

Productivity Measurement, R&D Assets and Mark-ups in OECD Countries

Paul Schreyer and Belen Zinni
OECD Statistics and Data Directorate *

Version 4 October 2018,
forthcoming as *OECD Statistics Working Paper*

Abstract

A key feature of the 2008 revision of the System of National was the treatment of R&D expenditure as investment. The question arises whether the standard approach towards 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 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 1985-2015.

*We thank participants at the World KLEMS Conference, Harvard University May 2018 for useful remarks. Opinions expressed in this paper are those of the authors and do not necessarily reflect those of the OECD or its Members

1 Introduction

One of the central achievements of the 2008 revision of the System of National Accounts (SNA 2008 – European Commission et al. 2009) was the treatment of research and development expenditure (R&D) as investment that gives rise to knowledge assets. With the completed implementation of the SNA 2008 among OECD countries by end-2016, users of statistics now dispose of sets of estimates for the investment in R&D as well as software (already present in the 1993 revision) along with estimates of more traditional non-financial produced assets (machinery, equipment, structures).

As all these assets provide inputs 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, Hulten and Sichel (2005) who measured intangible capital for the United States and employed 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 et al 2016).

There are, however, several issues when it comes to using R&D assets in productivity measurement. First is that R&D projects often involve sunk costs and upfront 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. Sunk costs thus imply increasing returns to scale at the firm level. Increasing returns to scale may also arise at the aggregate level due to externalities and spill-overs that R&D assets generate¹: “. . . 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 that aggregate returns to scale may not be constant but increasing. A first objective of the analysis here is to test for the presence of increasing returns to scale when R&D assets and to measure the evolution of mark-ups and to distinguish those associated with returns to

⁰The 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).

¹For an overview of the literature see Senna (2004).

scale. We shall conclude that the hypothesis of increasing returns cannot easily be rejected and there is a pattern of rising mark-ups in nearly all countries of the sample.

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 work by Parham (2007), Pitzer (2004), and Diewert and Huang (2011). Pitzer (2004) observed that R&D capital functions as a source of 'recipes'. Diewert and Huang (2011) start 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 licences 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 given level of other, '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, Schenkerman 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 minimising behaviour 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-minimising 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 have to be estimated econometrically.

We use data for 20 OECD countries over the period 1985-2015 and estimate cost elasticities of R&D capital to test whether these diverge significantly from the standard non-parametric elasticities. While there are variations across

countries and over time, it turns out that on average the econometric point estimate lines up rather well with the index number results. This is in particular the case when we allow for non-constant returns to scale at the same time. We will therefore 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 standard index number results that do not assume quasi-fixity of the R&D input. There is in particular the need to revert to econometric techniques which reduces 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 get 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 price change of the inputs used in their creation, much of it being the wage rate of R&D personnel. This is an added reason for testing whether cost shares are reflective of cost elasticities of R&D, as explained above. Measurement problems do not stop with valuation of the asset, however. There is also an issue of how to determine the rate of depreciation which, in the case of R&D is driven by obsolescence rather than wear-and-tear as with other capital goods. Lastly, because R&D assets are intangible, they can easily be transferred, including across national borders.² R&D assets can therefore appear and disappear in lumps, 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 System of National Accounts and fit also with the broader blueprint of productivity measurement in a national accounts

²A widely discussed example is Ireland where trans-border movements of intellectual property assets and the associated production and income flows gave rise to a staggering 25 percent rise in real GDP in 2015 and a similar unusual two-digit growth in labour productivity. While Ireland may have brought the issue of measuring and production and productivity into sharp focus, this constitutes by no means a unique case.

framework as developed by Jorgenson and Landefeld (2004).

The paper at hand is organised as follows. Section 2 deals with productivity measurement under non-constant 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 non-constancy 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. We do, however, 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 into a part that reflects autonomous productivity change. The Section finishes with the dual picture to the MFP decomposition, mark-ups over marginal and average costs.

2 IP assets in productivity measurement

2.1 Technology

We characterise technology by a production function where labour and traditional capital inputs are combined with services from a knowledge asset to produce aggregate output:

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

In (1), Q is the volume of aggregate output; $X \equiv (X_1, X_2, \dots)$ is the vector of labour and various types of non-R&D capital; R is the stock of R&D and t is a time variable to capture autonomous productivity change. $f_Q(X, R, t)$ is continuous and non-decreasing in inputs X , R and t . No constant returns are imposed here. This is motivated by the desire to maintain a general approach but also by the nature of R&D: its creation typically entails large, fixed upfront 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 below.

In addition to allowing for non-constant 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 pre-determined variable that cannot be adjusted instantaneously and in a cost-minimising 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^3 . For non-R&D inputs X the usual assumption of instantaneous cost-minimising adjustment is maintained.

The production function above characterises technology and can be used as the framework for measuring autonomous technical change. The latter is then measured 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 characterised by constant returns to scale, when production is efficient and when producers minimise costs. Primal and dual measures will divert, however, when one or several of these conditions fail to hold ⁴. 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) point to the strong intuitive appeal of a cost-based measure of scale elasticity as the percentage change in total cost due to a one percent increase in the quantity of output, for a given level of input prices ⁵. Further, cost-based productivity

³Formally, 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 due to Sono (1961) and Leontief (1947). Separability is a rather restrictive assumption but Diewert (1980) offers a way forward with his *Method III* (p.455 ff.) where he shows 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 realisations in comparison periods.

⁴See Balk (1998) for a comprehensive overview of the various primal and dual productivity measures and their relationship.

⁵While 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.

measures allow for a simple set-up of producer behaviour on output markets when competition is imperfect. We shall therefore make use of the following restricted (variable) cost function:

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

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, Christensen and Swanson (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 minimisation by producers only in regards to X , and is conditional on a level of R and t . The second equality in (2) states that minimised variable costs equal observed variable costs $\sum_i w_{X_i} X_i$. 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.

Shepard’s (1953) Lemma holds for the variable cost function: for non R&D inputs X_i , ($i = 1, 2, \dots$) factor demand equals marginal cost changes associated with a change in input prices: $\partial C(Q, w_X, R, t) / \partial w_{X_i} = X_i(Q, w_X, R, t)$. For the R&D input, we define a shadow price w_{RS} as the marginal reduction in variable costs due to 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:

$$\begin{aligned}
\frac{d \ln C(Q, w_X, R, t)}{dt} = & \frac{\partial \ln C(Q, w_X, R, t)}{\partial \ln Q} \frac{d \ln Q}{dt} + \sum_i \frac{\partial \ln C(Q, w_X, R, t)}{\partial \ln w_{X_i}} \frac{d \ln w_{X_i}}{dt} \\
& + \frac{\partial \ln C(Q, w_X, R, t)}{\partial \ln R} \frac{d \ln R}{dt} + \frac{\partial \ln C(Q, w_X, R, t)}{\partial t}.
\end{aligned} \tag{3}$$

The cost elasticity of output is the definition of (inverted) returns to scale and we shall denote $\partial \ln C(Q, w_X, R, t) / \partial \ln Q \equiv 1/\epsilon$. 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), $\partial \ln C(Q, w_X, R, t) / \partial t$, 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 re-written as:

$$\frac{d \ln C}{dt} = \frac{1}{\epsilon} \frac{d \ln Q}{dt} + \sum_i \frac{w_{X_i} X_i}{C} \frac{d \ln w_{X_i}}{dt} - \frac{w_{RS} R}{C} \frac{d \ln R}{dt} + \frac{\partial \ln C}{\partial t}. \tag{4}$$

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

$$\frac{d \ln X}{dt} \equiv \sum_i \frac{w_{X_i} X_i}{C} \frac{d \ln X_i}{dt} = \frac{d \ln C}{dt} - \sum_i \frac{w_{X_i} X_i}{C} \frac{d \ln w_{X_i}}{dt}. \tag{5}$$

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

$$\begin{aligned}
\frac{d \ln X}{dt} &= \frac{1}{\epsilon} \frac{d \ln Q}{dt} - \frac{w_{RS} R}{C} \frac{d \ln R}{dt} + \frac{\partial \ln C}{\partial t}; \\
\frac{d \ln Q}{dt} &= \epsilon \left(\frac{d \ln X}{dt} + \frac{w_{RS} R}{C} \frac{d \ln R}{dt} - \frac{\partial \ln C}{\partial t} \right)
\end{aligned} \tag{6}$$

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 ($\partial \ln C / \partial t \leq 0$) – fewer inputs are needed for a given output when technology and R&D inputs increase. The second line in (6) reverts 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, 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

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

The full expression for the unrestricted cost function is

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

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):

$$\begin{aligned} \frac{\partial \ln C^*}{\partial \ln Q} &= \frac{\partial \ln C}{\partial \ln Q} \frac{C}{C^*} = \frac{1}{\epsilon} \frac{C}{C^*} \equiv \frac{1}{\epsilon^*}; \\ \frac{\partial \ln C^*}{\partial t} &= \frac{\partial \ln C}{\partial t} \frac{C}{C^*}; \\ \frac{\partial \ln C^*}{\partial \ln w_{X_i}} &= \frac{\partial \ln C}{\partial w_{X_i}} \frac{C}{C^*} = \frac{w_{X_i} X_i}{C} \frac{C}{C^*} \quad i = 1, 2, \dots \\ \frac{\partial \ln C^*}{\partial \ln w_R} &= \frac{w_R R}{C^*} \end{aligned} \quad (9)$$

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 straight forward 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:

$$\begin{aligned} \frac{d \ln Q}{dt} &= \frac{\epsilon}{C/C^*} \left(\frac{C}{C^*} \frac{d \ln X}{dt} + \frac{w_{RS} R}{C} \frac{C}{C^*} \frac{d \ln R}{dt} - \frac{C}{C^*} \frac{\partial \ln C}{\partial t} \right) \\ &= \epsilon^* \left(\frac{C}{C^*} \frac{d \ln X}{dt} + \frac{w_{RS} R}{C^*} \frac{d \ln R}{dt} - \frac{\partial \ln C^*}{\partial t} \right) \\ &= \epsilon^* \left(\frac{d \ln Z}{dt} - \frac{\partial \ln C^*}{\partial t} \right) \end{aligned} \quad (10)$$

Here we have defined the short-run Divisia quantity aggregate of all inputs as $\frac{d\ln Z}{dt} \equiv \left(\frac{C}{C^*} \frac{d\ln X}{dt} + \frac{w_{RS}R}{C^*} \frac{d\ln R}{dt}\right)$. Similarly, we can define an unrestricted, long-run Divisia quantity aggregate of inputs as $\frac{d\ln Z^*}{dt} \equiv \left(\frac{C}{C^*} \frac{d\ln X}{dt} + \frac{w_R R}{C^*} \frac{d\ln R}{dt}\right)$.

The OECD measures MFP growth as the difference between output and aggregate input growth (OECD 2017, 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:

$$\begin{aligned}
MFP &\equiv \frac{d\ln Q}{dt} - \frac{d\ln Z^*}{dt} \\
&= \epsilon^* \left(\frac{d\ln Z}{dt} - \frac{\partial \ln C^*}{\partial t} \right) - \frac{d\ln Z^*}{dt} \text{ using (10)} \\
&= \epsilon^* \frac{d\ln Z}{dt} - \epsilon^* \frac{\partial \ln C^*}{\partial t} - \frac{d\ln Z^*}{dt} - \epsilon^* \frac{d\ln Z^*}{dt} + \epsilon^* \frac{d\ln Z^*}{dt} \\
&= \epsilon^* \left(\frac{d\ln Z}{dt} - \frac{d\ln Z^*}{dt} \right) + (\epsilon^* - 1) \frac{d\ln Z^*}{dt} - \epsilon^* \frac{\partial \ln C}{\partial t} \\
&= \epsilon^* \left(\frac{w_{RS}R}{C^*} - \frac{w_R R}{C^*} \right) \frac{d\ln R}{dt} + (\epsilon^* - 1) \frac{d\ln Z^*}{dt} - \epsilon^* \frac{\partial \ln C}{\partial t}. \quad (11)
\end{aligned}$$

When shadow elasticities of R&D equal computable user cost shares ($\frac{w_{RS}R}{C^*} = \frac{w_R R^*}{C^*}$, $\frac{d\ln Z}{dt} = \frac{d\ln Z^*}{dt}$), 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) below 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:

$$\begin{aligned}
MFP &= (\epsilon^* - 1) \frac{d\ln Z^*}{dt} - \epsilon^* \frac{\partial \ln C}{\partial t} \text{ for } \frac{w_{RS}R}{C^*} = \frac{w_R R^*}{C^*} \\
&= \left(1 - \frac{1}{\epsilon^*}\right) \frac{d\ln Q}{dt} - \frac{\partial \ln C}{\partial t} \\
&= -\frac{\partial \ln C}{\partial t} \text{ for } \epsilon^* = 1. \quad (12)
\end{aligned}$$

2.2 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 average costs, i.e., 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 of aggregate output Q :

$$P_Q Q = \sum_i w_{X_i} X_i + w_{RS} R + M = \sum_i w_{X_i} X_i + w_R R + M^* \quad (13)$$

$P_Q Q$ represents total value-added (GDP at the economy-wide level), and $\sum_i w_{X_i} X_i$ 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 are measured through $w_R R$ and M^* , the longer-run mark-up over average costs, can be measured 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:

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

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 ‘pure’ 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 ‘pure’ profits.

3 Empirical implementation

3.1 R&D cost shares – too low, too high, about right?

While the relationships above were derived in continuous time, actual data comes 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 time⁶:

$$\begin{aligned}\Delta \ln X^t &= \frac{1}{\epsilon} \Delta \ln Q^t - 0.5 \left(\frac{w_{RS}^t R^t}{C^t} + \frac{w_{RS}^{t-1} R^{t-1}}{C^{t-1}} \right) \Delta \ln R^t - \Delta \pi^t \\ \Delta \ln Q^t &= \epsilon \left[\Delta \ln X^t + 0.5 \left(\frac{w_{RS}^t R^t}{C^t} + \frac{w_{RS}^{t-1} R^{t-1}}{C^{t-1}} \right) \Delta \ln R^t + \Delta \pi^t \right] \quad (15)\end{aligned}$$

In (15), $\Delta \ln X^t \equiv \ln X^t - \ln X^{t-1}$ 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 ϵ , $0.5 \left(\frac{w_{RS}^t R^t}{C^t} + \frac{w_{RS}^{t-1} R^{t-1}}{C^{t-1}} \right)$ 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 $\Delta \ln X^t$ is measured via index numbers, derived from the restricted cost function. This hybrid approach is due to 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 behaviour that is fully in line with economic theory, it would suffice to estimate either the first or the second equation of (15). But measurement errors will lead to different results depending on whether the direct or the reverse formulation is estimated as further discussed below. Re-formulating (15) for estimation gives:

⁶This can be justified more rigorously by assuming that the restricted cost function is of the translog form (introduced by Christensen et al. 1971 and generalised 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.

⁶Note that 15 is not a system of simultaneous equations.

$$\begin{aligned}
\Delta \ln X^t &= \alpha_{a0} + \alpha_{a1} \Delta \ln Q^t + \alpha_{a2} \Delta \ln R^t + \mu_a^t \\
\Delta \ln Q^t &= \alpha_{b0} + \alpha_{b1} \Delta \ln X^t + \alpha_{b2} \Delta \ln R^t + \mu_b^t.
\end{aligned}
\tag{16}$$

In (16) we have assumed that time autonomous technical 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 – the R&D input 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 towards addressing the endogeneity problem. At the same time, they tend to give rise to other problems. Diewert and Fox (2008) provide an in-depth discussion of estimation in a similar context and note in regards 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) find 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 which leads 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). We therefore follow Diewert and Fox (2008), Basu and Fernald (1997, 2002) and Roeger (1995) and rely on OLS estimates.

Another, related point is that all variables – and in particular the R&D

variable - are likely measured with error ⁷ . When there is a measurement error in the regressor and it is of the classical type, i.e., independent of the true value of the variable, OLS estimates have been shown to under-estimate the magnitude of the regression coefficient (see, for instance Hyslop and Imbens 2001). ⁸ Klepper and Leamer (1984) have demonstrated that with classical measurement error in the two-variable case the true value of the regression coefficient lies between the estimated coefficients ⁹ 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 1985-2015 are shown in (17) where fixed effects for countries and years have been applied and standard errors are shown in brackets:

$$\begin{aligned}
 \Delta \ln X^t &= 1.008 + 0.533 \Delta \ln Q^t - 0.045 \Delta \ln R^t; \text{adj} R^2 = 0.65; DF = 564 \\
 &\quad (0.316) \quad (0.026) \quad (0.008) \\
 \Delta \ln Q^t &= 1.011 + 0.797 \Delta \ln X^t + 0.115 \Delta \ln R^t; \text{adj} R^2 = 0.77; DF = 564. \\
 &\quad (0.388) \quad (0.039) \quad (0.009)
 \end{aligned}
 \tag{17}$$

All coefficients are significant and show the right sign. However, as expected, direct and reverse regression 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 $1/0.530 = 1.88$ when based on the first result in (17) or 0.797 when based on the second result in (17). The cost elasticity of the R&D asset as implied by the first regression equals $w_{RS}R/C^* = (w_{RS}R/C)(C/C^*) =$

⁷The econometric issues with 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 analysing R&D in a production context, pioneered by Griliches (1973) and recently reviewed by Ugur et al (2016).

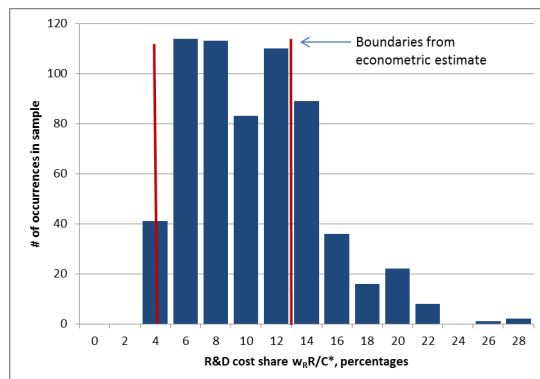
⁸When there is classical measurement error in both the regressor and the dependent variable, the OLS bias cannot in general be signed, unless it is assumed that the measurement errors of the regressor and the dependent variable are independent in which case the downward bias in regression coefficients remains.

⁹Klepper and Leamer (1984) also demonstrate 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 – i.e., 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.

$(w_{RS}R/C^*)/(1 + w_{RS}R/C^*) = 0.045/(1 + 0.045) = 0.043$. The cost elasticity of R&D as implied by the second regression equals $w_{RS}R/C^* = [(\epsilon w_{RS}R/C)/\epsilon][1 + \epsilon w_{RS}R/\epsilon C] = (0.115/0.797)/(1 + 0.115/0.797) = 0.126$. Thus, our lower bound for the cost share as recovered by the estimation is around 4% and the upper bound is around 13%. We thus find a rather large possible range of cost elasticities for R&D ¹⁰.

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 9.7%, with a minimum value of around 2%, and a maximum value of 66% ¹¹. Figure (1) below shows the frequency distribution of all $w_R R/C^*$, along with the upper and lower boundaries from the regression results. About 2/3 of all computed values lie within these bounds and we conclude that the econometric results do not offer significant additional insight over the unconstrained index number results.

Figure 1: Cost-elasticities of R&D: distribution of unrestricted measures and econometric results



Source: authors' calculations, based on *OECD Productivity Database June 2018*

¹⁰If 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 41%

¹¹This 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

Ugur et al (2016) conduct 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 $\Delta\pi^t = \alpha_{c0} + \mu_c^t$:

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

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. 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 Haines et al., 2017) and, in a somewhat different context, as an estimate for spillovers from ICT and intangibles (Stiroh 2002, Corrado et al. 2014). 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 (15) 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 (15) reads as follows:

$$\begin{aligned}\Delta \ln Z^{*t} &= \frac{1}{\epsilon^*} \Delta \ln Q^t - \Delta \pi^t, \\ \Delta \ln Q^t &= \epsilon^* (\Delta \ln Z^{*t} + \Delta \pi^t); \end{aligned} \quad (19)$$

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

$$\begin{aligned}\Delta \ln Z^{*t} &= \alpha_{d0} + \alpha_{d1} \Delta \ln Q^t - \mu_d^t \\ \Delta \ln Q^t &= \alpha_{e0} + \alpha_{e1} \Delta \ln Z^{*t} + \mu_e^t. \end{aligned} \quad (20)$$

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), 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 mean of 1.19. 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 which corresponds to $\epsilon^* = 1.2$.

This is in line with related research. For instance, Diewert and Fox (2008) find a scale elasticity of between 1.2 and 1.5 for U.S manufacturing industry. Basu and Fernald (1997) produce 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.

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 MFP growth into an element that reflects returns to scale, $(1 - 1/\epsilon^*)\Delta \ln Q^t$, and into an element of ‘residual’ productivity growth, $\Delta \pi_S^t$. The qualification ‘residual’ is important because there are almost certainly other forces than pure technical change that affect this measure.

$$MFP^t = \Delta Q^t - \Delta Z^{*t} = (1 - 1/\epsilon^*)\Delta \ln Q^t + \Delta \pi_S^t. \quad (21)$$

Figure (2) exhibits results of this decomposition for 20 OECD countries over the period 1985-2016 based on our preferred average value $\epsilon^* = 1.2$. 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.

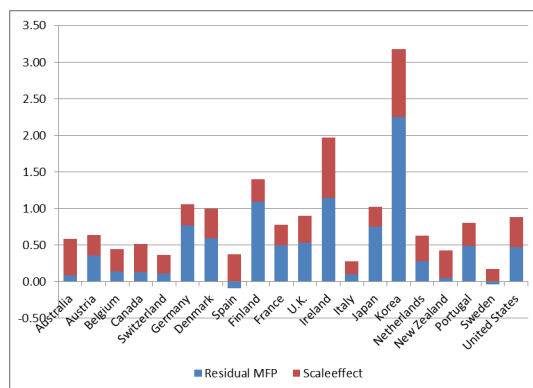
A scale effect of some magnitude has policy-relevant consequences.

- 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 declines (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). Further, some of the procyclical nature of productivity growth can be explained when demand affects productivity, as has been suggested by Hall (1988) and Basu and Fernald (1997). We do find, however, that $\Delta\pi_S^t$ remains a series of high variance.

- 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.
- A third consequence is that increasing returns to scale imply the existence of mark-ups over marginal costs and therefore some monopolistic elements. Whether or not 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.

Figure 2: Scale effects and residual MFP
Annual average percentage changes, 1985-2015*



*Portugal, Spain and Sweden: 2015.

Source: authors' calculations, based on *OECD Productivity Database June 2018*

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 $M^{*t}/(P_Q^t Q^t)$ or ‘residual’ mark-up rates M^{*t}/C^{*t} over marginal costs. M^{*t} is the difference between labour 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 years is around 1.3 or a 30% addition to marginal costs. This is broadly consistent with early work by Oliveira-Martins et al. (1996), and Christopoulou and Vermeulen (2012), although the authors assume constant returns and consider the private sector rather the total economy. Diewert and Fox (2008) derive mark-ups between 1.4 and 1.7 for U.S. manufacturing, Devereux et al. (1996) review the literature and estimate that mark-ups of up to 1.5 constitute a plausible value for use in modelling. De Loecker and Warzynski (2012), in a firm-level study of Slovenian manufacturing firms, obtain mark-ups in the range of 1.17–1.28. As in other studies, mark-up levels across countries vary significantly, as can be seen from Figure 3. This reflects a host of factors, including the degree of competition and regulation, differences in the presence and in the returns to other assets such as natural resources or intangibles that have not been explicitly captured; and measurement issues.

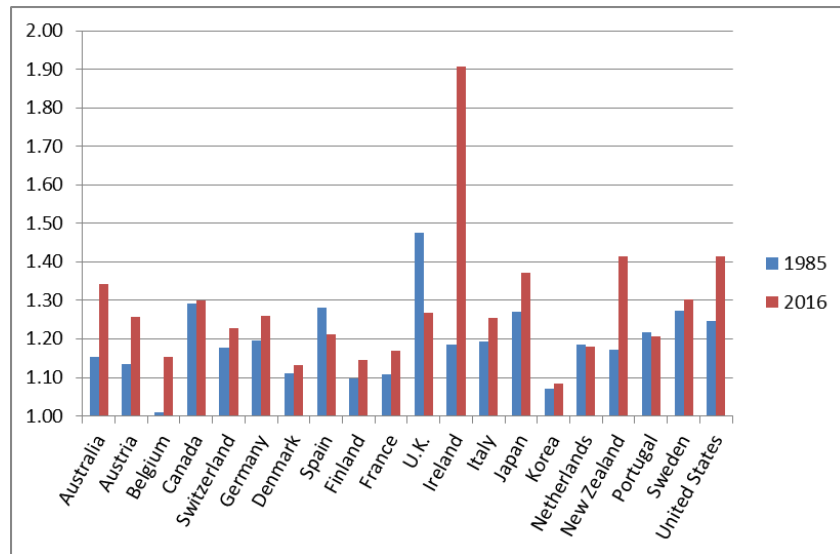
It should be recalled here that the level of residual mark-ups M^{*t} also reflects assumptions about the longer-run real rate of return to capital that have entered the computation of user costs (Annex A). Indeed, the standard way to proceed (Jorgenson 1985, Jorgenson and Landefeld 2004) is letting 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 equals exactly total costs. Absent M^{*t} , the mark-up rate over marginal costs equals exactly the degree of returns to scale as can be seen 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.

Karabarbounis and Neiman (2018) explore several hypotheses about the sources of ‘factorless income’, which corresponds to our measure of economic profits, M_{*t} . Their favoured 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” (p55). In other words, they hypothesise 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).

Figure 4 shows how mark-up rates over marginal costs develop over time, measured as $1.2[1 + M^{*t}/C^{*t}]$. The same pattern holds for residual mark-up rates $1 + M^{*t}/C^{*t}$ but scaled down by the (constant) degree of returns to scale $1/\epsilon^* = 0.83$. One notes that with a time-invariant ϵ^* , all changes in overall mark-up rates ($1 + m^{*t}$) are triggered by changes in residual mark-up rates 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 1985-2016, 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 dataset also observe upward trending average mark-ups in OECD countries, mostly driven by firms in market services sectors. Analysis of causes of this secular increase in mark-ups over marginal costs is beyond the scope of this paper but several possibilities suggest themselves:

- Rising returns to produced assets, as a reflection of rising risk premia. Karabarbounis and Neiman (2018) explore several hypotheses about the sources of ‘factorless income’, which corresponds to our measure of economic profits, M^*t . Their favoured 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 hypothesise 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 are 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

Figure 3: Mark-ups over marginal costs - by country



*Portugal, Spain and Sweden: 2015

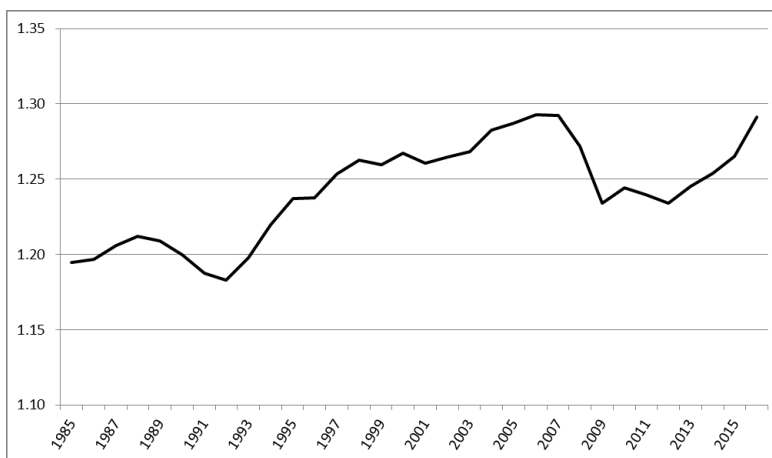
Source: authors' calculations, based on *OECD Productivity Database June 2018*

with non-financial investment would be an important subject of future research.

- 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. This is the argument pursued in Calligaris et al. (2018), who show 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 intellectual property assets.
- 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 recognised in the present computations. When of the intangible kind, these assets include human capital, organisational capital, or marketing assets as investigated by Corrado et al. (2005), OECD (2013) or Goodridge et al (2016). When of the tangible kind, these assets include in particular land

whose real price (and real return) has registered an upward trend over the past decades in many OECD countries.

Figure 4: Mark-ups over marginal costs - average across countries



*Unweighted average. Portugal, Spain and Sweden: 2015

Source: authors' calculations, based on *OECD Productivity Database June 2018*

4 Conclusions

With the implementation of the 2008 System of National Accounts, R&D capital stock measures are now widely available in OECD countries. While 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 provide a specific type of service, they are replicable and easily transferable 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 has to rely on more assumptions than is normally the case for other assets. We investigate whether the usual assumption of period-to-period cost-minimising 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 non-constant 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’ 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.

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 upwards 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 This chimes well with effects associated with globalisation and digitalisation 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

- [1] Andrews, D., C. Criscuolo and P. N. Gal (2016), “The Global Productivity Slowdown, Divergence Across Firms and the Role of Public Policy”, OECD Productivity Working Papers No 5.
- [2] Andrews, D., P. Gal and W. Witheridge (2018), “A Genie in a Bottle? Globalisation, Competition and Inflation”, OECD Economics Department Working Papers, No. 1462.
- [3] Auclert, M. and M. Rognlie (2018), ‘Inequality and Aggregate Demand’, NBER Working Paper No. 24280.
- [4] Balk, Bert M. (1998); *Industrial Price, Quantity and Productivity Indices, The Micro-economic Theory and an Application*, Kluwer Academic Publishers.
- [5] Basu, S., Fernald, J.G., (1997) ”Returns to scale in US production: Estimates and Implications”; *Journal of Political Economy* 105, 249–283.

- [6] Basu, S., Fernald, J.G., (2002), “Aggregate productivity and aggregate technology”, *European Economic Review* 46, 963–991.
- [7] Berndt, E. R. and M. A. Fuss (1986), “Productivity Measurement with Adjustments for Variations in Capacity Utilisation and Other Form of Temporary Equilibria,” *Journal of Econometrics*, Vol. 33, 7-29.
- [8] Burnside, C., (1996), “Production function regressions, returns to scale, and externalities”, *Journal of Monetary Economics* 37, 177–201.
- [9] Caballero, R., E. Farhi, and P.-O. Gourinchas (2017); “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), pp 614-620.
- [10] Calligaris, S., C. Criscuolo and L. Marcolin (2018), “Mark-ups in the digital era”, *OECD Science, Technology and Industry Working Papers*, No. 2018/10, OECD Publishing, Paris, <https://doi.org/10.1787/4efe2d25-en>.
- [11] Caves, D. W., L. R. Christensen and J.A. Swanson, (1981), ‘ Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955-74’, *The American Economic Review*, Vol. 71, No. 5 (Dec., 1981), pp. 994-1002.
- [12] Christensen, L.R., Jorgenson, D.W., Lau, L.J., (1971), “Conjugate duality and the transcendental logarithmic production function”, *Econometrica* 39, p. 255–256.
- [13] Christopoulou, R. and Ph. Vermeulen (2012), “Markups in the Euro area and the US over the period 1981–2004: a comparison of 50 sectors”, *Empirical Economics*, vol. 42, issue 1, 53-77.
- [14] Corrado, C., J. Haskel and C. Jona-Lasinio (2014), “Knowledge Spillovers, ICT and Productivity Growth”, IZA DP No. 8274, <http://ftp.iza.org/dp8274.pdf>.
- [15] Corrado, C., Ch. Hulten and D. Sichel (2005); “Measuring Capital and Technology: An Expanded Framework”; in Corrado, Carol, Charles Hulten and Daniel Sichel (eds.); *Measuring Capital in the New Economy*; National

Bureau of Economic Research, Studies in Income and Wealth Volume 65, pp. 11-46, University of Chicago Press.

- [16] Diewert, W.E. (1980), “Aggregation Problems in the Measurement of Capital”, in D. Usher (ed.) *The Measurement of Capital*, University of Chicago Press, pp. 433 – 538.
- [17] Diewert, W. E. (1974). “Applications of duality theory”. In: Intriligator, M.D., Kendrick, D.A. (Eds.), *Frontiers of Quantitative Economics*, vol. II. North-Holland Publishing Co., Amsterdam, pp. 106–171.
- [18] Diewert, W. E. (1976), “Exact and Superlative Index Numbers”; *Journal of Econometrics* 4, 115-145.
- [19] Diewert, W. E. and K. J. Fox (2008), “On the estimation of returns to scale, technical progress and monopolistic markups”, *Journal of Econometrics* 145 (2008) 174–193.
- [20] Diewert, W. E. and N. Huang (2011); ‘Capitalizing R&D Expenditures’, *Macroeconomic Dynamics*, 15, 2011, 537–564.
- [21] Diewert, W. E., T. Nakajima, A. Nakamura, E. Nakamura and M. Nakamura (2011); “Returns to scale: concept, estimation and analysis of Japan’s turbulent 1964–88 economy”; *Canadian Journal of Economics / Revue canadienne d’Economie*, Vol. 44, No. 2; May, p.451-485.
- [22] De Loecker, J. and F. Warzynski (2012), “Markups and Firm-Level Export Status”, *American Economic Review* 102(6): 2437–2471.
- [23] Devereux, M.B., Head, A.C., and Lapham, B.J., (1996), “Monopolistic competition, increasing returns, and the effects of government spending”, *Journal of Money, Credit, and Banking* 28, pp 233–254.
- [24] European Commission, IMF, OECD, World Bank, United Nations (2009), *2008 System of National Accounts*, New York, United Nations; Available under <http://unstats.un.org/unsd/nationalaccount/sna2008.asp>
- [25] Galindo-Rueda, F., Ker, D., Moris, F. and Jankowski, J. (2018); “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) 9-10 March 2018, Washington DC.

- [26] Goodridge, P. J. Haskel, G. Wallis (2016), “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.
- [27] Griliches, Z., 1973. “Research expenditures and growth accounting”. In: Williams, R.B. (Ed.), *Science and Technology in Economic Growth*. John Wiley and Sons, New York.
- [28] Griliches, Z. (1998), “R&D and Productivity: The Unfinished Business”, in Z. Griliches (ed.), *R&D and Productivity: The Econometric Evidence*, University of Chicago Press.
- [29] Griliches, Z. (1995), ‘R&D and Productivity’, in P. Stoneman (ed.), *Handbook of the Economics of Innovation and Technical Change*, 52-89, Oxford.
- [30] Haines, W., L. Marcolin, L. and M. Squiccarini (forthcoming), “Intangibles and Investment – A New Assessment”; Report to the OECD Working Party on Industrial Analysis 2017.
- [31] Hall, R. E. (1988), “The Relationship between Price and Marginal Cost in U. S. Industry”, *Journal of Political Economy* 96, 921-947.
- [32] Huang, N. and W. E. Diewert (2011), ‘Estimation of R&D Depreciation Rates: A Suggested Methodology and Preliminary Application’, *The Canadian Journal of Economics*, Vol. 44, No. 2 (May), pp. 387-412.
- [33] Hulten, Ch. R. (1990), “The Measurement of Capital”, *Fifty Years of Economic Measurement*, E.R. Berndt and J.E. Triplett (eds.), *Studies in Income and Wealth*, Volume 54, The National Bureau of Economic Research, The University of Chicago Press, , Chicago pp. 119-152.
- [34] Hyslop, D. R. and G. W. Imbens (2001),”Bias From Classical and Other Forms of Measurement Error”, *Journal of Business & Economic Statistics*, October 2001, Vol. 19, No. 4, pp. 475-481.

- [35] Jorgenson, D. W. and J. S. Landefeld (2004); “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 Account*, NBER, Studies in Income and Wealth, University of Chicago Press.
- [36] Jorgenson, D. W. (1995), *Productivity; Volumes I and II*, MIT Press.
- [37] Klepper, S. and E. E. Leamer (1984), “Consistent Sets of Estimates for Regressions with Errors in All Variables”, *Econometrica*, Vol. 52, No. 1 (Jan), 163-184.
- [38] Lau, L. J. (1976), ”A Characterization of the Normalized, Restricted Profit Function,” *J. Econ. Theory*, Feb. 1976, 12, 131-63.
- [39] Leontief, W. W. (1947), “Introduction to a theory of the internal structure of functional relationships”, *Econometrica* 15; pp. 361-73.
- [40] Li, W. C.Y. (2012), “Depreciation of Business R&D Capital”, Bureau of Economic Analysis/National Science Foundation R&D Satellite Account Paper.
- [41] Karabarbounis, L and B. Neiman (2018), “Accounting for Factorless Income”, NBER Working Paper 24404.
- [42] McFadden, D. (1978); ‘Cost, Revenue and Profit Functions’ in *Production Economics: A Dual Approach to Theory and Applications*, Vol. 1, ed. by M. Fuss and D. McFadden (Amsterdam: North-Holland), pp. 3–109.
- [43] Morrison, Catherine J. (1992). “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
- [44] Niebel, T., M. O’Mahony and M. Saam (2013), “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>
- [45] OECD (2001), *Measuring Productivity*; OECD Publishing.
- [46] OECD (2009), *Measuring Capital*; OECD Publishing.

- [47] OECD (2010), Handbook on Deriving Capital Measures of Intellectual Property Products; OECD Publishing; available under <http://www.oecd.org/std/na/44312350.pdf>.
- [48] OECD (2013), Supporting Investment in Knowledge Capital, Growth and Innovation, Paris <http://dx.doi.org/10.1787/9789264193307-en>
- [49] OECD (2015), 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>.
- [50] OECD (2017), OECD Compendium of Productivity Indicators 2017, OECD Publishing Paris.
- [51] Oliveira Martins, J., S. Scarpetta and D. Pilat (1996); “Mark-up Pricing, Market Structure and the Business Cycle”; OECD Economic Studies, No . 27 (2), 71-105.
- [52] Parham, D. (2007); 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, available at <https://www.oecd.org/std/productivity-stats/37511005.pdf>
- [53] Pitzer, J. S. (2004) ‘Intangible produced assets,’ Paper presented at the London Meeting of the Canberra 2 Group: On the Measurement of Non-Financial Assets, 1-3 September.
- [54] Roeger, W. (1995), “Can imperfect competition explain the difference between primal and dual productivity measures? Estimates for US manufacturing”, *Journal of Political Economy* 103, 316–330.
- [55] Roth, F. and A-E. Thum (2013), “Intangible Capital and Labor Productivity Growth: Panel Evidence for the EU from 1998–2005”, *Review of Income and Wealth*, Vol. 59 (3), pp. 486-508.
- [56] Schankerman, M. and M. I. Nadiri (1984); “Restricted Cost Functions and the Rates of Return to Quasi-Fixed Factors, with an Application to R&D in the Bell System”; NBER Working Paper No 1259.

- [57] Schreyer, P. (2010); “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 Volumes 1 and 2*, Trafford Press.
- [58] Schreyer, P., P-E. Bignon and J. Dupont, (2003); “OECD Capital Services Estimates: Methodology and a First Set of Results”; OECD Statistics Working Paper 2003/6.
- [59] Sena, V. (2004), “The return of the Prince of Denmark: a survey and recent developments in the economics of innovation”, *Economic Journal*, 114 (2004), pp. F312-F332.
- [60] Shephard, R.W. (1953). *Cost and Production Functions*. Princeton University Press, Princeton, NJ. Sono, M. (1961). “The effect of price changes on the demand and supply of separable goods”; *International Economic Review* 2: 239-75.
- [61] Stiroh, K. (2002), ‘Are ICT Spillovers Driving the New Economy?’, *Review of Income and Wealth* 48(1), 33–57.
- [62] Summers, L. H.(2015) “Demand Side Secular Stagnation,” *American Economic Review*, May, 105 (5), 60–65.
- [63] Ugur, M., E. Trushinb, E. Solomona and F. Guidi (2016), “R&D and productivity in OECD firms and industries: A hierarchical meta-regression analysis”, *Research Policy*, Volume 45, Issue 10, December 2016, Pages 2069-2086.

Annex A Measurement and Data Sources

R&D assets

The measurement of capital requires a number of methodological choices that, more often than not, suffer from weak empirical support and require more or less well-founded assumptions by statisticians. Examples include the choice of service lives (or depreciation rates), retirement distributions and the form of the age-efficiency function (OECD 2009). Some of these choices matter little for the final productivity measure. But with large and lumpy shifts of the asset base as observed recently with intellectual property products, they may become important.

We start with a representation of the production of the IP asset itself. In line with the 2008 SNA, R&D is an investment activity that adds to final demand and GDP. Investment may happen as a result of own-account production in the functional unit of a larger enterprise or in a separate corporation. Statistical practice now introduces several simplifications to deal with missing information in regards to R&D.

Absent market observations on the value of own-account research, a first constraint is that the gross value of research output at current prices has to be measured by summing costs – compensation of employees, user costs of capital employed in R&D and (other) taxes on production. Current price value-added (gross output net of intermediate inputs) is then measured by summing the value of primary inputs labour and capital. Thus, the value-added created in R&D firms or production units in period t equals

$$P_R^t I_R^t = \sum_i w_{X_i}^t X_{Ri}^t. \quad (\text{A.1})$$

In (A.1), I_R^t is the volume of R&D output (in value-added terms) in period t and $X_R \equiv [X_{R1}, X_{R2}, \dots]$ captures volumes of labour and capital services purchased at prices $w_X \equiv [w_{X1}, w_{X2}, \dots]$. Although we specify price and volume components P_R^t and I_R^t for R&D output, these are not in general separately observable. A second assumption is necessary here, namely that the volume change of research output is measured by the volume change of its inputs. By implication, productivity growth in R&D production is zero and the price index of research output moves in tandem with the price index of research inputs ¹².

¹²The implied production function is $I_R = f_R(X_R)$. Note that this reflects a statistical

Absence of an independent measure of the price P_R and its movements over time also implies that the usual assumption that P_R is the equilibrium price generated on the market for capital goods does not necessarily hold. Such an equilibrium price connects the (marginal) cost of producing a unit of R&D investment with the discounted stream of future revenues that is expected from using R&D in production. There is no guarantee that the input-based price that is imputed by statisticians reflects such an equilibrium price. However, a valuation by private asset owners may be observable when assets are sold or transferred. As will be seen below, this raises an issue of consistency of valuation of capital measures.

A third element of statistical practice - indeed, needed for most types of assets and not only for R&D - is that measures of stocks are constructed by cumulating measures of flows of investment volumes over time after correcting for depreciation and retirement:

$$P_R^t R^t = \lambda_0 P_R^t I_R^t + \lambda_1 P_R^t I_R^{t-1} + \lambda_2 P_R^t I_R^{t-2} + \dots \quad (\text{A.2})$$

The sequence $1 \geq \lambda_0 \geq \lambda_1 \geq \lambda_N > 0$ captures the depreciation, retirement and obsolescence patterns for a service life of N periods. One issue is the choice of service lives N and the implied rates of depreciation. We shall devote little space to this question here although we note that depreciation of IP assets reflects obsolescence or patent expiration rather than physical wear-and-tear. This complicates the estimation of depreciation rates. Diewert and Huang (2011) and Li (2012) show how N and the sequence of λ can be derived.

In (A.2), the sequence of $[P_R^t I_R^{t-i}]$ was somewhat loosely referred to as investment flows. This requires some precision. Capital formation does not only consist of newly produced investment products but may also include existing or second-hand assets that are being acquired. Another, less frequent, source of additions to capital is the ‘appearance’ of assets. This may arise with discoveries of natural resources or with the transfer of an asset within a (multinational) corporation. Both acquired existing assets and appearing assets need to be added to a country’s or industry’s capital stock if they generate capital services. Thus, for any period t , the addition to the capital stock is $\lambda_i (P_R^t I_R^{t-i} + P_R^t I_{RA}^{t-i})$

constraint rather than economic reasoning. If independent volume measures or deflators for research output are available, the zero productivity growth assumption is not needed as the growth rate of IR can be estimated independently from the volume of inputs. In this case, the production function would read as $I_R = f_{R1}(X_R, t)$.

where $P_R^t I_{RA}^{t-i}$ is the volume of the appearing stock I_{RA}^{t-i} , valued in prices P_R^t of year t . If the source of information for the appearing asset is a company balance sheet, this creates a potential inconsistency as companies may have applied a different valuation from P_R^t , call it P_{RA}^t . Unless a revaluation is undertaken, there is a danger of inconsistency if $P_{RA}^t I_{RA}^{t-i}$ rather than $P_R^t I_{RA}^{t-i}$ enters the computation of the capital stock. Consistent revaluation requires also that information is available about the remaining service life of the appearing asset. Such a revaluation and adjustment for the age of the appearing asset is not always possible absent relevant information. The analyst faces a trade-off between an inconsistency in valuation as well as an inaccurate depreciation profile and not accounting for the appearing (or disappearing) asset at all. It would seem that the latter likely constitutes a worse choice than the former.

There is also the selection of the depreciation pattern. A common choice is a geometric pattern where a cohort of assets loses value and productive capacity at a constant rate. Another, widely used sequence is the hyperbolic age-efficiency profile for $\lambda : \lambda_i = \frac{N-i}{N-bi}; i = 0, 1, 2, \dots, N; 0 < b \leq 1$ implying that the service flows from assets decline little at first and more rapidly towards the end of the service life. An extreme case of the hyperbolic profile arises with $b = 1$ for $i = 1, 2, \dots, N$ so that $\lambda = 1$ throughout the asset's service life and dropping to zero thereafter ('one-hoss shay'). In the case of knowledge assets it stands to reason that service flows follow a hyperbolic or one-hoss shay profile: absent any wear and tear, there is a non-diminished flow of services during the asset's service life coupled with a rapid decline at the end of the service life. However, things may be different if one reasons in terms of cohort of assets rather than a single asset. For whole cohorts, it is necessary to introduce a retirement distribution unless it is assumed that N is identical for all individual assets within the same cohort. The sequence of service flows for an entire cohort may look quite different from that for an individual asset (Hulten 1990).

The treatment of appearing assets that are lumpy and large requires also careful attention to infra-annual patterns (assuming that observations are annual) so that large additions to the capital stock appear when they actually provide capital services and affect output. Note that in line with national accounts conventions, investment flows or appearance of assets (I_R^t, I_{RA}^t) are measured in terms of average values of the period. Whether they affect productive stocks R^t and associated service flows at the beginning, in the middle or at the end of

period t does not normally matter but may become important when I_R^t or I_{RA}^t are large, discrete flows.

In summary, then, while the principles of measuring R&D capital are aligned with other types of assets (OECD 2010), there are some major complications that are specific to R&D (and other knowledge-based assets):

- It is often difficult to obtain independent observations on the value and price of R&D investment, which requires applying an input-based approach. There is also greater uncertainty about the accuracy of rates of depreciation – or obsolescence – than with many other fixed assets.
- As intellectual property assets can easily be transferred across borders, there is the possibility of large appearances of such assets on countries' balance sheets. These additions to the capital stock should be recognised in the measurement of capital services although they raise further issues of valuation and estimation of their remaining service lives.

The *OECD Productivity Database* uses the perpetual inventory method as in (A.2) to compute stocks of R&D capital. The age-efficiency pattern is hyperbolic with a service life of 10 years and the retirement function follows a normal distribution with a standard deviation of 25% of the average service life. Investment data on R&D is augmented by the value of appearing assets where this plays a sizable role, for example in Ireland. National deflators for R&D investment are applied which in general reflect price changes of inputs in R&D activity.

Flows of R&D investment expenditure are sourced from countries' national accounts as compiled in the OECD's Annual National Accounts database. These are broadly consistent with data on R&D performance as compiled in line with the OECD Frascati Manual (2015) although differences arise in particular where R&D assets are traded or transferred internationally. As Galindo-Rueda et al (2018) point out "Notwithstanding practical differences across R&D performance measures and SNA IPP investment statistics [...], the globalisation of R&D appears to be, as expected, a first order factor underpinning observed differences between Frascati-based statistics on R&D performance and the SNA view of how much countries invest in R&D. In most countries, the value of R&D assets capitalised annually has been fairly similar to the value of domestic R&D performance, with the ratio of R&D investment to performance sitting in a band

between roughly 80% and 110% in many cases and being relatively stable over time. However, divergence has been more marked in countries characterised by large international R&D related flows. In Ireland, R&D investment has grown much more quickly than GERD since around 1997. This difference is driven by large imports of R&D assets [...] By contrast, in Israel R&D investment is estimated to be less than half of R&D performance in 2014, having declined from nearer 100% in the 1990s. ” For purposes of capturing capital input in productivity measurement, the R&D stocks adjusted for imports and exports would appear to be the preferred concept and have been used in the work at hand.

4.1 Capital services

More generally, in the OECD Productivity Database capital services provided to production by each type of capital good are estimated by the rate of change of their productive capital stocks. Estimates of productive capital stock are computed using the perpetual inventory method on the assumption that the same service lives and retirement functions are applicable for any given asset irrespective of the country. Productive capital stocks and the respective flows of capital services are computed separately for eight non-residential fixed assets. The following average service lives are currently assumed for the different assets: 7 years for computer hardware, 15 years for telecommunications equipment, transport equipment, and other machinery and equipment and weapons systems, 40 years for non-residential construction, 3 years for computer software and databases, 10 years for R&D and 7 years for other intellectual property products. The approach further uses harmonised deflators for computer hardware, telecommunications equipment and computer software and databases, for all countries, to sort out comparability problems that exist in national practices for deflation for this group of assets (Schreyer, 2002; Colecchia and Schreyer, 2002). The overall volume measure of capital services is computed with a Törnqvist index by aggregating the volume change of capital services of all individual assets using asset specific user cost shares as weights.

For R&D assets, the value of capital services is measured as $\sum_s u_0^t R_s^t$ where u_0^t is the user cost per unit of a new asset and $R_s^t = \lambda_s I_R^{t-s}$ is the volume of the s-year old asset expressed in ‘new equivalent’ units. User costs are defined as $u_0^t = P_R^{t-1}(r^t + d_0^t - \zeta^t + d_0^t \zeta^t)$ where P_R^{t-1} is the purchase price of an asset

at the end of period $t - 1$, d_0^t is the rate of depreciation for a new asset, ζ^t is the rate of price change of a new asset, $d_0^t \zeta^t$ is an interaction term and r^t is the net rate of return. We obtain a value for the expected nominal rate of return r^t by first computing a long-run average of observed real interest rates in countries (nominal financial market interest rates deflated with a consumer price index). The so-obtained real interest rate is then reflated with a smoothed consumer price index. ζ^t is measured as a smoothed series of nominal asset price changes. Similar procedures are applied to other types of assets, each with an asset-specific depreciation and asset-price measure. Further details can be found in Schreyer et al (2003) and OECD (2009).

4.2 Labour inputs

The preferred measure of labour input in the OECD Productivity Database, and hence the labour input measure used in this paper, is the total number of hours worked by all persons engaged in production (i.e. employees plus self-employed). While the preferred source for total hours worked in the database is countries' national accounts, in the case of Japan and New Zealand, for which national accounts data on hours worked are not available at the time of writing this paper, other sources have been used, i.e. data from labour force surveys as published in the OECD Employment and Labour Market Statistics.

Annex B Tables by Country

Available at <http://www.oecd.org/sdd/Annex-B-tables-by-country.xlsx>