the review of income and wealth

Review of Income and Wealth
Series 67, Number 4, December 2021
DOI: 10.1111/roiw.12492

PRODUCTIVITY MEASUREMENT, R&D ASSETS, AND MARK-UPS IN OECD COUNTRIES

BY PAUL SCHREYER* and MARÍA BELÉN ZINNI

OECD

A key feature of the 2008 revision of the System of National Accounts was the treatment of research and development (R&D) expenditure as investment. The question arises whether the standard approach toward accounting for growth contribution of assets is justified, given the special nature of R&D that provides capital services by affecting the working of other inputs as a whole—akin to technical change and often requires up-front investment with sunk costs. We model R&D inputs with a restricted cost function and compare econometric estimates with those derived under a standard index number approach but find no significant differences. However, we cannot reject the hypothesis of increasing returns to scale. The standard multi-factor productivity (MFP) measure is then broken down into a scale effect and a residual productivity effect, each of which explains about half of overall MFP change. The scale effect points to the importance of the demand side and market size for productivity growth. We also compute mark-up rates of prices over marginal cost and find widespread evidence of rising mark-ups for the period 1985–2017.

JEL Codes: D24, E01

Keywords: aggregate productivity, mark-ups, R&D

1. INTRODUCTION

One of the central achievements of the 2008 revision of the System of National Accounts (2008 SNA, European Commission *et al.*, 2009) was the treatment of research and development (R&D) expenditure as an investment that gives rise to knowledge assets.[1]

With the completed implementation of the 2008 SNA among OECD countries by the end of 2016, users of statistics now dispose of sets of estimates for the investment in R&D as well as software (already present since the 1993 revision of the SNA) along with estimates of other, more traditional nonfinancial, produced assets (machinery, equipment, and structures). As all these assets provide inputs

Note: This paper represents the opinions of the authors and not necessarily those of the OECD or its members. The authors thank Erwin Diewert, Kevin Fox, Susanto Basu, two anonymous referees, and participants at the 2018 Economic Measurement Group at the UNSW Sydney and the 2019 meeting of the Society for Economic Measurement for many helpful comments. All shortcomings remain those of the authors.
*Correspondence to: Paul Schreyer (Paul.SCHREYER@oecd.org).

[1]The specifics of measuring R&D expenditure are laid down in detail in the Frascati Manual (OECD, 2015). How the intellectual property assets that are the fruit of R&D investment should be measured in practice has been elaborated in OECD (2010).

into production in the form of capital services, it is only natural to base productivity estimates on the whole set of assets. Indeed, the economics literature has preceded national accounts standards and embraced an even broader set of intangibles in an attempt to account for new sources of economic growth and competitiveness. The work by Corrado *et al.* (2005), who measured intangible capital for the USA and used it in a new set of productivity estimates, was a seminal piece that spawned other work, applying similar or refined concepts to other countries and time periods (OECD, 2013; Goodridge and Haskel, 2016).

However, there are several issues when it comes to using R&D assets in productivity measurement. First, R&D projects often involve sunk costs and up-front investment. These sunk costs need to be recuperated over the economic service life of the R&D asset, requiring a mark-up over marginal costs of production.[2] Sunk costs thus imply a certain degree of increasing returns to scale at the firm level. Increasing returns may also arise at the aggregate level because of externalities and spill-overs generated by R&D assets:[3] "The level of productivity achieved by one firm or industry depends not only on its own research efforts but also on the level or pool of general knowledge accessible to it" (Griliches, 1995, p. 63). The implication for measurement is again that aggregate returns to scale may not be constant but increasing. Another reason for non-constant returns to scale at the aggregate-level first objective of the analysis here is to test for the presence of increasing returns to scale and to measure the evolution of mark-ups. We shall conclude that the hypothesis of increasing returns at the aggregate level cannot easily be rejected, and there is a pattern of rising mark-ups in nearly all countries of the sample. We hasten to add that the presence of rising returns to scale at the aggregate level is not necessarily the proof of rising returns at the firm or industry level (Basu and Fernald, 1997).

A second issue associated with R&D capital is how its services enter the production process and the consequences for productivity measurement. This was highlighted in the work by Parham (2006), Pitzer (2004), and Huang and Diewert (2011). Pitzer (2004) observed that R&D capital functions as a source of "recipes." Diewert and Huang (2011) started their discussion of R&D assets by explaining that "we do not treat the stock of R&D capital as an explicit input factor. Rather, we define the stock of R&D capital to be a technology index that locates the economy's production frontier. An increase in the stock of R&D shifts the production frontier outwards" (p. 389). R&D assets thus provide capital services by *enabling* production, for example, through licenses that permit usage of knowledge or intellectual property (IP) in production. This suggests treating capital services from R&D assets as a technology index that affects the working of all other inputs *as a whole* so that R&D capital services operate akin to autonomous neutral technical change.

If one adopts this reasoning, production takes place with services from non-R&D inputs conditional on a given stock of R&D assets (and conditional on a

[2]A similar issue arises with other assets. Diewert and Fox (2016) treated the case of land where, once a building is constructed on it, initial fixed costs must be distributed across the lifetime of the structure.

[3]For an overview of the literature, see Sena (2004).

given level of "autonomous" technical change). This amounts to treating R&D capital as a quasi-fixed input. The theoretical tools to deal with quasi-fixity have long been developed in the form of restricted profit and restricted cost functions (Lau, 1976; McFadden, 1978; Berndt and Fuss, 1986; Schankerman and Nadiri, 1984). When an input is quasi-fix, it cannot be adjusted instantaneously—a plausible notion for R&D assets with sometimes-long gestation periods. One consequence is that the assumption of period-to-period cost minimizing behavior of producers with regard to the quasi-fixed factor of production no more holds. Then, the user costs for R&D assets as constructed under standard cost-minimizing assumptions cannot be used to approximate production elasticities of R&D (or cost elasticities in a dual formulation). Exclusive reliance on an index number approach is no more possible, and R&D production elasticities must be estimated econometrically.

We use data for 20 OECD countries over the period 1985–2017 and estimate cost elasticities of R&D capital to test whether these diverge significantly from the standard nonparametric elasticities. Although there are variations across countries and over time, it turns out that on average the econometric point estimate lines up rather well with the standard index number results that do not assume quasi-fixity of the R&D input. This is in particular the case when we allow for nonconstant returns to scale at the same time. Therefore, we will conclude that the theoretical case for treating R&D assets as quasi-fixed inputs does not outweigh the practical disadvantages that it entails and that can be avoided with an index number approach. There is in particular the need to revert to econometric techniques that reduce reproducibility of results, and the need to accept constancy of R&D elasticities over time and across countries—at least in a case where the number of observations is limited.

A third—and related—issue is how exactly to construct R&D capital stocks. Unlike other assets, market prices for R&D investment are hard to come by, given that much R&D activity is undertaken within firms ("own account investment") with the consequence that R&D investment is valued at cost. Similarly, prices of the capital services from R&D assets are essentially reflective of the prices of the inputs used in their creation, much of it being the wage rate of R&D personnel. Another tricky point is how to determine depreciation. As Diewert and Fox (2016) pointed out, the pattern of depreciation allocations for assets that constitute partially or entirely sunk costs is largely determined by the cash flows that the asset generates over its lifetime, rather than a reduction in the value of capital services as is the case for many other assets. Traditional depreciation models for structures may therefore not provide adequate descriptions of true user costs. This is an added reason for testing whether cost shares as traditionally measured are reflective of cost elasticities of R&D, as explained earlier.

However, measurement problems do not stop with valuation of the asset. Finally, because R&D assets are intangible, they can easily be transferred, even across national borders.[4] R&D assets can therefore appear and disappear in lumps,

[4]A widely discussed example is Ireland where there are movements of IP assets within multinational enterprises into Ireland, and the resulting production and income flows resulted in a staggering 25 percent rise in real GDP in 2015 and a similar unusual two-digit growth in labor productivity for Ireland. Although Ireland may have brought the issue of measuring production and productivity into sharp focus, this constitutes by no means a unique case.

leading to corresponding changes in measured capital stocks and services. Large additions or subtractions from stocks require careful construction of the measures of R&D stocks with attention paid to infra-annual movements: whether an asset appears at the beginning or at the end of an accounting year is no more an ancillary measurement question. Annex A describes at some detail how we proceeded with the measurement of R&D stocks. All our measurement proposals are consistent with the 2008 SNA and fit also with the broader blueprint of productivity measurement in a national accounts framework as developed by Jorgenson and Landefeld (2004).

This paper is organized as follows. Section 2 deals with productivity measurement under nonconstant returns to scale and a quasi-fixed R&D input. In Section 3, we follow Diewert *et al.* (2011) and combine index number and econometric approaches to derive a parsimonious way of testing for quasi-fixity of the R&D input and nonconstancy of returns to scale. As our results regarding quasi-fixity are inconclusive, and in light of many practical considerations, we opt for a treatment of R&D as a standard flexible input. However, we do maintain the finding of increasing returns to scale, and the last part of Section 3 uses these results to decompose the OECD multi-factor productivity (MFP) index into a part that reflects scale effects and a part that reflects residual productivity change. The section finishes with the dual picture to the MFP decomposition and mark-ups over marginal and average costs. Section 4 concludes.

2. IP Assets in Productivity Measurement

2.1. *R&D in Production*

We characterize technology by a production function where labor and non-R&D capital inputs are combined with services from a knowledge asset to produce aggregate output:

$$Q = f_Q(X, R, t), \tag{1}$$

where Q is the volume of aggregate output; $X \equiv [X_1, X_2, ...]$ is the vector of labor and various types of non-R&D capital; R is the stock of R&D; and t is the time variable to capture autonomous productivity change. $f_Q(X, R, t)$ is continuous and nondecreasing in inputs X,R, and t. No constant returns are imposed here. This is motivated by the desire to maintain a general approach and also by the nature of R&D: its creation typically entails large, fixed up-front investment expenditure that needs to be recuperated over the economic service life of the asset. The implication is that prices will not be set at short-run marginal costs of production. There may also be mark-ups on marginal costs above and beyond those needed for cost recovery—a point to which we shall return in greater detail later.

In addition to allowing for nonconstant returns to scale, we treat R as a quasi-fixed input in the sense of McFadden (1978), Schankerman and Nadiri (1984), or Berndt and Fuss (1986). As a quasi-fixed input, R takes the role of a predetermined variable that cannot be adjusted instantaneously and in a cost-minimizing manner as is usually assumed in productivity measurement. By treating the quantity of R&D as a predetermined, exogenous variable, it can also be interpreted as a "shifter" to non-R&D input requirements, similar to autonomous productivity change that is captured by the time variable t.[5] For non-R&D inputs X, the usual assumption of instantaneous cost-minimizing adjustment is maintained.

The production function above characterizes technology and can be used as the framework for measuring autonomous technical change as the shift of the production function or the extra output that a given input bundle can produce with the passage of time. Alternatively, a cost function can be used to characterize a production unit's technology. Then, autonomous technical change is measured as the shift of the cost function, or the reduction in costs to produce a given output, for given input prices. Primal (production function)-based and dual (cost function)-based productivity measures coincide when production is characterized by constant returns to scale, when production is efficient and when producers minimize costs. However, primal and dual measures will diverge when one or several of these conditions fail to hold.[6] Similarly, the degree of returns to scale in production can be measured based on the production or on the cost function. Diewert *et al.* (2011) pointed to the strong intuitive appeal of a cost-based measure of scale elasticity as the percentage change in total cost because of a 1 percent increase in the quantity of output, for a given level input prices.[7] Furthermore, cost-based productivity measures allow for a simple setup of producer behavior on output markets when competition is imperfect. Therefore, we shall make use of the following restricted (variable) cost function:

$$(2) \qquad C(Q, w_X, R, t) = \min_X \left(\sum_i w_{Xi} X_i \,|\, f_Q(X, R, t) \geq Q \right) = \sum_i w_{Xi} X_i.$$

The general properties of the restricted cost function were established by Lau (1976) and McFadden (1978). Early empirical references that used the variable cost function include, in particular, Caves *et al.* (1981), Schankerman and Nadiri (1984), Berndt and Fuss (1986), and Morrison (1992). $C(Q, w_X, R, t)$ reflects the minimum variable cost of producing Q, given a vector of input prices w_X, and a level of knowledge assets R as well as autonomous, "costless" technology t. One notes that (2) assumes cost minimization by producers only regarding X and is conditional on a level of R and t. The second equality in (2) states that minimized variable costs equal observed variable costs $\sum_i w_{Xi} X_i$. We thus abstract from cases of waste or inefficient production where actual costs exceed minimum costs. $C(Q, w_X, R, t)$ captures short-run variable costs.

[5]Formally, this requires treating R (or t) as separable from X so that the rate by which a change in R (or t) affects output is independent of the rates of substitution between the elements of X. The concept of weak separability is because of Sono (1961) and Leontief (1947). Separability is a rather restrictive assumption, but Diewert (1980) offered a way forward with his *Method III* (p. 455 ff.), where he showed that price and quantity indices can be constructed using observable prices and quantities only if one is ready to accept that these aggregates are conditional on reference values of variables outside the aggregate (R or t in the case at hand) that are averages of their realizations in comparison periods.

[6]See Balk (1998) for a comprehensive overview of the various primal and dual productivity measures and their relationship.

[7]Although not relevant for the present case where we consider an aggregate measure of output, a cost function-based measure of the returns to scale has the advantage of easily allowing for changes in the composition of output.

Shephard's (1953) Lemma holds for the variable cost function: for non R&D inputs X_i ($i = 1, 2, \ldots,$) factor demand equals marginal cost changes associated with a change in input prices: $\partial C(Q, w_X, R, t)/\partial w_{Xi} = X_i(Q, w_X, R, t)$. For the R&D input, we define a shadow price w_{RS} as the marginal reduction in variable costs because of a marginal increase in R: $\partial C(Q, w_X, R, t)/\partial R \equiv -w_{RS}$. This shadow price (or rather, shadow user cost) of R&D is unknown and may or may not be close to the computable user cost of R&D, w_R whose measurement is isomorphic to the user costs of other produced assets. The shadow price w_{RS} can only be evaluated econometrically, whereas w_R lends itself to an index number approach.

To derive a measure of technical change, we start by differentiating (2) totally and obtain a continuous time expression for the growth rate of short-run variable costs.

$$\begin{aligned}(3)\qquad &\frac{dlnC(Q, w_X, R, t)}{dt} = \\ &\frac{\partial lnC(Q, w_X, R, t)}{\partial lnQ}\frac{dlnQ}{dt} + \sum_i \frac{\partial lnC(Q, w_X, R, t)}{\partial lnw_{Xi}}\frac{dlnw_{Xi}}{dt} \\ &+ \frac{\partial lnC(Q, w_X, R, t)}{\partial lnR}\frac{dlnR}{dt} + \frac{\partial lnC(Q, w_X, R, t)}{\partial t}.\end{aligned}$$

The cost elasticity of output is the definition of (inverted) returns to scale, and we shall denote $\partial lnC(Q, w_X, R, t)/\partial lnQ \equiv 1/\epsilon$. Thus, there are increasing, constant, or decreasing returns to scale in short-term variable costs if ε exceeds, is equal to, or is smaller than one. The last expression in (3), $\partial lnC(Q, w_X, R, t)/\partial t$, captures the short-run measure of autonomous technical change or the shift of the restricted cost function over time. With Shepard's Lemma and the definition of the R&D shadow price, and using simplified notation by setting $C(Q, w_X, R, t) = C$, (3) is rewritten as

$$(4)\qquad \frac{dlnC}{dt} = \frac{1}{\epsilon}\frac{dlnQ}{dt} + \sum_i \frac{w_{Xi}X_i}{C}\frac{dlnw_{Xi}}{dt} - \frac{w_{RS}R}{C}\frac{dlnR}{dt} + \frac{\partial lnC}{\partial t}.$$

Next, define a Divisia quantity index of non-R&D inputs, $dlnX/dt$, that equals the Divisia index of deflated variable input costs:

$$(5)\qquad \frac{dlnX}{dt} \equiv \sum_i \frac{w_{Xi}X_i}{C}\frac{dlnX_i}{dt} = \frac{dlnC}{dt} - \sum_i \frac{w_{Xi}X_i}{C}\frac{dlnw_{Xi}}{dt}.$$

Combining (4) and (5) gives rise to the following two, equivalent expressions:

$$\begin{aligned}(6)\qquad &\frac{dlnX}{dt} = \frac{1}{\epsilon}\frac{dlnQ}{dt} - \frac{w_{RS}R}{C}\frac{dlnR}{dt} + \frac{\partial lnC}{\partial t}; \\ &\frac{dlnQ}{dt} = \epsilon\left(\frac{dlnX}{dt} + \frac{w_{RS}R}{C}\frac{dlnR}{dt} - \frac{\partial lnC}{\partial t}\right).\end{aligned}$$

The first line in (6) states that non-R&D input growth depends positively on output growth, and negatively on the growth of R&D and time-autonomous technical change ($\partial lnC/\partial t \leq 0$)—fewer inputs are needed for a given output when technology and R&D inputs increase. The second line in (6) converts this into a growth accounting equation, where output growth is explained by the combined growth of non-R&D inputs, R&D inputs, and time-autonomous technical change. Combined inputs and technical change are augmented by the degree of short-run returns to scale.

To compare the short-run (restricted) relationships in (6) with their long-run (unrestricted) counterparts,[8] we define an unrestricted cost function $C^*(Q, w_X, w_R, t)$. Here, the shadow price of R&D equals its computable user costs ($w_{RS} = w_R$), and demand for the R&D input $R(Q, w_X, w_R, t)$ is always in equilibrium, implicitly defined via

$$\frac{-\partial C(Q, w_X, R, t)}{\partial R} = w_R. \tag{7}$$

The full expression for the unrestricted cost function is

$$C^*(Q, w_X, w_R, t) = C(Q, w_X, R(Q, w_X, w_R, t), t)) + w_R R(Q, w_X, w_R, t). \tag{8}$$

It is now possible to derive the relationship between restricted and unrestricted elasticities (Schankerman and Nadiri, 1984) by differentiating (8) and making use of (7).

$$\begin{aligned}
\frac{\partial lnC^*}{\partial lnQ} &= \frac{\partial lnC}{\partial lnQ}\frac{C}{C^*} = \frac{1}{\epsilon}\frac{C}{C^*} \equiv \frac{1}{\epsilon^*};\\
\frac{\partial lnC^*}{\partial t} &= \frac{\partial lnC}{\partial t}\frac{C}{C^*};\\
\frac{\partial lnC^*}{\partial lnw_{Xi}} &= \frac{\partial lnC}{\partial w_{Xi}}\frac{C}{C^*} = \frac{w_{Xi}X_i}{C}\frac{C}{C^*} \quad i = 1, 2, ...\\
\frac{\partial lnC^*}{\partial lnw_R} &= \frac{w_R R}{C^*}.
\end{aligned} \tag{9}$$

The passage between unrestricted and restricted cost functions and the associated measures of productivity, returns to scale, and cost elasticities of non-R&D inputs is thus rather straightforward and achieved by multiplying the short-term expressions by C/C^*, the share of non-R&D inputs in total costs. For instance, expanding the second line in (6) by C/C^* yields

[8]"Long-run" results here need not be understood in the sense of long-run steady-state results as developed, for instance, by Jones (1995) who challenges the scale effect of Schumpeterian endogenous growth theory but deals with a very long view, certainly exceeding the 32-year period in the present analysis.

$$
\begin{aligned}
\frac{dlnQ}{dt} &= \frac{\epsilon}{C/C^*}\left(\frac{C}{C^*}\frac{dlnX}{dt}+\frac{w_{RS}R}{C}\frac{C}{C^*}\frac{dlnR}{dt}-\frac{C}{C^*}\frac{\partial lnC}{\partial t}\right)\\
&= \epsilon^*\left(\frac{C}{C^*}\frac{dlnX}{dt}+\frac{w_{RS}R}{C^*}\frac{dlnR}{dt}-\frac{\partial lnC^*}{\partial t}\right)\\
&= \epsilon^*\left(\frac{dlnZ}{dt}-\frac{\partial lnC^*}{\partial t}\right).
\end{aligned} \tag{10}
$$

Here we have defined the short-run Divisia quantity aggregate of all inputs as $\frac{dlnZ}{dt}\equiv(\frac{C}{C^*}\frac{dlnX}{dt}+\frac{w_{RS}R}{C^*}\frac{dlnR}{dt}$). Similarly, we can define an unrestricted, long-run Divisia quantity aggregate of inputs as $\frac{dlnZ^*}{dt}\equiv(\frac{C}{C^*}\frac{dlnX}{dt}+\frac{w_RR}{C^*}\frac{dlnR}{dt})$.

The OECD measures MFP growth as the difference between output and aggregate input growth (OECD, 2019; Schreyer *et al.*, 2003; Schreyer, 2010). This MFP growth can now be broken down into three effects: one that captures the difference between restricted and unrestricted measures of inputs, one that captures the effect of returns to scale, and one that captures technical change.

$$
\begin{aligned}
MFP &\equiv \frac{dlnQ}{dt}-\frac{dlnZ^*}{dt}\\
&= \epsilon^*\left(\frac{dlnZ}{dt}-\frac{\partial lnC^*}{\partial t}\right)-\frac{dlnZ^*}{dt} \quad \text{using (10)}\\
&= \epsilon^*\frac{dlnZ}{dt}-\epsilon^*\frac{\partial lnC^*}{\partial t}-\frac{dlnZ^*}{dt}-\epsilon^*\frac{dlnZ^*}{dt}+\epsilon^*\frac{dlnZ^*}{dt}\\
&= \epsilon^*\left(\frac{dlnZ}{dt}-\frac{dlnZ^*}{dt}\right)+(\epsilon^*-1)\frac{dlnZ^*}{dt}-\epsilon^*\frac{\partial lnC^*}{\partial t}\\
&= \epsilon^*\left(\frac{w_{RS}R}{C^*}-\frac{w_RR}{C^*}\right)\frac{dlnR}{dt}+(\epsilon^*-1)\frac{dlnZ^*}{dt}-\epsilon^*\frac{\partial lnC^*}{\partial t}.
\end{aligned} \tag{11}
$$

When shadow elasticities of R&D equal computable user cost shares ($\frac{w_{RS}R}{C^*}=\frac{w_RR^*}{C^*}, \frac{dlnZ}{dt}=\frac{dlnZ^*}{dt}$), the first term in the last line of (11) vanishes, and MFP growth is reduced to a scale effect and to a technical change effect. Equation (12) presents the same MFP decomposition in a slightly different form and confirms that with constant returns to scale ($\epsilon^*=1$), MFP simply equals the shift in the cost function:

$$
\begin{aligned}
MFP &= (\epsilon^*-1)\frac{dlnZ^*}{dt}-\epsilon^*\frac{\partial lnC^*}{\partial t} \quad \text{for } \frac{w_{RS}R}{C^*}=\frac{w_RR^*}{C^*}\\
&= \left(1-\frac{1}{\epsilon^*}\right)\frac{dlnQ}{dt}-\frac{\partial lnC^*}{\partial t}\\
&= -\frac{\partial lnC^*}{\partial t} \quad \text{for } \epsilon^*=1.
\end{aligned} \tag{12}
$$

2.2. *Capturing Mark-Ups*

Output prices that are equal to marginal variable costs (of non-R&D inputs) are insufficient to recover the fixed costs that may have been needed to generate or purchase the R&D asset in the first place. Even prices that are equal to total marginal costs may not cover average costs in the presence of longer-term increasing returns to scale. Thus, there has to be a mark-up over total marginal costs. There may also be an additional mark-up M above and beyond the average costs, that is, what is needed to avoid losses. Its level will depend on market conditions and on the degree of competition under which Q is sold. This additional mark-up could also reflect returns to other, unmeasured assets. We shall return to the interpretation of mark-ups when presenting results.

To place M into context, we recall the accounting relationship for value-added and nominal aggregate output $P_Q Q$

$$P_Q Q = \sum_i w_{Xi} X_i + w_{RS} R + M = \sum_i w_{Xi} X_i + w_R R + M^*, \tag{13}$$

where $P_Q Q$ represents the total value-added (GDP at the economy-wide level) and $\sum_i w_{Xi} X_i$ is the value of non-R&D inputs. Both are measurable. In the short-term restricted case where R commands the shadow price w_{RS}, the sum $w_{RS} R + M$ can be observed but cannot be broken into its parts. In the unrestricted case, the cost of R&D services is measured through $w_R R$, and M^*, the longer-run mark-up over average costs, can be derived residually.

Let the mark-up rate m of prices over marginal costs in the restricted case and let the mark-up rate m^* of prices over marginal costs in the unrestricted case be defined by the following relationship:

$$\begin{aligned}
P_Q &= \frac{\partial C}{\partial Q}(1+m) \text{ from which it follows that} \\
\frac{P_Q Q}{C} &= \frac{\partial C}{\partial Q}\frac{Q}{C}(1+m) = \frac{1}{\epsilon}(1+m) \text{ for the restricted case; and} \\
\frac{P_Q Q}{C^*} &= \frac{1}{\epsilon^*}(1+m^*) \text{ for the unrestricted case such that} \\
(1+m^*) &= \epsilon^* \frac{P_Q Q}{C^*} = \epsilon^* [1 + M^* / C^*] = \epsilon^* \frac{1}{1 - M^* / P_Q Q}.
\end{aligned} \tag{14}$$

The last line in (14) reproduces a well-known identity: (one plus) the mark-up rate over marginal costs equals the degree of returns to scale times (one plus) the average mark-up rate M^*/C^* or an expression that rises with the profit rate $M^*/P_Q Q$. In the absence of residual profits, ($M^* = 0$), the mark-up rate over marginal costs will equal returns to scale. When $M^* > 0$ and there are constant returns to scale ($\epsilon^* = 1$), all mark-ups will reflect residual profits.

3. Empirical Implementation

3.1. *R&D Cost Shares: Too Low, Too High, about Right?*

Although the relationships above were derived in continuous time, actual data come in discrete form—annual observations in the case at hand—and the relevant relationships need to be expressed in discrete form. We use Törnqvist indices to express equations (6) in discrete time.[9]

$$\text{(15)}\quad \begin{aligned} \Delta lnX^t &= \frac{1}{\epsilon}\Delta lnQ^t - 0.5\left(\frac{w_{RS}^t R^t}{C^t} + \frac{w_{RS}^{t-1} R^{t-1}}{C^{t-1}}\right)\Delta lnR^t - \Delta\pi^t \\ \Delta lnQ^t &= \epsilon\left[\Delta lnX^t + 0.5\left(\frac{w_{RS}^t R^t}{C^t} + \frac{w_{RS}^{t-1} R^{t-1}}{C^{t-1}}\right)\Delta lnR^t + \Delta\pi^t\right]. \end{aligned}$$

In (15), $\Delta lnX^t \equiv lnX^t - lnX^{t-1}$ denotes the logarithmic growth rate of X between periods t and $t-1$, and the same notation is used for the other variables. The relations in (15) will constitute the main vehicle to assess shadow prices of R&D inputs, short-run returns to scale, and technical change. Note that in (15) the unknown terms are $\epsilon, 0.5\left(\frac{w_{RS}^t R^t}{C^t} + \frac{w_{RS}^{t-1} R^{t-1}}{C^{t-1}}\right)$ and $\Delta\pi^t$ that will need to be estimated. This requires assuming constancy of $0.5\left(\frac{w_{RS}^t R^t}{C^t} + \frac{w_{RS}^{t-1} R^{t-1}}{C^{t-1}}\right)$. The non-R&D input aggregate ΔlnX^t is measured via index numbers, derived from the restricted cost function. This hybrid approach is because of Diewert *et al.* (2011) who applied it for estimates of returns to scale in Japanese manufacturing, albeit with an unrestricted cost function. Main advantages of the hybrid approach are parsimony in the number of parameters to be estimated and a strong theoretical basis as relations are directly derived from flexible functional forms. In a world of perfect data and producer behavior that is fully in line with economic theory, it would suffice to estimate either the first or the second equation of (15). However, measurement errors will lead to different results depending on whether the direct or the reverse formulation is estimated as further discussed here.[10] Reformulating (15) for estimation purposes:

$$\text{(16)}\quad \begin{aligned} \Delta lnX^t &= \alpha_{a0} + \alpha_{a1}\Delta lnQ^t + \alpha_{a2}\Delta lnR^t + \mu_a^t \\ \Delta lnQ^t &= \alpha_{b0} + \alpha_{b1}\Delta lnX^t + \alpha_{b2}\Delta lnR^t + \mu_b^t. \end{aligned}$$

In (16), we have assumed that time autonomous productivity change follows a stochastic process around a long-term average: $-\Delta\pi^t = \alpha_{a0} + \mu_a^t$ in the first expression of (16) and $\Delta\pi^t/\epsilon = \alpha_{b0} + \mu_b^t$ in the second expression of (16) with productivity shocks μ_a^t and μ_b^t. A well-known and long-standing issue in the estimation of production or cost functions is that productivity shocks are correlated with factor inputs, thus creating an endogeneity problem when (16) is estimated. Estimation of the reverse regression does not solve the issue[11]—the R&D input

[9]This can be justified more rigorously by assuming that the restricted cost function is of the translog form (introduced by Christensen *et al.*, 1971 and generalized by Diewert,1974). As a flexible functional form, it approximates an arbitrary cost function to the second degree. As Diewert (1974, 1976) has shown, a Törnqvist index is then an exact representation of the change in the cost function.

[10]Note that (15) is not a system of simultaneous equations.

[11]See Berndt (1976) for an early discussion and application.

still figures as an independent variable with potential correlation with μ_b^t. We use time dummies and country-specific fixed effects in the error term to at least partially address this issue.

Instrumental variables are another avenue toward addressing the endogeneity problem. At the same time, they tend to give rise to other problems. Diewert and Fox (2008) provided an in-depth discussion of estimation in a similar context and note in regard to the use of instrumental variables: "Since different researchers will choose a wide variety of instrument vectors …, it can be seen that the resulting estimates … will not be reproducible across different econometricians who pick different instrument vectors" (p.186). Reproducibility and simplicity are major concerns in the present setting as our work aims at providing guidance for producing periodic productivity statistics, typically by National Statistical Offices. Instrumental variables may also introduce other problems, if they are not completely exogenous, and results may be very sensitive to the choice of instruments (Burnside, 1996). Basu and Fernald (1997) found that aggregation effects are important and that these effects are correlated with demand shocks. This may be exacerbated by relatively weak correlation of instruments with the explanatory variables that led Basu and Fernald (1997) to conclude that "instruments that are both relatively weak and potentially correlated with the disturbance term suggest that instrumental variables may be more biased than ordinary least squares" (p. 258). Therefore, we follow Diewert and Fox (2008), Basu and Fernald (1997, 2002), and Roeger (1995) and rely on ordinary least squares (OLS) estimates.

Another, related point is that all variables—and in particular the R&D variable—are likely measured with error.[12] When there is a measurement error in the regressor and it is of the classical type, that is, independent of the true value of the variable, OLS estimates have been shown to underestimate the magnitude of the regression coefficient (see, for instance, Hyslop and Imbens, 2001).[13] However, Klepper and Leamer (1984) reported that with classical measurement error in the two-variable case, the true value of the regression coefficient lies between the estimated coefficients[14] from the direct and the reverse regression. Our estimation strategy is to apply OLS to both expressions in (16) and so obtain bounds for the coefficients. Estimation results from a panel data set for 20 OECD countries and for the period 1985–2017 are shown in (17) where fixed effects for countries and years have been applied and standard errors are shown in brackets.

[12]The econometric issues because of using R&D in a production function have long been discussed (e.g. Griliches, 1998) but never been fully satisfactorily resolved. The work here harks back to a long tradition of analyzing R&D in a production context, pioneered by Griliches (1973) and recently reviewed by Ugur *et al*. (2016).

[13]When there is a classical measurement error in both the regressor and the dependent variable, the OLS bias cannot in general be signed, unless it is assumed that the measurement errors of the regressor and the dependent variable are independent in which case the downward bias in regression coefficients remains.

[14]Klepper and Leamer (1984) also reported that in the case of three variables, the true value of the coefficients lies inside the triangular area mapped out by these three regressions. We refrain from formally setting out all three regressions—that is, also including a specification where R&D is the dependent variable because such a specification would be very hard to justify on economic grounds. It is very unlikely that R&D capital services are driven by contemporaneous output and non-R&D inputs.

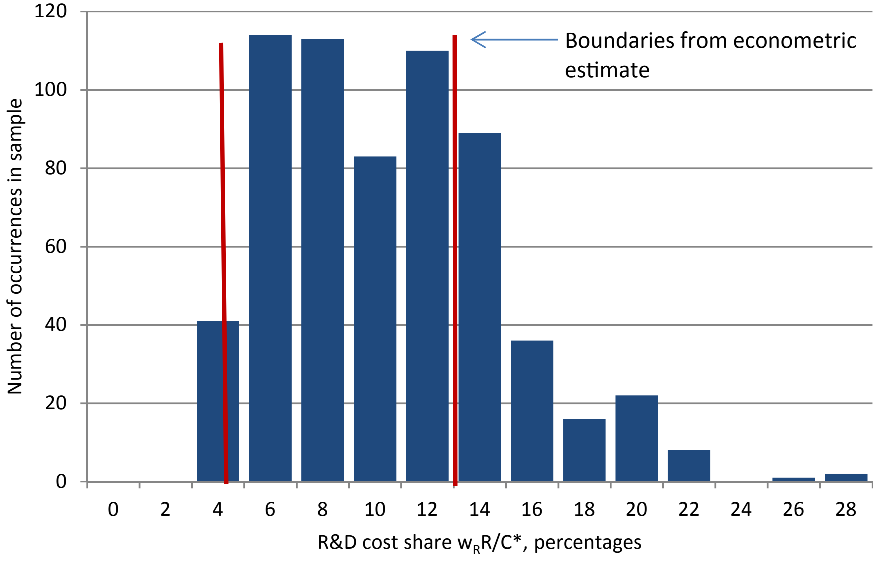

Figure 1. Cost-Elasticities of R&D: Distribution of Unrestricted Measures and Econometric Results
Source: Authors' calculations, based on OECD (2019), OECD Productivity Statistics (database), http://dx.doi.org/10.1787/pdtvy-data-en. [Colour figure can be viewed at wileyonlinelibrary.com]

$$
(17)\quad \begin{aligned} \Delta lnX^t &= \underset{(0.318)}{0.965} + \underset{(0.025)}{0.531}\Delta lnQ^t - \underset{(0.007)}{0.036}\Delta lnR^t;\ adjR^2 = 0.64;\ DF = 584 \\ \Delta lnQ^t &= 0.387)1.117 + \underset{(0.038)}{0.788}\Delta lnX^t + \underset{(0.008)}{0.103}\Delta lnR^t;\ adjR^2 = 0.76;\ DF = 584. \end{aligned}
$$

All coefficients are significant and show the right sign.[15] However, as expected, direct and reverse regressions lead to very different measures of returns to scale and of shadow prices for the R&D asset. In particular, short-run returns to scale are either 1/0.531 = 1.88 based on the first result in (17) or 0.788 when based on the second result in (17). The cost elasticity of the R&D asset as implied by the first regression equals $w_{RS}R/C^* = (w_{RS}R/C)(C/C^*) = (w_{RS}R/C)/[1+w_{RS}R/C] = 0.036/(1+0.036) = 0.035$. The cost elasticity of R&D as implied by the second regression equals $w_{RS}R/C^* = [(\epsilon w_{RS}R/C)/\epsilon]/[1+\epsilon w_{RS}R/\epsilon C] = (0.103/0.788)/(1+0.103/0.788) = 0.115$. Thus, our lower bound for the cost share as recovered by the estimation is around 4 percent and the upper bound is around 12 percent. We thus find a rather large possible range of cost elasticities for R&D.[16]

Compare these point estimates with the descriptive statistics for the cost shares $w_R R/C^*$ that have been computed with a standard index number approach: their mean and median are around 9.4 percent, with a minimum value of around 2 percent and a maximum value of 67 percent.[17] Figure 1 shows the density distribution

[15]Tests with various lagged R&D variables in the first equation of (17) produced only insignificant results.
[16]If the second reverse regression with R&D as the dependent variable is run despite its theoretical implausibility, the implied upper bound to the coefficient is even higher, around 48 percent.
[17]This unusually high share concerns Ireland in the year 2015 that saw a massive transfer of R&D assets into the country, leading to a leap in GDP growth and a singularly large cost share of R&D.

of all $w_R R/C^*$, along with the upper and lower boundaries from the regression results (red vertical lines). About two-thirds of all computed values lie within these bounds, and we conclude that the econometric results do not offer significant additional insights over the unconstrained index number results.

Ugur *et al*. (2016) conducted a meta-data analysis of 773 elasticity estimates of R&D capital on output at the firm level and 135 elasticity estimates at the industry level in OECD countries. Their median estimate ranges from 0.008 to 0.313 for elasticities at the industry level. Our own estimates appear to be well within this range, considering in particular that the authors also find that elasticity estimates tend to be higher when R&D capital is constructed with the perpetual inventory method and when output is measured as value added, which is the case in our data set.

With the help of equation (11), we can carry out another test for significant differences between estimated cost elasticities and those derived from the unrestricted model. We first express equation (11) in discrete time, and then assume that both restricted and unrestricted cost elasticities are constant, along with the assumption that technical change again follows a simple stochastic process $\Delta\pi^t = \alpha_{c0} + \mu_c^t$.

$$\begin{aligned} MFP^t &= \Delta lnQ^t - \Delta lnZ^{*t} \\ &= \epsilon^*[w_{RS}R/C^* - w_R R/C^*]\Delta lnR^t + (\epsilon^* - 1)\Delta lnZ^{*t} + \Delta\pi^t \\ MFP^t &= \alpha_{c0} + \alpha_{c1}\Delta lnR^t + \alpha_{c2}\Delta lnZ^{*t} + \mu_c^t. \end{aligned} \tag{18}$$

If restricted and unrestricted cost elasticities of R&D are constant and significantly different from each other, the coefficient $\alpha_{c1} = \epsilon^*(w_{RS}R/C^* - w_R R/C^*)$ should be significantly different from zero.[18] Estimation of (18) produces insignificant results for α_{c1}, and the same holds for the reverse regression.

In light of these outcomes and various other advantages of using unconstrained index numbers—full variability across countries and years, reproducibility, and greater ease of applicability in regular statistical production—we conclude that there is no strong reason to prefer the econometric approach over the index number approach. In what follows, we shall therefore rely on an unrestricted cost function as set out earlier.

3.2. *Scale Elasticity*

We next turn to the estimation of returns to scale. Our workhorse is the growth accounting equation (10) that presents the growth rate of output as a function of the growth rate of combined inputs and technical change, augmented by long-run returns to scale. Transformed into discrete time, the unrestricted cost function in equation (10) reads as follows:[19]

[18]A similar specification has been used to test whether output elasticities of knowledge-based capital exceed its factor shares (Roth and Thum 2013; Niebel *et al*., 2013) and, in a somewhat different context, as an estimate for spillovers from Information and Communication Technology (ICT) and intangibles (Stiroh, 2002 and Corrado *et al*., 2014).

[19]These expressions are similar to those derived from a generalized approach in Diewert and Fox (2010, equation 49) and Diewert and Fox (2017b, equation 58).

$$\begin{aligned} \Delta lnZ^{*t} &= \frac{1}{\epsilon^*}\Delta lnQ^t - \Delta\pi^{*t}, \\ \Delta lnQ^t &= \epsilon^*\left(\Delta lnZ^{*t} + \Delta\pi^{*t}\right), \end{aligned} \tag{19}$$

where $\Delta lnZ^{*t} \equiv 0.5\left(\frac{C^t}{C^{*t}} + \frac{C^{t-1}}{C^{*t-1}}\right)\Delta lnX^t + 0.5\left(\frac{w_R^t R^t}{C^{*t}} + \frac{w_R^{t-1}R^{t-1}}{C^{*t-1}}\right)\Delta lnR^t$ is the cost-share weighted Törnqvist index of inputs. We have again specified both the direct and the reverse forms of the growth accounting equation as the same points about errors in the variables apply that were discussed earlier. Equation (19) sets up the estimation where unrestricted productivity growth $\Delta\pi^{*t}$ is taken to follow a simple stochastic form with a constant expected value and randomly distributed variations around it: $\Delta\pi^{*t} = \alpha_{d0} + \mu_{dt}$.

$$\begin{aligned} \Delta lnZ^{*t} &= \alpha_{d0} + \alpha_{d1}\Delta lnQ^t - \mu_d^t \\ \Delta lnQ^t &= \alpha_{e0} + \alpha_{e1}\Delta lnZ^{*t} + \mu_e^t. \end{aligned} \tag{20}$$

Our baseline results are the direct and the reverse OLS estimate of (20). For each direct and reverse estimate, we add country-specific fixed effects and time-specific fixed effects, first separately and then combined. Two types of time effects are tested, one with dummies for all years (bar one) and the other with dummies for the crisis years 2008 and 2009 only. Overall, we end up with 12 estimates for long-run returns to scale. The corresponding evaluations of ϵ^* range from around 0.8 to around 1.6 with an unweighted geometric mean of 1.17 and a reliability-adjusted average of 1.26.[20] With (classical) measurement errors likely present in all variables, the arguments developed earlier apply again and suggest that the set of direct estimates around the first expression in (20) will produce estimates of $\epsilon^* = 1/\alpha_{d1}$ that are downward biased, whereas reverse estimates around the second expression in (20) will produce estimates of $\alpha_{e1} = \epsilon^*$ that are upward biased. As the true coefficient will lie in between each pair of estimates, we take as point estimate—and best guess—for ϵ^* the geometric average of the various results that corresponds to $\epsilon^* = 1.2$.

This is in line with related research. For instance, Diewert and Fox (2008) found a scale elasticity of between 1.2 and 1.5 for US manufacturing industry. Basu and Fernald (1997) produced evidence of scale elasticities of between 1.29 and 1.46 for a comparable aggregate, value-added based measure for the private sector of the US economy. Foster (2015) argued that economies of scale are likely to have been driving Australia's productivity growth and suggests that scale elasticity may be around 1.5.

3.3. *Productivity, Demand, and Market Size*

With an estimate for ϵ^* at hand, it is now possible to implement (19) empirically and de-compose the existing standard measure of MFP growth into an element that reflects returns to scale, $(1-1/\epsilon^*)\Delta lnQ^t$, and into an element of "residual"

[20]Each estimate was weighted by its inverted normalized standard deviation to assign higher weights to more reliable estimates.

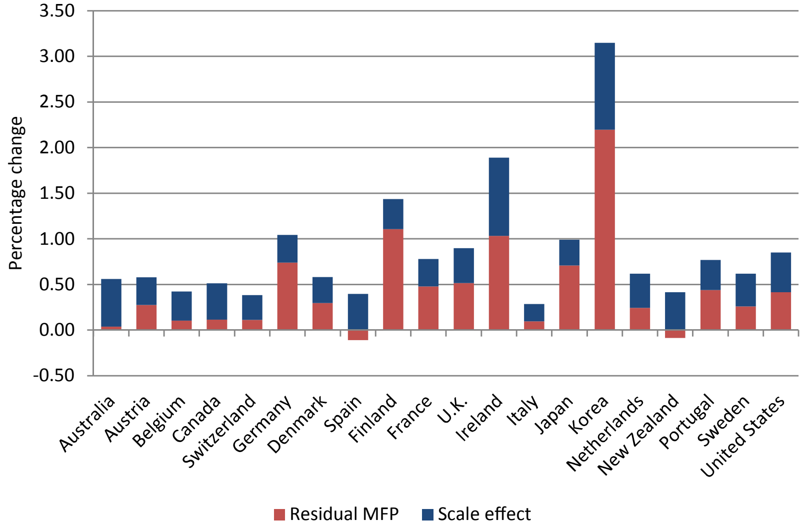

Figure 2. Scale Effects and Residual MFP

Note: Sample refers to 1985–2017 except: Austria: 1996–2017; Switzerland 1992–2017; Spain: 1985–2016; New Zealand 1987–2017; Portugal 1985–2016.

Source: Authors' calculations, based on OECD (2019), OECD Productivity Statistics (database), http://dx.doi.org/10.1787/pdtvy-data-en. [Colour figure can be viewed at wileyonlinelibrary.com]

productivity growth, $\Delta\pi_S^{*t}$. Unlike the measure of average productivity growth $\Delta\pi^{*t}$ evaluated as part of the estimation (20), there is no specific stochastic specification for $\Delta\pi_S^{*t}$. It picks up all those effects on output growth that are explained neither by inputs nor by scale effects. The qualification "residual" is thus important because there are certainly other forces than pure technical change that bear on $\Delta\pi_S^{*t}$.

(21) $$MFP^t = \Delta lnQ^t - \Delta lnZ^{*t} = (1 - 1/\epsilon^*)\Delta lnQ^t + \Delta\pi_S^{*t}.$$

Figure 2 exhibits results of this decomposition for 20 OECD countries over the period 1985–2017, based on our preferred average value of $\epsilon^* = 1.2$. Despite differences between countries, it is apparent that both effects are important, although a look at the annual data shows much greater volatility of the residual MFP component. Overall, and across all countries and periods, the scale effect and the residual MFP effect are approximately equally strong determinants of MFP growth. It is also apparent from Figure 2 that much of the cross-country variability comes from the residual MFP effect. Scale effects are more similar across countries (although this is partly a consequence of the country-invariant scale parameter) than residual MFP effects. This could imply that country characteristics such as differences in policies and institutions matter more for residual MFP than for scale effects.

An aspect not to be neglected (but glossed over in the current set-up) is aggregation in the presence of a changing number of firms as discussed by Basu and Fernald (1997, 2002): industry returns to scale may be driven by entry and exit and different from firm returns to scale. Thus, the non-rejection of aggregate increasing returns to scale is not necessarily proof of firm-level or industry-level increasing returns to scale.[21]

A scale effect of some magnitude has policy-relevant consequences.

1. One is the implied effect of demand on productivity—a causality that runs counter to the more standard supply-side interpretation where technology and efficiency improvements affect output. On the one hand, this concerns longer-term demand effects: for instance, rising income inequality may have a dampening effect on demand and consequently on productivity if the average propensity to consume decreases (Summers, 2015) or if lower-income households desire to accumulate precautionary savings in response to the higher income risk associated with persistent inequality (Auclert and Rognlie, 2018). Furthermore, some of the pro-cyclical nature of productivity growth can be explained when demand affects productivity, as has been suggested by Hall (1988) and Basu and Fernald (1997). However, we do find that $\Delta\pi_S^{*t}$ remains a series of high variance.
2. A second and related policy-relevant conclusion is that market size matters for MFP. With markets expanding globally, returns to scale come into force and reduce marginal costs. This is one of the positive effects of expanding trade and vice versa; shrinking market size will negatively affect productivity growth.
3. A third consequence is that increasing returns to scale—if present at the firm level—imply the existence of mark-ups over marginal costs and therefore some monopolistic elements. Whether these monopolistic elements give rise to "pure" mark-ups above and beyond what is needed to cover average costs is an important question for competition policy.

3.4. *Mark-Ups in the OECD Area*

Turning to mark-ups rates over marginal costs, these are measured with the help of equation (19):

$$1+m^{*t} = \epsilon^* \left(1-\frac{M^{*t}}{P_Q^t Q^t}\right)^{-1} = \epsilon^* \left(1+\frac{M^{*t}}{C^{*t}}\right). \quad (22)$$

To measure $1+m^{*t}$, we use the constant average value $\epsilon^* = 1.2$ and the time- and country-varying measure of "residual" profit rates $\frac{M^{*t}}{P_Q^t Q^t}$, or "residual" mark-up rates $\frac{M^{*t}}{C^{*t}}$. M^{*t} is the difference between labor compensation, user costs of capital, and the nominal value of output. The latter is measured at basic prices, so any (other) taxes and subsidies on production are excluded from the residual mark-up M^{*t}. In our sample, the average mark-up factor $1+m^{*t}$ across all countries and

[21]Suppose an oligopolistic industry with firms that have overhead costs and constant returns to scale thereafter. Then the degree of industry returns to scale depends on the extent to which firm-level output changes in the same direction as industry output.

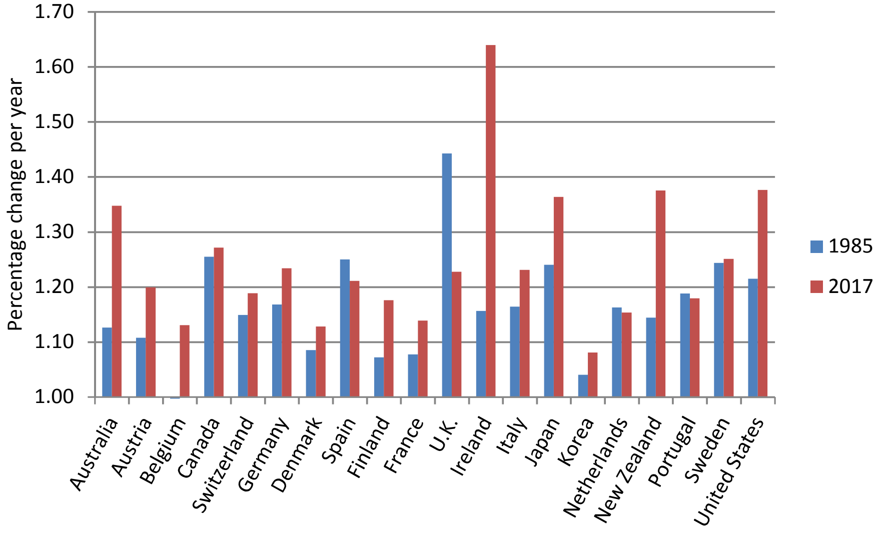

Figure 3. Mark-Ups Over Marginal Costs: By Country

Notes: Sample refers to 1985 and 2017 except: Austria: 1996 and 2017; Switzerland 1992 and 2017; Spain: 1985 and 2016; New Zealand 1987 and 2017; Portugal 1985 and 2016.

Source: Authors' calculations, based on OECD (2019), OECD Productivity Statistics (database), http://dx.doi.org/10.1787/pdtvy-data-en. [Colour figure can be viewed at wileyonlinelibrary.com]

years is around 1.25 or a 25 percent addition to marginal costs. This is broadly consistent with the early work by Oliveira Martins *et al.* (1996), and more recently, Christopoulou and Vermeulen (2012), although the authors assume constant returns and consider the private sector rather the total economy. Diewert and Fox (2008) derived mark-ups between 1.4 and 1.7 for US manufacturing, and Devereux *et al.* (1996) reviewed the literature and estimated that mark-ups of up to 1.5 constitute a plausible value for use in modeling. De Loecker and Warzynski, 2012, in a firm-level study of Slovenian manufacturing firms, obtained mark-ups in the range of 1.17—1.28. In a more recent work, De Loecker *et al.* (2020) obtained much higher mark-ups, but comparability with the work at hand is limited given the firm-level nature of their analysis and, importantly, their use of a gross output framework.

It should be recalled here that the level of residual mark-ups M^{*t} also reflects assumptions about the longer-run real rate of return to capital that have entered the computation of user costs (Annex A). Indeed, the standard way to proceed (Jorgenson, 1995; Jorgenson and Landefeld, 2004) is letting the rate of return to capital that enters user cost measures adjust so that M^{*t} vanishes ("endogenous rates of return") and the value of output exactly equals total costs. Absent M^{*t}, the mark-up rate over marginal costs equals exactly the degree of returns to scale as can be observed from (22). In this case, time-invariant returns to scale ε^* would imply time-invariant mark-ups $1+m^*$, and all variation in profits would show up as variations in the price of capital services.

Figure 3 shows how mark-up rates over marginal costs develop over time, measured as $1.2(1+\frac{M^{*t}}{C^{*t}})$. Figure 4 presents average mark-up rates across countries

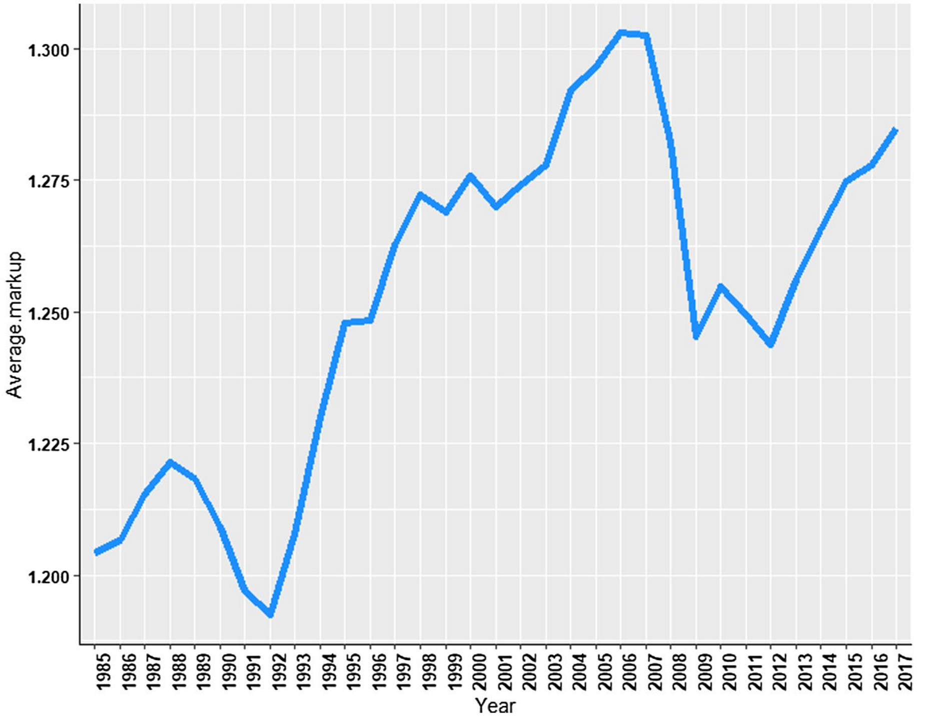

Figure 4. Mark-Ups Over Marginal Costs: Averages across Countries

Notes: Unweighted average; sample refers to 1985–2017 except: Austria: 1996–2017; Switzerland: 1992–2017; Spain: 1985–2016; New Zealand: 1987–2017; Portugal: 1985–2016.

Source: Authors' calculations, based on OECD (2019), OECD Productivity Statistics (database), http://dx.doi.org/10.1787/pdtvy-data-en. [Colour figure can be viewed at wileyonlinelibrary.com]

over the period 1985–2017. One notes that with a time-invariant $\epsilon^* = 1.2$ all changes in overall mark-up rates $1+m^{*t}$ are triggered by changes in $\frac{M^{*t}}{C^{*t}}$. If returns to scale were allowed to vary over time, the split of overall mark-ups over marginal costs into scale effects and residual profit effects might turn out differently. Over the period 1985–2017, overall mark-ups over marginal costs increased on average and in 16 of the 20 countries considered, which corroborates other findings in the literature. Calligaris *et al*. (2018) and Andrews *et al*. (2016), albeit with an entirely different firm-level data set, observed upward trending average mark-ups in OECD countries, mostly driven by firms in market services sectors. Conceptual and measurement problems underlying the increase in estimated mark-ups over marginal costs have recently been discussed in Basu (2019) and Syverson (2019). An analysis of the causes of this secular increase in mark-ups over marginal costs is beyond the scope of this paper but several possibilities suggest themselves:

1. Rising returns to produced assets, as a reflection of rising risk premia. Karabarbounis and Neiman (2018) explored several hypotheses about the sources of "factorless income," which corresponds to our measure of residual profits, M^{*t}. Their favored explanation is one whereby "simple measures of the rental rate of capital might deviate from the rate that firms face

when making their investment decisions." In other words, they hypothesize that the most plausible explanation for the existence of M^{*t} is that remuneration of measured capital is understated. This could, for instance, reflect risk premia, a conclusion in Caballero *et al*. (2017). If rising risk premia is the issue, the corresponding residual profits should be reallocated as factor income to the relevant assets. From an analytical and policy perspective, identifying the source of rising risk premia associated with nonfinancial investment would be an important subject of future research.

2. Monopoly rents: rising residual profits are certainly consistent with situations where the digital economy and associated network effects lead to "winner-takes-most" outcomes and reduced competition (see, for instance, Andrews *et al*., 2016). This is the argument pursued in Calligaris *et al*. (2018), who showed that average firm mark-ups are higher in more digital-intensive sectors, even after controlling for various factors. A particularly strong hike in residual mark-ups is measured for Ireland, possibly reflecting supra-normal returns to IP assets.
3. Rising mark-ups over marginal costs may also be a reflection of the rising importance or rising returns to those assets that have not been explicitly recognized in the present computations. When of the intangible kind, these assets include human capital, organizational capital, or marketing assets as investigated by Corrado *et al*. (2005), OECD (2013), or Goodridge and Haskel (2016). When of the tangible kind, these assets include in particular land whose real price (and real rate of return) has registered an upward trend over the past decades in many OECD countries.[22]

4. Conclusions

With the implementation of the 2008 System of National Accounts, R&D capital stock measures are now widely available in OECD countries. Although it is natural to include R&D capital services into the measurement of productivity, R&D assets are also somewhat special: conceptually, they shape production rather than providing a specific type of service, they are replicable and easily transferrable, and their production often entails long gestation and sunk costs; and measurement of the value and prices of R&D investment and R&D assets must rely on more assumptions than is the case for other assets. We investigate whether the usual assumption of period-to-period cost-minimizing choices of capital inputs is warranted for R&D inputs and conclude that on the whole the traditional index number method cannot be rejected.

We also test for nonconstant returns to scale and find econometric evidence for moderately increasing returns at the aggregate economy level, much in line with the available literature. This permits decomposing MFP growth rates into a component that is triggered by returns to scale and into a component of "pure"

[22]See Diewert Fox (2017a) for an analysis of US productivity growth and Cho *et al*. (2015) for Korea including land and inventories.

or "residual" technical change. Across the 20 countries examined and over three decades, the two components are approximately equally important. A dependence of MFP on the level of activity both helps explaining cyclical patterns of MFP growth and points to the importance of long-term demand, market size, and international trade as supporting factors of productivity. With the data at hand, we cannot, however, determine to which extent our finding of economy-wide rising returns to scale is driven by aggregation effects or genuine economies of scale at the firm level.

The dual picture of imperfect competition and increasing returns to scale is mark-ups over marginal costs. We find that mark-up rates have trended upward in nearly all countries investigated. As our measure of increasing returns to scale is time-invariant, this reflects a rise in residual profits, above and beyond what is needed to cover average costs. Such a picture chimes well with effects associated with globalization and digitalization where some markets may have become less competitive. Extra profits may also reflect returns to assets not measured in our set of inputs, including intangibles other than R&D, and tangibles such as land and natural resources. Future research will have to explore which of these explanations is most accurate.

References

Andrews, D., C. Criscuolo and P. N. Gal, "The Global Productivity Slowdown, Divergence Across Firms and the Role of Public Policy," OECD Productivity Working Papers No 5, 2016.

Auclert, M. and M. Rognlie, "Inequality and Aggregate Demand," NBER Working Paper No. 24280, 2018.

Balk, B. M., *Industrial Price, Quantity and Productivity Indices, The Micro-economic Theory and an Application*. Boston/Dordrecht/London: Kluwer Academic Publishers, 1998.

Basu, S., "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence," *Journal of Economic Perspectives*, 33, 3–22, 2019.

Basu, S., and J. G. Fernald, "Returns to Scale in US Production: Estimates and Implications," *Journal of Political Economy*, 105, 249–283, 1997.

______, "Aggregate productivity and aggregate technology," *European Economic Review*, 46, 963–991, 2002.

Berndt, E. R., "Reconciling Alternative Estimates of the Elasticity of Substitution," *The Review of Economics and Statistics*, 58(1), 59–68, 1976.

Berndt, E. R., and M. A. Fuss, "Productivity Measurement with Adjustments for Variations in Capacity Utilisation and Other Form of Temporary Equilibria," *Journal of Econometrics*, 33, 7–29, 1986.

Burnside, C., "Production function regressions, returns to scale, and externalities," *Journal of Monetary Economics*, 37, 177–201, 1996.

Caballero, R., E. Farhi, and P.-O. Gourinchas, "Rents, Technical change and Risk Premia Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares," *The American Economic Review*, 107(5), 614–20, 2017.

Calligaris, S., C. Criscuolo and L. Marcolin, "Mark-Ups in the Digital Era," OECD Science, Technology and Industry Working Papers, No. 2018/10, OECD Publishing, Paris, 2017, https://doi.org/10.1787/4efe2d25-en

Caves, D. W., L. R. Christensen, and J. A. Swanson, "Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955–74," *The American Economic Review*, 71(5), 994–1002, 1981.

Cho, T., J. Kim, and P. Schreyer, "Measuring the Evolution of Korea's Material Living Standards, 1980–2012," *Journal of Productivity Analysis*, 44(2), 157–73, 2015.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau, "Conjugate Duality and the Transcendental Logarithmic Production Function," *Econometrica*, 39, 255–56, 1971.

Christopoulou, R., and Ph. Vermeulen, "Markups in the Euro Area and the US Over the Period 1981–2004: A Comparison of 50 Sectors," *Empirical Economics*, 42(1), 53–77, 2012.

Corrado, C., J. Haskel and C. Jona-Lasinio, "Knowledge Spillovers, ICT and Productivity Growth," IZA Discussion Paper No. 8274, 2014.
Corrado, C., Ch. Hulten and D. Sichel, "Measuring Capital and Technology: An Expanded Framework," in C. Carol, C. Hulten and D. Sichel (Eds.), *Measuring Capital in the New Economy; National Bureau of Economic Research, Studies in Income and Wealth*, 65, University of Chicago Press, Chicago, 11–46, 2005.
De Loecker, J., and F. Warzynski, "Markups and Firm-Level Export Status," *The American Economic Review*, 102(6), 2437–71, 2012.
De Loecker, J., J. Eeckhout, and G. Unger, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 135(2), 561–644, 2020.
Devereux, M. B., A. C. Head, and B. J. Lapham, "Monopolistic Competition, Increasing Returns, and the Effects of Government Spending," *Journal of Money, Credit, and Banking*, 28, 233–54, 1996.
Diewert, W.E., "Aggregation Problems in the Measurement of Capital," in D. Usher (ed.) *The Measurement of Capital*. University of Chicago Press, Chicago, 433–538, 1980.
Diewert, W. E., "Applications of Duality Theory," in Intrilligator, M.D., Kendrick, D.A. (Eds.), *Frontiers of Quantitative Economics. II*. North-Holland Publishing Co., Amsterdam, 106–71, 1974.
______, "Exact and Superlative Index Numbers," *Journal of Econometrics*, 4, 115–145, 1976.
Diewert, W. E., and K. J. Fox, "On the Estimation of Returns to Scale, Technical Progress and Monopolistic Markups," *Journal of Econometrics*, 145(2008), 174–193, 2008.
______, "Malmquist and Törnqvist Productivity Indexes: Returns to Scale and Technical Progress with Imperfect Competition," *Journal of Economics*, 101(1), 73–95, 2010.
______, "Sunk Costs and the Measurement of Commercial Property Depreciation," *Canadian Journal of Economics/Revue canadienne d'économique*, 49(4), 1340–66, 2016.
______, "Alternative User Costs, Productivity and Inequality in US Business Sectors," *UNSW Business School Research Paper No. 2017 ECON 14*, https://ssrn.com/abstract=2978451, 2017a.
______, "Decomposing productivity indexes into explanatory factors," *European Journal of Operational Research*, 256(1), 275–91, 2017b, Elsevier.
Diewert, W. E., and N. Huang, "Capitalizing R&D Expenditures," *Macroeconomic Dynamics*, 15, 537–64, 2011.
Diewert, W. E., T. Nakajima, A. Nakamura, E. Nakamura, and M. Nakamura, "Returns to Scale: Concept, Estimation and Analysis of Japan's Turbulent 1964–88 Economy," *Canadian Journal of Economics / Revue canadienne d'Economique*, 44(2), 451–85, 2011.
European Commission, IMF, OECD, World Bank, United Nations, 2008 System of National Accounts, New York, United Nations; http://unstats.un.org/unsd/nationalaccount/sna2008.asp, 2009.
Foster, J., "The Australian Multi-factor Productivity Growth Illusion," *The Australian Economic Review*, 48(1), 33–42, 2015.
Galindo-Rueda, F., D. Ker, F. Moris, and J. Jankowski, "Capturing International R&D Trade and Financing Flows: What Do Available Sources Reveal About the Structure of Knowledge-Based Global Production?" Paper presented at the Conference on Research in Income and Wealth (CRIW), Washington DC, 9–10, 2018.
Goodridge, P. J., and G. Wallis Haskel, "UK Intangible Investment and Growth: New Measures of UK Investment in Knowledge Assets and Intellectual Property Rights," Imperial College Business School Discussion Paper 2016/08, 2016.
Griliches, Z., "Research Expenditures and Growth Accounting". In: Williams, R.B. (Ed.), *Science and Technology in Economic Growth*, John Wiley and Sons, New York, 1973.
______, "R&D and Productivity: Econometric Results and Measurement Issues," in P. Stoneman (ed.), *Handbook of the Economics of Innovation and Technical Change*, 52–89, Oxford, 1995.
______, "R&D and Productivity: The Unfinished Business," in Z. Griliches (ed.), *R&D and Productivity: The Econometric Evidence*. University of Chicago Press, Chicago, 1998.
Hall, R. E., "The Relationship between Price and Marginal Cost in U. S. Industry," *Journal of Political Economy*, 96, 921–947, 1988.
Huang, N., and W. E. Diewert, "Estimation of R&D Depreciation Rates: A Suggested Methodology and Preliminary Application," *The Canadian Journal of Economics*, 44(2), 387–412, 2011.
Hulten, Ch. R., "The Measurement of Capital," in E.R. Berndt and J.E. Triplett (eds.), *Fifty Years of Economic Measurement, Studies in Income and Wealth*, The National Bureau of Economic Research, The University of Chicago Press, Chicago, 54, 119–152, 1990.
Hyslop, D. R., and G. W. Imbens, "Bias From Classical and Other Forms of Measurement Error," *Journal of Business & Economic Statistics*, 19(4), 475–481, 2001.
Jones, C. I., "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 103, 759–84, 1995.
Jorgenson, D. W. and J. S. Landefeld, "Blueprint for Expanded and Integrated U.S. Accounts: Review, Assessment, and Next Steps," in Jorgenson, D., J. S. Landefeld, W. Nordhaus (Eds.), *A New*

Architecture for the U.S. National Accounts, NBER, Studies in Income and Wealth, University of Chicago Press, 2004.

Jorgenson, D. W., *Productivity*, I and II. MIT Press, Cambridge, 1995.

Klepper, S., and E. E. Leamer, "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52(1), 163–84, 1984.

Lau, L. J., "A Characterization of the Normalized, Restricted Profit Function," *Journal of Economic Theory*, 12, 131–63, 1976.

Leontief, W. W., "Introduction to a Theory of the Internal Structure of Functional Relationships," *Econometrica*, 15, 361–73, 1947.

Karabarbounis, L. and B. Neiman, "Accounting for Factorless Income," NBER Working Paper 24404, 2018.

McFadden, D., "Cost, Revenue and Profit Functions," in M. Fuss and D. McFaddeneds), *Production Economics: A Dual Approach to Theory and Applications*, Amsterdam: North-Holland,1, 3–109, 1978.

Morrison, C. J., "Unraveling the Productivity Growth Slowdown in the United States, Canada and Japan: The Effects of Subequilibrium, Scale Economies and Markups," *Review of Economics and Statistics*, 74(3), 381–93, 1992.

Niebel, T., M. O'Mahony and M. Saam, "The Contribution of Intangible Assets to Sectoral Productivity Growth in the EU," ZEW Discussion Paper No. 13-062; http://ftp.zew.de/pub/zew-docs/dp/dp13062.pdf, 2013.

OECD, *Handbook on Deriving Capital Measures of Intellectual Property Products*, OECD Publishing, Paris. http://www.oecd.org/std/na/44312350.pdf, 2010.

______, *Supporting Investment in Knowledge Capital, Growth and Innovation*, Paris, http://dx.doi.org/10.1787/9789264193307-en, 2013.

______, *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*; OECD Publishing; http://www.oecd.org/publications/frascati-manual-2015-9789264239012-en.htm, 2015.

______, *OECD Compendium of Productivity Indicators 2019*. OECD Publishing, Paris, 2019.

Oliveira Martins, J., S. Scarpetta, and D. Pilat, "Mark-up Pricing, Market Structure and the Business Cycle," *OECD Economic Studies*, 27(2), 71–105, 1996.

Parham, D., "Empirical Analysis of the Effects of R&D on Productivity: Implications for Productivity Measurement?" Paper presented at OECD Workshop on Productivity Measurement and Analysis Bern, Switzerland 16–18 October, 2006, https://www.oecd.org/std/productivity-stats/37511005.pdf, 2007.

Pitzer, J. S., "Intangible Produced Assets," Paper presented at the London Meeting of the Canberra 2 Group On the Measurement of Non-Financial Assets, 1–3 September, 2004.

Roeger, W., "Can Imperfect Competition Explain the Difference Between Primal and Dual Productivity Measures? Estimates for US Manufacturing," *Journal of Political Economy*, 103, 316–30, 1995.

Roth, F., and A.-E. Thum, "Intangible Capital and Labor Productivity Growth: Panel Evidence for the EU from 1998–2005," *Review of Income and Wealth*, 59(3), 486–508, 2013.

Schankerman, M. and M. I. Nadiri, "Restricted Cost Functions and the Rates of Return to Quasi-Fixed Factors, with an Application to R&D and Capital in the Bell System," NBER Working Paper No 1259, 1984

Schreyer, P., "Measuring Multi-factor Productivity when Rates of Return are Exogenous," in W. Erwin Diewert, Bert M. Balk, Dennis Fixler, Kevin J. Fox and Alice O. Nakamura (Eds.), *Price and Productivity Measurement*, 1 and 2, Trafford Press, 2010. http://www.indexmeasures.ca/Vol6_10,09,26.pdf

Schreyer, P., P-E. Bignon and J. Dupont, "OECD Capital Services Estimates: Methodology and a First Set of Results," OECD Statistics Working Paper 2003/6, 2003.

Sena, V., "The Return of the Prince of Denmark: A Survey and Recent Developments in the Economics of Innovation," *Economic Journal*, 114, F312–32, 2004.

Shephard, R. W., *Cost and Production Functions, Princeton University Press*. Princeton, NJ, 1953.

Sono, M., "The effect of price changes on the demand and supply of separable goods," *International Economic Review*, 2, 239–75, 1961.

Syverson, C., "Macroeconomics and Market Power: Context, Implications, and Open Questions," *Journal of Economic Perspectives*, 33, 23–43, 2019.

Stiroh, K., "Are ICT Spillovers Driving the New Economy?" *Review of Income and Wealth*, 48(1), 33–7, 2002.

Summers, L. H., "Demand Side Secular Stagnation," *The American Economic Review*, 105(5), 60–5, 2015.

Ugur, M., E. Trushinb, E. Solomona, and F. Guidi, "R&D and Productivity in OECD Firms and Industries: A Hierarchical Meta-Regression Analysis," *Research Policy*, 45(10), 2069–86, 2016.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Annex A Measurement and Data Sources
Annex B Tables by Country
Annex References