the latter actually exists.

The third issue in estimating the Egger et al. (1997) model is that the relationship between primary-study estimates and their standard errors may be non-linear. Indeed, Moreno et al. (2009) provide evidence that a quadratic specification is superior if 'genuine effect' exists beyond selection bias, i.e., if the PET in (6), (7) or (8) rejects the null hypothesis of zero effect. Then, the correct specification is obtained by weighting both sides of the Egger regression with precision-squared instead of precision. This inverse-variance weighting is referred to as precision-effect estimation corrected for standard errors (PEESE).

The random-intercept-only and random-intercept-and-slope versions of the hierarchical PEESE models are given below in (9) and (10) respectively; and all subscripts, random effects, error terms and parameters are as defined above.

$$t_{ij} = \alpha SE_{ij} + \beta\left(1/SE_{ij}\right) + v_{0j} + \lambda_{ij} \quad (9)$$

$$t_{ij} = \alpha SE_{ij} + \beta\left(1/SE_{ij}\right) + v_{0j} + v_{1j}\left(1/SE_{ij}\right) + w_{ij} \quad (10)$$

The 'average' R&D elasticity or rate-of-return (β) is estimated after controlling for selection bias. This is more reliable than simple or weighted means that do not take account of selection bias. However, its out-of-sample generalizability may be limited due to excessive heterogeneity in the evidence base. Therefore, we provide quantitative measures of heterogeneity that cannot be explained by within-study sampling variation drawing on the random-effect meta-regression proposed by Harbord and Higgins (2008). Then we use a multivariate meta-regression model (MRM) through which we estimate the effects of the potential sources of heterogeneity summarized in Table 1. The random-intercepts-only and random-slopes-and-intercepts versions of the MRM are given in (11) and (12), respectively:

$$t_i = \alpha + \beta(1/SE_{ij}) + \sum_k \theta_k Z_k(1/SE_{ij}) + v_j + \epsilon_{ij} \quad (11)$$

$$t_i = \alpha + \beta(1/SE_{ij}) + \sum_k \theta_k Z_k(1/SE_{ij}) + v_{0j} + v_{1j}(1/SE_{ij}) + \mu_{ij} \quad (12)$$

The *kx1* vector of covariates (Z_k) are dummy variables as defined in Table 1 above and summarized in Table A2 in the Appendix. They are all interacted with precision to capture their effects on elasticity or rates-of-return estimates reported in primary studies. Given that the expected effect of R&D on productivity is positive, the coefficients on the *Z* dummy variables in (11) and (12) are interpreted as follows: (i) a positive (negative) and significant coefficient indicates that primary-study estimates for which $Z_k=1$ are larger (smaller) than the reference category specified in Table 1; (ii) a non-significant coefficient indicates that estimates for which $Z_k=1$ are not systematically different than those associated with the reference category.

We estimate the PET/FAT/PEESE and multivariate models for 4 pools of evidence: (1) elasticity estimates at the firm level; (2) elasticity estimates at the industry level; (3) rate-

of-return estimates at the firm level; and (4) rate-of-return estimates at the industry level. Estimates from (1) and (3) measure *private returns,* whereas estimates from (2) and (4) measure *within-industry social returns* to R&D.

To avoid multicollinearity and overfitting, we follow a general-to-specific model-estimation routine. We present the general model results, together with relevant diagnostic statistics including the variance inflation factor (VIF) in Table A3 in the Appendix. The specific models are obtained by omitting the most insignificant covariates (those with the largest p-value) one at a time until all remaining covariates are statistically significant. The estimates from the specific model are presented in the main text and used for inference about sources of heterogeneity.

4. **Meta-analysis results: R&D effects and sources of heterogeneity**

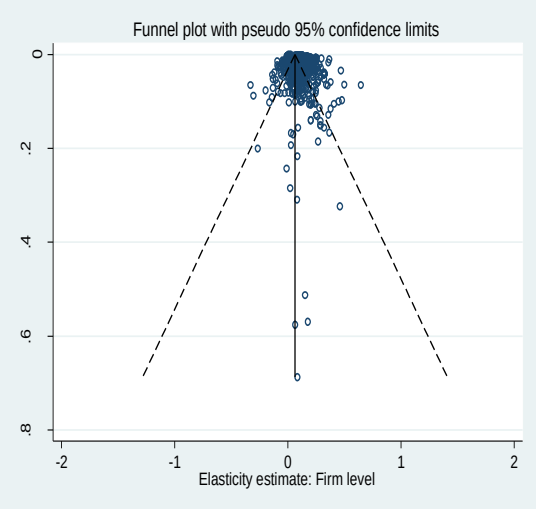
We report three sets of meta-analysis evidence: (1) descriptive statistics about the dimensions of the research field and funnel plots for visual inspection of heterogeneity and selection bias; (2) bivariate meta-regression estimates of publication selection bias and average 'effects' beyond selection bias; and (3) multivariate meta-regression estimates aimed at identifying the sources of variation in the evidence base.

4.1 Overview of the evidence base

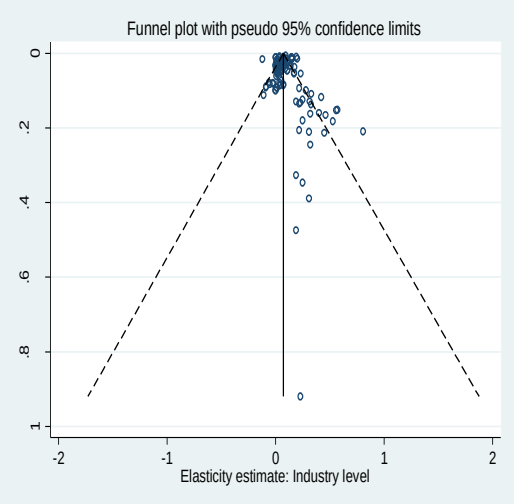
Tables A1.1 - A1.4 in the Appendix summarize the evidence base by study, reported elasticity or rate of return, and unit of analysis (firm or industry). Tables A1.1 and A1.2 include studies that report elasticity estimates at the firm and industry levels; whereas Tables A1.3 and A1.4 include studies that report rate-of-return estimates at the firm and industry levels respectively. Further study characteristics include publication type, country origin of the data, mid-point of the data period, dependent variable, estimation method, number of firms or industries, median of the estimates reported in the study, median t-value, and the number of estimates reported.

The majority of the studies consists of journal articles (59%) followed by working papers (25%). Also, a large proportion of the primary studies utilize US data (41%) followed by French, German and UK data (at about 8% each). The outcome variable is TFP in about 10% of the total observations, with the rest consisting of output, value added or sales. Whilst 77% of the studies utilise firm-level data, 23% are based on industry-level data. Of the latter, 43% (9 out of 21 studies) draw on US data and this is comparable with the overall share of US studies (41%) in the full sample.

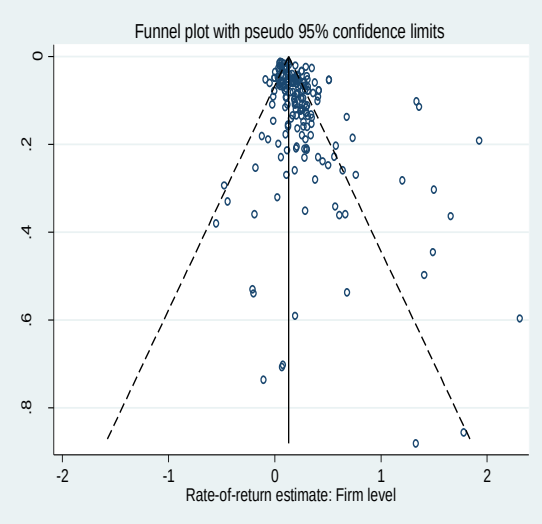
Figure 1: Funnel plots: potential selection bias and heterogeneity[10]



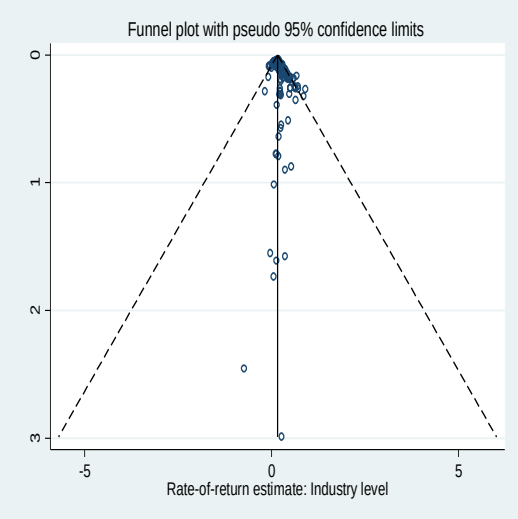
1. Elasticity estimates: Firm level
   Heterogeneity: 98%



2. Elasticity estimates: Industry level
   Heterogeneity: 86%



3. Rate of return estimates: Firm level
   Heterogeneity: 81%



4. Rate of return estimates: Industry level
   Heterogeneity: 17%

The number of estimates reported by each study varies between 2 and 82, depending on the evidence pool. Median values of the estimates by study are positive, with the exception of Odagiri (1983). The median t-value is greater than 1.645 in 90% of the studies. This overview points out to positive and significant median estimates, but the latter are fairly heterogeneous. The median estimate ranges from 0.03 to 0.255 for elasticities at the firm

[10] The heterogeneity measure is a generalization of Cochran's Q and indicates the proportion of residual between-study variation due to heterogeneity, as opposed to within-study sampling variability (Harbord and Higgins, 2008). Higgins et al. (2003) suggest that heterogeneity is low if the measure is between 25%–50%, moderate if it is between 50%–75%, and high if over 75%.

level; from 0.008 to 0.313 for elasticities at the industry level; from -0.110 to 0.380 for rates of return at the firm level; and from 0.08 to 0.683 for rates of return at industry level.

The funnel plots in Figure 1 above are centred on the fixed-effect 'average' of the productivity effect, which is positive for both elasticity and rate-of-return estimates at the firm and industry levels. However, the distribution of the estimates around the mean (the vertical line) indicates a tendency to report larger estimates more often than otherwise. This is an indication of potential selection bias, which must be verified by funnel-asymmetry tests (FAT). Moreover, a large number of estimates lie beyond the 95% pseudo confidence intervals, indicating heterogeneity that cannot be explained by sampling errors (Sterne and Harbord, 2004). The random-effect meta-regression estimator proposed by Harbord and Higgins (2008) indicate that residual heterogeneity is 98% and 86% for the elasticity estimates at the firm and industry levels respectively (1 and 2); and 81% and 17% for rate-of-return estimates at the firm and industry levels (3 and 4). Given these indicators of selection bias and heterogeneity, summary measures may lead to incorrect inference, the risk of which is exacerbated when the summary statistics are based on selected estimates chosen by primary-study authors or reviewers.

4.2 Elasticities and rates of return beyond selection bias

Hierarchical meta-regression model (HM) estimates of mean productivity effects beyond selection bias are reported in Table 2 below. The HMs are fitted with random slopes and intercepts (models 1-3 and 5-7) or random intercepts only (models 4 and 8) in accordance with LR test results. Furthermore, LR tests favour the HM specification for all models as the null hypothesis that the least-squares models are nested within the HMs is rejected strongly. Further justification for the choice of HMs is provided by the log-likelihood values, which are smaller in magnitudes in the HMs compared to least-squares models. Finally, the standard errors are clustered at the study level to ensure correct inference.

Publication selection bias (α) is significant and substantial in the evidence pool that consists of rate-of-return estimates at the firm level (model 3); and moderate in the evidence pool for rate-of-return estimates at the industry level (model 4). The test results are in line with funnel graphs (3) and (4) above, where the proportion of individual estimates above the fixed-effect average is higher. The selection bias is positive but insignificant in the evidence pools for elasticity estimates at the firm and industry levels (1 and 2).[11] Selection bias does not invalidate the 'genuine' effect (β), which is positive and significant in all evidence pools after controlling for selection. Therefore, the consistent PEESE estimates are reported in columns 5 – 8, where we control for non-linear relationship between primary-study estimates and their standard errors (Moreno et al., 2009).

[11] The absence of statistically-significant bias in these pools may reflect the low power of the funnel-asymmetry test as indicated in note 9 above.

Table 2: Productivity effects of R&D: Average effects and selection bias

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: t-value	PET/FAT	PET/FAT	PET/FAT	PET/FAT	PEESE	PEESE	PEESE	PEESE
β	0.073***	0.066***	0.089***	0.115***	0.077***	0.074***	0.140***	0.144***
	(0.015)	(0.022)	(0.018)	(0.037)	(0.012)	(0.021)	(0.017)	(0.027)
α	0.479	0.501	1.404***	0.746***				
	(0.531)	(0.392)	(0.290)	(0.270)				
Std. error					0.779	-0.232	2.244**	-0.157
					(2.639)	(0.710)	(0.890)	(0.162)
Std. dev. of random slopes (log)	-2.813***	-2.964***	-3.543***		-2.808***	-2.994***	-3.200***	
	(0.173)	(0.354)	(0.265)		(0.165)	(0.334)	(0.492)	
Std. dev. of random intercepts (log)	0.466**	-0.435	-0.867***	-0.512	0.471***	-0.106	-0.457	-0.151
	(0.215)	(0.344)	(0.307)	(0.324)	(0.178)	(0.437)	(0.278)	(0.313)
Std. dev of residuals (log)	1.473***	0.480	0.700***	-0.069	1.474***	0.477	0.719***	-0.077
	(0.229)	(0.297)	(0.255)	(0.146)	(0.229)	(0.299)	(0.265)	(0.164)
Observations	773	135	192	153	773	135	192	153
Studies	37	9	21	12	37	9	21	12
LR Test chi^2	23.454	9.364	23.249	9.425	39.025	12.086	145.096	30.146
P> chi^2	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.000
Log-likelihood (HM)	-2293.001	-270.072	-456.764	-188.969	-2293.556	-270.694	-464.735	-191.501
Log-likelihood (Comp. model)	-2601.318	-323.428	-466.604	-199.970	-2610.669	-323.061	-479.624	-217.009
Heterogeneity#	98%	86%	81%	17%				

Notes: **(1)** & **(5)**: Elasticity estimates at the firm level; **(2)** & **(6)**: Elasticity estimates at the industry level; **(3)** & **(7)**: Rate-of-return estimates at the firm level; **(4)** & **(8)**: Rate-of-return estimates at the industry level. Robust standard errors (in brackets) are clustered at the study level. Significance of random effects is based on the natural logarithms of the standard deviations. Observations with undue influence are excluded, using the DFBETA influence statistics. LR Test Chi-squares indicate that the hierarchical models are preferred to least-squares estimators. LR tests for the specifications of the hierarchical models (not reported to save space) favour random-intercepts-and-slopes specification in (1, 2, 3) and (5, 6, 7); and random-intercepts-only specification in (4) and (8). Log likelihood values for the hierarchical and comparator models provide additional evidence in favour of HMs. # indicates the proportion of residual between-study variation due to heterogeneity, as opposed to within-study sampling variability. *, **, *** indicate significance at 10%, 5% and 1%, respectively.

In columns 5 and 6, the mean elasticity estimate is around 0.07 – with no significant difference between the elasticities at the firm level (column 5) compared to the industry level (column 6). The mean elasticity at the firm level is smaller than most comparable elasticities reported in prior reviews. Wieser (2005: 596) reports an average elasticity of 0.118 in the level and temporal dimensions; whereas Møen and Thorsen (2015) report mean elasticities of 0.109 before and 0.098 after correcting for selection bias in the level dimension. The closest to our finding is that of Hall et al. (2010), who report a median elasticity of 0.08 in the level and temporal dimensions combined.

The gross rate-of-return estimates in columns (7) and (8) are around 14% and smaller than those reported in prior reviews. The latter range from 20% to 30% in Hall et al. (2010) and from16% to 28% in Weiser (2005). The closest mean values (13% with and 18% without correction for selection bias) have been reported by Møen and Thorsen (2015).

The results indicate that the informational content of the primary-study estimates should be evaluated critically for three reasons. First, the gross private rate of return at the firm level (14.4%) is lower than the depreciation rate for R&D capital, usually assumed at 15%. This finding indicates that the net rate of private returns may be negative or very small – depending on the true depreciation rate. As indicated earlier, this is likely to be due to downward bias in the rate-of-return estimates based on the Cobb-Douglas production function. The sources of downward bias are twofold: (i) rate-of-return estimates are derived from contemporaneous estimations that do not take account of time-lags in the completion of R&D projects or conversion of the latter into innovative products and processes (Añón Higón, 2007; Doraszelski and Jaumandreu, 2013); and (ii) the use of gross R&D intensity in the model could bias the rate-of-return estimate downward between 5% and 50% (Hall and Mairesse, 1995; Griliches and Mairesse, 1991a).

Secondly, we find no difference between the *firm-level private returns* and the *within-industry social returns* to R&D – irrespective of whether the returns are estimated as elasticities (column 5 and 6) or as rates of return (column 7 and 8).[12] This is again in contrast to what the underlying theoretical model implies. As shown in footnote 5, the within-industry social returns are expected to be higher than private returns at the firm level provided that within-industry knowledge spillovers exist.

[12] This is in line with Hall et al. (2010), who also report that estimates with industry data are close to those based on firm data. However, Hall et al. (2010) discuss neither the lack of difference between private and within-industry social returns to R&D nor the compatibility/incompatibility of this finding with the underlying theoretical model.

Our finding indicates that either the methods of estimating the underlying theoretical model or the measurement of within-industry knowledge spillovers or both may be inadequate for capturing the difference between private and within-industry social returns to R&D. This interpretation ties in with Eberhardt et al. (2013: 25), who concludes that "… search for a more appropriate specification of the knowledge production function that … allows identification of private and social returns to R&D" is an important challenge for future research. It is also in line with Bloom et al. (2013) who report larger technology spillover effects when the weight matrix used to aggregate the R&D capital is based on a Mahalanobis distance measure of technological proximity.

This finding, however, should be interpreted in the light of relatively small number of industry-level studies. This aspect of the research field is already documented in prior reviews. Although our sample (21 studies) is 61% larger than the most comprehensive sample of 13 studies in Hall et al. (2010), it is difficult to ascertain whether the industry-level social returns would be different in a hypothetically larger sample. Therefore, we suggest that an increase in industry-level studies is desirable because industry-level data with a sufficiently long time dimension allows for taking account of the lag structure in R&D capital (Añón Higón, 2007; Hall et al., 2010) and of the cross-sectional dependence in panel data (Erberhardt et al., 2013).

The third reason is that the proportion of between-study variation due to heterogeneity as opposed to within-study sampling variability is usually high (between 81% and 98%), with the exception of the evidence pool on rate-of-return estimates at the industry level (17%). As indicated above, high levels of heterogeneity do not invalidate the average elasticity or rate-of-return estimates we report. However, they imply that all summary measures, including meta-regression estimates, may conceal a high degree of heterogeneity that should be revealed and quantified. The implication for evidence-based policy is that public support for R&D investment should be contingent and selective, paying attention to evolving evidence on firms/industries/technology classes with higher social returns to R&D investments.

The limitations to the informational content of the private and social returns to R&D are essentially due to data quality and identification problems. As Griliches and Mairesse (1995: 22) have already observed, much of the work "has been guided … by what 'econometrics' as a technology might be able to do … rather than focusing on the more important but technically less tractable problems of data quality and model specification."

We suggest that the informational content of the productivity estimates can be enhanced by: (i) availability of firm-level deflators and depreciation rates; (ii) identifying the factors that affect firms differently as they choose their inputs, including R&D capital; and (iii) using richer models preferably with industry-level data to (a) take account of the lag structure in the relationship between R&D capital and productivity, (b) disentangle social returns to R&D from cross-sectional dependence due to unobserved common factors, and

(c) incorporate technology class and market power into the technological progress component of the model. Until further progress along these dimensions, all we can infer is that R&D investment is associated with positive private and social returns, but the magnitude of the estimated effects is likely to fall short of reflecting the 'true' productivity or rate-of-return concepts implied by the underlying theory.

4.3 Multivariate meta-regression results

In what follows, we investigate how a range of moderating factors (analytical and empirical dimensions of the research field) affect the elasticity and rate-of-return estimates reported in the primary studies. The moderating factors are captured through dummy variables that reflect a specific feature of the research field *vis-a-vis* a reference category as specified in Table 1 above. Summary statistics for the covariates are given in Table A2 in the Appendix. For estimation, we use the hierarchical model specification justified by LR tests and follow the general-to-specific model routine discussed above.

The general-model estimates are presented in Table A3 in the Appendix. Across evidence pools, there is evident variation in the mix of moderating variables that explain heterogeneity in the evidence base. In addition, the residual heterogeneity remains high in the evidence pools related to firm-level (private) returns to R&D (97% in column 1 and 77% in column 3). Compared to Table 2 where the only moderating factor is the precision of the primary-study estimates, the moderating variables in the general model explain only a small percentage (1% and 4%) of the residual heterogeneity in the evidence base on private returns to R&D. However, the moderating variables in the general model explain a substantial percentage (17% and 21%) of the residual heterogeneity in the industry-level (social) returns to R&D. The implication here is that heterogeneity in the firm-level estimates is inherently larger and less likely to be explained by study characteristics.

As anticipated in the methodology section, the general models suffer from high levels of multicollinearity – with variance inflation factors (VIFs) ranging from 4.79 to 166.16. Therefore, we follow a general-to-specific model estimation routine, whereby we exclude the covariates with the highest p-value one at a time until all remaining covariates are significant. The results are presented in Table 3 below.

The routine has led to lower VIF statistics between 2.02 and 3.36, apart from evidence pool (3) where the VIF value is 7.03. Although this is relatively high, it is below the conventional threshold of 10 adopted in applied econometrics studies. Secondly, the specific models do not lead to any loss with respect to the level of heterogeneity explained by the remaining covariates. Therefore, our inference will be based on the specific-model results in Table 3.

Table 3: Sources of heterogeneity: Specific model estimations

Dependent variable: t-value	(1)	(2)	(3)	(4)
Precision	0.039*	0.116***	0.162***	0.215***
	(0.023)	(0.012)	(0.046)	(0.052)
Journal article	-0.040**	pmc	-0.127***	-0.114**
	(0.020)	pmc	(0.045)	(0.058)
Control for spillovers	pmc		-0.106***	
	pmc		(0.037)	
Control for capacity utilization		0.039***		
		(0.012)		
Industry dummies included		-0.051***	0.076***	-0.106***
		(0.012)	(0.023)	(0.035)
Time dummies included	0.022***		-0.122**	
	(0.004)		(0.048)	
R&D capital constructed with	0.050**	pmc	n.o.	n.o.
Perpetual inventory method	(0.022)	pmc	n.o.	n.o.
Output measured as value added	0.048***			0.027*
	(0.005)			(0.016)
Small firms	-0.017**	n.o.		n.o.
	(0.008)	n.o.		n.o.
French data				0.122***
				(0.041)
US data	0.030***			
	(0.009)			
R&D-intensive firm/industry	0.021**	0.082***		
	(0.009)	(0.015)		
Publicly-funded R&D	-0.142***	n.o.	-0.165***	-0.296***
	(0.032)	n.o.	(0.043)	(0.030)
First-differenced estimations	-0.053***		0.076***	
	(0.005)		(0.018)	
General method of moments	-0.020*			n.o.
estimators (GMM)	(0.012)			n.o.
Instrumental variable estimators (IV)		-0.150***	n.o.	0.125***
		(0.018)	n.o.	(0.040)
Long-differenced estimations	-0.021**		Reference	Reference
	(0.010)		category	category
Within estimators	-0.008*	-0.031***	n.o.	n.o.
	(0.004)	(0.011)	n.o.	n.o.
Constant	0.373	-0.406	1.117***	0.791*
	(0.392)	(0.469)	(0.302)	(0.426)
Std. dev. of random slopes (log)	-3.033***			
	(0.154)			
Std. dev. of random intercepts (log)	0.311	0.030	-0.283	0.310
	(0.200)	(0.370)	(0.262)	(0.251)
Std. dev. of residuals (log)	1.316***	0.383***	0.611***	-0.480***
	(0.026)	(0.064)	(0.050)	(0.066)
Observations	773	135	192	153
Number of studies	37	9	21	12
Log-likelihood (HM)	-2169.635	-251.690	-439.125	-147.384
LR Test Chi^2	342.366	179.625	102.019	236.519
P> Chi^2	0.000	0.000	0.000	0.000
Log-likelihood (comparator model)	-2365.473	-256.691	-445.803	-175.866
VIF	2.94	2.02	7.03	3.36
Heterogeneity	97%	64%	77%	0%

Notes: **pmc**: dropped due to perfect multicollinearity; **n.o.**: no observations for the covariate; **blank**: insignificant in the specific model. **(1 and 2)**: Elasticity estimates at the firm and industry levels, respectively; **(3 and 4)**: Rate-of-return estimates at the firm and industry levels, respectively. Significance of random effects is based on the natural logarithms of the standard deviations. Observations with undue influence are excluded, using the DFBETA influence statistics. LR tests for the specifications of the hierarchical models (not reported to save space) favour random-intercepts-and-slopes specification in (1), but random-intercepts only in (2), (3) and (4). *, **, *** indicate significance at 10%, 5% and 1%, respectively.

Given the potential trade-offs between the benefits of reducing the VIF and the cost of potentially omitting relevant covariates, in what follows we will qualify our inference by comparing the results from the specific models with those of the general counterparts. If the sign and significance of the coefficients on a covariate remain the same in both, we infer strong consistency. If the covariate is significant only in the specific model, we infer moderate consistency. Using this decision rule, we report the following findings.

4.3.1 Strongly-consistent evidence on sources of heterogeneity

1. *Journal articles* tend to report smaller estimates compared to other publication types (e.g., book chapters, reports, theses and working papers) for both private and within-industry social returns to R&D. Given that journals are 'reputable gatekeepers' concerned with research quality, we infer that the *winner's curse* reported in Costa-Font et al. (2013) is not observed in this research field: journal editors do not seem to exploit the benefits of reputation to publish 'more selected' evidence.

2. Studies that construct R&D capital with the *perpetual inventory method (PIM)* tend to report larger elasticity estimates compared to others that model R&D capital accumulation as a multiplicative process. The PIM accords the same weight to each unit of additional R&D investment in the current period irrespective of the R&D capital stock in the preceding period. As such, it may be a source of upward bias in elasticity estimates if the contribution of new R&D investment to R&D capital is a positive function of the latter in the preceding year(s) (Klette, 1994). Our finding indicates that modelling of the R&D capital accumulation process is a potential area for further research.

3. *Small firm data* is associated with lower elasticity estimates at the firm level. From a Schumpeterian perspective (Aghion et al., 2014), this finding can be interpreted as indicator of lower market power enjoyed by small firms. However, it may also be due to higher risks of measurement errors and self-selection in small-firm R&D data. Small firms are usually not subject to the same level of auditing rigour as large firms, which tend to be listed. Secondly, the response rate of small firms to statistical agency surveys is usually lower than large firms. Therefore, the smaller elasticity estimates associated with small-firm data may reflect not only lower market power but also a mixture of measurement errors and sample selection.

4. *Elasticity estimates for R&D-intensive firms or industries* are larger to non-R&D-intensive or mixed firms/industries. This finding is in line with Hall et al. (2010) and reflects the larger and more precise estimates reported in several primary studies (Griliches, 1980; Griliches and Mairesse, 1981; Cunéo and Mairesse, 1984; Odagiri, 1983; Bartelsman, 1990; and Hall, 1993). The explanation is that R&D-intensive firms/industries have better capacities to exploit the benefits of innovation.

5. However, the *rate-of-return estimates do not differ between R&D-intensive firms/industries and their counterparts* (columns 3 and 4). We checked whether this may be due to differences in the R&D intensity that underpins the elasticity and rate-of-return samples. The sample average of R&D intensity in studies estimating elasticity (e.g., Aldieri et al., 2008; Cincera, 1998; and Cunéo and Mairesse, 1984) ranges between 5% and 10%; whereas the interval is from 3% to 5% in the rate-of-return studies (e.g., Harhoff, 1994; Verspagen, 1995; and Wakelin, 2001). Our interpretation is that rates of return to R&D may vary not only between R&D-intensive firms and their reference category, but also with the level of R&D intensity itself. Therefore, we suggest that it is appropriate to control for scale effects in the R&D-productivity relationship.

   In the neoclassical theory of innovation, the marginal product of knowledge capital exhibits increasing returns to scale (Arrow, 1962, 1996; Romer, 1986). However, the only study that controls for scale effects in the R&D-productivity relationship in our sample reports decreasing returns to scale (Lokshin et al., 2008). The latter is in line with case-study findings (Pammolli et al., 2011) and review evidence (DiMasi and Grabowski, 2012) on the pharmaceutical industry. It is also in line with Schumpeterian models (Aghion et al., 2014), which establish an inverted-U relationship between R&D intensity and productivity. The inverted-U pattern is explained by the type of competition in technology (leap-frogging or neck-and-neck competition) and the distance to the technology frontier. In these models, firms closer to the technology frontier have to maintain high levels of R&D investment just to maintain their positions relative to the technology leader(s). Given the conflicting theoretical insights and empirical evidence, we argue that more work is required to ascertain whether the productivity effects of R&D (measured as elasticities or rates of return) are subject to scale effects and whether the latter are increasing or decreasing with the level of R&D investment.

6. *Publicly-funded R&D* is associated with lower estimates for both private and social returns to R&D. Although this finding is based on a relatively small number of estimates from five primary studies (Bartelsman, 1990; Lichtenberg and Siegel, 1991; Mansfield, 1980; Terleckyj, 1980 and Wolf and Nadiri, 1993), it remains significant after we control for any outlier effect. The existing literature suggests three reasons as to why elasticity and rate-of-return estimates based on publicly-funded R&D may be lower: (a) public support for business R&D may be rightly concentrated in firms/industries that generate higher levels of R&D (knowledge) spillovers and hence lower levels of appropriability (e.g., health and defence); (b) public funds may be concentrated in industries with lower returns due to the large scale of the R&D investments at the capacity building phase (e.g., aircraft and communications sectors); and (c) firms may be less efficient in the use of public subsidies in general, or subsidies may be misdirected (Hall et al., 2010; Lichtenberg and Siegel, 1991).

Nevertheless, we think a fourth explanation is also possible and perhaps more relevant. As Griliches (1979) has indicated, decomposing the R&D capital into public and private components is admissible if both types are complements and enter the production function in a multiplicative form. If they are substitutes, they should be subsumed under total R&D. Given that the theory is ambiguous on this issue, *ad hoc* disaggregation may be associated with model specification bias. Therefore, we argue for further modelling and estimation work to identify the correct functional form and obtain more reliable estimates for the productivity effects of different R&D types, including publicly-funded R&D.

7. The effects of different *estimations methods on elasticities* (columns 1 and 2 in Table 3) are relative to the excluded category of pooled/total OLS estimators. This specification enables us to compare time-differenced estimators for temporal elasticities with OLS estimators in the level dimension. We report that estimators based on *time-differencing* (including *first-differenced, long-differenced* and *within estimators*) yield lower elasticity estimates compared to pooled OLS. This finding confirms the attenuation bias in time-differenced estimators reported in primary studies and existing reviews. However, we are also able to rank the downward bias associated with time-differenced data, which appears to be the largest in the first-differenced estimations followed by long-differenced and within estimations. This is in line with Draca et al. (2007), who report that the attenuation bias associated with longer differences is smaller than that associated with shorter differences as the transitory shocks are averaged out in the former.

8. We find that *GMM* and *other IV estimators* yield smaller elasticity estimates on average compared to OLS estimators. This finding suggests that R&D investments and productivity may be responding to unobserved shocks in the same direction, leading to upward bias in the OLS estimates of elasticity.

4.3.2 Moderately-consistent or insignificant sources of heterogeneity

1. The effects of *time and/or industry dummies* on reported estimates are uncertain. The effect is positive when industry dummies are included in the private rate-of-return models, but negative in within-firm social rate-of-return models. In addition, the effect of including time dummies is positive in private elasticity models, but negative in private rate-of-return models. These conflicting findings resonate with Hall et al. (2010) who report that the effect of industry dummies on primary-study estimates is ambiguous in the temporal dimension. In addition, time dummies may capture unobserved variations across time, but these variations may be related to actual relationship between R&D and productivity.

2. The effect of *controlling for spillovers* through an additional term in the production function is significant only when the sample consists of firm-level (private) rates of return in column 3 of Table 3. This finding reinforces our concern about the absence of difference between private and social returns – particularly when the latter are based on rate-of-return estimations. Despite the attenuation effect of controlling for spillovers through an additional term, the private rates of returns are not smaller than social rates returns - as predicted by the theoretical model. Hence, we reiterate that the absence of difference between private and within-industry social returns calls for further modeling, identification and estimation effort to disentangle private and social returns to R&D.

3. In the rate-of-return pools (columns 3 and 4), the data is time-differenced and the reference category is long-differenced estimators. There is evidence of moderate consistency that *IV estimators* are associated with larger rate-of-return estimates compared to estimations based on long-differenced data. The attenuation (downward) bias in rate-of-return estimates is smaller when the data is long-differenced, but it may not be eliminated altogether because time-differencing is based on the assumption that a given investment has a constant weight over its estimated life. This assumption does not hold if there are gestation lags beyond the period over which time-differencing is carried out. Hence long-differencing can still be a source of downward bias when there is mismatch between the gestation lags and the period over which long-differencing is carried out. Therefore, the positive coefficient on the IV estimators indicate that the latter may be correcting for the residual downward bias in the long-differenced estimations (the reference category). Nevertheless, the positive coefficient on the IV dummy may also reflect upward bias in IV estimators, which are reported to perform less satisfactorily with time-differenced data (Griliches and Mairesse, 1995). Hall et al. (2010) indicate that only system GMM estimates based on both level and difference equations yield precise estimates, but the GMM dummy turns out to be insignificant in the rate-of-return pool. Therefore, we reiterate the need to identify and model the lag structure of the returns to R&D with a obtain estimates of long-term returns. This is more feasible with industry or finely-grained product line data over long time horizons compared to firm-level data (Hall et al., 2010).

4. The meta-regression results also indicate that *a range of moderating factors are insignificant* in explaining the variation in the evidence base. One such factor concerns correction for double counting. Schankerman (1981) demonstrates that elasticities and rates-of-return calculated indirectly are biased downwards if primary studies do not correct for double counting. The latter requires deducting the capital-related part of the R&D expenditures and the number of R&D personnel from capital and labour. However, the downward bias occurs mainly in the level dimension (Hall et al., 2010).

In our samples, correction for double counting is insignificant and this may be due to pooling together of the elasticity estimates in the level and temporal dimensions.[13]

Another set of moderating variables that remain inconsistent or insignificant in this meta-analysis relate to sample characteristics. For example, we do not find systematic differences between elasticities or rates of return based on different country data. Although US data is associated with larger elasticity and French data with larger rate-of-return estimates, the relationship does not carry across evidence pools. Similarly, we find that relatively more recent data with a mid-point after 1980 is not associated with a significant effect on either elasticities or rates return, compared to earlier data.[14]

5. Conclusions

The work on R&D and productivity has made significant contributions to existing knowledge by producing a wealth of evidence and addressing a wide range of measurement, modelling and estimation challenges. However, our analysis suggests that some of the challenges identified by the pioneering contributors still lie ahead. As Zvi Griliches concludes in a posthumously published volume:

> "There is much that remains to be learned about productivity, especially in understanding its economic determinants and its economic and social consequences... But in the pursuit of this knowledge, we should always remember that we can see farther than our predecessors because we stand on their shoulders." (Griliches, 2001: 120).

Embracing this perspective, we offer the following conclusions.

Meta-analysis is a useful method for synthesizing the evidence on returns to R&D investment and for identifying the sources of variation in the evidence base. The method enables us to report that the private and within-industry social returns to R&D are positive but smaller and more heterogeneous than what is reported in prior reviews. The upward bias in the latter is due to reliance on selected samples. This potential source of bias in prior reviews is combined with limited attention to the extent of heterogeneity in the evidence base.

Two of the findings in this study call for caution in the interpretation of the estimates for returns to R&D investment. The first concerns the lack of difference between *private* and

[13] Weiser (2005) also pools both dimension and reports that control for double counting corrects for downward bias in elasticity estimates. However, this result is based on studies that only use value added as the measure of output.

[14] We have conducted 16 more estimations with different dummies for different data mid-points, including more recent data mid-points after 1985 and 1990 and older data mid-points before 1975 and 1965. The choice of the data mid-point has no significant effect on primary-study estimates in 14 estimations. In the remaining two estimations, the effect of using data with mid-point before 1965 was negative in one and positive in the other. Hence, we reiterate our argument that the time frame for the data is not a source of systematic variation in the evidence base. These estimations are not reported here to save space, but can be provided on request.

within-industry social returns to R&D. The second finding indicates that the gross private returns are smaller than the commonly-assumed depreciation rate for R&D capital. Whereas the first runs against the prediction of higher within-industry social returns to R&D from the underlying theoretical model, the latter implies negative or very small net private returns to R&D in the face of continued firm investments in R&D and public policy support for the latter. We are of the view that these anomalies reflect the range of modeling, data and estimation challenges that the leading contributors to the field have been aware of. Although the existing reviews acknowledge these issues as sources of variation in the evidence base, they stop short of spelling out the extent to which they may constrain the informational content of the estimates reported in primary studies.

Our reading of the literature and the synthesized evidence suggests four possible causes for the wedge between existing estimates and 'true' returns to R&D. First, the lack of firm-level price deflators implies that private returns to R&D may reflect a mixture of both market power and 'true' productivity effects. Secondly, the elasticity and rate-of-return estimates are likely to be biased downward as they are usually based on contemporaneous estimations that do not take account of lagged effects. Third, the private rate-of-return estimates are likely to be biased downward as they are based on gross R&D intensity. Finally, the perpetual inventory method used to construct R&D capital may lead to upward bias in the elasticity estimates as it accords the same weight to a given increase in R&D investment irrespective of the levels of R&D capital in the preceding period(s).

Further insights we distil from the multivariate meta-regression includes the following: (a) it is necessary to control for scale effects in the R&D-productivity relationship as there is evidence that returns to R&D may vary with different levels of R&D intensity; (b) further modelling and estimation work is required to identify the correct functional form and obtain more reliable estimates for the productivity effects of different R&D types, including publicly-funded R&D; (c) the downward bias in time-differenced estimates is the largest in first-differenced estimations followed with long-differenced and within estimations; and (d) IV estimations may not be effective in tackling endogeneity when the estimation is based on time-differenced data.

We offer some suggestions for future research and evidence-based policy. First, further work is required to clarify whether different R&D types such as privately-funded *versus* publicly-funded business R&D, intramural *versus* extramural R&D, basic *versus* applied R&D, etc. are complements or substitutes; and whether the productivity effects of R&D are subject to scale effects. Further work is also required to model the firm-specific technology as a function of market structure and distance to the technology frontier with a view to differentiate between 'true' and revenue productivity effects. The Schumpeterian models of competition, innovation and growth discussed in Aghion et al. (2014) are pertinent in this respect.

A third set of suggestion ties in with recent contributions by Bloom et al. (2013), who draw attention to two issues that may distort the informational content of the estimates for private and social returns to R&D capital. One relates to the 'reflection problem' discussed in Manski (1993), who draw attention to the fact that social returns to R&D may

be upward-biased if all firms in the industry increase their R&D investment in reaction to an exogenous increase in new research opportunities. In this case the social return estimate picks up not only the effect of spillovers but also that of own R&D on productivity. To address this issue and estimate causal effects either matching techniques or external instruments are required. Another issue concerns the separation of the positive technology spillover effects from the negative effects on private returns due to product-market rivalry (or market-stealing effect) among firms sharing similar product-market positions. Bloom et al (2013) address this issue by identifying the firm's position in the product market space using information on the distribution of its sales activity across different industries. This specification allows for distinguishing between technology and product-market spillovers.

Fourth, further econometric work with long industry data can help enhancing the informational content of the estimated private or social returns to R&D for several reasons.

- It is desirable to increase the number of industry-level studies as the latter are fewer than those based on firm-level data.
- Industry-level data is found to be associated with lower levels of residual heterogeneity and the latter has been reduced significantly when we control for the dimensions of the research field.
- Industry-level data with a long time dimension allows for identifying the lag structure of the R&D capital and for estimating both short and long-run effects of R&D on productivity. Autoregressive distributed lag (ARDL) estimations (see Añón Higón, 2007) or Markov chain approaches (Doraszelski and Jaumandreu, 2013) can be applied in this line of research. Pesaran and Shin (1999) demonstrate that the ARDL modeling framework can yield consistent estimates of the long-run effects even when the regressors are potentially endogenous. On the other hand, the Markov chain modelling of productivity allows for recovering the full distribution of the elasticity estimates, but need explicit tests for homogeneity and Markov property, which are often missing in empirical studies of the field.
- As demonstrated in Eberhardt et al. (2013), industry-level data with a sufficiently long time dimension also allows for accounting for cross-sectional dependence between industries and reduce the risk of upward bias in the estimates of social returns based on weighted measures of R&D (knowledge) spillovers.

Fifth, we would like to indicate that the transition to capitalisation of R&D expenditures has the potential of reducing measurement errors and/or differences between data quality. It also has the potential of generating industry-level data with long time periods and with finer disaggregation along technology classes and product lines.

Finally, we think that the prevalent policy stance in favour of public support for R&D investment may be too sanguine given the extent of heterogeneity in the evidence base and the limitations to the informational content of the existing evidence. Therefore, we

suggest that public support for R&D investment should be conditional and time-variant in order to: (a) prioritise R&D projects with better scope for generating social returns; and (b) take account of new evidence from the evolving modeling, estimation and evidence synthesis techniques.

References (excluding primary studies listed separately)

Aghion, P., U. Akcigit, and P. Howitt (2014). What do we learn from Schumpeterian growth theory? In P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, *Vol. 2*. Amsterdam: Elsevier, pp. 515-563.

Arrow, K.J. (1962), Economic Welfare and the Allocation of Resources for Invention, in R.R. Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*, National Bureau of Economic Research, Conference Series, Princeton: Princeton University Press, 609-625.

Arrow, K.J. (1996), The Economics of Information: An Exposition, *Empirica* 23(2), 119-128.

Bartelsman, E.J. (1990). R&D spending and manufacturing productivity: An empirical analysis. Washington, DC*: Federal Reserve Board of Governors Finance and Economics Discussion* Series 122.

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

Bollen, K. A. and R. W. Jackman (1990). Regression diagnostics: An expository treatment of outliers and influential cases, in J. Fox and J. S. Long *(eds.), Modern Methods of Data Analysis*, Newbury Park, CA: Sage, 257–291.

Card, D. and A. B. Krueger (1995). Time-series minimum-wage studies: a meta-analysis. *American Economic Review*, 85(2), 238–243.

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

Costa-Font, J., A. McGuire and T. Stanley (2013). Publication selection in health policy research: the winner's curse hypothesis. *Health Policy,* 109(1), 78–87.

Crépon, B., E. Duguet, and J. Mairesse (1998). Research, innovation, and productivity: An econometric analysis at the firm level. *Economics of Innovation and New Technology* 7, 115-156.

Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Hoboken, NJ: Wiley.

Dickersin, K., and Y. I. Min (1993). Publication bias: the problem that won't go away. *Annals of the New York Academy of Sciences*, *703*(1), 135-148.

DiMasi, J. A. and H. G. Grabowski (2012). R&D costs and returns to new drug development: a review of the evidence. In Tufts Centre for the Study of Drug Development. *March/April CSDD Impact Report. The Oxford handbook of the economics of the biopharmaceutical industry.* Oxford: Oxford University Press, 21-46.

Doraszelski, U. and J. Jaumandreu (2013). R&D and productivity: Estimating endogenous productivity, *The Review of Economic Studies*, 80(3), 1338-1383.

Doucouliagos, H. (2011). How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics. Deakin Faculty of Business and Law, Working Papers, no. SWP 2011/5. https://www.deakin.edu.au/__data/assets/pdf_file/0003/408576/2011_5.pdf

Doucouliagos, H., and P. Laroche (2009). Unions and profits: A meta-regression analysis. *Industrial Relations: A Journal of Economy and Society*, 48(1), 146-184.

Doucouliagos, H., and T. D. Stanley (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47(2), 406-429.

Doucouliagos, H., and T. D. Stanley (2013). Theory competition and selectivity: Are all economic facts greatly exaggerated? *Journal of Economic Surveys*, 27, 316-39.

Draca, M., R. Sadun and J. Van Reenen (2007). Productivity and ICT: A review of the evidence. In R. Mansell (Ed.), *The Oxford Handbook of Information and Communication Technologies*. Oxford University Press, Oxford and New York, pp. 100-147.

Egger, M., G. D Smith, M. Schneider, and C. Minder (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 316, 629-34.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1), 394-425.

Griliches, Z. (1973). Research expenditures and growth accounting. In: *Science and Technology in Economic Growth*, R.B. Williams (ed.). John Wiley and Sons, New York.
Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*, 10(1), 92-116.
Griliches, Z. (1992). The Search for R&D Spillovers. *The Scandinavian Journal of Economics*, 94(Supplement), 29-47.
Griliches, Z. (1994). Productivity, R&D and the data constraint. *American Economic Review*, 84(1), 1-23.
Griliches, Z. (1995). R&D and productivity: Econometric results and measurement issues, in P. Stoneman (ed.), *Handbook of the Economics of Innovation and Technical Change*, Blackwell Handbooks in Economics.
Griliches, Z. (2001). *R&D, Education, and Productivity: A Retrospective*. Cambridge, Mass. And London: Harvard University Press.
Griliches, Z. and J. Mairesse (1995). Production functions: The search for identification. *NBER Working Papers* No. 5067. National Bureau of Economic Research.
Hall, B. H. (1993). Industrial research during the 1980s: Did the rate of return fall? *Brookings Papers on Economic Activity*, Micro (2), 289-344.
Hall, B. H. (1996). The private and social returns to research and development. In Bruce L.R. Smith and Claude E. Barfield (eds.), *Technology, R&D, and the Economy*, Brookings Institution, Washington D.C.
Hall, B. H. (2007). Measuring the returns to R&D: The depreciation problem. *NBER Working Papers*, No. w13473. National Bureau of Economic Research.
Hall, B. H. (2011). Innovation and productivity. *NBER Working Papers*, No. w17178. National Bureau of Economic Research.
Hall, B. H., J. Mairesse and P. Mohnen (2010). Measuring the returns to R&D. In B. H. Hall and N. Rosenberg (eds.), *Handbook of the Economics of Innovation*, vol. 2, Elsevier, New York.
Harbord, R. M and J. P. Higgins (2008). Meta-regression in Stata. *Stata Journal*, 8(4), 493-519.
Higgins, J. P., S. G. Thompson, J. J. Deeks and D. G. Altman (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557-560.
Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Jama*, *294*(2), 218-228.
Jaffe, A. B. (1988). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value. *American Economic Review*, 76(5), 984–1001.
Keller, W. (2004). International technology diffusion. *Journal of Economic Literature*, 42 (3), 752–782.
Klette, T. J. (1994). R&D, scope economies, and company structure: a 'not so fixed effect' model of plant performance. Oslo, Norway: Central Bureau of Statistics *Discussion Paper* No. 120.
Klette, T. J. (1996). The accumulation of R&D-Capital and the dynamic performance of manufacturing firms. Oslo, Norway: Central Bureau of Statistics.
Mairesse, J. and Z. Griliches (1988). Heterogeneity in panel data: Are there stable production functions? NBER Working Paper 2619, National Bureau of Economic Research, Inc.
Mairesse, J., and J. Jaumandreu (2005). Panel-data estimates of the production function and the revenue function: What difference does it make? *Scandinavian Journal of Economics,* 107(4), 651-672.
Mairesse, J., and P. Mohnen (1994). R&D and productivity growth: What have we learned from econometric studies? *Eunetic Conference on Evolutionary Economics of Technological Change: Assessment of Results and New Frontiers*, 817-888, Strasbourg.
Mairesse, J. and M. Sassenou (1991). R&D and productivity: a survey of econometric studies at the firm level. *STI Review*, 8, 9-46.
Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60 (3), 531–542.
McCulloch, C. E., S. R. Searle, and J. M. Neuhaus (2008). *Generalized, Linear, and Mixed Models (2nd ed)*. Hoboken, NJ: Wiley.

Minasian, J. R. (1969). Research and development, production functions, and rates of return. *American Economic Review*, 59(2), 80-85.

Møen, J., and H.S. Thorsen (2015). Publication bias in the returns to R&D literature. *Journal of the Knowledge Economy*. http://link.springer.com/article/10.1007/s13132-015-0309-9

Moreno, S.G., A. J. Sutton, A. Ades, T. D. Stanley, K. R. Abrams, J. L. Peters, and N. J. Cooper (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology*, 9(2), 1–17.

Pammolli, F., L. Magazzini and M. Riccaboni (2011). The productivity crisis in pharmaceutical R&D. *Nature Reviews: Drug Discovery* 10(6), 428-438.

Pesaran, M.H. and Y. Shin (1999). An autoregressive distribution lag modelling approach to cointegration analysis. In: S. Strom (Ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. Cambridge University Press, Cambridge.

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

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359-1366.

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys,* 19(3), 309-45.

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

Stanley, T.D. and H. Doucouliagos (2007). Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression, Deakin University Economics *Working Paper* No. 2007_11; 2007.

Stanley, T.D. and H. Doucouliagos (2012). *Meta-regression Analysis in Economics and Business*. London and New York: Routledge.

Stanley, T.D. and H(C) Doucouliagos (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60-78.

Stanley, T.D. and S. B. Jarrell (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys,* 3(2), 161-170.

Stanley, T. D., H. Doucouliagos, M. Giles, J.H. Heckemeyer, R. Johnston, , P. Laroche, . . . K. Rost (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, 27(2), 390-394.

Sterling, T.D., W.L. Rosenbaum and J.J. Weinkam (1995). Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician* 49(1): 108–112.

Sterne, J.A. and R.M. Harbord (2004). Funnel plots in meta-analysis. *The Stata Journal*, 4(2), 127–141.

Terleckyj, N. E. (1974). *Effects of R&D on the productivity growth of industries: An exploratory study* (No. 140). Washington, DC: National Planning Association.

Wieser, R. (2005). Research and development, productivity and Spillovers: Empirical evidence at the firm level. *Journal of Economic Surveys*, 19(4), 587–621.

Primary studies included in the meta-analysis

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

Aldieri L, M. Cincera, A. Garofalo, and C. P. Vinci (2008). Micro evidence of the effects of R&D on labour productivity for large international R&D firms, *International Journal of Manpower*, 29(3), 198-215.

Añón Higón, D. (2007). The impact of R&D spillovers on UK manufacturing TFP: A dynamic panel approach, *Research Policy*, 36(7), 964-979.

Ballot, G., F. Fakhfakh and E. Taymaz (2006). Who benefits from training and R&D: The firm or the workers?, *British Journal of Industrial Relations*, 44(3), 473-495.

Bartelsman, E. J. (1990). R&D spending and manufacturing productivity: An empirical analysis. Washington, DC*: Federal Reserve Board of Governors Finance and Economics Discussion* Series 122.

Bartelsman E. J., G. van Leeuwen, and H. Nieuwenhuijsen (1996). R&D and productivity growth: evidence from firm-level data for the Netherlands. Netherlands Official Statistics, 11(Autumn), 52-69.

Blanchard, P., J-P Huiban and P. Sevestre (2005). R&D and productivity in corporate groups: An empirical investigation using a panel of French firms, *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, No. 79/80, 461-485.

Boler, E. A., A. Moxnes and K-H Ulltveit-Moe (2012). Technological change, trade in intermediates and the joint impact on productivity, *C.E.P.R. Discussion Papers.* No. 8884.

Bond S., D. Harhoff and J. van Reenen (2002). Corporate R&D and productivity in Germany and the United Kingdom, *Centre for Economic Performance Discussion Paper*s, LSE. No. 595.

Bonte, W. (2003). R&D and productivity: Internal vs. external R&D - Evidence from West German manufacturing industries, *Economics of Innovation & New Technology*. 12(4): 343-360.

Branstetter, L. (1996). Are knowledge spillovers international or intra-national in scope? Microeconometric evidence from the U.S. and Japan, NBER *Working Paper Series*. No. 5800.

Cameron, G., J. Proudman and S. Redding (2005). Technological convergence, R&D, trade and productivity growth, *European Economic Review*, 49(3), 775-807.

Cincera, M. (1998). *Economic and Technological Performances of International Firms.* PhD Thesis Presented to the Faculty of Social, Political and Economic Sciences, Free University of Brussels, 1998.

Clark, K. and Z. Griliches (1998). Productivity growth and R&D at the business level: Results from the PIMS data base. In: *R&D and Productivity: The Econometric Evidence*. University of Chicago Press, 134 - 156.

Cunéo, P. and J. Mairesse (1984). Productivity and R&D at the firm level in French manufacturing. In: Z. Griliches (ed.), *R&D, Patents and Productivity*. Chicago, IL: University of Chicago Press, 375- 392.

Doraszelski, U. and J. Jaumandreu (2013). R&D and productivity: Estimating endogenous productivity, *The Review of Economic Studies*, 80(3), 1338-1383.

Eberhardt, M., C. Helmers and H. Strauss (2013). Do spillovers matter when estimating private returns to R&D? *Review of Economics and Statistics*, 95(2), 436-448.

Frantzen, D. (2002). Intersectoral and international R&D knowledge spillovers and total factor productivity, *Scottish Journal of Political Economy*, 49(3), 280-303.

Goto, A. and K. Suzuki (1989). R &D capital, rate of return on R&D investment and spillover of R&D in Japanese manufacturing industries, *The Review of Economics and Statistics*, 71(4), 555-564.

Griffith, R., R. Harrison, and J. van Reenen (2006). How special is the special relationship? Using the impact of U.S. R&D spillovers on U.K. firms as a test of technology sourcing, *American Economic Review*, 96(5), 1859-1875.

Griffith, R., S. Redding, and J. van Reenen (2004). Mapping the two faces of R&D: Productivity growth in a panel of OECD manufacturing industries, *Review of Economics and Statistics,* 86(4), 883-895.

Griliches, Z. (1980a). Returns to research and development expenditures in the private sector. In John W Kendrick; Beatrice N Vaccara (eds), *New Developments in Productivity Measurement*. University of Chicago Press, 419-462.

Griliches, Z. (1980) R&D and the Productivity Slowdown. *American Economic Review,* 70(2), 343-348.

Griliches, Z. (1998). Productivity, R&D, and basic research at the firm level in the 1970s. In Z. Griliches (ed.), *R&D and Productivity: The Econometric Evidence*. University of Chicago Press, 82 - 99.
Griliches, Z. and F. R. Lichtenberg (1984). Interindustry technology flows and productivity growth: A re-examination, *Review of Economics & Statistics* 66, 325-329.
Griliches, Z. and J. Mairesse (1981). Productivity and R&D at the firm level, NBER *Working Papers*, no. 826.
Griliches, Z. and J. Mairesse (1991a). R&D and productivity growth: Comparing Japanese and U.S. manufacturing firms. In Hulten, C. R. (ed.), *Productivity Growth in Japan and the United States*. University of Chicago Press, 317 - 348.
Griliches Z. and J. Mairesse (1991b). Comparing productivity growth: An exploration of French and U.S. industrial and firm data. In de Ménil, G. and R. J. Gordon (eds), *International Volatility and Economic Growth: The First Ten Years of The International Seminar on Macroeconomics*. Elsevier Science Publishers B.V., 45 - 82.
Hall, B. H. and J. Mairesse (1995). Exploring the relationship between R&D and productivity in French manufacturing firms, *Journal of Econometrics* 65, 263-293.
Hall, B. H., E. Mansfield and A. B. Jaffe (1993). Industrial research during the 1980s: Did the rate of return fall?, *Brookings Papers on Economic Activity*. Microeconomics. 1993(2), 289-344.
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.
Harhoff, D. (1994). R&D and productivity in German manufacturing firms. ZEW *Discussion Papers*, No. 94 – 01.
Harhoff, D. (2000). R&D Spillovers, Technological Proximity, and Productivity Growth - Evidence From German Manufacturing Firms, *Schmalenbach Business Review*, 52(3), 238-260.
Heshmati, A. and K. Hyesung (2011). The R&D and productivity relationship of Korean listed firms, *Journal of Productivity Analysis*, 36(2), 125-142.
Hsing, Y. and W-J Lin (1998). R&D spending and employee productivity in the United States, *International Journal of Management* 15: 79.
Kafouros, M. I. (2005). R&D and productivity growth: Evidence from the UK, *Economics of Innovation & New Technology*, 14(6), 479-497.
Klette, T. J. (1991). On the importance of R&D and ownership for productivity growth. Evidence from Norwegian micro-data 1976-85, *Central Bureau of Statistics Discussion Paper*. No. 60.
Kwon, H. U. and T. Inui (2003). R&D and productivity growth in Japanese manufacturing firms, *Economic and Social Research Institute* (*ESRI) Discussion Paper Series*. No. 44.
Lehto, E. (2007). Regional impact of Research and Development on productivity, *Regional Studies,* 41(5), 623-638.
Lichtenberg, F. R. and D. Siegel (1991). The impact of R&D investment on productivity: New evidence using linked R&D-LRD data, *Economic Inquiry*, 29(2), 203.
Link, A. N. (1981). Basic research and productivity increase in manufacturing: Additional evidence, *American Economic Review*, 71(5), 1111-1112.
Link, A. N. (1983). Inter-firm technology flows and productivity growth, *Economics Letters,* 11(1), 179-184.
Lokshin, B., R. Belderbos and M. Carree (2008). The productivity effects of internal and external R&D: Evidence from a dynamic panel data model, *Oxford Bulletin of Economics & Statistics*, 70(3), 399-413.
Los, B. and B. Verspagen (2000). R&D spillovers and productivity: Evidence from U.S. manufacturing microdata, *Empirical Economics*, 25(1), 127-148.
Mairesse J., and B. H. Hall (1996). Estimating the productivity of research and development: An exploration of GMM methods using data on French and United States manufacturing firms. NBER *Working Paper Series*. No. 5501.
Mate-García J. J. and J. M. Rodriguez-Fernandez (2008). Productivity and R&D: An econometric evidence from Spanish firm-level data, *Applied Economics*, 40(14), 1827-1837.
Medda, G., C. Piga and D. Siegel (2003). On the relationship between R&D and productivity: A treatment effect analysis. : Fondazione Eni Enrico Mattei, *Nota Di Lavoro* 34.

O'Mahony, M. and M. Vecchi (2009). R&D, knowledge spillovers and company productivity performance, *Research Policy*, 38(1), 35-44.
Odagiri, H. (1983). R & D expenditures, royalty payments, and sales growth in Japanese manufacturing corporations. *The Journal of Industrial Economics*, 32(1), 61-71.
Odagiri H, and H. Iwata (1986). The impact of Research and Development on productivity increase in Japanese manufacturing companies, *Research Policy*, 15(1), 13-19.
Ortega-Argiles, R., M. Piva, L. Potters and M. Vivarelli (2010). Is corporate R&D investment in high-tech sectors more effective?, *Contemporary Economic Policy*, 28(3), 353–365.
Rogers, M. (2010). R&D and productivity: Using UK firm-level data to inform policy, *Empirica*, 37(3), 329-359.
Schankerman, M. (1981). The effects of double-counting and expensing on the measured returns to R&D, *Review of Economics and Statistics,* 63(3), 454-458.
Scherer, F. M. (1982). Interindustry technology flows and productivity growth, *Review of Economics and Statistics,* 64(4), 627-634.
Scherer, Frederic M. Concentration, R&D and productivity change. *Southern Economic Journal,* 50(1), 221-225.
Smith, V., M. Dilling-Hansen, T. Eriksson, E. S. Madsen (2004). R&D and productivity in Danish firms: Some empirical evidence, *Applied Economics*, 36(16), 1797-1806.
Sorensen, A., H. C. Kongsted and M. Marcusson (2003). R&D, public innovation policy, and productivity: The case of Danish manufacturing, *Economics of Innovation and New Technology*, 12(2), 163-78.
Sterlacchini, A. (1989). R&D, innovations, and total factor productivity growth in British manufacturing, *Applied Economics*, 21(11): 1549.
Sveikauskas, L. (1981). Technology inputs and multifactor productivity growth, *Review of Economics and Statistics*, 63(2), 275-282.
Terleckyj, N. (1980). Direct and indirect effects of industrial research and development on the productivity growth of industries. In J. Kendrick and B. Vaccara (eds.), *New Developments in Productivity Measurement and Analysis*. UMI/NBER publication, 357-386.
van Meijl, H. (1997). Measuring intersectoral spillovers: French evidence, *Economic Systems Research*, 9(1), 25-46.
Verspagen, B. (1995). R&D and productivity: A broad cross-section cross-country look, *Journal of Productivity Analysis*, 6(2), 117-35.
Verspagen, B. (1997). Estimating international technology spillovers using technology flow matrices, *Weltwirtschaftliches Archiv*, *133*(2), 226-248.
Wakelin, K. (2001). Productivity growth and R&D expenditure in UK manufacturing firms. *Research Policy*, 30(7), 1079-1090.
Wolff, E. N., and M. I. Nadiri (1993). Spillover effects, linkage structure, and Research and Development, *Structural Change and Economic Dynamics,* 4(2), 315-331.

Table A1.1: Overview of study characteristics: Elasticity estimates at the firm level

Study	Publication type	Country	Dependent variable	Estimation method	Data period	Number of firms / industries	Median estimate	Median t-value	Reported estimates
Aiello and Cardamone (2005)	Journal article	OECD Other	Value Added (VA)	First Diff., GMM	1995-2000	1017	0.069	5.410	4
Aldieri et al (2008)	Journal article	OECD Other, US	Sales	First Diff., GMM, Within	1988-1997	116-465	0.255	1.945	16
Ballot et al (2006)	Journal article	OECD Other, France	VA	GMM, Pooled OLS	1987-1993	101-268	0.059	8.890	10
Bartelsman (1990)	Working paper	US	Output	Pooled OLS, Within	1956-1988	n.a.	0.052	1.954	12
Bartelsman et al (1996)	Report	OECD Other	Output, VA	Total OLS, Long Diff.	1985-1993	159-436	0.055	2.526	22
Blanchard et al (2005)	Journal article	France	VA	GMM, Pooled OLS, Within	1994-1998	793-3141	0.115	2.245	7
Boler et al (2012)	Working paper	OECD Other	Sales	GMM, Pooled OLS	1997-2005	850-850	0.030	3.000	5
Bond et al (2002)	Working paper	UK, Germany	Sales	GMM, Pooled OLS, Within	1987-1996	234-239	0.053	1.810	12
Branstetter (1996)	Working paper	OECD Other, US	Sales	Long Diff.	1985-1989	205-209	0.188	1.528	2
Cincera (1998)	Thesis	OECD Other	Sales	GMM, Pooled OLS, Within, Long Diff., First Diff., Between	1987-1994	101-2445	0.230	7.442	58
Cuneo and Mairesse (1984)	Working paper	France	VA	Pooled OLS, Within	1972-1977	84-182	0.130	3.000	20
Doraszelski and Jaumandreu (2013)	Journal article	OECD Other	Output, VA	GMM	1991-1999	131-304	0.018	1.563	18
Griffith et al (2006)	Journal article	UK	VA	GMM, Pooled OLS, IV	1990-2000	89-188	0.024	2.116	14
Griliches (1980)	Book chapter	US	Sales, VA	Total OLS, Between	1957-1965	31-883	0.075	2.875	59
Griliches (1998)	Journal article	US	Sales, VA	Total OLS, First Diff.	1966-1977	386-911	0.117	5.235	17
Griliches and Mairesse (1981)	Working paper	US	Sales	Pooled OLS, Within, Between	1966--1977	20-133	0.143	4.848	32

Griliches and Mairesse (1991b)	Book chapter	OECD Other	Sales	First Diff.	1973-1978	528-528	0.025	0.834	2
Hall (1993)	Journal article	US	Sales	Pooled OLS, Long Diff, First Diff.	1964-1990	200-1600	0.030	1.375	85
Hall and Mairesse (1995)	Journal article	US	VA	Pooled OLS, Long Diff, First Diff., Within	1980-1987	197-340	0.093	2.431	56
Harhoff (1994)	Working paper	Germany	Sales	Pooled OLS, Long Diff, First Diff., Within	1977-1989	188-443	0.116	4.737	59
Harhoff (2000)	Journal article	Germany	Sales	Long Diff.	1977-1989	439-439	0.068	2.429	5
Hsing (1998)	Journal article	US	Sales	Total OLS	1994	30-30	0.204	1.892	2
Kafouros (2005)	Journal article	UK	Sales	Pooled OLS	1989-2002	19-78	0.040	4.316	17
Kwon and Inui (2003)	Working paper	OECD Other	VA	Pooled OLS, Long Diff, First Diff., Within	1995-1998	400-3830	0.052	2.650	82
Lehto (2007)	Journal article	OECD Other	VA	Pooled OLS, Within, Between, IV	1987-1998	1362-2171	0.031	5.000	18
Los and Verspagen (2000)	Journal article	US	Sales	First Diff., Within, Between	1977-1991	211-680	0.014	1.220	11
Mairesse and Hall (1996)	Working paper	US, France	Sales, VA	Pooled OLS, First Diff., Within, GMM, IV	1981-1989	381-1232	0.031	1.000	63
Ortega-Argiles et al (2010)	Journal article	OECD Other	VA	Pooled OLS, Between	2000-2005	532-532	0.110	5.359	8
O'Mahony and Vecchi (2000)	Book chapter	OECD Other, US	Sales	Pooled OLS, First Diff.	1993-1997	160-783	0.168	5.026	9
O'Mahony and Vecchi (2009)	Journal article	OECD Other, US	Sales	GMM, First Diff.	1988-1997	285-6803	0.124	2.182	9
Rogers (2010)	Journal article	UK	VA	Pooled OLS	1989-2000	86-719	0.131	11.340	11
Schankerman (1981)	Journal article	US	VA	Total OLS	1963	31-419	0.082	3.592	18
Smith et al (2004)	journal article	OECD Other	VA	Total OLS, First Diff.	1995-1997	109-378	0.090	2.701	10
Total									**773**

Table A1.2: Overview of study characteristics: Elasticity estimates at the industry level

Study	Publication type	Country	Dependent variable	Estimation method	Data period	Number of firms / industries	Median estimate	Median t-value	Reported estimates
Añón Higón (2007)	Journal article	UK	Output	Panel Coin., CF Framework	1970-1997	18-18	0.313	2.617	4
Bonte (2003)	Journal article	Germany	VA	First Diff, Long Diff., Within, Between	1980-1993	26-26	0.008	0.224	6
Eberhardt et al (2013)	Journal article	OECD Other	VA	Pooled OLS, First Diff, GMM, CF Frame.	1980-2005	84-144	0.037	0.960	17
Frantzen (2002)	Journal article	OECD Other	Output	Pooled OLS	1972-1994	308-308	0.152	13.170	7
Goto and Suzuki (1989)	Journal article	OECD Other	Output	First Diff.	1976-1984	13-99	0.250	1.960	21
Griliches (1980a)	Journal article	US	Output, VA	Within	1959-1977	39-39	0.044	1.692	5
Ortega-Argiles et al (2010)	Journal article	OECD Other	VA	Pooled OLS, Between	1987-2002	15-15	0.062	3.936	8
Verspagen (1995)	Journal article	US, UK, France, Germany, OECD Other	Output	IV	1973-1988	15-15	0.019	0.430	55
Verspagen (1997)	Journal article	OECD Other	Output	Within, Between	1974-1992	22-22	0.081	4.190	12
Total									**135**

Table A1.3: Overview of study characteristics: Rate-of-return estimates at the firm level

Study	Publication type	Country	Dependent variable	Estimation method	Data period	Number of firms / industries	Median estimate	Median t-value	Reported estimates
Bartelsman et al (1996)	Report	OECD Other	Output, VA	Long Diff.	1985-1993	159-368	0.173	2.110	9
Cincera (1998)	Thesis	OECD Other	Sales	First Diff.	1987-1994	625-625	0.380	6.333	1
Clark and Griliches (1998)	Book chapter	US	Sales	First Diff.	1970-1980	924-924	0.190	3.800	6
Griliches and Mairesse (1991a)	Book chapter	OECD Other, US	Sales	First Diff.	1973-1980	406-525	0.285	2.519	6
Griliches and Mairesse (1991b)	Book chapter	OECD Other, US, France	Sales	First Diff.	1973-1978	185-528	0.120	1.727	13
Hall and Mairesse (1995)	Journal article	US	VA	Long /First Diff.	1980-1987	197-340	0.213	2.028	20
Harhoff (1994)	Working paper	Germany	Sales	Long / First Diff.	1977-1989	188-443	0.221	3.277	6
Heshmati and Hyesung (2011)	Journal article	OECD Other	VA	First Diff.	1986-2002	1200-1200	0.129	2.210	2
Klette (1991)	Working paper	OECD Other	Output	First Diff.	1976-1985	218-1268	0.108	4.154	20
Kwon and Inui (2003)	Working paper	OECD Other	VA	First Diff.	1995-1998	516-3830	0.232	5.300	2
Lichtenberg and Siegel (1991)	Journal article	US	Output	First Diff.	1972-1985	2000-2000	0.189	3.930	33
Link (1981)	Journal article	US	Output	Long Diff.	1973-1978	51-51	1.250	2.850	2
Link (1983)	Journal article	OECD Other	Output	Long Diff.	1975-1979	302-302	0.055	1.810	2
Lokshin et al (2008)	Journal article	OECD Other	VA	GMM, Within, Between	1996-2001	304-304	0.302	2.988	4
Mansfield (1980)	Journal article	US	VA	Long Diff.	1960-1976	16-16	0.105	1.850	25
Mate-García and Rodriguez-Fernandez (20	Journal article	OECD Other	VA	GMM	1993-1999	1312-1312	0.266	2.163	1
Medda et al (2003)	Working paper	OECD Other	Output	Long Diff.	1988-1997	2215-2268	0.327	4.316	2
Odagiri (1983)	Journal article	OECD Other	Sales	First Diff.	1966-1980	370-370	-0.110	0.521	2
Odagiri and Iwata (1986)	Journal article	OECD Other	Output	First Diff.	1966-1973	135-135	0.170	1.876	4
Rogers (2010)	Journal article	UK	VA	First Diff., GMM	1989-2000	86-719	0.205	2.060	18
Wakelin (2001)	Journal article	UK	Sales	First Diff.	1988-1992	85-170	0.265	1.275	14
Total									**192**

Table A1.4: Overview of study characteristics: Rate-of-return estimates at the industry level

Study	Publication type	Country	Dependent variable	Estimation method	Data period	Number of firms / industries	Median estimate	Median t-value	Reported estimates
Cameron et al (2005)	Journal article	UK	TFP	First Diff.	1970-1992	14-14	0.638	2.438	9
Griffith et al (2004)	Journal article	OECD Other	TFP	First Diff.	1974-1990	12-12	0.473	2.621	15
Griliches and Lichtenberg (1984)	Journal article	US	TFP	Long Diff.	1959-1978	193-193	0.233	2.515	20
Hanel (2000)	Journal article	OECD Other	TFP	First Diff.	1974-1989	22-22	0.152	1.735	8
Scherer (1982)	Journal article	US	TFP	Long Diff.	1948-1978	20-20	0.192	1.625	4
Scherer (1983)	Journal article	US	TFP	Long Diff.	1964-1978	87-87	0.364	1.605	4
Sterlacchini 1989	Journal article	UK	TFP	First Diff.	1954-1984	15-15	0.125	1.775	6
Sveikauskas (1981)	Journal article	US	TFP	Long Diff.	1959-1969	69-144	0.194	2.270	20
Terleckyj (1980)	Book chapter	US	TFP	Long Diff.	1948-1966	20-20	0.225	2.130	12
Verspagen (1995)	Journal article	OECD Other, France, Germany, UK, US	Output	IV	1973-1988	15-15	0.226	0.455	26
Wolff and Nadiri (1993)	Journal article	US	TFP	Long Diff.	1958-1977	19-50	0.181	2.310	14
van Meijl (1997)	Journal article	France	TFP	Long Diff.	1978-1992	30-30	0.080	1.640	15
Total									**153**

Table A2: Summary statistic for moderating variables by evidence pool

Evidence pools	Elasticity at firm level: 773 observations			Elasticity at industry level: 135 observations			Rate of return at firm level: 192 observations			Rate of return at industry level: 153 observations		
Moderating variables												
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Effect indicators												
t-value	5.778	-6.067	52.290	2.130	-7.059	18.970	2.781	-1.611	13.090	1.873	-0.810	5.270
Precision	76.341	1.455	999.900	30.352	1.087	142.854	16.088	1.135	77.037	9.735	0.335	27.737
Publication type												
Journal article	0.435	0	1	1	1	1	0.693	0	1	0.910	0	1
Model specification												
Control for spillovers	0.128	0	1	0.259	0	1	0.146	0	1	0.301	0	1
Control for capacity utilisation	n.o.	n.o.	n.o.	0.052	0	1	0.137	0	1	0.188	0	1
Industry dummies included	0.290	0	1	0.496	0	1	0.571	0	1	0.278	0	1
Time dummies included	0.577	0	1	0.170	0	1	0.344	0	1	0.278	0	1
Variable return to scales allowed	0.396	0	1	0.719	0	1	0.382	0	1	0.233	0	1
Correction for double counting	0.301	0	1	0.200	0	1	0.175	0	1	0.323	0	1
Measurement												
R&D capital is constructed with perpetual inventory method	0.695	0	1	1.000	1	1	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.
Output is measured as value added	0.199	0	1	0.126	0	1	0.245	0	1	n.o.	n.o.	n.o.
Sample characteristics												
Mid-point of data period is post-1980	0.719	0	1	0.400	0	1	0.495	0	1	0.361	0	1
Small firms	0.018	0	1	n.o.	n.o.	n.o.	0.014	0	1	0.000	0	0
French data	0.094	0	1	0.030	0	1	0.024	0	1	0.128	0	1
German data	0.091	0	1	0.089	0	1	0.028	0	1	0.023	0	1
UK data	0.062	0	1	0.074	0	1	0.151	0	1	0.135	0	1
US data	0.423	0	1	0.081	0	1	0.542	0	1	0.429	0	1
R&D-intensive firm/industry	0.195	0	1	0.215	0	1	0.033	0	1	0.045	0	1
Publicly-funded R&D	0.008	0	1	n.o.	n.o.	n.o.	0.042	0	1	0.045	0	1
Estimation method												
Common factor frame estimators	n.o.	n.o.	n.o.	0.104	0	1	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.
First-differenced estimation	0.135	0	1	0.170	0	1	0.613	0	1	0.286	0	1
General method of moments (GMM)	0.120	0	1	0.007	0	1	0.033	0	1	0.000	0	0
Instrumental variable (IV) estimators	0.010	0	1	0.407	0	1	n.o.	n.o.	n.o.	0.195	0	1
Long-differenced estimation	0.085	0	1	0.015	0	1	0.344	0	1	0.519	0	1
Within estimation	0.190	0	1	0.119	0	1	n.o.	n.o.	n.o.	n.o.	n.o.	n.o.

Note: n.o. indicates no observations for the moderating variable in the evidence pool.

Table A3: Multivariate meta-regression results: General models

Dependent variable: t-value	(1)	(2)	(3)	(4)
Precision	0.026	0.098**	0.153***	0.221***
	(0.027)	(0.044)	(0.055)	(0.084)
Journal article	-0.043**	pmc	-0.124**	-0.103
	(0.021)	pmc	(0.052)	(0.065)
Control for spillovers	0.017	-0.145	-0.044*	-0.024
	(0.016)	(0.181)	(0.024)	(0.016)
Control for capacity utilization	n.o.	0.010	-0.072	0.009
	n.o.	(0.043)	(0.047)	(0.030)
Industry dummies included	0.006	0.042**	-0.012	-0.006
	(0.004)	(0.017)	(0.035)	(0.041)
Time dummies included	0.012	-0.035	0.070	-0.093**
	(0.011)	(0.033)	(0.061)	(0.038)
Variable returns to scale allowed	-0.006	-0.001	-0.044	0.022
	(0.004)	(0.025)	(0.055)	(0.018)
Control for double counting	0.022***	0.001	-0.106*	0.008
	(0.004)	(0.012)	(0.058)	(0.126)
R&D capital constructed with	0.054**	pmc	n.o.	n.o.
Perpetual inventory method	(0.023)	pmc	n.o.	n.o.
Output measured as value added	0.049***	-0.164	0.012	0.029
	(0.005)	(0.178)	(0.031)	(0.021)
Data mid-point is post-1980	-0.004	0.167	0.005	0.004
	(0.008)	(0.180)	(0.019)	(0.055)
Small firms	-0.020**	n.o.	0.062	n.o.
	(0.008)	n.o.	(0.060)	n.o.
French data	0.008	0.012	0.020	0.105
	(0.012)	(0.021)	(0.112)	(0.082)
German data	0.014	-0.021	0.024	0.004
	(0.041)	(0.020)	(0.109)	(0.063)
UK data	0.004	0.015	0.038	-0.006
	(0.031)	(0.021)	(0.080)	(0.065)
US data	0.038***	0.004	0.004	-0.015
	(0.014)	(0.020)	(0.036)	(0.060)
R&D-intensive firm/industry	0.021**	0.089***	-0.010	0.023
	(0.009)	(0.016)	(0.084)	(0.087)
Publicly-funded R&D	-0.140***	n.o.	-0.169***	-0.299***
	(0.032)	n.o.	(0.043)	(0.030)
Common factor frame estimators	n.o.	0.147	n.o.	n.o.
	n.o.	(0.179)	n.o.	n.o.
First-differenced estimations	-0.053***	-0.032	0.091***	-0.007
	(0.005)	(0.032)	(0.032)	(0.094)
General method of moments	-0.026**	0.406	0.060	n.o.
estimators (GMM)	(0.013)	(0.323)	(0.061)	n.o.
Instrumental variable estimators (IV)	-0.010	-0.135***	n.o.	0.064
	(0.013)	(0.052)	n.o.	(0.175)
Long-differenced estimations	-0.018*	-0.011	Reference	Reference
	(0.011)	(0.038)	category	category
Within estimators	-0.011**	-0.035**	n.o.	n.o.
	(0.005)	(0.014)	n.o.	n.o.
Constant	0.384	-0.345	1.126***	0.797*
	(0.384)	(0.442)	(0.312)	(0.441)
Std. dev. of random slopes (log)				
	-3.019***			
	(0.160)			

Std. dev. of random intercepts (log)				
	0.262	-0.188	-0.336	0.265
	(0.207)	(0.416)	(0.320)	(0.278)
Std. dev. of residuals (log)				
	1.310***	0.368***	0.597***	-0.490***
	(0.026)	(0.065)	(0.051)	(0.067)
Observations	773	135	192	153
Studies	37	9	21	12
Log-likelihood (HM)	-2165.458	-248.138	-435.830	-145.683
Chi^2	353.067	220.976	111.872	243.983
$p>Chi^2$	0.000	0.000	0.000	0.000
Log-likelihood (comparator model)	-2347.730	-250.780	-438.984	-160.747
VIF	4.79	166.16	9.74	18.09
Heterogeneity#	97%	65%	77%	0%

Notes: **pmc**: dropped due to perfect multicollinearity; **n.o.**: no observations for the covariate. **(1 and 2)**: Elasticity estimates at the firm and industry levels, respectively; **(3 and 4)**: Rate-of-return estimates at the firm and industry levels, respectively. Significance of random effects is based on the natural logarithms of the standard deviations. Observations with undue influence are excluded, using the DFBETA influence statistics. LR Test Chi-squares indicate that the hierarchical models are preferred to least-squares estimators. LR tests for the specifications of the hierarchical models (not reported to save space) favour random-intercepts-and-slopes specification in (1), but random-intercepts only in (2), (3) and (4). # indicates the proportion of residual between-study variation due to heterogeneity, as opposed to within-study sampling variability. *, **, *** indicate significance at 10%, 5% and 1%, respectively.