Statistisk sentralbyrå
Statistics Norway

A new approach to estimating private returns to R&D

TALL SOM FORTELLER

Ådne Cappelen, Pierre Mohnen, Arvid Raknerud and Marina Rybalka

DISCUSSION PAPERS

1005

Discussion Papers: comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

The Discussion Papers series presents results from ongoing research projects and other research and analysis by SSB staff. The views and conclusions in this document are those of the authors.

Published: August 2023

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:

https://www.ssb.no/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Abstract

This paper revisits the estimation of private returns to R&D. In an extension of the standard approach, we allow for endogeneity of production decisions, heterogeneity of R&D elasticities, and asymmetric treatment of intramural and extramural R&D. Our empirical analyses are based on an extended Cobb-Douglas production function that allows for firms with zero R&D capital, which is especially useful for studying firms' transition from being R&D-non—active to becoming R&D-active. Using a large panel of Norwegian firms observed in the period 2001-2018, we estimate the average private net return to be in the range 0-5 percent across a variety of model specifications if we treat intra- and extramural R&D symmetrically. If in compliance with the Frascati manual, we treat intramural R&D as investment and extramural R&D as intermediate input, the estimated net return increases to 5-10 percent.

Keywords: Returns to R&D, Intramural R&D, Extramural R&D, Capitalization of R\&D, Dynamic panel data models, GMM

JEL classification: C33, C52, D24, O38

Acknowledgements: The project is financed by the Research Council of Norway, grant no. 256240 (OSIRIS). The authors are grateful to Magnus Gulbrandsen, Erlend Osland Simensen and Terje Skjerpen for many helpful comments.

Address: Ådne Cappelen, Statistics Norway, Research Department. E-mail: cap@ssb.no

Pierre Mohnen, Maastricht University and UNU Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), E-mail: p.mohnen@maastrichtuniversity.nl

Arvid Raknerud, Statistics Norway, Research Department. E-mail: rak@ssb.no

Marina Rybalka, Statistics Norway, Division for R&D, technology and business dynamics.
E-mail: ryb@ssb.no

Sammendrag

I denne artikkelen estimeres den private avkastningen av investeringer i FoU. I en utvidelse av standardtilnærmingen i faglitteraturen tillater vi endogen produksjonsbeslutning, heterogenitet i FoU-elastisiteter og asymmetrisk behandling av intern og ekstern FoU. Våre empiriske analyser er basert på en utvidet Cobb-Douglas produktfunksjon som tillater at vi kan inkludere foretak uten FoU i analysen. Det er spesielt nyttig for å studere overganger fra å være FoU-inaktiv til å bli FoU-aktiv. Ved hjelp av et omfattende panel av norske foretak observert i perioden 2001-2018, estimerer vi gjennomsnittlig privat nettoavkastning av investeringer i FoU til 0-5 prosent dersom vi behandler intern og ekstern FoU symmetrisk. Hvis vi derimot, i samsvar med Frascati-manualens anbefalinger, behandler intern FoU som investeringer og ekstern FoU som produktinnsats, øker den estimerte nettoavkastningen til 5-10 prosent. Disse resultatene indikerer høyere avkastning av intern enn ekstern FoU og reflekterer et tydelig mønster i dataene: at veksten i aggregerte FoU-utgifter i Norge i 2001-2018 kun skyldes vekst i intern FoU. Uansett er våre estimater av avkastningen av investeringer i FoU lave sammenlignet med estimater rapportert i den ofte siterte undersøkelsen av Hall m.fl. (2010), men av samme størrelsesorden som meta-regresjonsestimatene rapportert av Ugur et al. (2016), basert på et større og nyere utvalg av artikler enn 2010-studien.

1 Introduction

Private returns to R&D are of key interest to many public and private agents, such as investors, businesses and policy-makers. The returns are often found to be higher for R&D investments than for other investments, a fact which is used by various interest groups to promote broad R&D policies or own business interests more specifically (see Bührer et al., 2020). Evidence of high general private returns to R&D might convince the public that R&D support policies are worth their costs to taxpayers, despite evidence that some schemes generate little in terms of spillover effects (see Nilsen et al., 2020), or attract customers to firms that sell R&D services.

The amount of public R&D support is large and increasing in most developed countries. In Norway, for example, total public R&D support (tax credits plus grants) to business enterprise R&D (BERD) increased from 0.11 to 0.22 percent of GDP from 2006 to 2018, whereas BERD itself increased from 0.78 to 1.05 percent. In comparison, the average public support to BERD in the OECD area increased from 0.15 to 0.18 percent of GDP (see OECD, 2022). From a public policy view, it is therefore more important than ever to have valid and reliable methods for assessing the private returns to R&D.

While there are many approaches to estimating returns to R&D in the literature – whether primal or dual (see the survey by Hall et al., 2010) – they almost all have in common that they derive return estimates from the productivity impact of R&D under a *ceteris paribus* assumption. In the primal approach, the marginal return to R&D is equivalent to the increase in output (Y) as a result of an increase in R&D (F). For example, Doraszelski and Jaumandreu (2013) refer to returns to R&D when presenting the elasticity of output with respect to R&D expenditures. However, it is more common to transform elasticities (β) into marginal returns (R) using the definition of R&D elasticity: $\beta = R(F/Y)$. In the dual approach, R refers to the corresponding reduction in costs for a given output. For example, Bernstein and Nadiri (1988) define R as the real cost reduction due to a unit increase in F for a given Y.

In the primal approach, the most common way of specifying the underlying production function is by way of Cobb–Douglas with equal (homogeneous) elasticities across firms. However, according to R&D surveys, most firms report that they do not undertake any R&D at all, implying infinite returns at the extensive margin when Cobb–Douglas is applied. The standard solution is to estimate average returns using only firms with positive R&D, i.e. excluding observations of firms before they become R&D-active. This creates a sample selection problem that may bias the results. Moreover, it does not remove the problem that returns will tend to be very high at the margin for firms with low R&D intensity, which is likely to inflate estimates of average returns purely because of a functional form assumption. A more flexible assumption might be that firms have heterogeneous elasticities with respect to R&D, which is consistent with the high heterogeneity of R&D intensity observed in the data.

The effect of increasing R&D should arguably also incorporate the indirect effects of optimally adjusting flexible inputs in response to increased R&D. A value-added function, derived by treating (tangible and intangible) capital as quasi-fixed and intermediate inputs and labour as fully flexible factors, might therefore be better suited to assessing the private returns to R&D than simply measuring the increase in productivity *ceteris paribus*. A value-added function would capture both the direct and indirect effects of a (partial) change in R&D, reflecting both increased profits to owners and increased earnings to employees.

We address the limitations of the existing literature in four ways. First, we propose an extended Cobb–Douglas production function, where output depends on a *translation* of the R&D capital stock with an unknown, but estimable, translation parameter to allow positive output from firms with zero R&D. This functional form may be particularly useful for analyzing the transition from being an R&D non-performer to being R&D-active. Second, we accommodate heterogeneity in observed R&D intensities through firm-specific R&D elasticities to help explain the very wide variation in R&D intensities across firms and why most firms do not engage in R&D at all. Third, we derive return estimates from a value-added function, where firms simultaneously optimize intermediate and labour inputs for any level of capital – as opposed to a return measure that only incorporates the (first order) impact of increased R&D on output or production costs. Fourth, we identify a potential problem of double counting related to extramural R&D, which we address by capitalizing only intramural R&D and considering extramural R&D as intermediate input.

We analyze a panel of Norwegian firms in all industries from 2001 to 2018. Treating intra- and extramural R&D symmetrically, we obtain estimates of the weighted average private *net* return in the range of 0-5 percent, using the amounts of R&D investment as weights. These estimates are low compared to estimates reported in the often cited survey by Hall et al. (2010), but of the same magnitude as the meta-regression estimates obtained by Ugur et al. (2016), based on a larger set of articles than the 2010 study. However, if we only treat intramural R&D as investment, as recommended by the Frascati manual, the estimated average net return increases to 5-10 percent. These results indicate much higher returns to intramural than to extramural R&D and indicate a distinct pattern in the data: that the growth in aggregate real R&D spending in Norway in 2001-2018 is due to growth in intramural R&D alone. We also show that allowing heterogeneity in R&D elasticities is key to obtain robust and plausible estimates of returns to R&D within a family of model variants, including Cobb-Douglas.

The structure of the rest of the paper is as follows: Section 2 discusses main concepts and presents some studies relevant to our investigation. Section 3 describes our theoretical framework for analyzing private returns to R&D, Section 4 presents the econometric model, Section 5 presents the data, Section 6 shows the results and Section 7 offers some concluding comments.

2 Main concepts and approaches to studying the relationship between R&D and productivity

Several models of the relationship between R&D investment and productivity at firm level have been proposed in the empirical literature. One general model structure, usually referred to as the CDM model, was proposed by Crepon, Duguet and Mairesse (1998) following a conceptual model by Pakes and Griliches (1998). In this model, firm output (Y) can be expressed as a function of total factor productivity

(A^*) and input factors labour (L), intermediate inputs (M) and capital (K):

$$Y = A^* f(L, M, K).,\tag{1}$$

where A^* is assumed to depend on several variables relating to R&D, market factors, industry, and other variables. One way of extending this model is to include additional variables in Equation (1) to capture the effect of intangible investment – both internal and external to the firm. One such factor is R&D, which is not directly treated as primary input in the CDM framework, but is instead assumed to influence A^* through product and process innovations.

There is also a long tradition in economics of specifying R&D as a primary input factor (F) in a standard Cobb-Douglas production function– rather than indirectly via product- and process innovations as in the CDM model (see Hall *et. al.*, 2010) . The factor F (R&D capital) is then explicitly included in Equation (1) in addition to L, M, K. F is ususally generated by accumulating R&D spending according to the perpetual inventory method (PIM):

$$F_t = (1 - \delta)F_{t-1} + I_{t-1},\tag{2}$$

where δ is the depreciation rate of the R&D capital stock and I is (real) R&D investment. The capital stocks (both K and F) are assumed to be quasi-fixed, in the sense that investments in period $t - 1$ affect output in period t and beyond. If an estimate (or qualified guess) of the depreciation rate, δ, is available, one can: i) calculate the R&D capital stock, F_t, using (2), ii) estimate the parameters of the production function, and iii) derive an estimate of the average returns, R, to R&D investments, I_{t-1}.

The Frascati manual (2005, Section 4.12) thoroughly discusses the problem of double counting. A stylized example is a firm (A) that employs a researcher and reports the person's wage costs, wL, as intramural R&D, $int = wL$. Double counting would occur if L is also counted as labour inputs (see Schankerman, 1981). We will address this issue by deriving a value added function that depends, not on labour (L), but on the wage rate (w).

A double-counting problem that has received far less attention in the literature is related to extramural R&D. Assume that L in the above example could either produce output for firm A, generating profits πL (for some factor π), or provide R&D services to another firm (B) at the hiring price $x = (\pi + w)L$. This would not change the aggregate R&D input into the economy or the profits of firm A. In both cases, *int* is intramural R&D for firm A, but in the latter case, x is extramural R&D for firm B. It would be double counting if both *int* and x were capitalized as R&D investment. This does not preclude x from generating *excess* return for B, or increased value added in the aggregate economy. However, the increased value added would be due to a more efficient allocation of the available R&D services in the economy, not to increased aggregate R&D capital stock.

To avoid double counting of R&D, the Frascati Manual recommends using R&D performed and not R&D financed. In other words, the extramural R&D of firm B should be counted as internal R&D by the performing firm, A. In practice, double counting may be a bigger problem at the aggregate level than at the firm level. For example, it is far from obvious that firm A would actually report in-house R&D financed by firm B as intramural R&D. Asymmetric treatment of intramural and extramural R&D also raises the question of how to treat extramural R&D. A simple (and natural) solution would be to treat it as intermediate input (M), in which case it would increase the firm's value added only by generating returns in excess of its costs, x (which are intermediate costs).

One could also argue that the extramural R&D benefits the firm that has ordered this R&D and should enter its knowledge (R&D) capital stock in the same way as its intramural R&D.[1] In practice, it may also be difficult to distinguish between the two types: the same project may be carried out partly internally and partly externally, and then it makes sense to treat the external and internal R&D as one homogeneous input. The data also show that *int* and x are highly correlated, both in levels and in (first) differences. For example, the raw correlation coefficient between the differenced variables Δint and Δx is 0.23. The implication is that, to the extent that x and *int* finance the same R&D projects, the estimated return on *int* would also

[1] In fact, this is how R&D is treated in the Norwegian National Accounts: Extramural R&D is treated as an investment in the purchasing firms, but as production in the (research) institute sector (see Sørensen, 2016). This would lead to a double counting problem if the investment was not properly consolidated with the accounts data of the R&D producing firms, as discussed in the Frascati manual.

include some excess return on x. We conduct both a *symmetric* and an *asymmetric* treatment of extramural and intramural R&D below, where, in the symmetric case, $I = int + x$, and, in the asymmetric case, $I = int$, with x treated as an intermediate input (i.e. included in M).

Increased worker quality is also an important source of productivity growth. For example, Piekkola (2020) makes a distinction between R&D-labour, management- and advertising-related labour, and general labour, and constructs a measure of labour input quality based on the share of employees in each category and their relative wages. Our approach is broadly in line with Piekkola (2020), with the distinction that we measure labour quality by educational attainments rather than by dividing employees into professional or task-related categories.

An important feature of the (standard) Cobb-Douglas production function framework is that it cannot be applied to all firms without modifications, as it predicts zero output from firms with zero R&D capital. In the literature, there are several options available to circumvent that problem.

An important feature of the (standard) Cobb-Douglas production function framework is that it cannot be applied to all firms without modification, as it predicts zero output from firms with zero R&D capital. In the literature, there are several options available for circumventing this problem. One "solution" is simply to study those firms that report positive R&D and neglect the others. This strategy is problematic with regard to firms that become R&D-active in the observation period. Including these firms from the year in which they become active creates a sample-selection problem which may bias estimates of returns to R&D at the extensive margin. Since *potential* R&D performers are often a target of public R&D policies, the bias is potentially important from a policy perspective. The problem of sample selection can be addressed *ad hoc* by adding a small amount of R&D investment to firms with zero reported R&D, which makes it technically possible to include them in the analysis. A refinement of this solution is suggested by Griffith et al. (2006) and Hall et al. (2013). Relying on the CDM approach, they replace observed R&D spending with imputed R&D using data for all firms. In this way, zero R&D is replaced by non-zero imputed R&D. While this approach may perhaps be justified for firms that report zero R&D in *some* years, it does not allow us to study returns to R&D on the extensive margin, i.e. the returns on *becoming* R&D-active.

Finally, one can specify a more flexible functional form that allows zero R&D, as suggested already by Griliches (1979). The advantage of this solution, which we favour, is that it takes all observations at face value, without any need to alter the data.

3 Empirical model

Our starting point is an extended Cobb-Douglas production function with labour, intermediates, tangible capital and R&D capital as inputs. The first extension is that we assume that the production function has output elasticity of β in a *translation*, $\lambda + F$, of the R&D capital stock, F, for some value of $\lambda > 0$. The marginal returns to R&D will then be finite even with zero R&D. In fact, our extension is a special case of the Stone-Geary production function (see Beattie and Aradhyula, 2015), where only one factor, F, is translated. The second extension is that the production function is homogenous of degree ε in an aggregate function $g(L)$ of the vector $L = (L^{(1)}, L^{(2)}, L^{(3)})$ of man-years from three skill classes. Although there are examples of studies that control for the quality of labour input (e.g. Doraszelski and Jaumandreu, 2013), our approach is among the most elabourate in this respect (Doraszelski and Jaumandreu, *op. cit.*, only distinguish between temporary and permanent employees). Since R&D-active firms generally hire more educated and more highly paid workers than other firms, the assumption of homogeneous labour across education groups risks confounding the productivity effect of R&D with that of the skill composition. Under the standard assumption that the production function is homogenous with respect to tangible capital and intermediates (of degree γ and ρ, respectively), we can write

$$Y = Ag(L)^{\varepsilon} M^{\rho} (\lambda + F)^{\beta} K^{\gamma} \tag{3}$$

where A is total factor productivity (unexplained "neutral efficiency"). Moreover, sales revenue equals $S = PY$, where P is the (potentially endogenous) output price, and value added equals $V = S - q_M M$, where q_M is the price of intermediates.

Importantly, the specification (3) allows R&D input, F, to be zero without implying $Y = 0$. In particular, the marginal output of R&D with respect to F is:

$$Y'_F = \beta \frac{Y}{\lambda + F}$$

and the R&D elasticity is:

$$\mathrm{El}_F Y = \beta \frac{F}{\lambda + F}$$

An important property of (3) is that Y'_F, does not increase towards infinity as F tends to zero. When $\lambda = 0$, we have the Cobb-Douglas case with $\mathrm{El}_F Y = \beta$ and $Y'_F \to \infty$ as $F \to 0$. In this case, it will always be profitable to invest in R&D as the return to R&D at the extensive margin is infinity.

In the empirical model, we assume that ε, ρ and γ are common parameters, but allow β to be firm-specific, for the purpose of the analysis below.

3.1 Economic behavior

We assume that producers are price takers in all factor markets, but not in product markets, and that both types of capital, K and F, are fixed in the short run. Thus the short-run optimization of the firm is with respect to L and M for *pre-determined* R&D input, F_{it}, and tangible capital, K_{it}. The corresponding labour cost function, i.e. given the level of aggregate labour input, $g_i(L)$, is

$$C_{it}(\mathbf{q}_{it}, g_i(L_{it})) = c_{it} g_i(L_{it}) \tag{4}$$

where $\mathbf{q}_{it} = (q_{it}^{(1)}, q_{it}^{(2)}, q_{it}^{(3)})$ is the vector of firm-specific wage rates of low-, medium- and high-skilled labour, respectively, and c_{it} is the firm-specific unit price of labour (the aggregate wage rate). In Appendix A, we derive the formulas in (4) for the case of a CES aggregation function of labour inputs, L. We also allow $g_i(\cdot)$ to be firm-specific, to be consistent with firms choosing $L_{it}^{(m)} = 0$, for example, not employing workers in the highest skill category ($L_{it}^{(3)} = 0$).

We next consider the partial optimization problem of firm i at the beginning of period t conditional on the *predetermined* variables F_{it} and K_{it}, assuming that the firm knows $\mathbf{q}_{it}$, q_{Mt} and A_{it}. The problem is then to choose the price that maximizes operating profits. Making the usual assumption of monopolistic competition with demand given by:

$$Y_{it} = \Phi_{it} P_{it}^{-e}$$

where $\Phi_{it} > 0$ is a stochastic demand shifter and $e > 1$ is the elasticity of demand with respect to P_{it}, profit maximization gives the following equation for log value added:

$$\ln V_{it} = -\widetilde{\varepsilon} \ln c_{it} + \widetilde{\beta}_i \ln r_{it}(\lambda) + \widetilde{\gamma} \ln K_{it} - \widetilde{\rho} \ln q_{Mt} + \widetilde{a}_{it} \tag{5}$$

where $\widetilde{\varepsilon} = \varepsilon\vartheta$, $\widetilde{\beta}_i = \beta_i\vartheta$, $r_{it}(\lambda) = \lambda + F_{it}$, $\widetilde{\gamma} = \gamma\vartheta$, $\widetilde{\rho} = \rho\vartheta$, and $\widetilde{a}_{it} = \vartheta(\ln A_{it} + \ln \Phi_{it}/(e-1)) + \widetilde{\theta}$, with

$$\vartheta = \frac{(e-1)}{(\varepsilon + \rho + e - e(\varepsilon + \rho))} \in (0, (1 - \varepsilon - \rho)^{-1}). \tag{6}$$

See Appendix A for the definition of $\widetilde{\theta}$ and proof of Equations (5)-(6). The left and right limits correspond to $e \to 1$ and $e \to \infty$, respectively. Equation (5) will be the key equation for estimating private return to R&D.

We have no information on firm-specific intermediate input prices, so the term involving $\ln q_{Mt}$ in Equation (5) cannot be distinguished from time dummies in our empirical specification – and therefore $\widetilde{\rho}$ cannot be identified. On the other hand, we do observe firm-specific wages. The problem that the aggregate wage rate, c_{it}, is an unknown function of $\mathbf{q}_{it}$, is overcome by using the *Sato-Vartia* index:

$$\frac{c_{it}}{c_{i,t-1}} = \prod_{k=1}^{3} \left(\frac{q_{it}^{(k)}}{q_{i,t-1}^{(k)}} \right)^{\omega_{it}^{(k)}}$$

where the weights, $\omega_{it}^{(k)}$, are proportional to the geometric average of the (observable) cost shares $\alpha_{it}^{(k)}$ and $\alpha_{i,t-1}^{(k)}$ of skill class k:

$$\alpha_{it}^{(k)} = \frac{q_{it}^{(k)} L_{it}^{(k)}}{\sum_{k=1}^{3} q_{it}^{(k)} L_{it}^{(k)}} \tag{7}$$

The Sato-Vartia index is exact in the case of the CES aggregator function, where $g_i(L) = g(L; \mathbf{a}_i)$ for weight parameters, $\mathbf{a}_i = (a_i^{(1)}, a_i^{(2)}, a_i^{(3)})$ (see Appendix A for formulas). In that case we have the well-known result that $\alpha_{it}^{(k)} = a_i^{(k)} \left(q_{it}^{(k)} / c_{it} \right)^{1-\sigma}$ (see Diewert, 1978 and Brasch et al., 2022). More important for our purpose is that the Sato-Vartia index is consistent to the second order with any exact, twice differentiable aggregator function $g_i(\cdot)$. We can therefore apply the Sato-Vartia index to obtain:

$$\Delta \ln c_{it} = \sum_k \omega_{it}^{(k)} \Delta \ln q_{it}^{(k)} \tag{8}$$

with the initial condition $\ln c_{i1} = \sum_k \omega_{i2}^{(k)} \ln q_{i1}^{(k)}$.

3.2 Returns to R&D

We define

$$R_{it} = \frac{\partial V_{it}}{\partial F_{it}} = \frac{\widetilde{\beta}_i V_{it}}{F_{it} + \lambda} \tag{9}$$

as our proposed value added-based measure of private returns to R&D investment, with a potential distinction between intramural and extramural R&D, as discussed in the introduction. In the tradition of Hall et al. (2010), it is often assumed that R_{it} varies randomly about a common mean, R, where R is the constant marginal *cost* of R&D. To apply this assumption in our context, where F and K are pre-determined – and therefore based on ex ante expected returns – we define $V_{it}^e = E(V_{it} \mid F_{it}, K_{it})$ and assume the existence of a steady state defined as follows:

$$E(R_{it} | F_{it}, K_{it}) = \frac{\widetilde{\beta}_i V_{it}^e}{F_{it} + \lambda} = R \tag{10}$$

The first equality follows from (9), assuming:

$$V_{it} = V_{it}^e + e_{it}$$

for a genuine error term, e_{it}, whereas the second equality says that in a steady state with $F_{it} > 0$ (the firm has $I_{is} > 0$ for some $s < t$), expected returns equal the marginal cost of R&D. Equation (10) can be interpreted as an equilibrium correction, where a firm over time adjusts F_{it} towards a *firm-specific* equilibrium R&D intensity (but not necessarily R&D level). As we discuss below, the adjustment may be sluggish and hampered by adjustment costs and uncertainty, so that in general $R_{it} \neq R$. Given the above formal assumptions:

$$\widetilde{\beta}_i V_{it} = (F_{it} + \lambda) R + \widetilde{\beta}_i e_{it} \tag{11}$$

If direct data on R_{it} were available and λ was known, an unbiased estimator of R would be the following weighted average:

$$\sum_{t=1}^{T} \omega_{it} R_{it} = R + \widetilde{\beta}_i \frac{\sum_{t=1}^{T} e_{it} 1_{(F_{it}>0)}}{\sum_{t=1}^{T} (F_{it} + \lambda) 1_{(F_{it}>0)}} \tag{12}$$

where $1_{(A)}$ denotes the indicator function which is equal to one if the statement A is true and the weights $\omega_{it} = (F_{it} + \lambda)1_{(F_{it}>0)} / \sum_{t=1}^{T} (F_{it} + \lambda)1_{(F_{it}>0)}$ give equal weight to each NOK of R&D. Taking the expectation conditional on $F_{it} > 0$ on both sides of (11) yields:

$$\widetilde{\beta}_i = R\psi_i(\lambda) \tag{13}$$

where

$$\psi_i(\lambda) = \frac{E(F_{it}|F_{it} > 0) + \lambda}{E(V_{it}|F_{it} > 0)} \tag{14}$$

The function $\psi_i(\lambda)$ represents an equilibrium R&D intensity where the firm has positive R&D capital stock ($F_{it} > 0$). Note that $\psi_i(\lambda)$ varies across i only because $\widetilde{\beta}_i$ does so. We will refer to Equations (13)-(14) as the constant marginal cost (CMC) model. In particular, the CMC model explains why some firms never invest in R&D, namely if $\widetilde{\beta}_i < R\psi_i(\lambda)$ for any positive investment.

Estimating R as an average of the R_{it} would require that all the $\widetilde{\beta}_i$ are estimable, which is impossible because of the incidental parameter problem. An alternative strategy is the following: (i) replace $\psi_i(\lambda)$ with

$$\overline{\psi}_i(\lambda) = \frac{\sum_{t=1}^{T} 1_{F_{it}>0}(F_{it} + \lambda)}{\sum_{t=1}^{T} 1_{F_{it}>0} V_{it}} \tag{15}$$

which depends on only one unknown parameter (λ); (ii) substitute (15) into (13) to eliminate the incidental parameters $\widetilde{\beta}_i$ (replacing $\psi_i(\lambda)$ with $\overline{\psi_i}(\lambda)$); and iii) estimate R and λ using GMM (see Section 5).

In the literature, the usual assumption is that $\beta_i = \beta$ (no heterogeneity in the elasticity of Y with respect to F), implying $\psi_i(\lambda) = \psi(\lambda)$ for all i under the CMC assumption. We will refer to this special case as the restricted CMC model (R-CMC), which can be stated as:

$$\widetilde{\beta} = R\psi(\lambda) \tag{16}$$

The function $\psi(\lambda)$ represents an equilibrium R&D intensity that does *not* depend on i. Therefore, there should be no systematic or persistent differences in R&D intensities across firms. To estimate the average return, R, in the R-CMC-model, we can do the following: (i) estimate $\widetilde{\beta}$ and λ using GMM, (ii) obtain individual R_{it} estimates from Equation (9), and (iii) calculate the average of R_{it} as in equation (12), but summing over all firm-years (i, t).

By definition, $\psi_i(\lambda)$ or $\psi(\lambda)$ refers to an equilibrium state where the ratio between V_{it} and F_{it} is stable. However, in the presence of adjustment costs, firms with a short R&D history are likely to be far from their equilibrium R&D intensity. It is well documented in the literature on tangible capital that large changes in the level of capital generate disruption costs, for example due to learning-by-doing. As a consequence, investment is not fully productive until after some time has passed since it was made. Cooper and Haltiwanger (2006) find that a combination of disruption costs and irreversibilities, where the selling price of capital is lower than the purchasing price, best fit the key features of observed investment series. A recent, but sparse, literature on the implications of adjustment costs for the time series properties of R&D investment suggests that fixed costs will cause higher rates of return for firms that invest relatively more in R&D (see Resutek 2022).

We suspect that learning-by-doing and disruption costs may be the most important grounds for adjustment costs for firms that start doing R&D. This supposition is based on Brasch et al. (2020), who show that start-up firms have much lower revenue labour productivity, V_{it}/L_{it}, than incumbent firms, and we conjecture that the same factors will cause "R&D productivity", V_{it}/F_{it}, to be low for firms with a short history of R&D compared to more R&D-experienced firms, implying $\psi_i(\lambda) < \overline{\psi}_i(\lambda)$ in the former group of firms. Regardless, $\overline{\psi}_i(\lambda)$ may be severely biased as an estimator of $\psi_i(\lambda)$ for firms with a short R&D history.

A simple remedy would be to assume that:

$$\psi_i(\lambda) \simeq \overline{\psi}_i(\lambda) \left(1 + \tau_{begin} 1_{(T_i \leq \overline{T}_{begin})} + \tau_{exper} 1_{(\overline{T}_{begin} < T_i \leq \overline{T}_{exper})}\right) \tag{17}$$

where

$$T_i = \sum_{t=1}^{T} 1_{(F_{it}>0)} \tag{18}$$

is the number of years with positive R&D ($F_{it} > 0$) at the *last* observation year, T, whereas τ_{begin} and τ_{exper} represent, respectively, the average bias of $\overline{\psi}_i(\lambda)$ as an estimator of $\psi_i(\lambda)$ for *R&D-beginners* (defined as $T_i \leq \overline{T}_{begin}$) and *R&D-experienced* (defined as $\overline{T}_{begin} < T_i \leq \overline{T}_{exper}$) relative to firms with $T_i > \overline{T}_{exper}$ (*R&D-incumbents*). In our operationalization we define $\overline{T}_{begin}$ and $\overline{T}_{exper}$ as the first and third quartiles of the distribution of T_i at the last observation year, T. Thus, R&D-beginners are the 25 percent of firms with least R&D experience in the sample recorded at the last observation year, and R&D-incumbents are the 25 percent with the most R&D experience.

If $\psi_i(\lambda) \simeq \overline{\psi}_i(\lambda)(1 + \tau_{begin})$, a negative estimate of τ_{begin} would indicate that the equilibrium R&D intensity of firms with little R&D experience is overestimated. Then, the weighted average return is:

$$\sum_{t=1}^{T} \omega_{it} R_{it} \simeq R(1 + \tau_{begin})$$

where we used Equations (9), (13), (15) and (17). A negative parameter τ_{begin} would capture low returns to R&D in firms with little R&D experience.

4 Variable construction and descriptive statistics

For our analysis, we have constructed a panel of annual firm-level data for Norwegian firms with at least three consecutive observations in the period 2001–2018. The basis for the sample is the R&D statistics, which are survey data collected by Statistics Norway. These data comprise detailed information on firms' R&D activities, such as total R&D expenses (divided into internally performed R&D and externally purchased R&D), the number of employees engaged in R&D activities, and the number of man-hours worked in R&D. Only firms with more than 50 employees are automatically included in the survey. For smaller firms (with 5–49 employees) a stratified sampling scheme is employed. The stratification is based on industry classification (NACE codes) and firm size. However, these smaller firms are not representative of firms of their size and industry, because they have a higher probability of engaging in R&D. We use data for 1993, 1995, 1997, 1999, and *annual data* from 2001 to 2018. To supplement the regular R&D census, we obtained questionnaire data from the tax credit scheme *Skattefunn* on each applicant's R&D expenditure for three years prior to their applying for tax credits. These data are collected by the Research Council of Norway, which must approve in advance any project that is to form the basis for tax credits. The information from all available surveys is used for the construction of R&D capital stocks.

The survey data on R&D are supplemented with data from four different registers: the accounts statistics, the Register of Employers and Employees (REE), the National Education Database (NED) and the R&D Support Database.[2] This last contains information about each firm's R&D support in the period 2001-2018 – both direct support and tax credits. Descriptive statistics are provided in Appendix B.

Value added, V, is gross value added at factor cost computed as operating income (pY) less intermediate factor costs (qM): $V = pY - qM$. If both intramural and extramural R&D are included in F, intermediate factor costs, qM, are calculated as total operating costs less labour costs, depreciation and extramural R&D, as it is common practice to classify extramural R&D as an intermediate input ("other operating costs"). We do *not* subtract intramural R&D from total operating costs to obtain qM, as intramural R&D costs consist mostly of labour costs (and, to a lesser degree, costs of tangible capital), which have already been subtracted. Any intramural R&D costs labelled "other operating costs" (rather than labour or capital costs) in the accounts may cause a downward bias in the return to R&D estimates by being incorrectly included in qM. On the other hand, if only intramural R&D is included in F, extramural R&D is assigned to the intermediate inputs, qM.

All prices are deflated by the price index for R&D investments, so that, in any time period, one NOK of any cost component has the same value as one NOK of a revenue component (this is equivalent to normalizing the price of R&D to NOK 1). The price index is based on the price indices from the national accounts for the various components making up total R&D. According to Hall et al. (2010) the choice of deflator usually does not matter much for the econometric results for the main parameters of interest.

Intramural and extramural R&D expenditures are annual data as reported in the

[2]In Norwegian: Virkemiddeldatabasen

R&D statistics. Following Hall and Mairesse (1995), the (real) R&D capital stock (F_{it}) at the beginning of year t is computed by the perpetual inventory method (2) using the depreciation $\delta = 0.15$. The benchmark for the R&D capital stock at the beginning of the observation period for a given firm i, F_{i1}, is calculated as if it was the result of an infinite R&D investment series, $I_{i,-t}$, for $t = 0, 1, 2, ...$, with a fixed presample growth rate g=0.05 (cf. Equation (5) in Hall and Mairesse, 1995).

The properties of some aggregate R&D series are shown in Figure 1, where R&D capital *services* are defined as $(r + \delta)F_{it}$ for two definitions of I as explained in Section 2: $I = int + x$ or $I = int$, and interest rate $r = 0.05$.[3] Total R&D expenditures (intramural plus extramural) follow roughly the same trend growth as total R&D services when $I = int + x$. However, R&D capital services are lower than total expenditures in most of the period 2001-2018, in particular in the period from the financial crisis in 2008 until 2015, where total R&D services continue to grow, although with increased volatility, whereas total R&D expenditures drop markedly before rebounding in 2015. On the other hand, the series for intramural R&D services and intramural expenditures follow each other closely with regard to both level and growth, where intramural R&D services are calculated by setting $I = int$ in the PIM formula (2). The series for extramural R&D is almost flat throughout the period, suggesting that intramural R&D is the only source of growth in R&D at the aggregate level.

To construct the physical capital stock, K, we used information from the business accounts statistics , which distinguish between several groups of physical assets. To obtain consistent definitions of asset categories over the whole sample period, all assets have been divided into only two types: equipment, which includes machinery, vehicles, tools, furniture and transport equipment, and buildings and land (real property). The expected lifetimes of equipment (of about 3–10 years) are considerably lower than those of buildings and land (about 40–60 years). Total aggregate capital stock, K, is an aggregate of the book value of equipment capital and real property (see Nilsen et al. 2009, Section 2.2 for details).

Man-hours in skill group k, $L_{it}^{(k)}$, is the sum of all individual man-hours worked

[3]This is approximately equal to the average real lending rate of Norwegian banks in the period 2001-2018 (see Figure 1.6 in Wettre, 2021)

by employees in that group in the given firm according to their contract. For each firm, we distinguish between three skill groups: employees with primary, secondary and post-secondary education (see Table 7 in Appendix C). Man-hours worked by persons in skill group k are aggregated to firm level to construct $L^{(k)}$. When calculating the (average) wage in each skill group, $q_{it}^{(k)}$, we use predicted wages resulting from a wage regression with random individual effects, in which we include dummies for skill category (k) and dummies for industry (NACE 2), region, gender and calendar year as regressors. The average of the *predicted* wages for all firm employees in the given skill category generated by this regression forms the basis for calculating $q_{it}^{(k)}$. Using matched employer-employee data we are able to match each firm with its registered employees over time. This prediction-based method is chosen to reduce the problem of errors in reported hours in the employer-employee register. Errors are often related to part-time employees and/or timeliness problems, because hours are reported by the week and wage costs by the year. When estimating the wage equation we therefore restrict the sample to full-time employees in the given calendar year. The series for average log-wage by skill class and the log-wage index are shown in Figure 2. The index closely follows the average wage of skill class 2, although with a slightly steeper trend as a result of increased shares of workers in skill classes 2 and 3 over time. See Table 8 for detailed descriptive statistics for all the key variables used in the model.

5 Empirical analyses

The dependent variable in the empirical analysis is $\ln V_{it}$ and the stochastic specification of the structural Equation (5) is:

$$\ln V_{it} = -\widetilde{\varepsilon} \ln c_{it} + \widetilde{\gamma} \ln K_{it} + \widetilde{\beta}_i \ln r_{it}(\lambda) + a_i + \mu_t^* + \zeta_{it} \tag{19}$$

where a_i is a fixed firm effect, μ_t^* is the fixed time-effect (which incorporates the term $\widetilde{\rho} \ln q_{Mt}$), and ζ_{it} is an error term assumed to follow a first-order autoregressive process:

$$\zeta_{it} = \phi \zeta_{i,t-1} + e_{it} \tag{20}$$

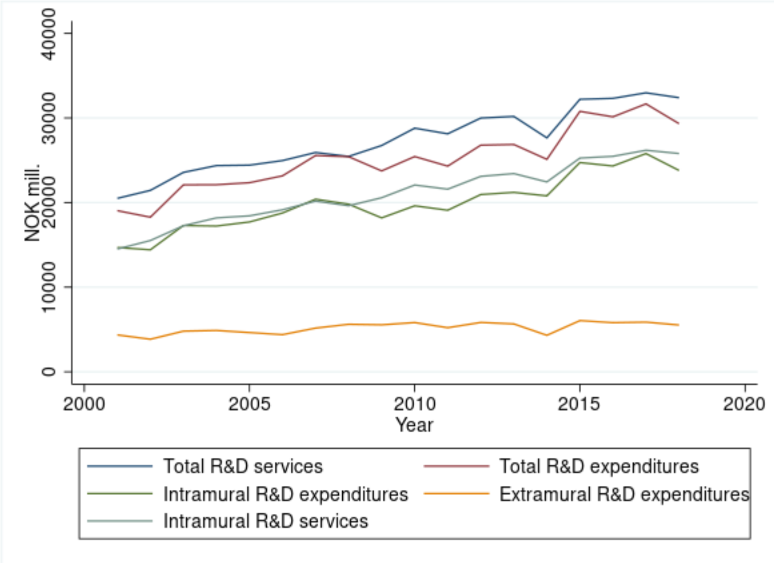

FIGURE 1: Aggregate R&D services and expenditures, by type

with

$$|\phi| \leq 1, \; E[e_{it}] = 0, \; E[e_{it}^2] = \sigma_e^2, \; E[\zeta_{it}^2] = \sigma_\zeta^2$$

and

$$\text{Cov}[e_{it}, e_{js}] = 0 \text{ if } t \neq s \text{ or } i \neq j.$$

Multiplying (19) by ϕ and quasi-differencing, yield:

$$\begin{aligned}\ln V_{it} = {} & \phi \ln V_{i,t-1} - \widetilde{\varepsilon} \ln c_{it} + \phi \widetilde{\varepsilon} \ln c_{i,t-1} + \widetilde{\beta}_i \ln r_{it}(\lambda) - \phi \widetilde{\beta}_i \ln r_{i,t-1}(\lambda) \\ & + \widetilde{\gamma} \ln K_{it} - \phi \widetilde{\gamma} \ln K_{i,t-1} + v_i + \mu_t + e_{it} \qquad (21)\end{aligned}$$

where $\mu_t = \mu_t^* - \phi \mu_{t-1}^*$ and $v_i = (1 - \phi) a_i$. Next, we difference to eliminate the fixed firm effect, v_i:

$$\begin{aligned}\Delta \ln V_{it} = {} & \phi \Delta \ln V_{i,t-1} - \widetilde{\varepsilon} \Delta \ln c_{it} + \phi \widetilde{\varepsilon} \Delta \ln c_{i,t-1} + \widetilde{\beta}_i \Delta \ln r_{it}(\lambda) - \phi \widetilde{\beta}_i \Delta \ln r_{i,t-1}(\lambda) \\ & + \widetilde{\gamma} \Delta \ln K_{it} - \phi \widetilde{\gamma} \Delta \ln K_{i,t-1} + \Delta \mu_t + \Delta e_{it} \qquad (22)\end{aligned}$$

Equation (22) constitutes the basis for GMM estimation.

5.1 The GMM Estimator

For given λ, the structural parameters are estimated by applying two-step GMM to Equation (22), where the initial (first-step) weight matrix is optimal under the assumption of i.i.d. errors e_{it} (implying that Δe_{it} is MA(1)). We use lagged levels of

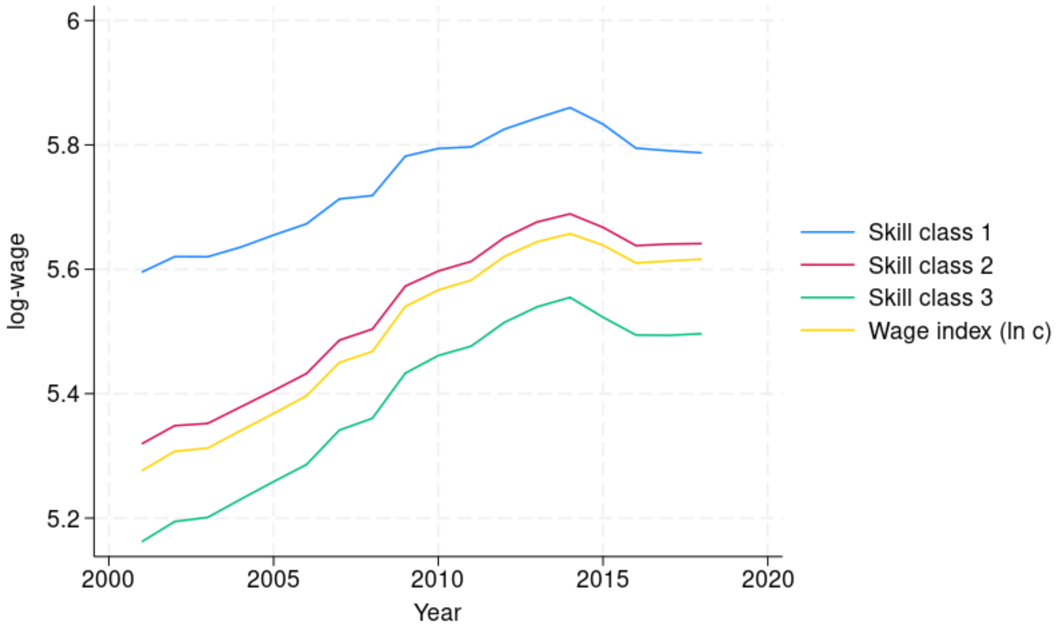

FIGURE 2: Average log-wage by skill class and average log-wage index ($\ln c_{it}$)

the endogenous variables as instruments, as proposed by Arellano and Bond (1991). To estimate λ, we perform a grid search in λ-space to minimize the sum of squares of the prediction errors of the structural Equation (22) (referred to as *residuals*), which is equivalent to maximizing the generalized R^2 model selection criterion proposed by Pesaran and Smith (1994) in the context of IV estimation.[4]

Following the general methodology of Arellano and Bond (1991) for dynamic panel data models, the GMM-estimator uses the following moments:

$$E(\ln V_{i,t-s}\Delta e_{it}) = 0$$

$$E(\ln c_{i,t-s+1}\Delta e_{it}) = 0$$

$$E(\ln r_{i,t-s+1}(\lambda)\Delta e_{it}) = 0$$

$$E(\ln K_{i,t-s+1}\Delta e_{it}) = 0$$

for $s \geq 2$ (see Equation (22)). That is, we treat *all* the right-hand side variables in Equation (22) as pre-determined endogenous variables. Our method does not rely on the often used, but strong, initial condition assumptions of Blundell and Bond (1998) to obtain additional moment conditions. A testable identifying assumption

[4]The prediction errors are obtained by replacing the endogenous explanatory variables in Equation (22) with their predicted values from the first-stage of 2SLS; see Pesaran and Smith (1994) for details.

is that Δe_{it} is an MA(1) noise term. Our results displayed in Figure 3 show that a very small λ ($\widehat{\lambda} = 0.37$) minimizes the residual sum of squares, i.e. an estimate on par with the lowest positive F_{it} observation in our estimation sample. In fact, this estimate supports the common practice of adding a "small" number to R&D capital to avoid zero input in a Cobb-Douglas production function.

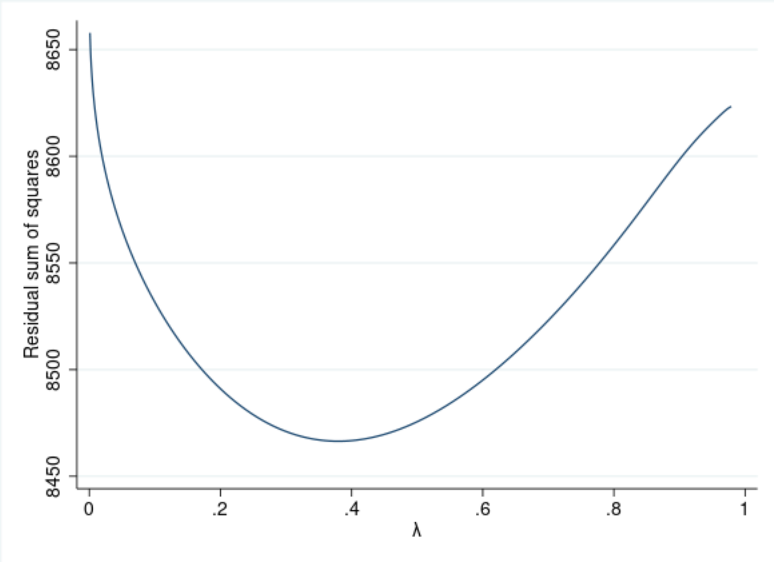

FIGURE 3: Residual sum of squares as a function of λ in the CMC model

As discussed in Section 3, we examine two specifications with respect to the marginal cost of R&D. The first is the CMC-model given by Equations (13)-(14), where we assume a firm-specific elasticity, $\widetilde{\beta}_i$, and a common cost of R&D, R, with a firm-specific steady-state R&D intensity, $\psi_i(\lambda)$. In this model, $\widetilde{\beta}_i = R\psi_i(\lambda)$. Using Equation (17) to approximate ψ_i, we consider R, τ_{begin} and τ_{exper} as GMM estimands. The second case is a restricted version of this model, referred to as R-CMC, where we assume a common elasticity for all firms: $\widetilde{\beta}_i = \widetilde{\beta}$ and consider $\widetilde{\beta}$ to be the GMM estimand. Given $\widetilde{\beta}$ and λ, it is straightforward to derive an estimate of R as an R&D-weighted average of R_{it}, as explained in Section 3.2.

5.2 Parameter estimates

The parameter estimates for the model with symmetric treatment of intramural and extramural R&D: $I = int + x$, are presented in Table 1. As a benchmark we also include a fixed-effects (FE) estimator of the R-CMC model. The FE estimator is the conventional within-estimator applied to the level Equation (19). This method yields biased estimates in the presence of endogeneous *time-varying* explanatory variables. In general there are only small difference between FE-estimates and GMM-estimates in Table 1, indicating there there is little endogeneity bias related to the time varying covariates.

TABLE 1: **Estimates of the coefficients of the value added equation with symmetric treatment of intramural and extramural R&D ($I = int + x$). Robust standard errors (SE)**

Indep. variables in structural equation	**Coeff.**	**GMM-estimates CMC**		**GMM-estimates R-CMC**		**FE-estimates R-CMC**	
		Est.	**SE**	**Est.**	**SE**	**Est.**	**SE**
$\ln V_{i,t-1}$	ϕ	.306	.011***	.311	.021***		
$-\ln c_{it}$	$\widetilde{\varepsilon}$	.502	.164***	.509	.167***	.639	.110***
$\ln K_{it}$	$\widetilde{\gamma}$	.195	.016***	.207	.016***	.166	.006***
$\ln r_{it}(\lambda)$	$\widetilde{\beta}$			.045	.003***	.042	.002***
$\overline{\psi}_i(\lambda) \ln r_{it}(\lambda)$[1]	R	.181	.049***				
$\overline{\psi}_i(\lambda) 1_{(T_i \leq 3)} \ln r_{it}(\lambda)$[2]	τ_{begin}	-.180	.049***				
$\overline{\psi}_i(\lambda) 1_{(T_i \in [4,12])} \ln r_{it}(\lambda)$	τ_{exper}	-.029	.053				
σ_e^2		.11		.11		—	
σ_ζ^2		—		—		.41	
λ		.38		.38		.38	
Number of firm-years		40,344		40,344		40,344	
Number of firms		4,590		4,590		4,590	
R-squared (R^2)[3]		.10		.10		.41	

Note: Windmeijer (2005) robust standard errors (SE); *,**,*** refer, respectively, to significant estimates at the 10, 5, and 1 percent level.
[1] $\overline{\psi}_i(\lambda)$ refers to the firm's average R&D intensity, as defined in Equation (15).
[2] T_i is the number of years with $F_{it} > 0$ in the years 2001-2018.
[3] R^2 refers to (the differenced) Equation (22) in the case of GMM and (the level) Equation (19) in the case of FE.

In the R-CMC model, where it is assumed that $\beta_i = \beta$ for all i, the coefficient of $\ln r_{it}(\lambda)$ is the common elasticity with respect to R&D, $\widetilde{\beta}$. It is estimated to be slightly higher than 0.04 and significantly positive regardless of estimation method (GMM or FE). In the CMC model, the coefficient of $\psi_i(\lambda) \ln r_{it}(\lambda)$ equals R. It is estimated to be 0.181, i.e. 18.1 percent. The terms $\overline{\psi}_i(\lambda) 1_{(\cdot T_i \cdot)} \ln r_{it}(\lambda)$, where $1_{(\cdot T_i \cdot)}$ refers to an indicator function involving T_i, can be interpreted as the gross return *in addition* to R for the group defined by $1_{(\cdot T_i \cdot)} = 1$. The estimated average return is significantly lower than R for R&D-beginners ($T_i \leq 3$), in fact only 18.1-18.0 = 0.01 percent. Moreover, the average gross return is estimated to be 2.9 percentage points lower for R&D-experienced firms (between 4 and 12 years of R&D experience) compared to R&D-incumbents (the reference category), but the difference is not significant. Thus, with the exception of R&D-beginners, the average gross return to R&D appears to be about 18 percent.

As expected, we find a significant positive relationship between tangible capital and value added: the estimated elasticity with respect to K obtained using different estimation methods lies in the range 0.17-0.21, and is three to four times larger than the estimated (common) elasticity with respect to R&D capital, $\widetilde{\beta}$. The GMM estimates of the autoregressive coefficient ϕ in Table 1 – the coefficient of $\ln V_{i,t-1}$ – are 0.306 and 0.311 and highly significant. Since the FE estimator uses the equations in levels, before quasi-differencing, ϕ is not identified in this case.

The coefficient of $-\ln c_{it}$ is $\widetilde{\varepsilon}$ and is estimated to be in the range 0.5-0.6 across specifications and estimation methods. The results show that labour is the production factor with the highest output elasticity.

The last row of Table 2 reports R^2 statistics. The much lower R^2 corresponding to the GMM-estimates of Equation (22) (0.10) compared to the FE-estimates of Equation (19) (0.41) is due to the loss of level information in the former equation: Equation (22) is a quasi-differenced version of (19). In fact, within-R^2 equals 0.10 in the case of the FE-estimator.

As seen from Table 2, the Arellano–Bond test for zero first-order autocorrelation in the error term Δe_{it} leads to rejection, but not the test for second-order autocorrelation. This confirms that ζ_{it} follows a first-order autoregressive process, as assumed in Equation (20). We also applied a Hansen J-test of the validity of overidentifying restrictions with regard to the instrumental variables. With a χ^2-test statistic of 634.51 and 619 degrees of freedom, we cannot reject the overidentifying restrictions. All these specification tests, seen together, give strong support for our econometric specification.

5.3 Returns to R&D

Using the GMM estimates displayed in Table 1, we calculate the marginal returns to R&D investment, $R_{it} = \partial V_{it}/\partial F_{it}$, for each observation as explained in Section 3. It is important to realize that each R_{it} represents a *prediction* of the gross return on

Table 2: **Specification tests of CMC model**

	Observed value of test statistic (Z)	**Level of significance** $\Pr(Z > z)$
Test of zero autocorrelation in errors*		
order 1	-14.54	.000
order 2	1.58	.110
J-test of overidentifying restrictions**	634.51	.321

Notes: *t-test ** Hansen (1982) test statistics is distributed as $\chi^2(619)$

Table 3: **Distribution of estimated marginal gross returns to R&D (R_{it}) with symmetric treatment of intramural and extramural R&D ($I = int + x$). By subsample, conditional on $F_{it} > 0$**

Model	**All obs.** $F_{it} > 0$	**Subsample with** $F_{it} > 0$ **R&D-begin.**[1]	**R&D-exper.**[2]	**R&D-incumb.**[3]
CMC: heterogeneous elast.				
Weighted average[4)]	.173	.001	.146	.177
Median	.169	.001	.153	.190
Unweighted average	.270	.001	.276	.294
R-CMC: common elasticity				
Weighted average	.209	.628	.334	.191
Median	.422	.678	.520	.344
Unweighted average	6.47	8.56	6.17	6.45
Share of R&D in 2018 (share $\sum_i F_{i,2018}$)	1	.05	.17	.78
No. of firm-years with $F_{it} > 0$	30,331	2,370	15,507	27,822
No. of firms with $F_{it} > 0$	4,238	1,046	2,146	1,046

Note: Derived using the GMM estimates displayed in Table 1

[1] Firms that were R&D-active (i.e., with $F_{it} > 0$) for maximum 3 years in the period 2001-2018

[2] Firms that were R&D-active for between 4 and 12 years in the period 2001-2018

[3] Firms that were R&D-active for more than 12 years in the period 2001-2018

[4] Weighted by share of R&D (F_{it})

an additional NOK investment in R&D capital in one year – not the actual return on the last amount invested. Of course, the ex post return cannot be meaningfully assessed as we cannot identify the marginal investment in isolation from the stock of capital. To evaluate the magnitude of R_{it}, we should keep in mind that both the interest rate and the depreciation rate of 15 percent, as well as a risk premium, should be covered.

Different average and median gross return values by amount of R&D experience (number of years with $F_{it} > 0$ in 2001-2018) are displayed in Table 3, with 1/4 of the firms classified as R&D-beginners and 1/4 as R&D-incumbents. The estimated marginal gross returns have R&D-weighted averages of 17.3 and 20.9 percent in the CMC and R-CMC models, respectively. We see that the R&D-beginners account for only 5 percent of the total R&D stock in 2018 (the last observation year); the R&D-experienced firms account for 17 percent, and the R&D-incumbents account for 78 percent. Thus quite a small share of the firms accounts for a very large share of R&D investment. As we anticipated, the results show highest returns for R&D-experienced firms and lowest returns for R&D-beginners (in fact, close to zero), likely reflecting high adjustment costs, as discussed in Section 3.2. The 3 percentage points difference in average returns between R&D-experienced and R&D-incumbents is driven by the negative, but insignificant, τ_{exper} estimate in Table 1.

Looking at the returns in the different groups of firms and across models, we find some striking patterns. In the CMC model, average and median returns are much more homogeneous across the groups than in the R-CMC model. In contrast, the R-CMC model shows a pattern of much higher unweighted average returns in all groups (more than 5) compared to the CMC model (less than 0.3). These results indicate that the R-CMC model is too rigid. However, if we only care about the weighted average return to R&D, the estimates are quite similar, from 17-21 percent gross return.

Our weighted average gross return of 17-21 percent, which is equivalent to a net return of 2-6 percent, is low compared to the rate of return commonly observed in the international literature, see Hall et al. (2010). In a more recent comprehensive survey of the literature, Ugur et al. (2016) found that estimated rates of return are much smaller and more heterogeneous than reported in the earlier literature. Our estimates are, in fact, close to the average gross return of 14 percent found by Ugur et al. (2016), using regression methods of meta-analysis. They found that returns are lower for small firms, which is consistent with our finding of lower returns for R&D-beginners. They also report that estimates obtained using GMM and IV are lower compared to OLS, pointing to an endogeneity problem which is carefully addressed in our study. We have also addressed many other issues that they raise, such as double-counting of R&D, firm heterogeneity and the need to take dynamics into account.

5.4 Asymmetric treatment of intramural and extramural R&D ($I = int$)

The results for the model with asymmetric treatment of intramural R&D (int) and extramural R&D (x) are shown in Table 4 (coefficient estimates) and Table 5 (average gross returns). In this model version, only int is counted as R&D investment, while x is included as a part of intermediate input, M, as discussed in Section 2.

A comparison of Tables 1 and 4, reveals significant differences in the returns to R&D estimates in the CMC model, but very small differences otherwise. Most importantly, in Table 4 the average gross return to R&D (the parameter R) is estimated to be 27.4 percent (with SE 2.6 percent), compared to 18.1 percent in Table 1 (with SE 4.9 percent). The higher returns are similarly reflected in Table 5, where, for example, the estimated weighted average return is 25.6 in the CMC model, where GMM estimation is used, and 22-24 percent in the R-CMC model where FE estimation is used. These results should not come as a surprise: since the growth of extramural R&D is virtually zero (see Section 4), it is hard to attribute any return to extramural R&D. Consequently, the average return is higher when only intramural R&D is counted as an investment, and extramural R&D is treated as intermediate inputs.

Interestingly, our results contrast strongly with those of Bönte (2003), in his study of external and internal R&D on aggregate (2-digit) West German industry data in the period 1980-1993, with a focus on high-tech industries. He finds an inverse U-relationship between productivity growth and the share of external R&D relative to total R&D. In contrast to Norway in 2001-2018, external R&D constitutes an increasing share of R&D in Germany in 1980-1993, pointing perhaps towards fundamental structural differences between the two countries and time periods.

5.5 The Cobb-Douglas model

As a final robustness check, we have estimated all the model variants considered above for the “workhorse ” Cobb-Douglas model, i.e. the special case with $\lambda = 0$. The estimation sample then have to be restricted to firm-years with $F_{it} > 0$. The derived return estimates are shown in Table 6 and the corresponding parameter

TABLE 4: **Coefficient estimates when only intramural R&D are treated as investments ($I = int$). Robust standard errors (SE)**

Indep. variables in structural equation	**Coeff.**	**GMM-estimates**				**FE-estimates**	
		CMC		**R-CMC**		**R-CMC**	
		Est.	**SE**	**Est.**	**SE**	**Est.**	**SE**
$\ln V_{i,t-1}$	ϕ	.299	.011***	.340	.010***		
$-\ln c_{it}$	$\widetilde{\varepsilon}$	.678	.164***	.686	.172***	.618	.112***
$\ln K_{it}$	$\widetilde{\gamma}$	.165	.015***	.190	.016***	.169	.006***
$\ln r_{it}(\lambda)$	$\widetilde{\beta}$			.045	.003***	.040	.002***
$\overline{\psi}_i(\lambda) \ln r_{it}(\lambda)^1$	R	.274	.026***				
$\overline{\psi}_i(\lambda) 1_{(T_i \leq 3)} \ln r_{it}(\lambda)^2$	τ_{begin}	-.272	.055***				
$\overline{\psi}_i(\lambda) 1_{(T_i \in [4,12])} \ln r_{it}(\lambda)$	τ_{exper}	-.082	.057				
σ_e^2		.11		.11		—	
σ_ζ^2		—		—		.40	
λ		.35		.35		.35	
Number of firm-years		40,344		40,344		40,344	
Number of firms		4,590		4,590		4,590	
R-squared $(R^2)^3$		.10		.10		.44	

Note: Windmeijer (2005) robust standard errors (SE); *,**,*** refer, respectively, to significant estimates at the 10, 5, and 1 percent level.

[1] $\overline{\psi}_i(\lambda)$ refers to the firm's average R&D intensity, as defined in Equation (15).

[2] T_i is the number of years with $F_{it} > 0$ in the years 2001-2018.

[3] R^2 refers to (the differenced) Equation (22) in the case of GMM and (the level) Equation (19) in the case of FE.

estimates in Table 9 in Appendix C. Table 6 reveals some striking patterns. First, the model with heterogeneous elasticities (the CMC model), yields return estimates of the same magnitude as those obtained with optimally chosen $\lambda > 0$, i.e. the weighted average gross return is about 15 percent with symmetric treatment of intramural and extramural R&D ($I = int + x$) and 25 percent when only intramural R&D is treated as investment ($I = int$). Also the estimated median and unweighted average returns are of the same magnitude in Table 3 and Table 6 – regardless of the chosen definition of I. However, the results become remarkably different when homogeneous elasticities are imposed. In that case, the weighted average return estimates are of magnitude 50-100 percent – depending on the estimation method and definition of I. In any case, the latter estimates should be disregarded as they are volatile and implausibly high. A main insight of this paper is therefore that allowing for heterogeneous elasticities with respect to R&D is key to obtaining robust and plausible estimates of returns to R&D across model specifications. The

Table 5: **Distribution of estimated marginal gross returns to R&D (R_{it}) when only intramural R&D are treated as investments ($I = int$). By subsample, conditional on $F_{it} > 0$**

Model	All obs. $F_{it} > 0$	Subsample with $F_{it} > 0$ R&D-begin.[1]	R&D-exper.[2]	R&D-incumb.[3]
CMC: heterogeneous elast.				
Weighted average[4)]	.256	.001	.184	.268
Median	.238	.001	.153	.194
Unweighted average	.339	.002	.276	.287
R-CMC: common elasticity				
Weighted average	.241	.900	.334	.221
Median	.420	.655	.493	.357
Unweighted average	3.71	5.03	3.05	4.01
Share of R&D in 2018 (share $\sum_i F_{i,2018}$)	1	.05	.17	.78
No. of firm-years with $F_{it} > 0$	30,331	2,370	15,507	27,822
No. of firms with $F_{it} > 0$	4,238	1,046	2,146	1,046

Note: Derived using the GMM estimates displayed in Table 4.
[1] Firms that were R&D active (i.e., with $F_{it} > 0$) in maximum 3 years in 2001-2018
[2] R&D active firms between 4 and 12 years in 2001-2018
[3] R&D active firms for more than 12 years in 2001-2018
[4] Weighted by share of R&D (F_{it})

definition of R&D capital is clearly also of key importance.

6 Conclusions

This paper has revisited the estimation of the private returns to R&D. We have proposed an extended Cobb-Douglas production function, which allows for firms with zero R&D capital, in order to study the transition from being R&D non–active to active, without restricting the sample to R&D performers. In the standard approach, the returns to R&D only incorporate the impact of increased R&D on productivity or production costs. In contrast, we have obtained return estimates from a value added function derived under the assumption of profit maximizing firms that optimize labour and intermediate factor inputs at any level of R&D capital. The value added function captures both increased profits to owners and increased earnings to employees resulting from R&D investment.

We further have accommodated the huge observed heterogeneity in R&D intensities by allowing R&D elasticities to be firm-specific, which we showed to be key

TABLE 6: **Distribution of marginal gross returns to R&D (R_{it}) estimated using Cobb-Douglas production function ($\lambda = 0$) and different definitions of R&D investment. By subsample, conditional on $F_{it} > 0$**

Model	**All obs.** $F_{it} > 0$	**Subsample with** $F_{it} > 0$ **R&D-begin.**[1]	**R&D-exper.**[2]	**R&D-incumb.**[3]
Symmetric treatment of intramural and extramural R&D ($I = int + x$)				
CMC: heterogeneous elast.				
Weighted average[4)]	.138	-.241	.241	.150
Median	.116	-.270	.052	.161
Unweighted average	.172	-.373	.276	.248
R-CMC: common elasticity				
Weighted average	.637	1.23	1.33	.598
Median	1.21	1.48	1.54	1.06
Unweighted average	24.9	63.5	20.82	25.78
Only intramural R&D treated as investment ($I = int$)				
CMC: heterogeneous elast.				
Weighted average[4)]	.249	-.295	.241	.253
Median	.267	-.330	.270	.271
Unweighted average	.357	-.429	.357	.390
R-CMC: common elasticity				
Weighted average	.991	2.84	1.33	.939
Median	1.65	2.04	1.96	1.49
Unweighted average	16.05	19.09	13.07	17.59

Note: Returns estimates derived from models estimated using GMM on subsample of observations with $F_{it} > 0$. See Table 9 in Appendix C.

[1] Firms that were R&D-active (i.e., with $F_{it} > 0$) for maximum 3 years in the period 2001-2018

[2] Firms that were R&D-active for between 4 and 12 years in the period 2001-2018

[3] Firms that were R&D-active for more than 12 years in the period 2001-2018

[4] Weighted by share of R&D (F_{it})

to obtain robust estimates of returns to R&D within a family of model variants, including Cobb-Douglas, and incorporated heterogeneity in labour quality by distinguishing between three levels of employee educational attainments. Estimating the model on a comprehensive panel of Norwegian firms observed in the period 2001-2018, we obtained estimates of the average private gross return in the range 15-20 percent, i.e. net return in the range 0-5 percent, with a symmetric treatment of intra- and extramural R&D. These estimates are similar to the average gross return of 14 percent reported by Ugur et al. (2016) in a meta analysis based on a very comprehensive literature – but low compared to estimates more commonly reported, e.g. by the often cited survey by Hall et al. (2010). However, if we count only intramural R&D as investments, as recommended by the Frascati manual, the estimated *net* returns increase to 5-10 percent.

Acknowledgements

The project is financed by the Research Council of Norway, grant no. 256240 (OSIRIS). The authors are grateful to Magnus Gulbrandsen, Erlend Osland Simensen and Terje Skjerpen for many helpful comments.

References

[1] Arellano, M. and S. Bond (1991): Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies,* 58, 277–297.

[2] Beattie, B.R. and S. Aradhyula (2015): A note on threshold factor level(s) and Stone-Geary technology. *Journal of Agricultural and Applied Economics*, 47(4), 482–493.

[3] Bernstein, J and M.I. Nadiri (1988): Interindustry R&D Spillovers, Rates of Return, and Production in High-Tech Industries. *American Economic Review*, 78(2), 429–434.

[4] Blundell, R. and S. Bond (1998): Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics*, 87, 115–143.

[5] Bönte, W. (2003): R&D and productivity: Internal vs. external R&D – Evidence from West-German manufacturing industries. *Economics of Innovation and New Technologies*, 12(4), 343–360.

[6] Brasch, T. von and A. Raknerud (2022): Productivity growth, firm turnover and new varieties. *Scandinavian Journal of Economics*, 124, 646–676.

[7] Bührer, S, R. Walz, S. Seus, M. Astor, T. Stehnken, F. Malik (2020): Evaluation der BMBF-Rahmenprogramme Forschung für die Nachhaltigkeit FONA 1 (2005-2009) und Forschung für Nachhaltige Entwicklungen FONA 2 (2010-2014). Karlsruhe, Berlin 2020.

[8] Cooper, R.W. and J. Haltiwanger (2006): On the Nature of Capital Adjustment Costs. *Review of Economic Studies*, 73(3), 611–633.

[9] Crepon, 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, 115–158.

[10] Diewert, W.E. (1978): Superlative Index Numbers and Consistency in Aggregation, *Econometrica*, 46(4), 883–900.

[11] Doraszelski, U. and J. Jaumandreu (2013): R&D and productivity: Estimating endogenous productivity, *Review of Economic Studies*, 80(4), 1338-1383.

[12] Frascati Manual (2015): Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD 2015.

[13] Golombek, R. and A. Raknerud (2018): Exit dynamics of start-up firms. Structural estimation using indirect inference. *Journal of Econometrics*, 205, 204–225.

[14] Griffith, R., E. Huergo, J, Mairesse and B. Peters (2006): Innovation and Productivity Across Four European Countries, *Oxford Review of Economic Policy*, 22, 483–498.

[15] 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.

[16] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029–1054.

[17] Hall, B. H., J. Mairesse and P. Mohnen (2010): Measuring the returns to R&D, in Chapter 24: *Handbook of the Economics of Innovation*, Vol. 2. North-Holland Publ. Comp., Amsterdam.

[18] Nilsen, Ø.A., A. Raknerud and D.-C. Iancu (2020): Public R&D Support and Firm Performance: A Multivariate Dose-Response Analysis. *Research Policy*, 49 (7).

[19] Nilsen, Ø.A., A. Raknerud, M. Rybalka and T. Skjerpen (2009): Lumpy investments, factor adjstments, and labour productivity *Oxford Economic Papers*, 104–127, 49(7).

[20] OECD (2022): *OECD R&D tax incentives database, 2021 edition.*

[21] Pakes, A. and Z. Griliches (1984): Patents and R&D at the firm level in French manufacturing: a first look. In Z. Griliches (ed.): *Research and Development, Patents and Productivity*, The University Press of Chicago, Chicago, pp. 55–72.

[22] Pesaran, M.H. and R.J. Smith (1994): A generalized R^2 criterion for regression models estimated by the instrumental variables method. *Econometrica*, 62(3), 705–710.

[23] Piekkola, H (2020): Innovation-labour biased technical change and growth – Evidence from Finnish innovation value chains. GOBALINTO (Deliverable D6.4).

[24] Resutek, R.J. (2022): Is R&D Really That Special? A Fixed-Cost Explanation for the Empirical Patterns of R&D Firms. *Contemporary Accounting Research*, 39(1), 721–749.

[25] Schankerman, M. (1981): The Effects of Double-Counting and Expensing on the Measured Returns to R&D. *The Review of Economics and Statistics*, 63, 454-58.

[26] Sørensen (2016). Forskning og utvikling i Nasjonalregnskapet: Dokumentasjon av arbeidet med FoU i hovedrevisjonen 2014. Notater 2016/32 (in Norwegian), Statistics Norway

[27] Ugur, M., E. Trushin, E. Solomon and F. Guidi (2016): R&D and productivity in OECD countries and industries: A hierarchical meta-regression analysis. *Research Policy*, 45, 2069–2086.

[28] Wettre, S (2021): Rentestatistikk 1820 - 2020. Utvikling i renter og dokumentasjon av rentestatistikken. Notater 2021/20 (in Norwegian), Statistics Norway.

[29] Windmeijer, F (2005): A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126, 25–51.

Appendix A. Derivation of Equation (5) and some related results

The production function is:

$$Y_{it} = A^*_{it} M^{\rho}_{it} g(L_{it}, H_{it})^{\varepsilon}$$

where $A^*_{it} = A_{it}(\lambda + F_{it})^{\beta} K^{\kappa}_{it}$ (which we in this appendix consider as fixed) and

$$g(L_{it}, H_{it}) = \left[a^{\frac{1}{\sigma}} L_{it}^{(\sigma-1)/\sigma} + (1-a)^{\frac{1}{\sigma}} H_{it}^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}$$

is the CES aggregate of (man hours from) high-skilled and low-skilled workers, H_{it} and L_{it} (the generalization to any number of types of workers is straightforward). Cost-minimization with respect to L_{it} and H_{it}, given factor prices for high and low skilled labour, w^l_t and w^h_t, conditional on M_{it} and Y_{it}, gives the conditional cost function

$$C(M_{it}, Y_{it}) = c_{it} \left(\frac{Y_{it}}{A^*_{it} M^{\rho}_{it}} \right)^{\frac{1}{\varepsilon}}$$

with

$$c_{it} = \left[a(w^l_{it})^{1-\sigma} + (1-a)(w^h_{it})^{1-\sigma} \right]^{\frac{1}{1-\sigma}} .$$

Now consider the problem of finding the cost minimizing M_{it}, given q_{Mt} and c_{it}:

$$\begin{aligned} M^*_{it} &= \arg\min_{M_{it}} \left(q_{Mt} M_{it} + C(M_{it}, Y_{it}) \right) \\ &= \arg\min_{M_{it}} \left(q_{Mt} M_{it} + c_{it} \left(\frac{Y_{it}}{A^*_{it} M^{\rho}_{it}} \right)^{\frac{1}{\varepsilon}} \right) \end{aligned}$$

The 1. order condition for cost minimization is:

$$\ln M^*_{it} = \frac{1}{\rho+\varepsilon} (\ln Y_{it} - \ln A^*_{it}) + \frac{\varepsilon}{\rho+\varepsilon} \ln c_{it} - \frac{\varepsilon}{\rho+\varepsilon} \ln q_{Mt} + \frac{\varepsilon}{\rho+\varepsilon} \ln \eta$$

where $\eta = \rho/\varepsilon$. This leads to the following cost function:

$$\begin{aligned} C_{it}(Y_{it}) &= c_{it} \left(\frac{Y_{it}}{A^*_{it} M^{*\rho}_{it}} \right)^{\frac{1}{\varepsilon}} + q_{Mt} M^*_{it} \\ &= \left(\frac{c^{\varepsilon}_{it} Y_{it}}{A^*_{it} \left[\eta^{\frac{\rho\varepsilon}{\rho+\varepsilon}} Y_{it}^{\frac{\rho}{\rho+\varepsilon}} A_{it}^{*\frac{-\rho}{\rho+\varepsilon}} c_{it}^{\frac{\rho\varepsilon}{\rho+\varepsilon}} q_{Mt}^{\frac{-\rho\varepsilon}{\rho+\varepsilon}} \right]} \right)^{\frac{1}{\varepsilon}} + q_{Mt} \left[\eta^{\frac{\varepsilon}{\rho+\varepsilon}} Y_{it}^{\frac{1}{\rho+\varepsilon}} A_{it}^{*\frac{-1}{\rho+\varepsilon}} c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}} q_{Mt}^{\frac{-\varepsilon}{\rho+\varepsilon}} \right] \\ &= q_{Mt}^{\frac{\rho}{\rho+\varepsilon}} c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}} A_{it}^{*\frac{-1}{\rho+\varepsilon}} Y_{it}^{\frac{1}{\rho+\varepsilon}} [\eta^{\frac{\varepsilon}{\rho+\varepsilon}} + \eta^{\frac{-\rho}{\rho+\varepsilon}}] = \theta q_{Mt}^{\frac{\rho}{\rho+\varepsilon}} c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}} A_{it}^{*\frac{-1}{\rho+\varepsilon}} Y_{it}^{\frac{1}{\rho+\varepsilon}} \end{aligned} \quad (23)$$

where

$$\theta = \eta^{\frac{\varepsilon}{\rho+\varepsilon}} + \eta^{\frac{-\rho}{\rho+\varepsilon}} = \frac{\rho+\varepsilon}{\rho}\eta^{\frac{\varepsilon}{\rho+\varepsilon}}$$

The factor demand functions can be derived from (23) by Shephards lemma:

$$\begin{aligned}
\ln H_{it}^* &= \frac{1}{\rho+\varepsilon}\left(\ln Y_{it} - \ln A_{it}^*\right) + \ln c_{it,h} \\
\ln L_{it}^* &= \frac{1}{\rho+\varepsilon}\left(\ln Y_{it} - \ln A_{it}^*\right) + \ln c_{it,l} \\
\ln M_{it}^* &= \frac{1}{\rho+\varepsilon}\left(\ln Y_{it} - \ln A_{it}^*\right) + \ln c_{it,M}
\end{aligned}$$

and

$$\begin{aligned}
c_{it,h} &= \theta \times q_{Mt}^{\frac{\rho}{\rho+\varepsilon}} \times \partial\left(c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}}\right)/\partial w_t^h \\
c_{it,l} &= \theta \times q_{Mt}^{\frac{\rho}{\rho+\varepsilon}} \times \partial\left(c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}}\right)/\partial w_t^l \\
c_{it,M} &= \theta \times c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}} \times \partial\left(q_{Mt}^{\frac{\rho}{\rho+\varepsilon}}\right)/\partial q_{Mt}
\end{aligned}$$

We see that

$$\begin{aligned}
\partial\left(c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}}\right)/\partial w_t^l &= \frac{\varepsilon}{\rho+\varepsilon}c_{it}^{\frac{-\rho}{\rho+\varepsilon}}\frac{\partial c_{it}}{\partial w_{it}^l} = \frac{\varepsilon}{\rho+\varepsilon}c_{it}^{\frac{-\rho}{\rho+\varepsilon}} a c_{it}^{1-r}(w_{it}^l)^{r-1} \\
\partial\left(c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}}\right)/\partial w_t^h &= \frac{\varepsilon}{\rho+\varepsilon}c_{it}^{\frac{-\rho}{\rho+\varepsilon}}\frac{\partial c_{it}}{\partial w_{it}^h} = \frac{\varepsilon}{\rho+\varepsilon}c_{it}^{\frac{-\rho}{\rho+\varepsilon}}(1-a)c_{it}^{1-r}(w_{it}^h)^{r-1} \\
\partial\left(q_{Mt}^{\frac{\rho}{\rho+\varepsilon}}\right)/\partial q_{Mt} &= \frac{\rho}{\rho+\varepsilon}q_{Mt}^{\frac{\rho-(\rho+\varepsilon)}{\rho+\varepsilon}} = \frac{\rho}{\rho+\varepsilon}q_{Mt}^{\frac{-\varepsilon}{\rho+\varepsilon}}
\end{aligned}$$

In particular, we obtain

$$\ln M_{it}^* = \frac{1}{\rho+\varepsilon}\left(\ln Y_{it} - \ln A_{it}^*\right) + \frac{\varepsilon}{\rho+\varepsilon}\ln c_{it} - \frac{\varepsilon}{\rho+\varepsilon}\ln q_{Mt} + \frac{\varepsilon}{\rho+\varepsilon}\ln\eta \tag{24}$$

Each firm is assumed to face the demand function:

$$Y_{it} = \Phi_{it}P_{it}^{-e} \Leftrightarrow P_{it} = \Phi_{it}^{\frac{1}{e}}Y_{it}^{\frac{-1}{e}} \tag{25}$$

with optimal Y_{it} given by:

$$\begin{aligned}
Y_{it}^* &= \arg\max_{Y_{it}}\left(P_{it}Y_{it} - C(Y_{it})\right) \\
&= \arg\max_{Y_{it}}\left(\Phi_{it}^{\frac{1}{e}}Y_{it}^{\frac{e-1}{e}} - \theta q_{Mt}^{\frac{\rho}{\rho+\varepsilon}} c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}} A_{it}^{*\frac{-1}{\rho+\varepsilon}} Y_{it}^{\frac{1}{\rho+\varepsilon}}\right)
\end{aligned}$$

The 1. order condition is:

$$\frac{e-1}{e}Y_{it}^{\frac{-1}{e}}\Phi_{it}^{\frac{1}{e}} = \theta q_{Mt}^{\frac{\rho}{\rho+\varepsilon}} c_{it}^{\frac{\varepsilon}{\rho+\varepsilon}} A_{it}^{*\frac{-1}{\rho+\varepsilon}} \frac{1}{\rho+\varepsilon} Y_{it}^{\frac{1-(\rho+\varepsilon)}{\rho+\varepsilon}}$$

$$\Updownarrow$$

$$\ln Y_{it}^* = \frac{-e(\rho+\varepsilon)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln\nu + \frac{\rho+\varepsilon}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln\Phi_{it} - \frac{\rho e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln q_{Mt}$$
$$- \frac{\varepsilon e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln c_{it} + \frac{e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln A_{it}^* \tag{26}$$

where

$$\nu = \frac{e}{e-1}\frac{\theta}{\rho+\varepsilon} = \frac{e}{\rho(e-1)}\eta^{\frac{\varepsilon}{\rho+\varepsilon}}$$

We do not observe output, but sales, S_{it}. Using $S_{it} = P_{it}Y_{it}^*$, we can rewrite (26) in terms of sales:

$$\begin{aligned}\ln S_{it} &= \frac{1}{e}\ln\Phi_{it} + \frac{e-1}{e}\ln Y_{it}^* \\ &= \ln\theta_S + \frac{1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln\Phi_{it} + \frac{e-1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln A_{it}^* \\ &- \frac{(e-1)\varepsilon}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln c_{it} - \frac{\rho(e-1)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln q_{Mt}\end{aligned}$$

where

$$\theta_S = \nu^{\frac{-(e-1)(\rho+\varepsilon)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}$$

From (24):

$$\begin{aligned}\ln M_{it}^* &= \frac{\varepsilon}{\rho+\varepsilon}\ln(\eta) + \frac{1}{\rho+\varepsilon}\left(\ln Y_{it}^* - \ln A_{it}^*\right) + \frac{\varepsilon}{\rho+\varepsilon}\ln c_{it} - \frac{\varepsilon}{\rho+\varepsilon}\ln q_{Mt} \\ &= \ln\theta_M + \frac{1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln\Phi_{it} + \frac{e-1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln A_{it}^* \\ &- \frac{(e-1)\varepsilon}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln c_{it} - \frac{\rho e+\varepsilon(e+\rho+\varepsilon-e(\rho+\varepsilon))}{(\rho+\varepsilon)(e+\rho+\varepsilon-e(\rho+\varepsilon))}\ln q_{Mt}\end{aligned}$$

where

$$\theta_M = \eta^{\frac{\varepsilon}{\rho+\varepsilon}}\nu^{-\frac{e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}.$$

We obtain:

$$\begin{aligned}\ln(q_{Mt}M_{it}^*) &= \ln\theta_M + \frac{1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln\Phi_{it} + \frac{e-1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln A_{it}^* \\ &- \frac{(e-1)\varepsilon}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln c_{it} - \frac{\rho(e-1)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln q_{Mt}\end{aligned}$$

and

$$V_{it} = A_{it}^{*\frac{e-1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}\Phi_{it}^{\frac{1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}} c_{it}^{-\frac{(e-1)\varepsilon}{e+\rho+\varepsilon-e(\rho+\varepsilon)}} q_{Mt}^{-\frac{\rho(e-1)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}(\theta_S - \theta_M)$$

Variable factor costs are given by:

$$\ln C(Y_{it}^*) = \ln\theta + \frac{\rho}{\rho+\varepsilon}\ln q_{Mt} + \frac{\varepsilon}{\rho+\varepsilon}\ln c_{it} - \frac{1}{\rho+\varepsilon}\ln A_{it}^* + \frac{1}{\rho+\varepsilon}\ln Y_{it}^* =$$
$$\ln\theta_C - \frac{(e-1)\varepsilon}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln c_{it} - \frac{\rho(e-1)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln q_{Mt} + \frac{e-1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln A_{it}^*$$
$$+\frac{1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}\ln\Phi_{it}$$

where

$$\theta_C = \theta\nu^{-\frac{e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}$$

Profit can be written as:

$$\Pi_{it} = A_{it}^{*\frac{e-1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}\Phi_{it}^{\frac{1}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}c_{it}^{-\frac{(e-1)\varepsilon}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}q_{Mit}^{-\frac{\rho(e-1)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}(\theta_S-\theta_C)$$

and profit as a share of V_{it} as:

$$\frac{\Pi_{it}}{V_{it}} = \frac{(\theta_S-\theta_C)}{(\theta_S-\theta_M)} = \frac{\nu^{\frac{-(e-1)(\rho+\varepsilon)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}} - \frac{\rho+\varepsilon}{\rho}\eta^{\frac{\varepsilon}{(\rho+\varepsilon)}}\nu^{\frac{-e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}}{\nu^{\frac{-(e-1)(\rho+\varepsilon)}{e+\rho+\varepsilon-e(\rho+\varepsilon)}} - \eta^{\frac{\varepsilon}{(\rho+\varepsilon)}}\nu^{\frac{-e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}} = \frac{\nu^{\frac{-e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}\eta^{\frac{\varepsilon}{\rho+\varepsilon}}\frac{e-(e-1)(\rho+\varepsilon)}{\rho(e-1)}}{\nu^{\frac{-e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}\eta^{\frac{\varepsilon}{\rho+\varepsilon}}\frac{e-(e-1)\rho}{\rho(e-1)}}$$

The manipulations in the above denominator shows that:

$$\widetilde{\theta} = \ln\left(\nu^{\frac{-e}{e+\rho+\varepsilon-e(\rho+\varepsilon)}}\eta^{\frac{\varepsilon}{\rho+\varepsilon}}\frac{e-(e-1)\rho}{\rho(e-1)}\right)$$

Moreover, Π_{it}/V_{it} can be simplified to:

$$\frac{\Pi_{it}}{V_{it}} = \frac{e-(e-1)(\rho+\varepsilon)}{e-(e-1)\rho} \in (\frac{1-\rho-\varepsilon}{1-\rho}, 1)$$

Appendix B. Data sources

Accounts statistics: All joint-stock companies in Norway are obliged to publish company accounts every year. The accounts statistics contain information obtained from the income statements and balance sheets of joint-stock companies, in particular, the information about operating revenues, operating costs and result, labour costs, the book values of a firm's tangible fixed assets at the end of a year, their depreciation, and write-downs.

The structural statistics: The term "structural statistics" is a general name for statistics of different industrial activities, such as manufacturing, building and construction, wholesale and retail trade statistics, etc. They all have the same structure and include information about production, input factors, and investments at the firm level. These structural statistics are organized according to the NACE standard and are based on General Trading Statements, which are given in an appendix to the tax return. The structural statistics contain data on purchases of tangible fixed assets and operational leasing. These data were matched with the data from the accounts statistics using the identification number given to the firm by the Register of Enterprises, which one of the Brønnøysund registers.

R&D statistics: R&D statistics are the survey data collected by Statistics Norway every second year up to 2001 and annually from then on. These data comprise detailed information about firms' R&D activities, in particular, about total R&D expenses with subdivision into internally performed R&D and externally performed R&D services, the number of employees engaged in R&D activities and the number of man-years worked in R&D. In each wave, the sample is selected using a stratified method for firms with 10–50 employees, while firms with more than 50 employees are all included. Strata are based on industry and number of employees. Each survey contains about 5,000 firms.

Register of Employers and Employees (REE): The REE contains information obtained from employers. All employers are obliged to send information to the REE about each individual employee's contract start and end, working hours, overtime and occupation. An exception is made only if a person works less than four hours per week in a given firm and/or was employed for less than six days. In addition, this register contains identification numbers for the firm and the employee, hence, the data can be aggregated to the firm level.

National Education Database (NED): The NED gathers all individually based statistics on education from primary to tertiary education and have been provided by Statistics Norway since 1970. We use this data set to identify the duration of education. This variable is constructed on the basis of the six-digit Norwegian Standard Classification of Education (NUS89), the leading digit of which is the code for the educational level of the person. There are nine educational levels in addition to the major group for "unspecified length of education." Education levels are given in Table 7 in Appendix C.

Appendix C: Supplementary tables and figures

Table 7: **Educational levels**

Subdivision of levels	**Level**	**Class level**
	0	Under school age
Primary education	1	1st – 7th
	2	8th – 10th
Secondary education	3	11-12th
	4	12th – 13th
	5	14th – 17th
Postsecondary education	6	14th – 18th
	7	18th – 19th
	8	20th+
	9	Unspecified

Source: Statistics Norway

TABLE 8: **Descriptive statistics for the main variables used in the final sample**

Variable	Firm-years	Mean	Median	IQ range	
L_{it}					
All firms	71,521	80.6	23	11	60
R&D-begin.[1]	15,522	43.7	15	7	32
R&D-incumb.[2]	27,944	129.4	43	17	105
$L_{it}^{(1)}/L_{it}$	71,521	.134	.038	0	.169
$L_{it}^{(2)}/L_{it}$	71,521	.583	.591	.453	.722
$L_{it}^{(3)}/L_{it}$	71,521	.283	.251	.104	.438
V_{it}/L_{it}[3]					
All firms	68,892	1803.0	853.8	618.7	1200.6
R&D-begin.	15,265	1055.7	786.3	560.3	1121.8
R&D-incumb.	27,408	2855.4	887.1	652.9	1249.9
I_{it}/L_{it}[4]					
All firms	45,968	117.4	8.64	0	94.9
R&D-begin.	3,886	203.1	39.4	4.13	187.3
R&D-incumb.	25,880	108.6	9.99	0	82.9
$I_{i,t-1}/F_{it}$[4]					
All firms	47,575	.224	.117	0	.316
R&D-begin.	4,052	.538	.530	0	1
R&D-incumb.	26,816	.163	.109	0	.217
f_{it}/V_{it}[4]					
All firms	46,442	.048[5]	.031	.008	.118
R&D-begin.	3,956	.029[5]	.029	.007	.110
R&D-incumb.	25,977	.050[5]	.034	.008	.123
k_{it}/V_{it}[6]					
All firms	65,825	.244[5]	.033	.009	.107
R&D-begin.	14,320	.167[5]	.027	.007	.099
R&D-incumb.	26,682	.271[5]	.042	.012	.122
Π_{it}/V_{it}[3]					
All firms	69,967	.455[5]	.392	.209	.358
R&D-begin.	15,513	.311[5]	.187	.064	.341
R&D-incumb.	27,861	.502[5]	.215	.086	.372

Note: V, I, F, f, k, Π denote, respectively, value added, $R\&D$ investments ($I = int+x$), R&D capital stock, R&D capital services (user cost equivalents of the stock variable F), tangible capital services (user cost equivalents of the stock variable K) and gross operating profits (before depreciation and amortization) in NOK 1000 fixed 2017 prices

[1] R&D active ($F_{it} > 0$) in max 3 years in the period 2001-2018 ($T_i \leq 3$)

[2] R&D active more than 12 years in the period 2001-2018 ($T_i > 12$)

[3] Conditional on $V_{it} > 0$

[4] Conditional on $F_{it} > 0$

[5] Weighted average with weight proportional to V_{it}, i.e., sum of nominator (over i and t) divided by sum of denominator

[6] Conditional on $K_{it} > 0$ and $V_{it} > 0$

Table 9: **Estimates of the coefficients of the value added equation with Cobb-Douglas production function ($\lambda = 0$) and different definitions of R&D investment. Robust standard errors (SE)**

Indep. variables in structural equation		**GMM-estimates CMC**		**GMM-estimates R-CMC**		**FE-estimates R-CMC**	
Symmetric treatment of intramural and extramural R&D ($I = int + x$)							
	Coeff.	**Est.**	**SE**	**Est.**	**SE**	**Est.**	**SE**
$\ln V_{i,t-1}$	ϕ	.253	.030***	.262	.032***		
$-\ln c_{it}$	$\widetilde{\varepsilon}$	.526	.322*	.212	.327	.642	.120***
$\ln K_{it}$	$\widetilde{\gamma}$	.119	.027***	.119	0.028***	.160	.008***
$\ln r_{it}(\lambda)$	$\widetilde{\beta}$			.141	0.045***	.099	.008***
$\overline{\psi}_i(\lambda) \ln r_{it}(\lambda)^1$	R	.154	.083***				
$\overline{\psi}_i(\lambda) 1_{(T_i \leq 3)} \ln r_{it}(\lambda)^2$	τ_{begin}	-.464	.221 **				
$\overline{\psi}_i(\lambda) 1_{(T_i \in [4,12])} \ln r_{it}(\lambda)$	τ_{exper}	-.101	.119				
Only intramural R&D treated as investment ($I = int$)							
	Coeff.	**Est.**	**SE**	**Est.**	**SE**	**Est.**	**SE**
$\ln V_{i,t-1}$	ϕ	.255	.030***	.275	.033***		
$-\ln c_{it}$	$\widetilde{\varepsilon}$	.636	.326*	.273	.329	.708	.129***
$\ln K_{it}$	$\widetilde{\gamma}$	.101	.027***	.104	.028***	.161	.008***
$\ln r_{it}(\lambda)$	$\widetilde{\beta}$			.191	.241***	.110	.009***
$\overline{\psi}_i(\lambda) \ln r_{it}(\lambda)$	R	.259	.094***				
$\overline{\psi}_i(\lambda) 1_{(T_i \leq 3)} \ln r_{it}(\lambda)$	τ_{begin}	-.641	.234***				
$\overline{\psi}_i(\lambda) 1_{(T_i \in [4,12])} \ln r_{it}(\lambda)$	τ_{exper}	.013	.126				

Note: Estimated on subsample of observations with $F_{it} > 0$. Windmeijer (2005) robust standard errors (SE); *,**,*** refer, respectively, to significant estimates at the 10, 5, and 1 percent level.

[1] $\overline{\psi}_i(\lambda)$ refers to the firm's average R&D intensity, as defined in Equation (15).

[2] T_i is the number of years with $F_{it} > 0$ in the years 2001-2018.