Measuring productivity: theory and British practice

Nicholas Oulton

Centre for Macroeconomics, LSE, NIESR and ESCoE

January 2020

Email: n.oulton@lse.ac.uk

Contents

Abstract

This paper lays out the basic theory behind productivity measurement, whether at the level of the country, region, industry or firm. The theory is illustrated using recent data from UK official publications. Productivity growth over time and differences in productivity levels between countries or regions at a point in time are both covered. Labour productivity and multi-factor productivity (MFP) are discussed. In the case of MFP special attention is paid to the measurement of capital inputs. Wherever possible, an accompanying spreadsheet supplies data from recent publications by the United Kingdom's Office for National Statistics so that readers can reproduce official estimates or even employ alternative assumptions to produce their own estimates. Limitations in the underlying theory are highlighted as are empirical difficulties in implementing the theory.

1. Introduction[1]

The purpose of this paper is to set out the theory of productivity measurement in an informal way and then to study how this theory is currently applied in British official statistics. A spreadsheet accompanying the paper, Data for Oulton (2020).xlsx, contains underlying recent data from which readers can reconstruct (some) official statistics on productivity. It also allows readers to construct their own estimates using alternative assumptions or formulas. The paper documents, though does not claim to explain, the so-called productivity puzzle: the remarkable stagnation in productivity (and living standards) since the start of the Great Recession in 2008.

We shall study productivity, defined broadly as output per unit of input. We shall start by looking at labour productivity which is output per unit of labour. First we shall consider why labour productivity is important. Then we go on to look at how it is measured both in theory and in practice. The growth of labour productivity over time and the differences between countries in productivity levels will be studied. The next topic is labour productivity at the industry level and the role of structural change. The growth rate of labour productivity at the whole economy level results from growth in individual industries but also from shifts in the structure of the economy. How do these two forces balance out in practice? After growth rates we consider levels. We study differences in labour productivity levels between the main regions of Britain and next differences between Britain and a range of other countries in size, living standards and labour productivity. International comparisons require an understanding of how different currencies can be converted to a common basis which is done by means of purchasing power parities (PPPs).

Then we shall broaden the discussion beyond labour productivity to analyse productivity in relation to all inputs at once, so-called multi-factor productivity (MFP), also known as total factor productivity (TFP). We shall show why this concept matters and discuss the techniques needed to measure it in practice, in particular the special difficulty involved in measuring capital inputs. A decline in multi-factor productivity growth turns out to be an important feature of the productivity puzzle.

[1] This paper is intended to form the basis of a chapter on productivity in the forthcoming ebook on measuring the economy from the UK's Office for National Statistics (https://measuringtheeconomy.uk/). I am grateful to Peter Sinclair for comments on an earlier draft.

Productivity can also be studied at the firm level. What can we learn from looking at how productivity varies between established and continuing firms, firms which are new entrants to an industry, and firms which are exiting their industry?

Along the way, some limitations of the current theoretical framework are pointed out as are some empirical difficulties in applying the framework. These are summarised in the conclusions.

2. Motivation

According to the much-quoted Nobel Prize winner Paul Krugman (1994, chapter 1), "Productivity isn't everything, but in the long run it is almost everything. A country's ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker."[2] If you are not convinced that productivity really is as important as Krugman says, take a look at Table 1 which shows long run growth in labour productivity in Britain as measured by real GDP per hour worked. Also shown is a common measure of living standards, real GDP per head, i.e. real GDP divided by the number of people in the population. Relative to the starting point in 1856, labour productivity in 2016 has been multiplied by a factor of 17.4 and living standards by a factor of 8.8.

If we divide real GDP per head by real GDP per hour we get hours per head, another interesting statistic. Hours per head have halved over this period. This change reflects many factors such as the near elimination of child labour, the increased percentage of the population that is retired due to longer life spans, the reduction in the average hours worked per year by those in employment, and (going the other way) the increased participation of women in paid work outside the home.[3] Rising life expectancy and increased leisure are aspects of welfare not captured by GDP.

Table 1 near here

[2] If you look up this famous quote on the Internet you will find a number of different versions, some of which don't make very good sense. I have taken my version from the printed book. As always, don't believe everything you read on the Internet.

[3] Remember that unpaid work in the home is not counted as part of GDP.

The full run of data for GDP per hour and GDP per head are shown in Chart 1. A casual glance at this chart might suggest that productivity growth was slow up to the end of the second world war but then increased sharply with this higher rate persisting up to 2007. It might seem too that the gap between the GDP per hour and GDP per head was quite small up to around 1918 but then exploded. *These conclusion are largely wrong. They result from the vertical scale used in this chart.* Exactly the same data are shown in Chart 2, but this time the vertical axis is a log scale. On a log scale, the slope of the line shows the growth rate: the steeper the line, the higher the growth rate. The vertical gap between two series shows the *percentage* difference between them. Now we see that productivity growth (the red line) was initially growing quite rapidly, then the growth rate fell and on average was fairly low from 1874 to 1939. After the second world war growth was more rapid, but declined quite a bit after 1973 though it was still faster than in the inter-war period, at least up until 2007. Since 2007 productivity growth has averaged close to zero, a fact known as the productivity puzzle on which more later. Also the gap between GDP per hour and GDP per head increases fairly steadily over the whole period.[4] This is made clear in Table 2 which shows average annual growth rates over various sub-periods within the overall span 1856-2016.

Charts 1 and 2 near here

Table 2 near here

For the study of quantities that are growing steadily over long periods it is essential to use a log scale in charts. So you should ignore Chart 1, which is just here as a warning, and consider only Chart 2. There are two ways to calculate growth rates, the geometric and the exponential. Table 2 uses the exponential method in which the growth rate is the log difference of the variable being analysed divided by the length of the time interval. The exponential method is favoured by economists since many economic relationships are specified in logs. Also the growth rate of variables like labour productivity is just the growth of output minus the growth of labour input. This is not the case with the alternative geometric method. The geometric method is however the one typically used by national statistical agencies like the ONS. For low growth rates the results are very similar but differences start to become noticeable when growth is at recent Chinese or Indian rates (8-10% per year). See Box 6.1 for some technical stuff on log scales and growth rates.

[4] See the spreadsheet accompanying this chapter for more detail: Data for Oulton (2020).xlsx.

Krugman's conclusion might still be wrong if the benefits of growth since 1856 had simply accrued to the rich. A nodding acquaintance with the novels of Charles Dickens, plus a mass of more systematic evidence on the lives of the poor in Victorian times, would suggest otherwise. Or you could just reflect on the fact that the typical inhabitant of Victorian Britain was not able to use the National Health Service, had no access to TV or the Internet, and had little opportunity for foreign travel (except perhaps for transportation to Australia). Economic historians have dug deeper. The Gini coefficient is a standard measure of inequality which in percentage terms varies from 0 (everyone gets the same income to 100 (one person gets everything, the rest get nothing). Estimates of the Gini coefficient have been made by Milanovic et al. (2011) for England and Wales in 1290, 1688, 1759 and 1801. They found that in England and Wales the Gini was 36.7 in 1290, 45.0 in 1688, 45.9 in 1759 and 51.5 in 1801. So inequality was rising in pre-industrial times and in the early stages of the industrial revolution. In 1999 the Gini for UK household income was 37.4 so inequality is lower now than two centuries ago (Milanovic et al, 2011, Table 2).

Another inequality measure is the share of the richest 1% in total income. This was 19% in 1918 (the earliest year currently available). Thereafter the share of the top 1% fell before starting to rise again in the 1980s. In 2012 it was 13%, still lower than in 1918 (Atkinson et al. 2017). So clearly the benefits of growth have been widely shared.

So are there any other ways to raise living standards apart from productivity growth? There seem to be only two. First, we could redistribute income from the rich minority to the poor majority. But the scope for this is quite limited in comparison to productivity growth. For example, if we confiscated the whole of the income accruing to the banking industry — wages, bonuses, and profits — and redistributed it to the population at large (while somehow maintaining the socially useful functions of a modern banking industry), average incomes of non-bankers would be raised by less than 10 per cent. Reason: the share of value added in the banking industry (approximately, wages plus bonuses plus profits) in GDP at current prices is about 7% (see Table 3). Prior to the Great Recession which started in 2008 labour productivity was growing at about 2% per year. So expropriating the bankers would be the equivalent of four to five years of productivity growth, at least as we experienced it before 2008. You might still want to expropriate the bankers but don't expect to make the average person rich as a result.

A second way to raise living standards is through an improvement in the terms of trade. If we could somehow get foreign buyers to pay more for our exports or to supply our imports at lower prices, then our living standards would improve. But it seems very unlikely that a country like Britain could count on this except as a temporary piece of good luck. And of course any benefit to Britain is a loss to other countries.

You might agree that productivity growth has had a huge effect on living standards in the past. You might also think that the average living standard of today is quite enough to meet any reasonable material needs so future productivity growth is pointless. But remember that productivity growth also makes possible additional leisure while keeping the same level of consumption of goods and services. Also, even if you think your own income is likely to be more than enough for your needs, remember that there are a few billion others in the world who probably take a different view.

3. The measurement of labour productivity

The figures in Table 1 and Chart 2 summarise the efforts of generations of economists, economic historians and official statisticians. But what does "real GDP " actually mean and how is it calculated? This and the next section describe the methods currently in use at the ONS.

Labour productivity is defined as output per unit of labour. But how should we measure output and labour? Output in current prices is relatively simple in principle. At the whole economy level it is GDP. GDP in current prices can be thought of as the sum of value added in all the industries which make up the economy. So if we are interested in productivity at the industry level then industry value added is the natural measure here too. Measuring real output is a lot more challenging: see below on this.

Measuring labour input is a bit more complicated than you might think. Taken literally the Krugman quote suggests labour input should be measured by the number of workers. But probably this is just shorthand on his part. Suppose a given level of output can be produced by two sets of equally numerous workers but the second set of workers is working only half the hours of the first set. Then we would probably say that the second set of workers has double the productivity of the first set. In other words, hours worked is a better measure of labour input than number of workers. Even hours worked is not a completely unambiguous concept. We need to distinguish between hours worked and hours paid. The standard definition of hours worked, the preferred concept for productivity analysis, is hours spent actually at work on the farm, or in the office, factory or shop, though it does allow for short breaks (meals and toilet visits) to be included too. "Hours paid for" is a wider concept and includes also paid absences through sickness, holidays, maternity leave, and time spent attending courses. In practice as we shall see, countries are not altogether consistent in the methods that they use to measure hours worked and this can distort international comparisons of productivity levels.

Care needs to be taken too even in measuring the number of workers. This is more than just the number of employees but should include also the self-employed and (possibly unpaid) family workers, still important in some industries like agriculture or in the family-run part of the retail sector. For many developing economies a series for hours worked is not available or is of poor quality so productivity growth has to be measured by output per worker.

We are usually interested either in measuring the growth of labour productivity over time or in comparing the level of labour productivity in one country (e.g. Britain) with the level in another. Take comparisons over time first. At the whole economy level, we are interested in *real* GDP or GDP corrected for inflation. For comparisons across countries the main problem is that German GDP for example is measured in euros and British GDP in pounds sterling. So we have to develop a conversion factor, known as a Purchasing Power Parity (PPP), to convert the currency of any one country into that of another. This is in effect a cross-country price index. More on this below.

For some purposes it is more useful to look at a sub-set of the whole economy, the market or business sector. This is because in many countries the *real* output supplied by the government — services like health, education, welfare provision or defence — is not well measured, or not measured at all. In other words we know how much was spent in providing these services in 2018 or 2017 in dollars or euros but we don't always know by how much the *real* output of these services changed. For this reason, productivity analysts often look at the market sector which excludes activities typically though not exclusively provided by government, even though some industries like finance which are in the market sector also come under the hard-to-measure category.[5]

4. The measurement of real output over time

Real GDP from the expenditure side

Earlier we spoke rather vaguely about real GDP or GDP "corrected for inflation". At an intuitive level the need for correcting for inflation is obvious. Suppose all prices double but no quantities change. Then on any sensible definition real GDP is unchanged even though nominal GDP has doubled. But how do we correct for inflation when prices are changing at different rates? Broadly speaking GDP in current prices (nominal GDP) from the expenditure side is an aggregate defined as follows:

$$GDP_t = \sum_{i=1}^{N} p_{it} q_{it}$$

where p_{it} is the price and q_{it} is the quantity of the i-th type of final expenditure and there are *N* such types (including imports in which case the quantity is negative, i.e. these must be subtracted). We want to be able to think of GDP in current prices as the product of an average price and an average quantity:

$$GDP_t = P_t Q_t = \sum_{i=1}^{N} p_{it} q_{it}$$

Here P_t is a price index, some sort of average of the individual prices p_{it}, and Q_t is a quantity index, some sort of average of the individual quantities q_{it}. Suppose we find a way of defining the quantity index in operational terms, i.e. in a way which can actually be calculated in practice. Then we have also defined the price index since

$$P_t = GDP_t / Q_t$$

Such a price index is called an *implicit deflator*, in this case the *GDP deflator*.

[5] Here I am implicitly defining the market sector by the type of industry involved. So education is excluded even though some educational establishments (e.g. universities and private schools) are not in the public sector. The alternative definition is by ownership. For example, the ONS publishes a regular series called "business investment" on this basis. Movements in this type of series can then be influenced by privatisation or nationalisation. The ONS's estimates of MFP use the ownership concept of the market sector.

There is more than one way of defining the quantity index but the method used by the ONS, following the rules of the 2010 European System of Accounts (Eurostat 2013). is called *annual chain linking*. It is easiest to think of this as defining the growth of real GDP between an adjacent pair of years, say year *t* and year *t*-1. First of all the ONS revalues all the components of GDP in year *t* to the prices of year *t*-1. (In practice this is done by deflating the value of each type of expenditure, $p_{it}q_{it}$, by an appropriate, low-level price index, i.e. by dividing the value by $p_{it} / p_{i,t-1}$). Then the ONS expresses this new aggregate as a ratio to current price GDP in year *t*-1.

$$\frac{Q_t}{Q_{t-1}} = \frac{\sum_{i=1}^{N} p_{i,t-1}q_{it}}{\sum_{i=1}^{N} p_{i,t-1}q_{i,t-1}} = 1 + g_t$$

where g_t denotes the (discrete) annual growth rate, Notice that all the prices are those of year *t*-1 ("previous years' prices"). In the numerator of the middle term the quantities are those of year *t* while in the denominator they are those of the previous year *t*-1. So this term is a quantity index, showing the average quantity in year *t* relative to the average quantity in year *t*-1. Hence this index defines the growth rate of real GDP between years *t*-1 and *t* (the last term).

Under annual chain linking the weights are changed every year. This contrasts with ONS practice prior to 2004 when the weights were typically changed only every five years. In some countries, such as the United States and France, the prices of a single year were used as weights for the whole run of their official national accounts (back to 1929 in the US case). Annual chain linking is now considered best practice. The main reason is that otherwise the weights get increasingly out of date and unrepresentative of current patterns of expenditure.

The last formula can be made a bit more intuitive by some simple algebra[6] so that it becomes

$$1 + g_t = \frac{Q_t}{Q_{t-1}} = \sum_{i=1}^{N} s_{i,t-1}\left(\frac{q_{it}}{q_{i,t-1}}\right) \qquad (1)$$

where $s_{i,t-1}$ is the share of the i-th type of expenditure in the total value of final expenditure (current price GDP) in the previous year:

[6] Multiply and divide the numerator of the middle term by $q_{i,t-1}$ and rearrange.

$$s_{i,t-1} = \frac{p_{i,t-1}q_{i,t-1}}{GDP_{t-1}}$$

Now we see that the growth factor $(1+g_t)$ is a weighted average of the *quantity relatives* $(q_{it}/q_{i,t-1})$ where the weights are the expenditure shares in the previous year.

The method is called annual chain linking since the growth factor we have just defined, $1+g_t$, is a single link in the chain. To get an index number of real GDP for e.g. year *T* relative to year 1, first set the index equal to 1 in year 1. Then we have to multiply all the links in the chain together:

$$\frac{Q_T}{Q_1} = (1+g_2)\times(1+g_3)\times\ldots\times(1+g_{T-1})\times(1+g_T)$$

(To put the index in the form year 1=100, multiply it by 100.) Since the index is 1 in year 1, year 1 is called the reference year. To change this so that year *R* for example is the reference year, divide the last equation by Q_R/Q_1. This will make the index equal to 1 in year *R*. Finally, after making year *R* the reference year, we can convert the index into what the ONS calls a *chained volume measure (CVM) in reference year prices*, by multiplying the index for each year by current price GDP in year *R*; the CVM value in year *R* is then GDP_R. This last operation of course just affects the levels while leaving growth rates unchanged.

Because it uses the previous year's shares as the weights, the ONS's chained volume measure is called an annually chained Laspeyres index. A possible alternative is to use the prices of year *t* instead of those of year *t*-1 in the index formula. The result would then be called an annually chained *Paasche* index. Because there seems no obvious reason why either one of the two possible years should be privileged over the other, a further step would be to take the geometric mean of the Laspeyres and the Paasche to get what is called the Fisher index.[7] The United States now uses the Fisher index in its national accounts.

With the ONS's and Eurostat's choice of a chained Laspeyres index, it can be shown that the GDP deflator is an annually chained Paasche index, i.e. the prices of the two years are weighted by the quantities of the current year, not those of the previous year. This contrasts with price indices like the CPI which is (skating over some complications) an annually

[7] Étienne Laspeyres (1834-1913), Hermann Paasche (1851-1925) and Irving Fisher (1867-1947) were all pioneers of index number theory. See Box 6.2 for a catalogue of formulas for index numbers in common use.

chained Laspeyres. With the US choice of the Fisher quantity index their GDP deflator is now an annually chained Fisher too (see Box 6.2).

Real GDP from the output side

We have just outlined the ONS method for deriving real GDP on the expenditure side. What about real GDP from the output side? We know that we can write current price GDP as the sum of current price value added in each industry:

$$GDP_t = \sum_{i=1}^{N} P_{it}^V V_{it}$$

where V_{it} is real value added in the i-th industry and P_{it}^V is the corresponding price, both at time *t*. In the first instance what we observe is just nominal value added, $P_{it}^V V_{it}$, in each industry but conceptually at least we can split this up into a price and a quantity. This suggests that, by analogy with (1) and using annual chain linking, the growth rate of real GDP should be a weighted average of the growth rates of real value added in each industry:

$$\frac{V_t}{V_{t-1}} = \sum_{i=1}^{N} v_{i,t-1} \left(\frac{V_{it}}{V_{i,t-1}} \right) \qquad (2)$$

Here V_t is aggregate real value added in year *t*, i.e. real GDP from the output side, and $v_{i,t-1}$ is the share of the i-th industry's nominal value added in aggregate nominal value added in the previous year (GDP_{t-1}):

$$v_{i,t-1} \equiv \frac{P_{i,t-1}^V V_{i,t-1}}{GDP_{t-1}} \qquad i = 1, \ldots, N$$

This assumes that we can give a coherent meaning to the price and quantity of value added and furthermore that the growth of GDP as measured by (2) will be the same as when measured by (1), at least in principle and abstracting from any errors and omissions in the data.[8]

There are two methods employed to derive real value added: single deflation and double deflation. The ONS currently uses a form of single deflation under which the growth of real

[8] When the ONS measures GDP from the expenditure side they use market prices but for the output side they use basic prices. The difference between the two is taxes less subsidies on products, e.g. excise duties and non-refundable VAT, and trade and transportation margins. The equality between GDP from the expenditure and output sides holds in principle when both measures have been adjusted to a common price basis.

value added in each industry is assumed to be the same as the growth of real output in that industry. Real output is nominal output deflated by whatever is considered the most appropriate price index.[9] So at the moment equation (2) is replaced in practice by

$$\frac{V_t}{V_{t-1}} = \sum_{i=1}^{N} v_{i,t-1} \left(\frac{Y_{it}}{Y_{i,t-1}} \right)$$

where Y_{it} is the real output of the i-th industry. One problem with this approach is that the growth of real GDP from the expenditure side is *not* now guaranteed to equal its growth from the output side, even in principle. Hence at the moment the growth rates of some private services industries such as finance are adjusted to force the output side measure to agree with the expenditure side one, to within a narrow tolerance (Lee 2011).

The alternative method, double deflation, requires that output and each of the inputs be deflated by its own price index. Real value added then emerges as the residual. Nominal output in any industry is given by the following accounting relationship:

Nominal output = Nominal value added

plus nominal value of inputs purchased from domestic suppliers

plus nominal value of inputs purchased from foreign suppliers

or in symbols

$$P_{it}Y_{it} = P_{it}^{V} V_{it} + \sum_{j=1}^{N} \tilde{P}_{jt} X_{ijt}^{D} + \sum_{j=1}^{N} \tilde{P}_{jt}^{M} X_{ijt}^{M} \qquad i = 1, \ldots, N \tag{3}$$

Here P_{it} is the basic price of industry i's output, $\tilde{P}_{jt}$ is the purchasers' price of the j-th input when purchased from a domestic supplier, $\tilde{P}_{jt}^{M}$ is the purchasers' price when purchased from a foreign supplier, X_{ijt}^{D} is the quantity of the j-th input purchased domestically, and X_{ijt}^{M} is the quantity purchased from foreign suppliers. Under the annual chain linking approach, each of the elements of this accounting relationship has to be revalued to previous years' prices, a formidable undertaking. More specifically, after revaluing output and all the inputs in the last equation to previous year's prices we can solve for value added at previous years' prices in the i-th industry:

$$P_{i,t-1}^{V} V_{it} = P_{i,t-1} Y_{it} - \sum_{j=1}^{N} \tilde{P}_{j,t-1} X_{ijt}^{D} \sum_{j=1}^{N} \tilde{P}_{j,t-1}^{M} X_{ijt}^{M}$$

Then we can form the ratio

[9] The ONS collects monthly price quotations for a large number of products under its Producer Price Indices (PPIs), Services Producer Price Indices (SPPIs), and Export Price Indices (EPIs) programmes from which it constructs price indices for individual industries.

$$\frac{P_{i,t-1}^V V_{it}}{P_{i,t-1}^V V_{i,t-1}} = \frac{V_{it}}{V_{i,t-1}}$$

for each industry. These ratios are just what we need to calculate double-deflated GDP from the output side using equation (2).

Double deflation is considered international best practice (European Commission et. 2009; Eurostat 2013) and has been adopted so far by about half of the leading G20 economies. One reason is that in principle under double deflation real GDP from the expenditure side equals real GDP from the output side. The ONS has made a start on introducing double deflation into the national accounts in the 2019 Blue Book though for the moment industry-level estimates of real value added are still on a single deflation basis. See Oulton et al. (2018) for more on the theoretical and practical issues involved in double deflation.

5. Labour productivity at the industry level and the regional level

Within industry productivity growth versus structural change

When confronted by statistics on the growth of labour productivity in the economy overall, people often want to look at what is happening in individual industries or sectors. If aggregate growth seems to be slowing, is the same true in manufacturing? Or are the results being overly influenced by what is happening in banking? Maybe growth is slow because growth in services is slow, but perhaps this is mis-measured, so overall growth is not really slow at all? So there is a lot of interest in breaking down aggregate growth into the contributions of individual industries.

Obviously, the growth rate of labour productivity at the aggregate level must be related to the growth rate of productivity in individual industries. But the relationship is not a simple one. Intuitively, it is clear that overall growth must be due to the balance of two forces: first, growth within individual industries and second, reallocation effects. Reallocation effects are of two types. The first type of effect, in this case favourable, is when labour is shifting from lower productivity industries to industries which have an initially higher-than-average *level* of labour productivity. The second type of favourable shift is when labour is shifting towards industries which have a higher-than-average *growth rate* of labour productivity. Both these types of reallocation may be unfavourable as well as favourable. And they may pull in opposite direction: an industry with a high level of labour productivity may also happen to have a low growth rate of productivity.

Box 6.3 develops formulae relating aggregate growth to growth in individual industries. In the general case aggregate growth depends on the individual growth rates and two sets of shares: the value added shares and the labour (hours) shares. The weight for each industry is a combination of its initial value added share and its labour shares in the two periods. Box 6.3 also derives formulas for two hypothetical special cases: first, when the initial level of productivity is the same in all industries and second, when the labour shares do not change over the period being analysed. Comparing the aggregate productivity growth rate in the general case and in the two special cases enables us to gauge the importance of structural change to overall growth.

The formulae derived in Box 6.3 are as follows.

General case: $$\frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} \left(\frac{v_{it-1}}{w_{it-1}} \right) w_{it} \left(\frac{Z_{it}}{Z_{i,t-1}} \right) \quad (4)$$

Constant labour shares $\left(w_{it} = w_{i,t-1}, \text{ all } t \right)$: $$\frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} v_{it-1} \left(\frac{Z_{it}}{Z_{i,t-1}} \right) \quad (5)$$

Equal initial producivity levels $\left(v_{i,t-1} = w_{i,t-1}, \text{ all } t \right)$: $$\frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} w_{it} \left(\frac{Z_{it}}{Z_{i,t-1}} \right) \quad (6)$$

Here v_{it} is the value added share, w_{it} is the labour share, and Z_{it} is the productivity level in the i-th industry. In the constant labour shares case aggregate growth is a weighted average of industry growth rates where the weights are the value added shares. In the equal initial productivity level case the weights are the labour shares.

Labour productivity growth before and after the Great Recession

We now put this apparatus to work by looking at the growth of labour productivity before and after the Great Recession. We know that overall growth declined sharply, in fact to near zero.

But did this result from a general fall found in all industries or was it concentrated in particular industries? How important a role did structural change play?

Table 3 shows the structure of the British economy in 2018 in terms of sections of the current Standard Industrial Classification (SIC 2007). The public services — Public administration and defence (O), Education (P) and Health (Q) — are amalgamated into one group here as are so-called “Other services” (S,T and U). The table shows for each of the 18 industry groups its importance in terms of its share of total value added in the prices of 2018, i.e. its share of GDP in 2018, and its share of total hours worked. Some of the numbers may surprise you. Manufacturing accounts for under 9% of GDP and 8% of hours, less than Wholesale and Retail. Finance and Insurance of which banking is a part accounts for less than 7% of GDP and less than 4% of hours worked.

The last column shows each group’s labour productivity relative to the average in 2018. Relative productivity (the value added share divided by the hours share) varies widely though there is usually a straightforward explanation. Mining and Quarrying which these days means mainly oil and gas production from the North Sea has more than five times the average level of labour productivity since it is very intensive in physical capital; the same is true of Energy. Agriculture and Construction are intensive in unskilled labour. Information and Communication and Finance and Insurance are intensive in human capital. The biggest number in the last column is for Real Estate with over 8 times the average productivity level. This is due to an anomaly. The bulk of value added here comes from the *imputed rent of owner-occupiers*.[10] But this does not give rise to any employment or hours since time spent by owner-occupiers in maintaining or repairing their homes is outside the production boundary of the national accounts.

Table 3 near here

Using the data underlying the ONS’s *Labour Productivity Bulletin* (ONS 2019d), Chart 4 shows labour productivity in the market sector since 1997Q1. The market sector is defined as the whole economy less the public sector (sections O, P and Q) and less Real Estate (section

[10] In the national accounts owner-occupiers are thought of as paying rent to themselves, the payment representing the value of the services they enjoy from living in their properties. This puts them on all fours with non-owners who pay actual rent to private or public landlords. These commercial rents are also included in GDP. If imputed rent were not counted as part of GDP a shift away from owner-occupation to renting would increase GDP.

L); i.e. in 2018 31% of value added and 24% of total hours is excluded. The grey bar in the chart marks the Great Recession, the period when real GDP was actually falling (2008Q2 to 2009Q2). The red line shows the actual course of labour productivity, corresponding to equation (4). The green line shows the result of weighting individual sections by value added (equation (5)) and the blue line the result of weighting by hours worked (equation (6)).

Focusing first on the red line, we see that labour productivity grew by 35% from 1997Q4 to the end of the boom in 2007Q4. Since the Y-axis is on a log scale we can see immediately that growth after 2007 was much slower than before. In fact labour productivity fell during the recession. It then started to recover very slowly. But it was only slightly higher than its pre-recession level by the end of 2018, 11 years later. This lengthy period of stagnation is the productivity puzzle. How unusual, indeed unprecedented, this is can be seen by comparison with the whole record since 1856 (Chart 2 and Table 2).

Structural change played only a small role in the period up to the end of 2007: the red, blue and green lines are very close. But since 2011 the lines have diverged, with the red line below the others. In other words structural change has been unfavourable to productivity growth as labour has been reallocated to industries with either a low level or a low growth rate of productivity. But even if all structural change had been avoided the growth rate over this period would still have been low by comparison with the pre-recession period.

Chart 4 near here

Another interesting question is whether any industries contributed disproportionately to the slowdown in productivity growth. An industry's contribution to the slowdown can be defined as its contribution to growth over the period of the boom (1997Q4-2007Q4) minus its contribution to growth over the subsequent period (2007Q4 to 2018Q4). Up to now we have used the chained Laspeyres approach to measuring output and labour productivity since this is consistent with the national accounts. But now it is convenient to use Törnqvist indices which are also used by the ONS in its multi-factor productivity estimates (see below). The reason is that the Törnqvist allows a much simpler decomposition of the contributions of individual industries.

In continuous time, aggregate labour productivity growth can be related to labour productivity in individual industries as follows:

$$\hat{Z} = \hat{V} - \hat{L}$$
$$= \sum_{i=1}^{N} v_i \hat{Z}_i + \sum_{i=1}^{N} (v_i - w_i)(\hat{L}_i - \hat{L}) \quad (7)$$

(See Box 6.3 for the derivation). The first term on the right hand side of (7) is the so-called "within" effect, coming from growth in individual industries. The second term is the "between" or reallocation effect coming from structural change. This second effect can only be non-zero when the growth of hours in an industry differs from the overall average growth of hours. Equation (7) gives us an additive decomposition, showing how the contributions of each industry add up to the growth in aggregate labour productivity. This continuous time formula can be approximated by the Törnqvist index (see Box 6.2) which replaces continuous time growth rates by log differences and point-in-time shares by averages of the shares over adjacent periods:

$$\Delta \log Z_t = \sum_{i=1}^{N} \bar{v}_{it} \Delta \log Z_{it} + \sum_{i=1}^{N} (\bar{v}_{it} - \bar{w}_{it})(\Delta \log L_{it} - \Delta \log L_t) \quad (8)$$

where the average shares of value added and hours are given by

$$\bar{v}_{it} \equiv \left(\frac{v_{it} + v_{i,t-1}}{2} \right)$$

and

$$\bar{w}_{it} \equiv \left(\frac{w_{it} + w_{i,t-1}}{2} \right)$$

The within and between effects are shown for the whole market sector in Table 4, based on equation (8). This table uses exactly the same data as in Chart 4. Reallocation (the between effect) is negative but accounts for very little of the story, either during the boom or after it. It is productivity growth in individual industries which accounts for nearly all productivity growth in the market sector, both during the boom and after it.

Table 4 near here

Table 5 shows the contributions of each section to growth during and after the boom and also to the slowdown after the end of the boom. Three sections — Manufacturing, Information and Communication, and Finance and Insurance — account for over two thirds of the slowdown, despite these sections accounting for only about a quarter of value added (Table 2). But two caveats should be noted. First, the picture might change if we disaggregated further by breaking the sections down into industries. Then we might find that reallocation played a bigger role in the slowdown. Second, the output measure here is single-deflated real value added. When double deflation is implemented in the national accounts the output measure will change. So the picture portrayed in Tables 4 and 5 may change too.

Table 5 near here

Interregional differences in labour productivity

There is considerable political and policy interest today in interregional differences in labour productivity and living standards. In principle the methods described above for studying labour productivity by industry can also be used for a breakdown by region. But there are also some issues specific to regional analysis as we shall see.

The ONS publishes estimates of labour productivity levels for various regional breakdowns: see ONS (2019f) for recent data and a discussion of the methodology. The regional level is the so-called "NUTS1"[11] geography which in the UK comprises the nine English regions and the three countries of Wales, Scotland and Northern Ireland. Below NUTS1 there are 41 NUTS2 subregions and below that 179 NUTS3 local areas.

The ONS recommends using GVA per hour worked rather than GVA per head as the productivity measure. As Prothero (2018) argues: "Firstly, people do not always live and work in the same place. Commuting can distort the picture if we measure economic output (GVA) per resident. Take the extreme example of the City of London with its resident population of less than 10,000 but where commuting means hundreds of thousands come to work every day. The resulting output per resident total for the City of London is huge, but essentially meaningless. It provides neither an accurate measure for the economic productivity of the area or for the household incomes of its residents."

At the national level three ways of measuring nominal GDP are available: the expenditure approach, the income approach and the output (or production) approach. At the national level these three approaches are harmonised through the supply-use balancing process. But the ONS does not have regional estimates of consumption, investment and the other components of the expenditure approach so at the regional level only the income and the output

[11] "NUTS" stands for (in French) "Nomenclature of Territorial Units for Statistics" . It is the geographical classification system employed across the EU.

approaches are available. These two approaches are used to produce a harmonised estimate of current price GDP by region. These estimates can be compared across regions at a point in time to measure relative productivity. Differences between regions in relative productivity depend in part on differences in industrial composition and we have already seen that labour productivity differs considerably across industries (Table 3). So if a region has low relative labour productivity it does not necessarily mean that its workers are less skilled or its firms less efficient though it might.

Chart 5 shows GVA per head for the UK's twelve regions and countries in 2016. The most striking feature is the huge gap in productivity between London and the rest of Britain which dwarfs differences between the other regions and countries. London's labour productivity level is 33% higher than the UK average. Within London and the South East there are subregions with even higher relative labour productivity (ONS 2019g).

ONS research suggests that industrial structure is not an important explanation of these disparities. Analysis of firm level data suggests that, within each industry, firms in the favoured regions have higher labour productivity than those in less favoured regions (ONS 2019g). Another reason for thinking that industrial structure is not the most important factor comes from comparing labour productivity with a measure of the standard of living, Household Income (Gross Domestic Household Income per head). Chart 5 shows that these two measures can diverge quite a lot. For example the South West has low productivity but close to an average standard of living. A likely explanation is that the workforce is concentrated in low productivity industries like agriculture and tourism while the population contains some prosperous retirees. But on the other hand for London the gap in labour productivity with the rest of the country is quite similar to the gap in living standards. Since wages are the most important source of household income this suggests that wages and therefore skills are higher in London.

Chart 5 near here

It is now possible to study also the growth rate of labour productivity by region. The ONS now publishes chained volume measure of labour productivity (GVA per hour worked) by region covering the period 2002 to 2017. In a given region these are derived by deflating current price value added in each industry by the appropriate price index. The growth rates of each industry's real GVA are weighted together using the share of each industry in total regional value added. So at the moment single rather than double deflation is being used as in the national level estimates. A second issue is that the price indices are national ones as separate regional price indices for each industry do not exist; in other words it is assumed implicitly that either the product composition of each industry is the same across regions or that if it differs the prices of all products within an industry move identically.

According to Table 1 of the February 2019 release (ONS 2019f), the total (not annual) growth of labour productivity from 2010 to 2017 across the 12 NUTS1 regions varied between zero and 5% with a UK average of 2%. In other words all regions shared in the productivity slowdown. So at this level of disaggregation regional analysis does not yield any further insights into the productivity puzzle.

6. International comparisons of output, living standards and productivity

Purchasing power parities (PPPs)

Given that living standards depend on labour productivity there is keen interest in comparing the labour productivity levels of different countries. To do this we need a way of converting the currencies of different countries into a common monetary unit. The obvious way of doing this is to use exchange rates. But it is generally agreed that this is not satisfactory for several reasons. First, even if the countries to be compared practised perfectly free trade amongst themselves (which of course they don't), the exchange rate would be a satisfactory converter only for traded goods and services. It would not necessarily give a good guide to the relative prices of the large number of goods and services which are not traded internationally. Second, exchange rates frequently make sizeable jumps in response to capital movements and these movements have little or nothing to do with current productivity and living standards. Because of this international comparisons are best done using what are called Purchasing Power Parities or PPPs.

National statistical agencies (NSAs) such as the ONS have long been engaged in periodic, large scale comparisons of prices in their respective countries. Since the 1960s these efforts have been coordinated by the World Bank in what is known as the International Comparison

Program (ICP). Successive rounds of the ICP haven taken place in 1970, 1975, 1980, 1985, 1996, 2005 and (the latest to be published) 2011. The number of countries participating has risen over time and the ICP now covers nearly the whole planet. In the 2011 round 199 countries took part, constituting 97% of the world population (a notable absentee was North Korea). Their GDP comprised 99% of the global total. The OECD and Eurostat carry out more frequent comparisons for their member countries plus some others such as Russia and these results feed into the ICP. The OECD-Eurostat group comprises 47 countries of which 10 are outside Europe (World Bank Group 2013).

The ICP breaks down GDP into expenditure categories called Basic Headings. In the 2011 round there were 155 Basic Headings though prices were collected only for 113 of these. The procedure for gathering prices is similar in principle to price collection procedures for a national Consumer Prices Index (CPI), except that the latter is a time series operation while the ICP is a cross section one. In a domestic CPI the price collectors are tracking the prices of *identical* items over time. For example, within the product category “Eggs” the ONS tracks the prices of various examples (large brown , medium white, etc) every month. It is important that in each case the *same* product is being tracked in the *same* shop, also that the shops sampled should be representative, e.g. not just large supermarkets in London. For each product the price *relative* is calculated: the price this month divided by the price last month. Then an unweighted average of these price relatives is taken; unweighted, because the ONS normally has no data on the amount spent on the individual products that it tracks, but only knows total expenditure on the item such as eggs. (Nowadays in the CPI the unweighted average is usually the geometric mean). Similarly in the ICP the price collectors in each country are trying to gather prices for products within a given Basic Heading which are identical in all relevant respects to the products being priced in every other country. In practice, this aim cannot be achieved completely since not every product is sold in every country. So much of the ICP’s work is concerned with filling in the missing prices by various statistical procedures.

An example of a Basic Heading is “Rice”. There is no such thing as *the* price of a Basic Heading, even an apparently homogeneous one like “Rice”. Rather there are prices for products which fall under the definition of the Basic Heading, just as in the domestic CPI. There is usually no information on expenditure below the Basic Heading level, so the “price” of the Basic Heading is an unweighted average of the prices of the products classified to that

Basic Heading; more precisely it is the prices in each country relative to the corresponding prices in the numeraire country, always the United States, which are averaged. To identify products suitable for pricing, the World Bank makes use of what they called "Specific product descriptions" (SPD): a description of a product which falls under a particular basic heading and for which a price could in principle be collected. A fictional example might be "Basmati rice, 500 gram bag", fictional since the actual products that are priced are not published. Several or even many SPDs may fall under any Basic Heading.[12] A product suitable for pricing is then one which falls under the SPD for a Basic Heading. A fictional example might be "Waitrose own brand Basmati rice, 500 gram bag, purchased in a Waitrose supermarket in Edinburgh". In practice prices are collected either monthly or quarterly and then averaged over the year.

PPPs, both the published, high level ones and the unpublished, Basic Heading level ones, are expressed as local currency units per US dollar, which serves as the numeraire currency. PPPs can be thought of in two ways. First, they are like exchange rates, indeed they are exchange rates for specific products or groups of products. But second, they can be thought of as prices. The corresponding quantity unit for any Basic Heading is the quantity which could have been purchased in the United States in the comparison year for one US dollar.

Once an average price has been developed for each Basic Heading in each country included in the comparison, the next step is to calculate PPPs for a high level aggregate such as household consumption or GDP. Initially, bilateral indices are calculated between each pair of countries. These indices are symmetrical, they give equal weight to the expenditure patterns of the two countries being compared; specifically, a Fisher index is used. After all, if the aim is to compare British and French prices you might think it is the expenditure patterns of these two countries which matter and not the expenditure pattern of some third country, say Albania, also a member of the OECD-Eurostat group. Albania is a poor country and its expenditure patterns are very different from those of Britain and France. But this requirement to focus on the most relevant expenditure patterns, known by the ugly name of

---

[12] In fact, the Specific Product Description for Rice in the 2005 round allowed for five types (long grain, medium grain, …), five varieties (white, brown, …), two types of preparation (pre-cooked or uncooked) and whether or not the product is organically certified, yielding a potential total of 100 products, with the possibility of individual countries adding to the list if other characteristics are regionally important (World Bank, 2005, chapter 1). Of course, not all countries will have been able to provide prices for all these 100+ products.

characteristicity, is in conflict with another requirement, transitivity. If country A's GDP is 20% larger than country B's, and country B's is 10% larger than country C's (in terms of log differences), then surely country A's is (20 + 10 =) 30% larger than country C's? Actually, not necessarily, if the comparison is made using bilateral indices. The bilateral PPP between A and B uses the expenditure patterns of countries A and B while that between B and C uses those of B and C. So there is no guarantee that transitivity will hold. Hence in the ICP a further step to enforce transitivity is applied, so that each published bilateral index is an average of the direct bilateral index and all possible indirect comparisons. So the published bilateral PPP between A and B is an average of the direct bilateral PPP between those two countries and all possible indirect comparisons, e.g. A with C and C with B (also A with D and D with B, etc): see Box 6.2 for details. In other words, transitivity is enforced by compromising on characteristicity. Put another way, the published bilateral PPP between Britain and France is influenced by the expenditure patterns of Albania, as well as by the expenditure patterns of all other countries in the OECD-Eurostat group.

Comparisons of size and living standards in the 2011 ICP

Table 6 shows the 30 largest economies in the world in 2011, as measured by GDP at Purchasing Power Parity. The United States was the largest, followed by China. Since China has continued to grow much more rapidly than the United States since 2011 it has very likely now overtaken the United States to become the largest economy. India and Japan were the third and fourth largest in 2011, but a long way behind the two leaders. Britain was the 9th largest economy in 2011, closely behind France. A general finding in this round of the ICP as in earlier ones is that the poorer the country, the larger is its GDP at PPP relative to its GDP measured using exchange rates. For example at PPP China's GDP was 87% of the US level in 2011 but only 47% using the exchange rate. The usual explanation is that the relative prices of labour-intensive services which do not enter into international trade (think haircuts) are lower in poor countries. So the exchange rate overstates the price level of poor countries and hence understates their real GDP.

If we look at living standards as measured by GDP per head at PPP, a quite different picture emerges. The leaders on this measure do not appear in the table at all since they are all very small economies with special features. They include oil states like Kuwait and Qatar or states like Luxembourg which serve at least in part as tax havens where multinationals like to book their profits. So the United States ranks only 12th on this measure while China is 100th, at only a fifth of the US level. Britain with 70% of the US level is 32nd, behind France again but ahead of Japan, The World Bank considers that what it calls "actual individual consumption" (AIC) per head is a better measure of living standards than GDP per head. Actual individual consumption consists of household consumption and that part of government expenditure which benefits identifiable individuals directly — most of health and education expenditure but not defence. On this measure the US now comes 2nd. Bermuda (not shown), an offshore insurance centre with some 71,000 inhabitants, comes top. Britain, at 70% of the US level, comes 18th, a little behind France but ahead of Japan once again.

Table 6 near here

International comparisons of labour productivity levels

The ICP considers GDP from the expenditure side so it is impossible to compare labour productivity at the industry level. The reason is that much of value added is generated in industries supplying mainly intermediate goods and services: examples are business services and finance. Because their products are intermediates their prices are not collected by the ICP. But cross-country comparisons of labour productivity at the whole economy level are still possible and a recent one carried out by the ONS for the G7 countries is illuminating.

Chart 6, taken from ONS (2018a), compares GDP (at market prices) per hour worked in the G7 group of leading industrial economies in two recent years, 2015 and 2016. Germany has the highest labour productivity, some 35% higher than in Britain. Britain is just ahead of Canada and around 7% ahead of Japan. The productivity gap between Britain and most other leading economies is of long standing. It widened after 2007 since growth in Britain has been slower than in these other countries.

Chart 6 near here

Chart 6 is based on each country's national accounts and labour market statistics, together with PPPs which are estimated on a collaborative basis as just described. Nevertheless it turns out that the underlying figures for labour input may not be wholly comparable. Labour input, which is supposed to be hours actually worked by all types of workers (full and part time employees, the self-employed, family workers), may in practice be measured differently in different countries (OECD 2018 and ONS 2019a). The OECD has found that if hours worked were calculated in a different but more comparable way across countries, then Britain's productivity gap with the United States would be reduced from 24% to 16%.

7. Multi-factor productivity: an overview[13]

Labour productivity will always be of interest because of its close connection to living standards. But why should we always focus on just labour as the sole input of interest? Sometimes in fact we focus on other inputs. For example agricultural economists are interested in yields (output per hectare). If we think more generally of productivity as measuring efficiency then we should be interested in all inputs: we should measure output per unit of all inputs at once. So we need to develop an index of total input. The natural way to do this if each input is weighted by its importance, that is, by how much is spent on it.

Let's pretend for a moment that there are only two inputs, labour and capital. Labour (*L*) is a flow of hours worked and capital (*K*) is a flow of services from some machine, also measured in hours. Labour is hired for an hourly wage (*W*) and the machine is also hired for an hourly rental (*R*). It is convenient now to measure growth in continuous terms though recognising that when we come to actually measure growth, it will have to be over a discrete interval such as a year or a quarter. Then the growth of a cost-weighted index of total inputs (*X*) is:

Growth of total input = (share of labour in total costs x growth of labour input)
plus (share of capital in total costs x growth of capital input)

or in symbols

$$\hat{X} = s_L \hat{L} + s_K \hat{K}$$

where s_L is the labour share (the wage bill as a proportion of total costs):

$$s_L \equiv \left(\frac{WL}{WL + RK} \right)$$

and s_K is the capital share

$$s_K \equiv \left(\frac{RK}{WL + RK} \right)$$

---

[13] For an in depth exposition of productivity measurement as applied in practice by the ONS see OECD (2001) and (2009).

and of course these shares sum to 1. Here the "hat" (^) notation is used to denote a continuous growth rate (see Box 6.1).

A measure of efficiency is output (*V*) per unit of input or *V*/*X*. (I represent output by the symbol *V* rather than the more usual *Y* to fit better with the earlier discussion of measuring real GDP). This is in fact the definition of multi-factor productivity or MFP, also known as total factor productivity (TFP).[14] Consider the growth of MFP over time. At a point in time (*t*) the growth of MFP (*MFPG*) is just the growth of output minus the growth of input:

Growth of MFP = growth of output *minus* growth of input

Or in symbols

$$\begin{aligned} MFPG(t) &= \frac{d\log MFP(t)}{dt} = \frac{d\log V(t)}{dt} - \frac{d\log X(t)}{dt} \\ &= \hat{V}(t) - \hat{X}(t) \\ &= \hat{V}(t) - s_K(t)\hat{K}(t) - s_L(t)\hat{L}(t) \end{aligned} \tag{9}$$

Note that the shares and the growth rates are written as functions of time *t* to emphasise that they need not be constant. Another way to interpret MFPG comes from rearranging the right hand side of (9):

$$MFPG(t) = s_K(t)\left[\hat{V}(t) - \hat{K}(t)\right] + s_L(t)\left[\hat{V}(t) - \hat{L}(t)\right] \tag{10}$$

Now we see that the growth of MFP is a share-weighted average of the growth rates of capital productivity (V/K) and of labour productivity (V/L).

Equation (9) is the foundation of *growth accounting* (Solow 1957). The left hand side of (9) can be calculated as a residual since the terms on the right hand side are known, at least in principle. Then the equation can be rearranged to place the growth of real value added on the left hand side:

$$\hat{V}(t) = s_K(t)\hat{K}(t) + s_L(t)\hat{L}(t) + MFPG(t) \tag{11}$$

A traditional approach is now to ask: what proportion of the growth of output over a given period is due to the growth of inputs and what proportion to the growth of MFP?

A deeper understanding of MFP, and the reason why economists, NSAs and policy makers are interested in it, comes from the Solow growth model (Solow (1956)). Suppose that aggregate output (real GDP) is a function of aggregate capital and aggregate labour:

[14] MFP and TFP are synonyms. NSAs tend to refer to MFP, economists to TFP. The American spelling is "multifactor". See Hulten (2001) for a history of the MFP or TFP concept.

$$V(t) = A(t)f(K(t), L(t)) \qquad A > 0 \tag{12}$$

Following Solow, we can interpret the factor A as the level of technology. To convert the aggregate production function (12) to growth rates, differentiate it logarithmically with respect to time to obtain (see Box 6.4 for the details):

$$\hat{V}(t) = \hat{A}(t) + \varepsilon_K(t)\hat{K}(t) + \varepsilon_L(t)\hat{L}(t) \tag{13}$$

Here ε_L is defined as the elasticity of output with respect to labour:

$$\varepsilon_L \equiv \frac{\partial V}{\partial L}\frac{L}{V} \tag{14}$$

and ε_K as the elasticity of output with respect to capital:

$$\varepsilon_K \equiv \frac{\partial V}{\partial K}\frac{K}{V} \tag{15}$$

So solving for the term in technical progress:

$$\hat{A}(t) = \hat{V}(t) - \varepsilon_K(t)\hat{K}(t) - \varepsilon_L(t)\hat{L}(t) \tag{16}$$

Equation (16) is starting to look rather like equation (9), our purely empirical, index number measure of efficiency which we have labelled MFP. In fact under the assumptions of the Solow model, the two equations are identical. This is because Solow assumed constant returns to scale and perfect competition, hence each input is paid the value of its marginal product. So the elasticities of capital and labour are equal to the input shares. To see this consider the definition of the labour elasticity in (14):

$$\varepsilon_L \equiv \frac{\partial V}{\partial L}\frac{L}{V} = \frac{WL}{P^V V} = s_L$$

because the real wage (W / P^V) equals the marginal product of labour ($\partial V / \partial L$). A parallel argument shows that

$$\varepsilon_K \equiv \frac{\partial V}{\partial K}\frac{K}{V} = \frac{RK}{P^V V} = s_K$$

Strictly speaking, we defined the shares as shares in total costs ($WL + RK$) not in total revenue ($P^V V$) but under the assumption of constant returns total costs are equal to total revenue (see Box 6.4 again).

The importance of MFP is demonstrated by the main result of the Solow model. If A is growing at the constant rate μ and if the supply of hours worked is growing at the constant rate n, then in the long run steady state labour productivity $Z\,(= V / L)$ grows at the rate:

$$\hat{Z}^* = \hat{V}^* - n = \frac{\mu}{1-\varepsilon_L^*} \tag{17}$$

as shown in macro textbooks. Here a star (*) indicates the steady state. Since Solow's model also predicts that in the steady state the labour and capital elasticities are constant, the labour elasticity gets a star too. So the long growth rate of labour productivity is determined entirely by technical progress: no technical progress, no growth. What about capital? The model predicts that capital per hour worked, often called capital deepening, grows at the same rate as labour productivity and capital deepening also raises labour productivity (and real wages). But in the long run all capital accumulation is induced by technical progress; so no technical progress, no capital deepening either. Things might be different in the short run, for example after a devastating war. Then capital accumulation can play a bigger role, with technical progress in the background. But once the economy has recovered from war-time damage, technical progress comes to the forefront again. The fundamental reason for the primacy of technical progress is that capital is assumed to be subject to diminishing returns, so that capital deepening in the absence of technical progress would eventually run out of steam .

To appreciate the power of the Solow model go back to equation (9). On the left hand side we have MFPG, now interpreted as the rate of technical progress, a mysterious and hard to measure phenomenon one might think. But on the right hand side we have straightforward economic quantities: GDP, capital and labour together with the labour and capital shares. So what seemed mysterious has become measurable.

Having said that, a few qualifications are in order. The true causes of MFP growth have been much debated. Solow emphasized technical and scientific progress, including improvements in management techniques, everything in fact that comes under the heading "useful knowledge". But as he also recognized, once we step outside the bounds of his model, MFP growth may arise from any or all of the following:

1. From economies of scale.
2. From learning effects, either learning by doing or learning from others.
3. By reallocation of inputs towards more (or less) productive uses, either at the firm or the industry level. Reallocation is discussed below.
4. From external effects. An example of a favourable effect is when the government improves the road system at no charge to users. Then vehicles and their drivers can deliver more tonne-kilometers than before in the same time. An example of an unfavourable effect might be

when climate change makes it necessary to install air conditioning to make working conditions tolerable.

5. As an artefact of measurement error. Examples are when increases in the quality of human or physical capital are wrongly ignored or output is mis-measured (as may be the case in finance). Or some types of asset (such as intangibles, see below) are wrongly omitted. This would lead to mis-measurement of the quantity of capital and so of MFP.

Sometimes MFP growth is described just as a measure of innovation. But here we must be careful. Much innovation is embodied in new or improved types of capital. But nowadays improved quality in capital goods is meant to be reflected in the price indices for these goods. The most conspicuous example of this is the computer price index (see below). So we would not expect a firm or industry which invests heavily in computers to see particularly rapid growth in MFP, at least not on this account. However the improvement of price indices to take better account of quality change is an ongoing process whose end is not yet in sight. So it is quite possible that the growth of capital services is still understated and so the growth of MFP growth overstated.

Interpreting MFP growth in terms of prices rather than quantities

There is an alternative way of measuring the growth of MFP that uses prices rather than quantities. Provided that the accounting system is consistent the answers should be the same. But looking at MFP growth through the lens of prices can add further insight. And sometimes data on prices is better than data on quantities, particularly for the pre-modern period. The prices approach starts with the accounting identity for the one-good, two-input model which states that the value of output equals expenditure on the inputs:

$$P^V V = WL + RK$$

Now take logs and totally differentiate this equation with respect to time, then rearrange to get terms in prices on the left and terms in quantities on the right:

$$-\left[\hat{P}^V - s_K(t)\hat{R}(t) - s_L(t)\hat{W}(t)\right] = \hat{V}(t) - s_K(t)\hat{K}(t) - s_L(t)\hat{L}(t)$$

But the right hand side is just our quantity measure of the growth of MFP, equation (9). So we have

$$MFPG(t) = -\left[\hat{P}^V - s_K(t)\hat{R}(t) - s_L(t)\hat{W}(t)\right] \qquad (18)$$

In words, the growth of MFP is the difference between the weighted average growth of input prices and the growth of the output price. MFP growth is positive if average input prices are growing more rapidly than the output price.

Finally, we can rearrange the right hand side of (18) to get

$$MFPG(t) = s_K(t)\left(\hat{R}(t) - \hat{P}^V\right) + s_L(t)\left(\hat{W}(t) - \hat{P}^V\right) \qquad (19)$$

The growth of MFP is now seen to equal the weighted average growth of the *real* input prices (the real wage and the real rental price).

8. Multi-factor productivity in more depth

Many inputs

So far we have made the highly unrealistic assumption that there are only two inputs, labour and capital. Obviously there are many types of both. And when we get down below the aggregate level there are also intermediate inputs like energy to consider. But it is straightforward, at least in principle, to expand the framework above to incorporate as many inputs as we like or at least that we can measure. Take labour first. Suppose we apply the principle of marginal productivity or just pragmatically accept that different types of labour input should be weighted by what they cost to hire. Then either way this justifies us in constructing an index of aggregate labour input as

$$\hat{L} = \sum_{l=1}^{N_L}\left(\frac{W_l L_l}{\sum_{l=1}^{N_L} W_l L_l}\right)\hat{L}_l$$

Here L_l is the l-th type of labour, e.g. hours worked by accountants or nurses, and N_L is the number of different types of labour. The term in round brackets is the share of the l-th type of labour in the aggregate wage bill. The growth of the index of labour input is then a share-weighted average of the growth rates of hours worked by the different types of labour. We must now distinguish between the growth of hours worked, not adjusted for quality, and the growth of the index *L*. Hours worked (*H*) are just the sum of hours worked across all types of workers:

$$H = \sum_{l=1}^{N_L} L_l$$

The growth of labour quality (also called labour composition) is the difference between the quality-adjusted growth rate and the crude growth rate of hours worked:

$$Growth\ of\ labour\ quality = \hat{L} - \hat{H}$$

The ONS now publishes a regular series called *quality adjusted labour input (QALI)*. According to the latest figures, between 1994 and 2018 hours worked (*H*) grew on average by 0.86% per year, labour quality (also called labour composition) by 0.45% per year and so quality adjusted labour input (*L*) grew by 1.31% per year.[15] Much of the growth in labour quality over these years has been due to the rising proportion of the labour force with university-level education. This reflects the fact that on average graduates earn more than non-graduates.

Capital input

We can try to apply the same approach to capital as we have just done to labour. The capital index would then be a weighted average of all the different types of capital. To construct it, we would need to know the physical quantities (the stocks) of each of the different types of tangible and intangible capital, And we would need to know too the prices at which each type can be hired. So if N_K is the number of asset types the index for capital would look just like the one for labour:

$$\hat{K}(t) = \sum_{k=1}^{N_K} \left(\frac{R_k(t)K_k(t)}{\sum_{k=1}^{N_L} R_k(t)K_k(t)} \right) \hat{K}_k(t) \qquad (20)$$

Here the weight for each type of capital is the share of that type in total rental payments:

$$\text{Weight for capital of type k} = \frac{\text{Total rental payments for type k capital}}{\text{Total rental payments for all types of capital}} = \frac{R_k(t)K_k(t)}{\sum_{j=1}^{N_L} R_k(t)K_k(t)}$$

And total rental payments are identified with profit (gross operating surplus). This treats capital in exactly the same way as we did labour. The weight attached to each type of labour

15 https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/datasets/qualityadjustedlabourinputexperimentalsummarydata. The ONS uses a Törnqvist index (see Box 6.2) to approximate the continuous indices I am using here.

is its share in the total wage bill. The weight attached to each type of capital is its share in total profit.

An expression like (20) defines what the ONS calls the *Volume Index of Capital Services* (VICS). Note the difference between the VICS and a possible alternative, a volume index of capital *stocks*:

$$\text{Growth of volume index of capital stocks } = \sum_{k=1}^{N_K} \left(\frac{P_k^A(t) K_k(t)}{\sum_{j=1}^{N_L} P_k^A(t) K_k(t)} \right) \hat{K}_k(t)$$

Here the weights are the share of each asset in the total value of all assets at a point in time and the prices are the *asset* prices, P_k^A, not the rental prices, R_k. (The asset price of a flat is what you have to pay to buy it, the rental price is the monthly rent you pay to occupy it for a month). For productivity analysis the relevant index is capital *services* (the VICS), equation (20), and not capital *stocks*.[16]

The difference between the stocks and services measures is not just of theoretical interest. It makes a substantial difference to the estimates. In recent decades the stocks of all types of machinery have been growing faster than the stock of buildings. Within machinery, the stocks of computers (mainframes, PCs, laptops, tablets, servers, etc) have been growing faster still. These are all cases where the ratio of the rental price to the asset price is high, particularly for computers. This means that the VICS grows faster than the stocks measure. But in trying to estimate the VICS we hit a major problem. Most of the required information does not exist. The situation here is quite different to the one the ONS faces with labour where it can draw on detailed surveys of the labour market giving wages and numbers employed broken down by industry, occupation and qualification level. Nothing like this exists for capital. True, firms publish accounts which show the value of their assets. But these are not usually broken down by type. Also, assets are generally valued at historic cost, i.e. the prices are those of when the assets were acquired, so the values represent an unknown mix of prices. The values are shown after depreciation which is influenced by tax considerations; as we will see, this is not the right basis for productivity analysis. Also the price at which an

[16] There is an ambiguity in the way the stocks index has been defined in that the same symbol K_k is being used to denote the flow of services from the stock of type k assets and also the stock of the type k asset itself. The ambiguity can be eliminated by assuming that one unit of type k capital delivers one unit of services per year. This means that the utilisation rate of capital is assumed constant: see below.

asset can be rented is frequently not observed at all since the asset is owned by the firm actually using it: the firm in effect rents the asset to itself.

Asset stocks

What the ONS does observe is *gross investment* (gross fixed capital formation) in current prices in different types of asset. These include various types of *tangible* asset such as buildings, vehicles, and machinery. These may be further broken down. For example, within machinery, computers are distinguished separately. Gross investment can also be in *intangible* assets. The most important types of intangible assets in Britain are software, R&D, and mineral exploration (mainly prospecting for oil and gas).

The ONS also constructs price indices for products which are purchased as capital assets. These price indices are (at least in principle) adjusted for quality change. The most significant quality adjustment has been for computers. This was first developed by the U.S. Bureau of Labor Statistics in the 1980s and has caused the quality-adjusted computer price to fall at the astonishing rate of around 20% per year for several decades. The US approach has since been adopted by other leading NSAs including the ONS.

In summary, the ONS has information about gross investment but not about asset stocks, and about asset prices but not rental prices. Nonetheless the data that are available can in fact be used to derive the missing information, by applying some economic theory.

Suppose that the ONS has information about asset lives.[17] Specifically, let the life of the k-th type of capital be T_k years, i.e. when an example of this asset becomes T_k years old it is scrapped. Then the *gross stock* of this asset at the beginning of year *t* (the total quantity still surviving at the beginning of year *t*) is just cumulated gross investment over the preceding $T_k - 1$ years

$$GK_{kt} = I_{k,t-1} + I_{k,t-2} + ... + I_{k,t-T_k+2} + I_{k,t-T_k+1}$$

(In practice the ONS assumes that, for a given asset type, scrapping, or retirement, does not all take place at exactly the same age but is distributed around the assumed asset life.) The gross stock is not a very interesting concept. Much more interesting is what the ONS calls the

[17] See ONS (2019i) for the ONS's current views on asset lives.

productive capital stock, where allowance is made for the possibility that an asset yields a lower level of services as it ages. Suppose that a one-year-old asset yields services which are only a fraction $(1-d_1)$ of the services yielded by a new asset; in general let a *t* year old asset yield only the fraction $(1-d_t)$ of the services from a new asset. The factor d_t is known as the rate of *decay* (or deterioration). Then the productive stock (equivalent to the total volume of services yielded by this asset) is

$$K_{kt} = I_{k,t-1} + (1-d_1)I_{k,t-2} + ... + (1-d_{T_k-1})I_{k,t-(T_k-2)} + (1-d_{T_k})I_{k,t-(T_k-1)}$$

This equation illustrates the method known as the *Perpetual Inventory Model* or PIM and is the basis for estimating asset stocks in practice.

A very interesting special case is when the rate of decay per period is constant so that $(1-d_1) = (1-d)$ and in general $(1-d_t) = (1-d)^t$. This effectively assumes that the asset lasts forever but in practical terms its services eventually become negligibly small. (Old assets never die, they just fade away). This assumption is known as *geometric decay*. Now the productive stock is

$$\begin{aligned} K_{kt} &= I_{k,t-1} + (1-d)I_{k,t-2} + (1-d)^2 I_{k,t-3} + ... \\ &= \sum_{r=0}^{\infty} (1-d)^r I_{k,t-r-1} \end{aligned}$$

This yields a simple way of estimating the productive stock. Lag the last equation one year, multiply by $(1-d)$, and subtract the result from the last equation:

$$K_{kt} = I_{k,t-1} + (1-d)K_{k,t-1} \tag{21}$$

If we can estimate a starting stock, the stock at some distant date in the past, or simply assume that prior to some year the stock was zero, then we can estimate the stock in all subsequent years by rolling forward equation (21).

But how are we going to estimate the rate of decay? This is where the geometric assumption becomes very useful since it creates a simple link between decay, a physical concept, and depreciation, a price concept. The prices of many types of second hand assets have been observed and are generally found, not surprisingly, to decline with age. The rate of decline is called depreciation and is denoted by δ. Empirical studies of asset prices generally find that the rate of decline is approximately geometric. So unlike decay, depreciation can be directly observed.

Suppose that a particular type of asset is subject to geometric decay. What would that imply about its price as it aged? Consider two examples of this asset, one new and the other one year old, both observed at the same point in time. The one year old asset yields just the same expected flow of future services as the new one, except reduced by a factor of $(1-d)$. Hence its price today must be $(1-d)$ times the price of the new asset today. So in this case the rate of decay and the rate of depreciation must be the same: $d = \delta$. In other words under the assumption of geometric depreciation we can estimate rates of decay from surveys of second hand asset prices which give us rates of depreciation.

An alternative assumption to geometric decay is the pattern known as "one-hoss-shay" or "lightbulb". Now the asset yields an undiminished flow of services over its lifetime before failing catastrophically at the end of its life. The rate of decay is zero (the productive stock and the gross stock are the same) but we would still see the asset price declining steadily before becoming zero at the moment of failure. This is because the asset price represents the present value of the stream of future returns from the asset up till the moment of failure. As the asset ages the stream of future profits is getting shorter so the price is falling.

The lightbulb pattern may seem superficially more realistic in many cases than the geometric one. If an asset is maintained why should its services decline with age? But there are strong arguments on the other side too. We are trying to estimate the contribution of an asset type with many individual members. Suppose the rate of failure is independent and random. Then even if any one example has the "lightbulb" pattern the type as a whole will have what looks like geometric decay. Also repair and maintenance cannot make up for all age-related damage, at least at an acceptable cost: the second-hand price of even well-maintained cars declines with age. Finally, we need to take account of the major cause of decay in the real world: obsolescence. Even though an asset may yield the same flow of physical services over time those services may well be valued less and less as the asset ages. Machinery and buildings may be specialised in the production of goods and services which nobody any longer wants to buy. Newspaper printing presses are an example as news provision moves increasingly online. Personal computers and laptops are a striking example of obsolescence. They are subject to very little physical wear and tear but nevertheless firms usually replace them every four or five years. This is because technical progress has been so rapid here that it is cheaper to scrap and replace rather than try to retrofit older models with better hardware capable of running new software.

Rental prices

So there are practical ways of estimating asset stocks. But to calculate the volume index of capital services of equation (20) we need to know rental prices too. The way to proceed is to find the relationship between asset prices, which the ONS can observe, and rental prices, which it usually can't.

Consider a firm which is thinking of buying some asset. The firm will have some nominal required rate of return in mind (r_t). If the funds are borrowed, the required rate is the borrowing rate. If the firms is using its own funds the required rate is what it could earn on a comparable investment. So part of the cost of buying and using the asset for one year, called the *user cost of capital*, is the required rate of interest times the price of the asset. But this is not the whole story. At the end of the year the firm will own an asset whose second hand price will be different from what it paid, very likely lower. The firm could then sell the asset second hand and probably make a capital loss (or conceivably a capital gain). This capital loss (or gain) has to be included in the user cost so that the firm can compare the prospective gain against the cost. The user cost of capital is now defined as follows:

User cost = Interest cost *plus* capital loss

Or in symbols

$$R_t = r_t P_{t,0}^A + P_{t,0}^A - P_{t+1,1}^A$$

Here R_t is the user cost at time *t* (which will turn out to be the same as the rental price we are looking for, so justifying use of the same symbol for both concepts), r_t is the required rate of return, $P_{t,0}^A$ is the asset price, i.e. the price of a new, age zero, asset, and $P_{t+1,1}^A$ is the price of a one year old asset one year later, i.e. the price at $t+1$ of an asset aged one year. (For clarity I have dropped the subscript *k* indicating the k-th asset type.) The aim now is to show that the user cost is actually the same as the rental price we are looking for.

The formula for the user cost can be put into more economically meaningful terms by defining the rate of growth of the price of a new asset, π_t:

$$\pi_t \equiv \frac{P^A_{t+1,0} - P^A_{t,0}}{P^A_{t,0}}$$

(note that all the prices in this formula relate to a new asset at different times) and the rate of depreciation δ by

$$\delta \equiv \frac{P^A_{t,0} - P^A_{t,1}}{P^A_{t,0}}$$

(note that all the prices in this formula are those observed at the beginning of year *t*). Box 6.5 shows that with these definitions the user cost of capital can be written as

$$R_t = \left[r_t + \delta(1+\pi_t) - \pi_t \right] P^A_{t,0} \qquad (22)$$

The rental price is high in relation to the asset price if the depreciation rate is high and/or if the price of a new asset is falling (when π_t is negative). This is the case with high-tech assets like computers and (probably) software where both these conditions apply.

The firm will decide to invest if the expected return is greater than the user cost. The expected return is the value of the marginal product of this type of capital. Under perfect competition the marginal product will be driven into equality with the real user cost. The required rate of return now becomes equal to the actual rate of return which in equilibrium must be the same for all assets types. The user cost is also the relevant consideration for a firm which actually owns the asset and is thinking of renting it out. Of course such a firm would like to get the highest possible price but competition will drive the rental price down to equal the user cost. (See below for an important qualification to this argument when competition is imperfect).

In practice rental prices are also influenced by taxes and subsidies. For example, depreciation can normally be set against corporation tax and governments often allow firms to depreciate their assets faster than true economic depreciation would warrant. This lowers the rental price. The degree of subsidy often varies by asset type. The rental price formula (22) can be adapted to take these factors into account since tax rates and depreciation allowances are known.

All the elements of the user cost of capital in equation (22) are now observable with the exception of the nominal rate of return r_t. The commonest way to proceed now is to use the fact that in equilibrium r_t is the same for all assets. Then the sum of the returns to all types of asset equals observed profit (gross operating surplus. *GOS*):

$$\sum_{k=1}^{N_K} R_{kt} K_{kt} = GOS_t$$

Taking account of (22), this is an equation in just one unknown, r_t, and we can therefore solve it to find the unknown rate of return and so calculate all the rental prices. Now we know the rental prices and the stocks we can calculate the VICS, equation (20).

Empirically, the ratio of the rental price to the asset price is found to vary widely between asset types. This is mainly due to differences in depreciation rates (asset lives). For computers researchers often use a depreciation rate of 33% while for buildings it is typically 1-2%.

In summary, a practical method has been outlined for estimating capital input. It relies on the ONS being able to observe gross investment by asset type, price indices for each asset type, and depreciation rates derived from knowledge of second hand asset prices. Geometric decay is a popular assumption in academic research for the reasons just given. In practice the ONS calculates rental prices using the method outlined above but asset stocks using different assumptions about decay. Technically this means that the rental prices are not consistent with the assumptions about decay. Whether the ONS's assumptions on decay are more realistic is hard to say since the empirical evidence is sparse.

Qualifications and extensions

The estimation of MFP can be a controversial matter. So a few qualifications are in order together with an indication of where more research may be needed.

1. The justification for the methods for estimating MFP set out here rests on the assumption of a competitive economy in which prices equal marginal costs, real wages equal marginal products, and there are no monopoly profits. This assumption makes a lot of people unhappy. The alternative is some form of imperfect competition, now the default assumption in short run macroeconomics. As Box 6.8 explains in more detail, under imperfect competition it is no longer true that the labour share equals the elasticity of output with respect to labour. Now the labour share has to be multiplied by the *markup*, the ratio of price to marginal cost. A similar adjustment has to be made to estimate the capital elasticity, with the added

complication that we must also distinguish between the total cost of capital and profit, the difference being monopoly profit which of course does not exist under perfect competition.

Why do NSAs not adopt the assumption of imperfect competition? Is it because they are tools of monopoly capitalism and infected with neoliberal ideology? Probably not. There is no very obvious and non-controversial way of estimating the size of markups. There have been several different approaches in the academic literature but no consensus has yet emerged (Basu 2019). The real difficulty in using imperfect rather than perfect competition may be data. One promising method requires log runs (over 30 years) of high quality data at the industry level. Currently this does not exist for Britain.

2. We noted above (footnote 16) that utilisation rates are assumed to be constant. This is a reasonable assumption in long period productivity analysis. But over the business cycle changes in utilisation may cause variations in measured MFP growth, with the latter falling in recessions and rising again in recoveries. This is because during a recession sales fall so capital may be left idle while labour is hoarded, at least for a while. During an expansion the opposite occurs: capital becomes more fully utilised and labour is more usefully employed. This pattern is frequently observed over the business cycle and we should recognise this in interpreting the short run movements in MFP.

3. In the case of labour we would want to include all types, in as much detail as possible. The same applies to capital, The number of types of capital included has in fact increased in recent decades. Software (including computerised databases) was the first to be included as a new asset type in the 1993 SNA. R&D was included as a new asset type in the 2008 SNA. These are examples of intangible assets. But recently it has been suggested that the number of recognised asset types should be further expanded. According to this view, various “business competencies” should be recognised as assets since firms spend considerable amounts on acquiring and maintaining them. These competencies which can be regarded as intangible assets include design, brands, organisational capital and firm-specific training (Corrado et al. 2005). It is possible that at some future date some or all of these new asset types may be included in the official estimates following a revision of the SNA.

4. Two other assets types which already feature in the national accounts but are not incorporated into the ONS’s productivity analysis are land and inventories. (Estimates of the

stock of buildings do not include the value of the land on which the buildings stand.) Some of gross operating surplus should be regarded as a return on these assets but at the moment all of gross operating surplus is attributed to the included assets. At the aggregate level land is almost constant in Britain. So the effect of including land would be to reduce the contribution of capital input (by including a zero-growth asset with a significant weight), and so to raise MFP growth.

5. Finally, how accurate are the estimates of MFP growth? MFP growth is the difference between two large numbers each of which is no doubt measured with error. Random errors would be easy to live with but there are reasons to expect systematic biases even at the aggregate level. The main concern here is the price indices used to deflate the prices of investment goods. As Bean (2016) emphasises the quality of these is in general lower than for consumer goods which come under the well-established consumer price index programme.

Suppose the prices of some investment goods are biased upwards due to inadequate allowance for quality change. This has certainly been true in the past for computers and may be the case still for other high tech goods like communications equipment and software. At the aggregate level there are now two forces pulling in opposite directions. On the input side the contribution of capital is understated so on this account MFP will be overstated. But since GDP includes investment the estimate of GDP growth will be too low, offsetting the first effect. However in the British case the first effect is likely to dominate since most high tech goods are imported. If all such goods were imported then any error in the price indexes of these goods will wash out of GDP completely: there will be an error in measuring real investment balanced by an equal and opposite error in measuring real imports.

9. ONS estimates of MFP growth at the aggregate level and the productivity puzzle

In this section we focus on the ONS's estimates of MFP for the market sector as a whole. Later we look at industry-level estimates after we have developed the necessary theory.

The ONS now publishes MFP estimates on a quarterly as well as an annual basis for the market sector and for 13 industry groups within the market sector. (The definition of the market sector differs from the one we used earlier to analyse labour productivity. Now the market sector is defined by ownership, not type of industry). The ingredients for these estimates are real and nominal value added, payments to labour and capital, hours worked, quality-adjusted labour input (QALI), and the volume index of capital services (VICS). The QALI and the VICS are also published separately.

The official estimates of QALI go back to 1970 and are for the whole economy, the market sector, and are further broken down into 19 industry groups (ONS 2017). Labour is broken down into 36 types: by sex (2), age (3 groups), and educational level (6 groups).

The ONS estimates of the VICS are for the market sector as a whole and for 57 industries that make up the market sector. These estimates go back to 1950 (ONS 2018b and 2019c). The VICS is built up from 13 asset types (Table 7). Points to note here are:

1. Dwellings are not included since these are mainly purchased by households and the output of dwellings is the imputed rent of owner-occupiers (see above). Imputed rent is not part of market sector output.
2. Land improvements are included but not land as such.
3. Five of the thirteen asset types are intangibles.
4. As noted above, inventories are not included.

Table 7 near here

Currently the ONS uses the rate of return actually achieved in the market sector as its measure of the required rate of return. It uses this rate in its industry estimates of capital services as well as for its aggregate estimate. The estimated real rate of return (the rate of return after subtracting a measure of inflation) turns out to be very high, over 20 % per year, which is much higher than the estimates of independent researchers (e.g. Oulton and Wallis 2016).

The official estimates of MFP for the market sector go back to 1970 (ONS 2019b). The ONS makes use of the following decomposition, a refinement of the growth accounting equation we derived earlier, equation (11). Subtract the growth of hours worked (*H*) from both sides of this equation to get:

$$\hat{V}(t)-\hat{H}(t)=s_K(t)\left[\hat{K}(t)-\hat{H}(t)\right]+s_L(t)\left[\hat{L}(t)-\hat{H}(t)\right]+MFPG(t) \qquad (23)$$

(Here we use the fact that the shares sum to 1.) Putting this into words,

Growth of labour productivity (*V*/*H*) = Contribution of capital deepening (*K*/*H*)

plus Contribution of labour quality (*L*/*H*)

plus MFP growth

These calculations are made using the Törnqvist formula (see Box 6.2).

Chart 7 shows this decomposition for the period 2008Q1 to 2018Q1 and illustrates the productivity puzzle once again. Labour productivity just about regained the level reached at the end of the boom in 2008Q1 by the end of 2018, ten years later. The contributions of labour quality (composition) and capital deepening were positive but that of MFP was consistently negative. In fact the level of MFP in 2018 was 4% lower than it had been in 2007Q4, a decade earlier. This has led many to conclude that the productivity puzzle is an MFP puzzle (Goodridge et al. 2018).

Chart 7 near here

10. MFP at the industry level

The value added approach

At the aggregate level as we have seen the growth rate of MFP is

$$MFPG=\hat{V}-s_K\hat{K}-s_L\hat{L}$$

to repeat equation (9). (Here we must now understand *K* and *L* to be aggregates of many types of capital and labour respectively, just as real value added is an aggregate of real value added in many different industries). MFP growth in the aggregate must arise somehow from MFP growth at the industry level. This suggests that we should measure MFP growth in each industry by

$$MFPG_i^{VA}=\hat{V}_i-s_{iK}\hat{K}_i-s_{iL}\hat{L}_i \qquad i=1,\ldots,N$$

I am using the notation $MFPG_i^{VA}$, with the superscript "VA", to indicate that this is the *value-added* approach to industry-level MFP. The alternative *gross output* approach is discussed later. The question now arises, what is the relationship between growth at the industry level and growth at the aggregate level? A good first guess is that the aggregate growth rate is a weighted average of the industry growth rates with the weights being the value added shares:

$$MFPG = \sum_{i=1}^{N} v_i MFPG_i^{VA} \tag{24}$$

After all, as we saw above the growth of GDP is a weighted average of the growth rates of real value added in all the industries, with the weights being these same value added shares (the v_i). Equation (16) can be thought of as the "top-down" approach to measuring MFP growth while equation (24) is the "bottom-up" approach.

However it turns out that equation (24) is only correct under a restrictive assumption. This assumption is that a given type of input, say 35 year old women with a law degree, is paid the same whichever industry they work in. If this is the case for all inputs then equation (24) is correct. Otherwise the full answer for the bottom up approach is

$$MFPG = \sum_{i=1}^{N} v_i MFPG_i^{VA} + \text{Realloaction effects} \tag{25}$$

To see how these reallocation effects arise imagine that a worker on a low wage gets a job in an industry where she is paid more. Since wages, by assumption, measure marginal products aggregate output has risen. True, output falls in the industry she leaves but at given prices it rises by more in the industry she joins. So aggregate output rises while aggregate inputs are the same, i.e. aggregate MFP increases. Reallocation may affect capital as well as labour and may be either positive or negative. The importance of reallocation can be gauged by calculating it as the difference between the top-down approach of equation (16) and the crude bottom-up approach of equation (24).

The example just given of what is apparently the same input — 35 year old women with a law degree — should give us pause. Are they really the same input? Suppose one woman works for a firm of solicitors doing conveyancing while the other for an investment bank structuring M&A deals. The latter is likely to earn much more. Her higher pay may reflect greater ability or a higher energy level. So they are not really the same input at all. Hence what appears as reallocation may be just the result of our inability to measure inputs with sufficient accuracy. Even so this hypothetical example uses a finer classification than in the official estimates which do not distinguish between degrees in different subjects; the age groups are also coarser.

On the other hand reallocation may have played a genuinely important role in some periods of rapid productivity growth. One example is the rapid development of a country like Italy in the twenty to thirty years after the second world war when large numbers of unskilled workers left agriculture to work in factories. Agriculture was very backward so productivity and wages were low. Wages in manufacturing were much higher, though still low by our current standards. The same phenomenon on an even larger scale is seen in today's China where hundreds of millions of peasants have moved from the countryside to cities. So reallocation may be a very important part of the story for some countries.

The gross output approach

A rise in MFP in the value added sense means that it has proved possible to produce more real value added with the same amount of capital and labour. But what does this really mean? Consider an industry like shirt-making. Rather than ask about real value added in shirt-making, surely it is more natural to ask: has it been possible to produce more *shirts* with the same quantity of *all* inputs — capital and labour yes, but also cloth, thread, electricity, accounting services, etc. In other words, real value added is a derived concept while real output (e.g. shirts) is the more fundamental one. Indeed we have derived real value added from real output and real inputs: see the discussion of double deflation in section 4.

At the industry level the natural approach is to start with the accounting framework we used in section 4, equation (3). In current prices, gross output[18] is the sum of payments to capital, payments to labour, and payments for intermediate inputs. Now suppose that there exists for each industry a gross output production function:

$$Y_i = A_i f(K_i, L_i, M_i) \qquad A_i > 0 \qquad i = 1,\ldots,N \tag{26}$$

where K_i is an index of capital input, L_i an index of labour input, and M_i an index of intermediate inputs. By an argument essentially the same as we used to derive equation (16) in Box 6.4, we can derive the gross output measure of MFP growth:

$$MFPG_i^{GO} \equiv \hat{A}_i = \hat{Y}_i - s_{Ki}^{GO}\hat{K}_i - s_{Li}^{GO}\hat{L}_i - s_{Mi}^{GO}\hat{M}_i \tag{27}$$

(Note the superscript "GO" to distinguish this concept of MFP growth from the value added one). Here the shares are shares in the value of gross output, not of value added: :

[18] NSAs now prefer to call this just "output".

$$s_{iK}^{GO} \equiv \frac{R_i K_i}{P_i Y_i}$$

$$s_{iL}^{GO} \equiv \frac{W_i L_i}{P_i Y_i}$$

$$s_{iM}^{GO} \equiv \frac{P_i^M M_i}{P_i Y_i}$$

and the shares sum to one. In words, the gross output measure of MFP growth is the growth of gross output minus the share-weighted growth rates of *all* the inputs, including intermediate ones.

The aggregate MFP growth rate is still given by equation (9):

$$MFPG = \hat{V} - s_K \hat{K} - s_L \hat{L}$$

So what is the relationship between this and the industry growth rates of MFP, now using the gross output concept of equation (27)? The answer is that, ignoring any reallocation effects, the aggregate growth rate is a weighted *sum* of the industry growth rates:

$$MFPG = \sum_{i=1}^{N} d_i MFPG_i^{GO} \qquad (28)$$

where the d_i are the so-called *Domar weights*:

$$d_i \equiv \frac{P_i Y_i}{GDP_t}$$

Notice that the Domar weights sum to more than 1, typically in practice to about 2. (This is another way of saying that to calculate GDP we add up value added because adding up gross output would involve double counting). The intuition here is that an industry's contribution to the aggregate is higher, the larger its sales to other industries. This is because under perfect competition technical progress in an industry reduces the price of its product and this benefits all other industries in proportion to its sales to them.

It still remains the case that according to equation (24) (ignoring once again reallocation effects):

$$MFPG = \sum_{i=1}^{N} v_i MFPG_i^{VA}$$

Box 6.7 shows that the relationship between the two concepts of MFP growth is then:

$$MFPG_i^{VA} = \left[\frac{P_i Y_i}{P_i^V V_i} \right] MFPG_i^{GO} \qquad (29)$$

So the value added measure of MFP always grows faster than the gross output one because gross output exceeds value added.

A problem with the value added measure and an advantage for the gross output measure is that the value added one can give a misleading signal as to the location of technical progress. It is possible for the value added MFP growth rate to increase while the more fundamental gross output one is constant. This can come about if the price of an intermediate input is falling. If this is happening on a sustained basis it is likely due to faster technical progress in the industry supplying the intermediate product. Then the industry purchasing the intermediate will buy more so enabling it to increase its own output without increasing purchases of other inputs or at least not by as much. So, looking at the last equation, output (Y_i) rises with everything else initially the same, so $MFPG_i^{VA}$ increases even though $MFPG_i^{GO}$ is constant. In other words the value added measure wrongly assigns technical progress which is really occurring in the supplying industry to the purchasing industry.

All the relationships discussed in this section have assumed that real value added is calculated by double deflation. In fact as we discussed in section 4, at the moment the ONS calculates real value added at the industry level by single deflation. So it is problematic to compare the two measures in practice. When double deflation is fully implemented in the national accounts this situation will improve.

11. Productivity at the firm level

Longitudinal databases of firm-level data

All NSAs in advanced industrial economies carry out regular, large scale surveys or censuses of businesses. Indeed these surveys provide an essential input into the construction of the national accounts. The data gathered from each business include industry, location, sales or turnover, wage bill, profit, purchases of inputs, employment and capital expenditure (gross investment). In the last 20-30 years there has been a movement to create longitudinal databases of data on firms and to open them up to researchers. These databases have been formed by linking together the (usually) annual data, potentially a challenging task.

Unfortunately, the data have been collected under very stringent confidentiality laws so that they cannot simply be placed on the Internet. However, ways have been found to give access to accredited researchers while preventing illegal disclosure.

The United States Bureau of the Census was the pioneer here but the ONS was close behind in setting up what is called the ARD (which now stands for Annual Respondents Database) in 1997 (Oulton 1997). This was originally based on the old Census of Production which covered only the production sector of the economy (manufacturing, mining, construction and the utilities); in practice the ARD originally covered only manufacturing. But it has now been expanded to include much of the private services sector.

The basic unit of analysis in these databases needs to be clearly defined. In Britain, it is the "reporting unit". For a small firm this is just the firm itself. For a large firm it is a unit which organises the collection of income tax under PAYE and National Insurance contributions; such a unit may be below the level of a legally incorporated business. The data reported by "reporting units" may cover one or more "local units". The latter are establishments (shops, offices, factories, etc) at a specific geographical location. Some data is also gathered at a "local unit" level. In other countries the basic unit may differ. In the United States for example it is the "establishment" which corresponds to the British local unit.

There are a number of measurement issues specific to firm-level studies:

1. Capital is hard to measure since the basic data are only for gross investment and usually not broken down by asset type. It is difficult to use the perpetual inventory method at the firm level since starting stocks are not usually known.
2. If capital is poorly measured then MFP will also be poorly measured.
3. To measure firm-level productivity we ideally need firm-level price indices. But these usually do not exist. So instead researchers deflate firm-level value added or sales by an industry-level price index. In the rare cases where firm-level price indices do exist it has been found that new entrants tend to charge lower prices than incumbent firms; if the entrants survive they tend to raise their prices over time. This completely changes the story as to why entrants have (apparently) low productivity. Taken literally their low productivity suggests that entrants have to learn best practice on the job over time. The other explanation is that they have to establish a reputation for

reliability with their customers and so can only win customers initially by charging low prices.

4. There is the issue of matching the micro and the macro data. If we add up firm-level value added in an industry we won't get the figure which appears in the national accounts. And movements over time in real value added won't necessarily match either. This is because the raw, firm-level data are subject to various adjustments to arrive at the national accounts number through the supply-use balancing process. So a theory which fits the facts at the micro level might not do so at the macro level.

An alternative to confidential data from business surveys is data from firms' accounts. These have the advantage of being in the public domain. But from a measurement point of view they present problems of their own. Firms report the value of their assets but these are usually at historic cost as noted above. If they report labour input it will usually be employees, not hours. Sales are usually available but not value added.

Nonetheless these databases have opened up a new window on the economy. A number of findings have emerged and seem broadly similar across countries (Syverson 2011; Andrews et al. 2015):

1. The variation across firms in productivity levels at a point in time is very high. This is true for both labour productivity and MFP, even for firms in the same industry.
2. The life of firms, even very successful ones, is quite short. Most UK firms, even large ones, are much younger than the average Premiership football club, many of which were founded in the 19th century.
3. Entry and exit rates of firms vary across industries. Both tend to be higher in more rapidly growing industries.
4. Entry rates have declined in recent years.
5. Within a given industry there is a tendency for the productivity level of lagging firms to converge towards that of leaders. However this tendency seems to have become weaker in the last 20 years.
6. New entrants into an industry tend to have low productivity. But provided they survive (and mortality is high) their productivity grows comparatively rapidly.
7. Exiting firms (exitors) tend to have low productivity. Since this is linked to low profitability, it helps to explain their decision to exit.

12. Conclusions

The ONS now publishes a wide range of statistics on productivity in the British economy. These cover labour productivity at both the level of the whole economy, by industry, and by region. Some international comparisons of whole economy labour productivity levels are also published for a restricted range of countries. Going beyond labour productivity, annual and quarterly estimates of multi-factor productivity (MFP) are now published for both the whole economy and by industry. MFP estimates rest on estimates of labour inputs which incorporate quality or compositional differences and of capital input which aggregate over a number of both tangible and intangible asset types. The theoretical framework is broadly neoclassical as set out in the OECD's two manuals on productivity and capital measurement (OECD 2001 and 2009) which in turn rest on the foundational work of Solow, Domar, Jorgenson, Hulten and others.

The official estimates clearly show that since 2007 both labour and multi-factor productivity growth have stagnated. Both levels are no higher today (the end of 2019) than they were at the height of the boom in 2007. This stagnation seems to be general across both industries and regions, with structural change playing only a minor role.

Despite impressive progress there are a number of both empirical and theoretical limitations to the estimates.

First, at the industry level real value added is still measured by single deflation rather than the theoretically preferred double deflation. When double deflation is fully implemented in the national accounts the picture of recent developments in productivity at the industry level (though not at the whole economy level) after 2007 might change dramatically.

Second, estimates of capital input may prove to be biased downwards due to the absence of good price indices for capital goods, particularly in the ICT area. This problem affects the estimates of GDP too but the effect on GDP is likely to be smaller than the effect on capital, leading to upwardly biased estimates of MFP growth.

Third, the list of asset types included in the capital input measures should be expanded to include land and inventories and possibly some further intangible assets.

Fourth, there are no official international comparisons of productivity (labour or MFP) at the industry level. The reason is that the International Comparison Program which delivers the PPPs used to compare living standards is done from the expenditure side of the national accounts, not the output side. So there is no readily available source for basic prices which is comparable across countries. Therefore there is no easy way to compare for example real value added per hour worked in the German and British steel industries let alone to make trickier comparisons such as real value added per hour worked in these countries' banking industries.

Fifth, even at the whole economy level international comparisons of labour productivity levels are hampered by differences in the way countries measure labour input, even when they are supposedly using the same concept (total hours actually worked by all types of labour).

Finally, the framework employed in all the MFP estimates assumes perfect competition. It is thus at variance with the dominant school of macroeconomics which assumes imperfect competition. If imperfect competition matters then the weights being used to calculate total input and hence MFP growth are systematically wrong. Perhaps more important, perfect competition allows no role for economies of scale. So the effects of economies of scale are being wrongly subsumed under MFP. How much any of this matters is not clear at the moment. Progress here requires estimating industry (or firm) production functions, a task not usually thought of as being within the scope of a national statistics agency.

BOX 6.1 Geometric, exponential and continuous growth rates, and log scales

Geometric growth rates

Consider some variable x which we observe at two adjacent periods t and t-1. In the *geometric* model, we say that

$$x_t = (1 + g_t)x_{t-1}$$

Then we can solve for g_t, the geometric growth rate of x from t-1 to t, to get

$$g_t = \frac{x_t - x_{t-1}}{x_{t-1}}$$

So the growth rate is the *change* in x, $x_t - x_{t-1}$, divided by the *level* of x at the start, x_{t-1}. Notice that the growth rate is a proportional concept and that it is independent of the units in which we are measuring x (pounds, euros, metres, tonnes, etc). If the two time periods are adjacent years, then $100g_t$ is the annual percentage growth rate of x from t-1 to t.

Consider some quantity like labour productivity (Z) which is defined as real value added (V) divided by labour input (L):

$$Z_t = V_t / L_t$$

Suppose that the geometric growth rate of V is g_t and that of labour is h_t. The growth rate of productivity from t-1 to t, denoted by q_t, is then given by

$$1 + q_t = \frac{Z_t}{Z_{t-1}} = \frac{V_t / L_t}{V_{t-1} / L_{t-1}} = \frac{V_{t-1}(1+g_t) / L_{t-1}(1+h_t)}{V_{t-1} / L_{t-1}}$$

$$= \frac{1+g_t}{1+h_t}$$

Hence

$$q_t = \frac{g_t - h_t}{1 + h_t}$$

Notice that the growth rate of productivity is *not* just the difference between the growth rates of output and labour input but a bit less than this if productivity growth is positive and labour input is increasing ($h_t > 0$).

Now consider how to calculate the average growth rate over more than one time period, say from year *R* to year *T*. . Our model now is

$$x_T = (1+g)^{T-R} x_R$$

i.e. it is as if every year *x* grows at the constant rate *g*. Then solving for *g*,

$$(1+g)^{T-R} = x_T / x_R$$

$$g = (x_T / x_R)^{1/(T-R)} - 1$$

In percentage terms, the average annual growth rate of *x* over this period is 100*g*.

Exponential growth rates

The exponential growth rate of *x* between *t*-1 and *t* is defined as

$$x_t = e^{\gamma_t} x_{t-1}$$

where the mathematical constant *e* (= 2.7182…) is the base of the natural logarithms. Taking logs and solving for the growth rate γ_t

$$\gamma_t = \log\left[\frac{x_t}{x_{t-1}}\right] = \log x_t - \log x_{t-1}$$

The growth rate is equal to the difference of the logs. To covert to percentage terms, multiply by 100.

As before, labour productivity is $Z_t = V_t / L_t$. Let the growth rate of productivity be μ, that of output be θ, and that of labour be λ. Then

$$e^{\mu} = \frac{Z(t)}{Z(t-1)} = \frac{V(t)/L(t)}{V(t-1)/L(t-1)} = \frac{V(t-1)e^{\theta} / L(t-1)e^{\lambda}}{V(t-1)/L(t-1)}$$
$$= e^{\theta} / e^{\lambda} = e^{\theta-\lambda}$$

Hence, taking logs,

$$\mu = \log Z(t) - \log Z(t-1) = \theta - \lambda$$

So the growth rate of labour productivity is just the difference between the growth rates of output and of labour.

Suppose we want to measure the average growth rate of x over several years, say from year R to year T, i.e. a time span of $T-R$ years. Then the model is

$$x_T = e^{\gamma(T-R)} x_R$$

So

$$\gamma = \frac{(\log x_T - \log x_R)}{T-R}$$

The average exponential growth rate is the log difference divided by the length of the time span ($T-R$).

What is the link between geometric and exponential growth rates? The growth of x between t and t-1 can be described by either $1+g_t$ or by e^{γ_t}. Equating these and taking logs

$$\gamma_t = \log(1+g_t) \approx g_t$$

by a basic result in calculus. How good is the approximation? Good when the growth rate is low, getting worse as it rises. For example:

Table B.6.1
Comparison of exponential and geometric growth rates (to 2 decimal places)

Exponential growth rate, γ % per year	*Corresponding geometric growth rate, g % per year* $[=100*(e^{\gamma}-1)]$
0.50	0.50
1.00	1.01
2.00	2.02
3.00	3.05
6.00	6.18
10.00	10.52

CAUTION: I have been using the usual convention of mathematicians and economists that the "log" symbol means the natural log, the logarithm to the base e (the number 2.7182…), e.g. the log of 10 is $\log_e 10$ = 2.3025…. If you use so-called common logs then the base is 10, i.e. the log of 10 is $\log_{10} 10$ = 1. In that case, the change in the log of x does *not* equal the growth rate of *x*. Unfortunately, Excel uses a different convention. If you put "=log(10)" into a cell the result will be displayed as "1". To get the natural log of 10, put "=ln(10)" into the cell and "2.3025" will be displayed.

Continuous growth rates

For theoretical purposes it is often useful to think of economic variables as changing *continuously*. So the growth rate is now the growth rate at a point in time. The analogy is with a car's speedometer. When your eye happens to glance at it, it shows a particular speed, say 30 mph. This does not mean that in an hour's time the car will have actually travelled 30 miles because the speed may change. All it tells us is that the car would travel 30 miles if the current speed were held constant for an hour. We now write some variable *x* as $x(t)$ to indicate the dependence on time which is assumed to flow continuously, not just in seconds or even nanoseconds. The growth rate of *x*, denoted by the "hat" symbol(^), is written as $\hat{x}$. Mathematically, the growth rate of *x* at time *t* is defined as the logarithmic derivative of *x* with respect to time, evaluated at time *t*:

$$\hat{x} = \frac{d \log x(t)}{dt} = \frac{1}{x(t)} \frac{dx(t)}{dt}$$

There is a close connection with exponential growth over discrete time periods. Suppose the behaviour of *x* is determined by the following law:

$$x(t) = Ae^{\gamma t}$$

Then $\hat{x}(t) = \gamma$ and in this case the growth rate is constant.

Log scales

If we use a log scale to plot some variable *x* measured on the Y-axis against time on the X-axis, then we are in effect plotting the log of *x* against time. But for readability we label the points on the Y-axis with the original values, not the log values. The slope of the curve between any two adjacent points is given by the change in $\log x$ divided by the change in time (the latter equal to one unit), i.e. the slope is

$$(\log x_t - \log x_{t-1}) / 1 = \log x_t - \log x_{t-1}$$

Using the exponential growth model

$$\log x_t - \log x_{t-1} = \gamma$$

and using the geometric model

$$\log x_t - \log x_{t-1} = \log(1+g)$$

So whichever model we use, the faster the growth rate, the steeper is the slope.

BOX 6.2 Index number formulas

The table below shows the index numbers in common use, together with their formulas. The two time periods are labelled *r* (the reference period) and *t* (the current period). In time-series comparisons we generally assume that $t > r$; the gap between the two periods could be greater than one period. In the context of annual chain-linking $r = t - 1$. These indices can also be applied to cross-country (or cross-regional) comparisons: just replace "period" by "country (or "region") in the description below.

Table B.6.2
Common two period index numbers

	Price index		*Quantity index*	
Name	*Symbol*	*Formula*	*Symbol*	*Formula*
Laspeyres	P_{tr}^{Lasp}	$\sum_i s_{ir}\left(\frac{p_{it}}{p_{ir}}\right)$	Q_{tr}^{Lasp}	$\sum_i s_{ir}\left(\frac{q_{it}}{q_{ir}}\right),\ q_{ir}>0$
Paasche	P_{tr}^{Paas}	$\left[\sum_i s_{it}\left(\frac{p_{ir}}{p_{it}}\right)\right]^{-1}$	Q_{tr}^{Paas}	$\left[\sum_i s_{it}\left(\frac{q_{ir}}{q_{it}}\right)\right]^{-1},\ q_{it}>0$
Fisher	P_{tr}^{Fish}	$\left[P_{tr}^{Lasp}\cdot P_{tr}^{Paas}\right]^{1/2}$	Q_{tr}^{Fish}	$\left[Q_{tr}^{Lasp}\cdot Q_{tr}^{Paas}\right]^{1/2}$
Törnqvist[a]	$P_{tr}^{Törn}$	$\sum_i\left[\frac{s_{ir}+s_{it}}{2}\right]\ln\left(\frac{p_{it}}{p_{ir}}\right)$	$Q_{tr}^{Törn}$	$\sum_i\left[\frac{s_{ir}+s_{it}}{2}\right]\ln\left(\frac{q_{it}}{q_{ir}}\right),$ $q_{ir}>0$

a. The formula for the Törnqvist index defines its exponential growth rate (i.e. the log difference), not its level; equivalently, the formula is for the log of the index, since the value in the reference period *r* is 1 (whose log is zero). To get the level in period *t*, raise *e* to the power of the expression in the table.

Note $s_{it}\,(s_{ir})$: share of good *i* in the value of total expenditure at time *t* (*r*); budget shares in the case of a consumer price index, shares in GDP in the case of a GDP index.

Points to note:

1. The expressions $(p_{it}/p_{ir}),\ (q_{it}/q_{ir})$ are known as *price relatives* and *quantity relatives*, respectively.
2. The formulas for the Laspeyres and Paasche indices are not the ones usually given in textbooks. The usual formulas for these indices are:

$$Q_{tr}^{Lasp}=\frac{\sum_i p_{ir}q_{it}}{\sum_i p_{ir}q_{ir}}$$

$$P_{tr}^{Lasp}=\frac{\sum_i p_{it}q_{ir}}{\sum_i p_{ir}q_{ir}}$$

$$Q_{tr}^{Paas}=\frac{\sum_i p_{it}q_{it}}{\sum_i p_{it}q_{ir}}$$

$$P_{tr}^{Paas}=\frac{\sum_i p_{it}q_{it}}{\sum_i p_{ir}q_{it}}$$

The formulas in Table B.6.2 are algebraically equivalent to these (if quantities are non-zero), but correspond better to how such indices are actually calculated. In practice, statistical agencies start with price (or quantity) relatives and weight these together using budget shares (for the CPI) or output weights (e.g. base year shares in GDP for constant price GDP). The text book formulas have the advantage of showing that Laspeyres, Paasche and Fisher quantity indices can still be calculated when some of the quantities are zero.

3. Note that in the Paasche formulas in the table, the price and quantity relatives are the other way up to the way they are in the Laspeyres formula. Hence we take the inverse "to get them the right way up".
4. By definition, the Fisher quantity (price) index is the geometric mean of the Laspeyres and Paasche quantity (price) indices.
5. The product of the Fisher quantity index and the Fisher price index is the ratio of the values of expenditure in the two periods, i.e. it is the expenditure index:

$$\frac{\sum_{i=1}^{N} p_{it} q_{it}}{\sum_{i=1}^{N} p_{ir} q_{ir}} = P_{tr}^{Fish} \cdot Q_{tr}^{Fish} .$$

This is *not* true of any other pair of price and quantity indices of the same type in B.6.2. However:

6. The product of the *Laspeyres* quantity index and the *Paasche* price index is the expenditure index. And the product of the *Paasche* quantity index and the *Laspeyres* price index is also the expenditure index.
7. Using the result that $\ln(1+x) \approx x$ for small *x*, we obtain an approximation for the Törnqvist:

$$Q_{tr}^{Törn} \approx \sum_{i} \left[\frac{s_{ir} + s_{it}}{2} \right] \cdot \left(\frac{q_{it}}{q_{ir}} \right)$$

which may be compared with the formula in the table for the Laspeyres. This shows that, approximately, the Törnqvist is just like a Laspeyres except that it uses an arithmetic average of the shares in the two periods, instead of just the reference period shares.

Making PPPs transitive: the GEKS method

The GEKS method (the initials commemorate the four theorists who devised it: Gini, Éltetö, Köves and Szulc) is the one used in much of the ICP including the OECD-Eurostat group. It is one way of making a set of bilateral indices transitive.

Let p_{ij} be the price level in country *j* relative to the price level in country *i*, i.e. the bilateral PPP between *i* and *j* calculated using the expenditure shares of just these two countries. These bilateral price indices are assumed to be symmetric so that $p_{ij} = 1/p_{ji}$ and $p_{ii} = 1$; examples of price indices with these properties are the bilateral Fisher and the bilateral Törnqvist. Then with C ($C \geq 3$) countries the GEKS multilateral price index (PPP) for country *j* relative to country *i* is defined as

$$PPP_{ij}^{GEKS} = \left[\prod_{k=1}^{k=C} p_{ik} p_{kj} \right]^{1/C}$$

It is a geometric mean of all possible direct and indirect comparisons between these two countries.

BOX 6.3 The growth of aggregate labour productivity: a decomposition

This Box shows the relationship between the aggregate growth of labour productivity and growth in individual industries, using first the chained Laspeyres approach and then the chained Törnqvist approach.

The chained Laspeyres approach

To repeat equation (2), the chained Laspeyres index of aggregate output (GDP) in year *t* relative to year *t*-1 is

$$\frac{V_t}{V_{t-1}} = \sum_{i=1}^{N} v_{i,t-1} \left(\frac{V_{it}}{V_{it-1}} \right)$$

Here the $v_{i,t-1}$ are the shares of each industry's nominal value added in aggregate nominal value added (GDP at basic prices) in the previous year:

$$v_{i,t-1} = \frac{P^V_{i,t-1} V_{i,t-1}}{GDP_{t-1}} \qquad i = 1,\ldots,N, \;\; \sum_{i=1}^{N} v_{i,t-1} = 1$$

It is helpful now to define the labour shares, the share of each industry's labour input in aggregate labour input:

$$w_{it} = \frac{L_{it}}{L_t} \qquad i = 1,\ldots,N, \;\; \sum_{i=1}^{N} w_{it} = 1$$

The aggregate productivity index for year *t* relative to year $t-1$ is then

$$\frac{Z_t}{Z_{t-1}} = \frac{V_t / L_t}{V_{t-1} / L_{t-1}} = \left(\frac{V_t}{V_{t-1}}\right)\left(\frac{L_{t-1}}{L_t}\right) = \sum_{i=1}^{N} v_{i,t-1}\left(\frac{V_{it}}{V_{it-1}}\right)\left(\frac{L_{t-1}}{L_t}\right)$$

How is aggregate productivity related to productivity in the individual industries? From the last equation, and using the definition of the labour shares,

$$\frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} v_{i,t-1}\left(\frac{V_{it}}{V_{i,t-1}}\right)\left(\frac{L_{t-1}}{L_t}\right)$$

$$= \sum_{i=1}^{N} v_{i,t-1}\left(\frac{V_{it} / L_{it}}{V_{i,t-1} / L_{i,t-1}}\right)\left(\frac{L_{it} / L_t}{L_{i,t-1} / L_{t-1}}\right) \qquad \text{(multiplying and dividing by } L_{it} / L_{i,t-1}\text{)}$$

$$= \sum_{i=1}^{N} v_{i,t-1}\left(\frac{w_{it}}{w_{i,t-1}}\right)\left(\frac{Z_{it}}{Z_{i,t-1}}\right)$$

We see that aggregate productivity is a weighted *sum,* not a weighted *average*, of productivity in individual industries. This is the case because the weight for the i-th industry is $v_{i,t-1}(w_{it} / w_{i,t-1})$ and these weights do not necessarily sum to 1. In fact, if there are significant shifts in the allocation of labour, the aggregate productivity growth rate could conceivably lie outside the interval bounded by the fastest and the slowest industry growth rates. To take a simple numerical example to demonstrate this possibility, assume two industries and that $v_{1,t-1} = v_{2,t-1} = 1/2, \quad w_{1t} = 3/4, \quad w_{1t-1} = 1/4, \quad Z_{1t} / Z_{1,t-1} = 1.2$, and $Z_{2t} / Z_{2,t-1} = 1.0$. Then plugging these numbers into the formula, the aggregate productivity growth rate is $Z_t / Z_{t-1} = 1.97 > 1.2$, i.e. the aggregate productivity growth rate exceeds that of the fastest growing industry (industry 1 in this example).

Another way to write the last equation is

$$\frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} \left(\frac{v_{it-1}}{w_{it-1}} \right) w_{it} \left(\frac{Z_{it}}{Z_{i,t-1}} \right)$$

We can show that the first factor on the right hand side, $v_{i,t-1} / w_{i,t-1}$, measures productivity in the i-th industry in the base period *t*-1 relative to productivity in the economy as a whole at *t*-1, both measured in the prices of year *t*-1. This follows from the definitions of the shares:

$$\frac{v_{i,t-1}}{w_{i,t-1}} = \frac{P^V_{i,t-1} V_{i,t-1} / GDP_{t-1}}{L_{i,t-1} / L_{t-1}}$$

$$= \frac{P^V_{i,t-1} V_{i,t-1} / L_{i,t-1}}{GDP_{t-1} / L_{t-1}}$$

If this factor exceeds 1 then the industry has above average productivity in the base period. So we can see that a shift in the allocation of labour (a rise in w_{it}) towards an industry with a high *level* of labour productivity will (other things equal) raise the overall productivity growth rate. That is, if the i-th industry has high productivity then a rise in the industry's labour share w_{it} raises the aggregate productivity growth rate. Equally a shift in the allocation of labour (a rise in w_{it}) towards an industry with a high *growth rate* of productivity (high $Z_{it} / Z_{i,t-1}$) will also raise aggregate productivity growth. Of course these two factors could work against each other: the high productivity level industry might have a low growth rate. (In the numerical example above, industry 1 has a higher productivity level and also a faster productivity growth rate, so the two factors work together).

This suggests we should compare the *actual* aggregate growth rate of labour productivity

$$\frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} \left(\frac{v_{it-1}}{w_{it-1}} \right) w_{it} \left(\frac{Z_{it}}{Z_{i,t-1}} \right)$$

with two hypothetical possibilities. First, what would the growth rate have been if there had been no change in the allocation of labour, i.e. if $w_{it} = w_{i,t-1}$? The answer is

$$\text{Constant labour shares: } \frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} v_{it-1} \left(\frac{Z_{it}}{Z_{i,t-1}} \right)$$

The other possibility is if all industries had started off with an equal level of labour productivity, i.e. if $v_{i,t-1} = w_{i,t-1}$. Then the aggregate growth rate would have been

$$\text{Equal initial producivity levels: } \frac{Z_t}{Z_{t-1}} = \sum_{i=1}^{N} w_{it} \left(\frac{Z_{it}}{Z_{i,t-1}} \right)$$

For the constant labour shares case we weight the individual productivity growth rates by the initial value added shares. For the equal initial productivity levels case we weight by the second period labour shares.[19]

Finally, it is useful to define the *contribution* of each industry to overall growth as its weight multiplied by its productivity growth rate. So in the general case the contribution is

$$v_{it-1}(w_{it} / w_{it-1})(Z_{it} / Z_{i,t-1})$$

and in the two special cases it is

$$v_{it-1}(Z_{it} / Z_{i,t-1})$$

and

$$w_{it}(Z_{it} / Z_{i,t-1})$$

The Törnqvist decomposition

In continuous time, aggregate labour productivity growth can be related to labour productivity in individual industries as follows:

$$\begin{aligned}\hat{Z} &= \hat{V} - \hat{L} \\ &= \sum_{i=1}^{N} v_i \hat{V}_i - \sum_{i=1}^{N} w_i \hat{L}_i \\ &= \sum_{i=1}^{N} v_i (\hat{V}_i - \hat{L}_i) + \sum_{i=1}^{N} (v_i - w_i) \hat{L}_i \\ &= \sum_{i=1}^{N} v_i \hat{Z}_i + \sum_{i=1}^{N} (v_i - w_i) \hat{L}_i \\ &= \sum_{i=1}^{N} v_i \hat{Z}_i + \sum_{i=1}^{N} (v_i - w_i)(\hat{L}_i - \hat{L})\end{aligned} \tag{30}$$

The first two lines in this derivation are from the definitions of productivity and aggregate real value added. The third line comes from adding and subtracting $v_i \hat{L}_i$. The fourth line applies the definition of productivity again. The fifth line makes use of the fact that $\sum_{i=1}^{N} (v_i - w_i) \hat{L} = \hat{L} \sum_{i=1}^{N} (v_i - w_i) = 0$ since the shares sum to 1.

The shares in this formula are point-in-time and the growth rates are continuous. If we replace the point-in-time shares by averages over adjacent periods and the continuous growth

[19] The ONS uses a more complex decomposition due to Diewert (2015).

rates by log differences we get the chained Törnqvist decomposition of equation (8) in the text.

BOX 6.4 The Solow model and growth accounting

Solow's (1956) aggregate production function was not quite the same as the one in the text, equation (12). Solow's production function was:

$$V(t) = f(K(t), E(t)) \tag{31}$$

Here E is the number of *efficiency units* of labour:

$$E(t) = e^{\lambda t} L(t)$$

where λ is the rate of *labour-augmenting technical progress*. To convert this to growth rate terms take the derivative of both sides of the aggregate production function with respect to time t and divide both sides by V:

$$\frac{1}{V(t)}\frac{dV(t)}{dt} = \frac{f_K(t)}{V(t)}\frac{dK(t)}{dt} + \frac{f_L(t)}{V(t)}\frac{dL(t)}{dt} + \frac{f_L L\lambda}{V(t)}$$

(Note that $f_L = f_E e^{\lambda t}$.) Now multiply and divide the term involving capital change by K and the term involving labour change by L to get:

$$\frac{1}{V(t)}\frac{dV(t)}{dt} = \frac{f_K(t)K(t)}{V(t)}\left(\frac{1}{K(t)}\frac{dK(t)}{dt}\right) + \frac{f_L(t)L(t)}{V(t)}\left(\frac{1}{L(t)}\frac{dL(t)}{dt}\right) + \frac{f_L(t)L(t)\lambda}{V(t)}$$

Noting that the elasticities of V with respect to K and L are given by

$$\varepsilon_K \equiv \frac{\partial V}{\partial K}\frac{K}{V} = \frac{f_K K}{V}$$

and

$$\varepsilon_L \equiv \frac{\partial V}{\partial L}\frac{L}{V} = \frac{f_L L}{V}$$

and adopting the "hat" notation for growth rates we obtain

$$\hat{V}(t) = \varepsilon_L(t)\lambda + \varepsilon_K(t)\hat{K}(t) + \varepsilon_L(t)\hat{L}(t)$$

Rearranging,

$$\varepsilon_L(t)\lambda = \hat{V}(t) - \varepsilon_K(t)\hat{K}(t) - \varepsilon_L(t)\hat{L}(t) \tag{32}$$

which is equivalent to equation (16) of the text if we equate the first term on the right hand side of each to get

$$MFPG = \hat{A}(t) = \varepsilon_L(t)\lambda \qquad (33):$$

Growth accounting

In the text the foundation of the growth accounting method is taken to be an aggregate production function of the form

$$V(t) = A(t) f(K(t), L(t)) \qquad A(t) > 0 \qquad (34)$$

Here technical progress is taken to be neutral, i.e. it reduces the requirement for both capital and labour to produce a given level of output and does so in the same proportion. Totally differentiating this aggregate production function with respect to time, we obtain

$$\hat{V}(t) = \hat{A}(t) + \varepsilon_K \hat{K} + \varepsilon_L \hat{L}$$

and solving for $\hat{A}(t)$

$$\hat{A}(t) = \hat{V}(t) - \varepsilon_K \hat{K} - \varepsilon_L \hat{L}$$

MFP growth is now identified with the growth rate of *A*.

Solow's aggregate production function and this one are not the same except in the special case of a Cobb-Douglas production function:

$$\begin{aligned} V &= K^{\alpha}(e^{\lambda t}L)^{1-\alpha} \\ &= e^{\lambda(1-\alpha)t}K^{\alpha}L^{1-\alpha} \\ &= AK^{\alpha}L^{1-\alpha} \end{aligned}$$

putting $A \equiv e^{\lambda(1-\alpha)t}$. Hence

$$\hat{A} = (1-\alpha)\lambda \qquad (35)$$

In the Cobb-Douglas case we have

$$\varepsilon_L = \frac{\partial V}{\partial L}\frac{L}{V} = 1 - \alpha$$

So equation (35) is then the same as equation (33). In the Cobb-Douglas case the labour share is constant. This corresponds roughly to the facts. In the years since the second world war the labour share in the corporate sector in the advanced industrial countries, after excluding housing and taking proper account of income from self-employment, has been roughly constant, though varying cyclically. The exception is the United States where the share has been falling since around 2000. But the same is not true in other advanced economies including Britain (Gutiérrez and Piton 2019).[20] To the extent that the labour share is constant the Solow and the other production function can be considered equivalent.

Total costs and total revenues

Since constant returns to scale are assumed, we have from the aggregate production function

$$V(t) = \frac{\partial V(t)}{\partial K(t)} K(t) + \frac{\partial V(t)}{\partial L(t)} L(t)$$

by Euler's Theorem. Then from marginal productivity theory we have

$$V(t) = \frac{W(t)}{P^V(t)} L(t) + \frac{R(t)}{P^V(t)} K(t)$$

or

$$P^V(t)V(t) = W(t)L(t) + R(t)K(t)$$

i.e. revenue equals total costs.

BOX 6.5 The user cost of capital and rental prices

Consider the equation in the text for the user cost of capital:

$$R_t = r_t P_{t,0}^A + P_{t,0}^A - P_{t+1,1}^A$$

Now add and subtract $P_{t,1}^A$, which is the price today of a one year old asset, on the right hand side:

$$R_t = r_t P_{t,0}^A + (P_{t,0}^A - P_{t,1}^A) + (P_{t,1}^A - P_{t+1,1}^A)$$

The first bracketed term on the right hand side is depreciation (the difference today between the price of a new asset and the price of a one year old asset) and under the geometric assumption can be written as $\delta P_{t,0}^A$. The second bracketed term is

$$(P_{t,1}^A - P_{t+1,1}^A) = (1-\delta)(P_{t,0}^A - P_{t+1,0}^A) = -(1-\delta)\pi_t P_{t,0}^A$$

where π_t is the growth rate of the price of a new asset between periods *t* and *t*+1:

[20] The labour share in the whole economy (GDP), including housing and taking proper account of self-employment income, has been rising in Britain since the mid-1990s (Sidhu and Dunn 2018).

$$\pi_t \equiv \frac{P^A_{t+1,0} - P^A_{t,0}}{P^A_{t,0}}$$

So the user cost of capital is

$$\begin{aligned} R_t &= r_t P^A_{t,0} + P^A_{t,0} - P^A_{t+1,1} \\ &= r_t P^A_{t,0} + \delta P^A_{t,0} - (1-\delta)\pi_t P^A_{t,0} \\ &= \left[r_t + \delta(1+\pi_t) - \pi_t \right] P^A_{t,0} \end{aligned}$$

after factoring out $P^A_{t,0}$.

The firm will decide to invest if the expected return is greater than the user cost. The expected return is the value of the marginal product of this type of capital. Under competition the marginal product will be driven into equality with the real user cost.

BOX 6.6 MFP at the industry level

Reallocation effects

In the text it was stated that the top-down and bottom up approaches to MFP growth produced the same answer provided that a given input is paid the same whatever the industry where it is employed. This was first proved in full generality by Jorgenson et al. (1987), chapter 2, Proving the point with many different types of labour and capital involves a lot of algebra so here I take a simpler approach and assume that there is only one type of labour and one type of capital. So assume that labour is paid a wage W_i and capital earns a return R_i in the i-th industry.

The top-down approach says that aggregate MFP growth is

$$MFPG = \hat{V} - s_K \hat{K} - s_L \hat{L}$$

MFP growth at the industry level is

$$MFPG_i^{VA} = \hat{V}_i - s_{iK} \hat{K}_i - s_{iL} \hat{L}_i \qquad i = 1, \ldots, N$$

and the bottom-up approach measures aggregate MFP growth by

$$MFPG = \sum_{i=1}^{N} v_i MFPG_i^{VA} = \sum_{i=1}^{N} v_i \hat{V}_i - \sum_{i=1}^{N} v_i s_{iK} \hat{K}_i - \sum_{i=1}^{N} v_i s_{iL} \hat{L}_i$$

The top-down and bottom-up measures of the contribution of value added obviously match since $\hat{V} = \sum_{i=1}^{N} v_i \hat{V}_i$ So consider the contribution of labour from the bottom-up approach:

$$\sum_{i=1}^{N} v_i s_{iL} \hat{L}_i = \sum_{i=1}^{N} \left(\frac{P_1^V V_i}{P^V V} \right) \left(\frac{W_i L_i}{P_1^V V} \right) \hat{L}_{ii} = \sum_{i=1}^{N} \left(\frac{W_i L_i}{P_1^V V} \right) \hat{L}_i$$

Now if the wage rate is the same in all industries, i.e. $W_i = W$ all *i*, and remembering that $L = \sum_{i=1}^{N} L_i$, the contribution of labour is

$$\sum_{i=1}^{N} \left(\frac{W_i L_i}{P_1^V V} \right) \hat{L}_i = \frac{W}{P_1^V V} \sum_{i=1}^{N} L_i \hat{L}_i = \frac{W}{P_1^V V} \sum_{i=1}^{N} \dot{L}_i = \frac{W}{P_1^V V} \dot{L}$$
$$= \left(\frac{WL}{P_1^V V} \right) \hat{L}$$

which is the same as the contribution of labour from the top-down approach. A parallel argument holds for the contribution of capital. Therefore if the wage rate is the same in all industries and the rental price is also the same in all industries, then the top-down and bottom-up approaches yield exactly the same answer.

BOX 6.7 The value added and gross output concepts of MFP growth

Start with some basic accounting relationships in an *N*-industry economy which uses *C* types of capital and *D* types of labour. For each industry the value of output equals payments for inputs (including profit):

$$GO_i := P_i Y_i = \sum_{l=1}^{D} P_{lL} L_{il} + \sum_{k=1}^{C} P_{kK} K_{ik} + \sum_{j=1}^{N} P_j M_{ij} \qquad (36)$$

Here GO_i is nominal gross output of the *i*-th industry, Y_i is real output, P_i is its price, K_{ik} is the quantity of the *k*-th type of capital used in industry *i*, L_{il} is the quantity of the *l*-th type of labour, M_{ij} is the quantity of the *j*-th type of intermediate input (we could easily extend this to include imported inputs), and P_{kK}, P_{lL}, P_{jM} are the corresponding prices of capital, labour and intermediate input respectively. Note that we are assuming that a given input is sold at a common price in all industries; if any intermediate inputs are imported then they are sold at the same price as their domestic counterparts. Nominal value added is defined as

$$VA_i \equiv P_{iV} V_i \equiv P_i Y_i - \sum_{j=1}^{N} P_j M_{ij} \quad (37)$$

where V_i is real value added and P_{iV} is the price of value added. These last two concepts are not directly observable but they become so by totally differentiating both sides of (37) with respect to time and collecting terms in prices and quantities:

$$\hat{P}_{iV} \equiv \frac{GO_i}{VA_i}\left[\hat{P}_i - \sum_{j=1}^{N} m_{ij}\hat{P}_j\right] \quad (38)$$

and

$$\hat{V}_i \equiv \frac{GO_i}{VA_i}\left[\hat{Y}_i - \sum_{j=1}^{N} m_{ij}\hat{M}_{ij}\right] \quad (39)$$

Here "hats" denote growth rates, e.g. $\hat{V}_i = d\ln V_i / dt$. As defined in the main text, m_{ij} is the share of intermediate input *j* in the total costs of industry *i*. Equation (39) defines double deflated real value added in continuous time and (38) defines the corresponding price index.

At the industry level there are two possible concepts of MFP growth, the value added one and the gross output one. The gross output concept is based on the existence of an industry production function:

$$Y_i = f_i(K_{i1,} ..., K_{iC}; L_{i1}, ..., L_{iD}; M_{i1}, ..., M_{iN}; t) \quad (40)$$

Now define μ_i^{GO} as the gross output concept of MFP growth in the *i*-th industry

$$\mu_i^{GO} \equiv \frac{\partial \ln Y_i}{\partial t} \quad (41)$$

We can readily find that

$$\mu_i^{GO} = \hat{Y}_i - \sum_{k=1}^{C} \alpha_{ik}\hat{K}_{ik} - \sum_{l=1}^{D} \beta_{il}\hat{L}_{il} - \sum_{j=1}^{N} m_{ij}\hat{M}_{ij} \quad (42)$$

where $\alpha_{ik}, \beta_{il}, m_{ij}$ are the elasticities of output with respect to the capital, labour and intermediate inputs respectively. Assuming competitive conditions these elasticities can be equated to the share of each input in the value of gross output (the cost shares) so that:

$$\beta_{il} = \frac{P_{lL} L_{il}}{P_i Y_i}, \; l = 1,...,D$$

$$\alpha_{ik} = \frac{P_{kK} K_{ik}}{P_i Y_i}, \; k = 1,...,C \qquad (43)$$

$$m_{ij} = \frac{P_j M_{ij}}{P_i Y_i}, \; j = 1,...,N$$

The cost shares sum to 1.

The value added concept of MFP growth is defined by:

$$\mu_i^{VA} \equiv \hat{V}_i - \sum_{k=1}^{C} \alpha_{ik}^{VA} \hat{K}_{ik} - \sum_{l=1}^{D} \beta_{il}^{VA} \hat{L}_{il} \qquad (44)$$

where α_{ik}^{VA}, β_{il}^{VA} are the shares of the capital and labour inputs in value added:

$$\beta_{il}^{VA} = \frac{P_{lL} L_{il}}{P_{iV} V_i} = \left[\frac{GO_i}{VA_i}\right] \beta_{il}, \; l = 1,...,D$$

$$\alpha_{ik}^{VA} = \frac{P_{kK} K_{ik}}{P_{iV} V_i} = \left[\frac{GO_i}{VA_i}\right] \alpha_{ik}, \; k = 1,...,C \qquad (45)$$

making use of (43). Now substitute (39) and (45) into (44) and use (42) to get

$$\mu_i^{VA} = \left[\frac{GO_i}{VA_i}\right] \mu_i^{GO} \qquad (46)$$

Hence $\mu_i^{VA} \geq= \mu_i^{GO}$ with equality if and only if $GO_i = VA_i$.

Note that we have given a theoretical justification for the gross output concept of MFP growth by invoking the industry production function. We have given no such justification for the value added concept. It is possible to base the value added measure more directly on theory by assuming the existence of a value added function:

$$V_i = g_i(K_{i1,},...,K_{iC}; L_{i1},...,L_{iD}; t) \qquad (47)$$

and by assuming that the gross output production function is separable in value added and intermediate input:

$$Y_i = f_i(V_i, M_{i1},...,M_{iN})$$

But this would be a very restrictive assumption since it says that technical progress can never reduce the requirement for intermediate inputs per unit of gross output. Note that it is always possible to *calculate* the value added measure by using either the direct formula, equation (44), or indirectly from the gross output measure, equation (42), even if this restrictive assumption does not hold. But then the interpretation of the measure becomes problematic.

At the aggregate level nominal GDP is the sum of value added in all industries:

$$GDP = \sum_{i=1}^{N} P_{iV} V_i \tag{48}$$

The growth rate of real GDP (*V*) is, using the continuous time approach (Divisia) approach,

$$\hat{V} = \sum_{i=1}^{N} v_i V_i, \quad v_i \equiv \frac{P_{iV} V_i}{GDP} \tag{49}$$

where the v_i are the value added shares of each industry in GDP.

Now define the aggregate growth rate of MFP (μ) as

$$\mu \equiv \hat{V} - \alpha \hat{K} - (1-\alpha)\hat{L} \tag{50}$$

where *K* is aggregate capital services, *L* is aggregate labour input, α is the capital (profit) share. Under the competitive assumptions made here μ can be shown to measure the rate at which the social production possibility frontier is shifting outwards as a result of technological progress (Hulten 1978; Gabaix 2011, Appendix B). In turn aggregate capital and aggregate labour of each type can be found from summing over the industries:

$$\begin{aligned} K_k &= \sum_{i=1}^{N} K_{ik}, \quad k = 1, \ldots, C \\ L_l &= \sum_{i=1}^{N} L_{il}, \quad l = 1, \ldots, D \end{aligned} \tag{51}$$

Continuous-time (Divisia) indices of aggregate capital and aggregate labour are then

$$\begin{aligned} \hat{K} &= \sum_{k=1}^{C} \left[\frac{P_{kK} K_k}{\sum_{k=1}^{C} P_{kK} K_k} \right] \hat{K}_k \\ \hat{L} &= \sum_{l=1}^{D} \left[\frac{P_{lL} L_l}{\sum_{l=1}^{D} P_{lL} L_l} \right] \hat{L}_l \end{aligned} \tag{52}$$

These last two equations embody the assumption that a given capital or labour input earns the same return in any industry. Finally, the aggregate capital share is

$$\alpha = \frac{\sum_{k=1}^{C} P_{kK} K_k}{GDP} \tag{53}$$

and the labour share is $1-\alpha$.

Equation (50) is the top-down approach to measuring aggregate MFP. The latter can also be measured by aggregating over industry-level MFP growth rates, the bottom-up approach. Straightforward algebra shows that the aggregate MFP growth rate as defined by (50) is identically equal to the following aggregation scheme.

$$\mu = \sum_{i=1}^{N} d_i \mu_i^{GO} \tag{54}$$

Here d_i is the Domar (1961) weight for the *i*-th industry, defined as

$$d_i \equiv \left[\frac{GO_i}{GDP} \right] \tag{55}$$

And using (46) we also have an alternative aggregation scheme based on the value added measure:

$$\mu = \sum_{i=1}^{N} v_i \mu_i^{VA} \tag{56}$$

For these equivalences to hold we just need to assume that a given input earns the same return wherever it is employed. If this is not the case then the aggregate formulas become more complex with additional terms reflecting the shift of resources to or from industries where they are more highly valued (see Jorgenson et al. (1987), chapter 2, page 66).

BOX 6.8 Measuring MFP under imperfect competition

Section 7 above set out the basic framework for measuring MFP under the assumption of perfect competition. Many economists find this assumption unrealistic preferring instead to assume some form of imperfect competition. In this Box I examine how imperfect competition could be incorporated into the measurement framework.

Let us start as before with the aggregate production function of equation (12):

$$V(t) = A(t) f(K(t), L(t)) \qquad A > 0$$

which as we saw can be transformed into growth rate terms as

$$\hat{V}(t) = \hat{A}(t) + \varepsilon_K(t)\hat{K}(t) + \varepsilon_L(t)\hat{L}(t)$$

Recall that ε_L is defined as the elasticity of output with respect to labour:

$$\varepsilon_L \equiv \frac{\partial V}{\partial L}\frac{L}{V}$$

and ε_K as the elasticity of output with respect to capital:

$$\varepsilon_K \equiv \frac{\partial V}{\partial K}\frac{K}{V}$$

We saw that under perfect competition these elasticities can be identified empirically with the shares of capital and labour in the value of output. This is no longer the case under imperfect competition. Suppose from now on that the typical firm has some degree of monopoly power in the product market but no such power in input markets. That is, it faces a downward-sloping demand curve for its output but in hiring labour and capital it must accept the going rates. Let us assume that firms seek to minimise costs whatever the level of output they choose to produce. This implies that the firm hires labour and capital up to the point where the marginal cost of additional output is the same whether it is achieved through expanding labour or expanding capital. The addition to total cost from employing an extra unit of labour is the wage rate *W*. This additional unit of labour produces additional output equal to the marginal product of labour, $\frac{\partial V}{\partial L}$. So the marginal cost (*MC*) of an additional unit of output produced by labour is

$$MC = \frac{W}{\partial V / \partial L}$$

By a parallel argument the marginal cost of producing an extra unit of output by expanding capital is

$$MC = \frac{R}{\partial V / \partial K}$$

where *R* is the rental price of capital. Cost minimisation requires that these two marginal costs be the same: if this were not the case the firm could reduce the total cost of producing a given level of output by reallocating its expenditure towards the input with the lower marginal cost. Hence solving for the two input prices we have

$$W = MC \frac{\partial V}{\partial L}$$

and

$$R = MC \frac{\partial V}{\partial K}$$

Take the equation for the wage rate first. Multiply both sides by $L / P^V V$ to get

$$\frac{WL}{P^V V} = \frac{MC}{P^V} \frac{\partial V}{\partial L} \frac{L}{V} = \frac{MC}{P^V} \varepsilon_L$$

using the definition of the labour elasticity. Solving for the labour elasticity,

$$\varepsilon_L = \left(\frac{P^V}{MC} \right) \left(\frac{WL}{P^V V} \right)$$

So in the presence of market power where price exceeds marginal cost ($P^V > MC$) the elasticity of output with respect to labour is greater than labour's share. The ratio P^V / MC is called the *markup*.

By a parallel argument

$$\varepsilon_K = \left(\frac{P^V}{MC} \right) \left(\frac{RK}{P^V V} \right)$$

This might make it seem that capital's elasticity is also greater than capital's share under imperfect competition. But this is not necessarily the case. *RK* is what the firm has to pay to hire the capital it wants but this is no longer equal to gross operating surplus because the firm now earns a monopoly profit equal to the difference between total revenue and total cost, $P^V V - WL - RK$.

Suppose for a moment that there are constant returns to scale which implies that $\varepsilon_K + \varepsilon_L = 1$. Now since under imperfect competition the labour elasticity is higher than the labour share, then the capital elasticity must be lower than the profit share (inclusive of monopoly profit). But if imperfect competition is really important empirically then we must question the assumption of constant returns. Fixed costs seem ubiquitous though at varying levels across industries and these are certainly one important cause of increasing returns. If a firm is to survive it must be able to recover its fixed costs and this requires a price in excess of marginal cost. So either the firm has to be granted a legal monopoly or its product is differentiated in some way from rivals, so it faces a downward-sloping demand curve.

From the point of view of measuring MFP, assuming imperfect rather than perfect competition requires us to estimate markups in every industry. One way to do this is through a regression approach first suggested by Hall (1988). Approximate the aggregate production function, equation (13), in discrete terms as

$$\begin{aligned}\Delta \ln V &= \varepsilon_K \Delta \ln K + \varepsilon_L \Delta \ln L + \Delta \ln A \\ &= m\left(\frac{RK}{P^V V}\right)\Delta \ln K + m\left(\frac{WL}{P^V V}\right)\Delta \ln L + \Delta \ln A \\ &= m\Delta X + \Delta \ln A\end{aligned} \tag{57}$$

where m is the markup (P^V / MC) and X is a measure of aggregate input which weights each input by its cost as a proportion of total revenue:

$$\Delta X \equiv \left(\frac{RK}{P^V V}\right)\Delta \ln K + \left(\frac{WL}{P^V V}\right)\Delta \ln L$$

Here we have to modify our previous approach to estimating R. The rate of return part of the rental price is now the required, not the actual, rate of return. One possibility is to use the corporate bond rate.

Now run equation (57) as a time series regression. The term $\Delta \ln A$ will show up as the constant plus a random error with mean zero. The constant can be interpreted as the mean rate of MFP growth over the period. Alternatively, we can calculate MFP growth as the growth of output minus the growth of aggregate input weighted by the estimated markup:

$$\Delta \ln \tilde{A} = \Delta \ln V - \tilde{m}\Delta \ln X$$

where a tilde (~) denotes an econometric estimate.

Unfortunately this approach faces many difficulties. First, it seems likely that markups vary across industries so it is necessary to run one regression for each industry. But not every country has good quality data at the industry level extending back for (say) 30 years, the minimum necessary for reasonably reliable econometric estimates. Second, there are econometric difficulties in estimating equation (57). A positive shock to MFP likely induces firms to expand and purchase more inputs. But this means that the error term in the regression equation is correlated with the right hand side variable so an instrumental variable approach is needed. But finding an appropriate instrument is not easy. Third, if the data is annual, then fitting the model in growth rate terms may be picking up mainly cyclical variation in utilisation which is itself poorly measured (see Section 7 above). Fourth, many economists believe that markups have been increasing in recent years. Allowing for this possibility is even more demanding of data.

Britain is certainly amongst the countries that currently lack the long runs of industry-level data needed to implement this approach. If only for this reason we are not likely to see the ONS employing it in the near future. See Basu (2019) for a review of the various methods that have been used to estimate markups.

References

Andrews, Dan, Chiara Criscuolo, and Peter N. Gal (2015). "Frontier firms, technology diffusion, and public policy: micro evidence from OECD countries". OECD Productivity Working Papers No. 02, OECD Publishing.

Atkinson, Anthony B., Joe Hasell, Salvatore Morelli and Max Moser (2017). "The chartbook of economic inequality". www.chartbookofeconomicinequality.com.

Bank of England (2017). "A millennium of macroeconomic data". https://www.bankofengland.co.uk/-/media/boe/files/statistics/research-datasets/a-millennium-of-macroeconomic-data-for-the-uk.xlsx.

Basu, Susanto (2019). "Are price-cost markups rising in the United States? A discussion of the evidence". *Journal of Economic Perspectives*, vol. 33 (3), Summer, pages 3-22. https://www.aeaweb.org/articles?id=10.1257/jep.33.3.3

Bean, Charles (2016). *Independent Review of Economic Statistics*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf.

Cette, Gilbert, John G. Fernald, and Benoit Mojon (2016). "The pre-Great Recession slowdown in productivity". *European Economic Review*, April, pp. 3-20.

Chadha, Jagjit S, Ana Rincon-Aznar, Sylaja Srinivasan and Ryland Thomas (2019). "A century of high frequency UK macroeconomics statistics: a data inventory". ESCoE Technical Report 03, April 2019. https://www.escoe.ac.uk/download/3643.

Corrado, C. A., Hulten, C. R. and Sichel, D. E. (2005). "Measuring Capital and Technology: An Expanded Framework". In *Measuring Capital in the New Economy*, Vol. 65 (Eds, Corrado, C.A., Haltiwanger, J.C. and Sichel, D.E.). Chicago: The University of Chicago Press.

Diewert, W. Erwin (2015). "Decompositions of productivity growth into sectoral effects". *Journal of Productivity Analysis*, vol. 43(3), pp. 367-387.

European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, World Bank (2009). *System of National Accounts 2008.* New York. http://documents.worldbank.org/curated/en/417501468164641001/pdf/574870PUB0Box353749B01PUBLIC10SNA2008.pdf

Eurostat (2013). *European System of Accounts: ESA 2010.* https://ec.europa.eu/eurostat/documents/3859598/5925693/KS-02-13-269-EN.PDF/44cd9d01-bc64-40e5-bd40-d17df0c69334

Goodridge, P., Haskel, J. and Wallis, G. (2018), "Accounting for the UK productivity puzzle: a decomposition and predictions." *Economica*, vol. 85 (July), No. 339, pp. 581-605.

Griliches, Zvi and Dale W. Jorgenson (1967). 'The explanation of productivity change', *Review of Economic Studies*, 34, 249-283.

Gutiérrez, Germán and Sophie Piton (2019). "Revisiting the global decline of the (non-housing) labor share". Centre for Macroeconomics Discussion Paper DP2019-13. http://www.centreformacroeconomics.ac.uk/Discussion-Papers/2019/CFMDP2019-13-Paper.pdf.

Hall, Robert E. (1988). "The relation between price and marginal cost in U.S. industry". *Journal of Political Economy*, vol. 96(5), pp. 921-947.

Hall, Robert E. and Jones, Charles I. (1999). "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics*, 114, 83-116.

Hulten, C. (1978). "Growth accounting with intermediate inputs". *Review of Economic Studies*, vol. 45(3), pp. 511-518.

Hulten, Charles (2001). "Total factor productivity: a short biography". in *New Directions in Productivity Analysis*, Charles R. Hulten, Edwin R. Dean, and Michael J. Harper, eds., Studies in Income and Wealth, The University of Chicago Press for the National Bureau of Economic Research, Chicago.

Jorgenson, Dale. W., Frank M. Gollop, and Barbara M. Fraumeni. (1987). *Productivity and U.S. Economic Growth*. Cambridge, MA: Harvard University Press.

Krugman, Paul (1994). *The Age of Diminished Expectations: U.S. Economic Policy in the 1990s*. Cambridge: The MIT Press.

Lee, Pete (2011). *UK National Accounts: A Short Guide.* Office for National Statistics. [http://www.ons.gov.uk/ons/rel/naa1-rd/national-accounts-concepts--sources-and-methods/august-2011/index.html]

Milanovic, Branko, Peter H. Lindert, and Jeffrey G. Williamson (2011). "Pre-industrial inequality". *Economic Journal*, 121 (551), pp. 255-272.

OECD (2001). *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth*. Paris: OECD. www.oecd.org/sdd//productivity-stats/2352458.pdf

OECD (2009). *Measuring Capital*. www.oecd.org/sdd/productivity-stats/43734711.pdf.

OECD (2018). "International productivity gaps: Are labour input measures comparable?" SDD Working Paper No. 99. Paris: OECD. www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=SDD/DOC(2018)12&docLanguage=En

ONS (2017). "Quality adjusted labour input: UK estimates to 2016". https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/articles/qualityadjustedlabourinput/ukestimatesto2016

ONS (2018a). *International Comparisons of UK Productivity (ICP), Final Estimates: 2016.* https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/bulletins/internationalcomparisonsofproductivityfinalestimates/2016

ONS (2018b). "Volume index of UK capital services (experimental): estimates to Quarter 2 (Apr to Jun) 2017". https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/volumeindexofukcapitalservicesexperimental/estimatestoquarter2aprtojune2017

ONS (2018c). "Estimating the impact of the self-employed in the labour share". https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/methodologies/estimatingtheimpactoftheselfemployedinthelabourshare

ONS (2019a). "Improving estimates of labour productivity and international comparisons". https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/articles/improvingestimatesoflabourproductivityandinternationalcomparisons/2019-01-09

ONS (2019b). "Multi-factor productivity estimates: Experimental estimates to October to December 2018". https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/articles/multifactorproductivityestimates/experimentalestimatestooctobertodecember2018#toc

ONS (2019c). "Capital services estimates (experimental)". https://www.ons.gov.uk/economy/economicoutputandproductivity/output/datasets/capitalservicesestimates

ONS (2019d). *Labour productivity, UK: October to December 2018.* https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/bulletins/labourproductivity/octobertodecember2018

ONS (2019e). "Public service productivity: total, UK, 2016" https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimatestotalpublicservices/2016

ONS (2019f). "Regional and sub-regional productivity in the UK: February 2019." https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/articles/regionalandsubregionalproductivityintheuk/february2019

ONS (2019g). "Understanding spatial labour productivity in the UK". https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/articles/understandingspatiallabourproductivityintheuk/2019-05-03

ONS (2019h). "Regional and subregional productivity comparisons, UK and selected EU countries: 2014". https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/compendium/economicreview/april2018/regionalandsubregionalproductivitycomparisonsukandselectedeucountries2014

ONS (2019i). "National Accounts articles: Changes to the capital stock estimation methods for Blue Book 2019". https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/articles/nationalaccountsarticles/changestothecapitalstockestimationmethodsforbluebook2019

ONS (no date). *UK SIC 2007*. https://www.ons.gov.uk/methodology/classificationsandstandards/ukstandardindustrialclassificationofeconomicactivities/uksic2007

Oulton, Nicholas (1997). "The ABI Respondents Database: a new resource for industrial economics research". *Economic Trends*, No. 528, November 1997.

Oulton, Nicholas, Ana Rincon-Aznar, Lea Samek and Sylaja Srinivasan (2018). "Double deflation: theory and practice". Economic Statistics Centre of Excellence, ESCoE DP 2018-17. https://www.escoe.ac.uk/wp-content/uploads/2018/12/ESCoE-DP-2018-17.pdf

Oulton, Nicholas, and Gavin Wallis (2016). "Capital stocks and capital services: integrated and consistent estimates for the United Kingdom, 1950-2013". *Economic Modelling*, vol. 54 (April), pages 117-125.

Prothero, Richard (2018). "Mind the gap: why the UK might not be the most regionally unequal country". https://blog.ons.gov.uk/2018/11/23/mind-the-gap-why-the-uk-might-not-be-the-most-regionally-unequal-country/.

Sidhu, Sunny and Marianthi Dunn (2018). "Estimating the impact of the self-employed in the labour share". https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/methodologies/estimatingtheimpactoftheselfemployedinthelabourshare.

Solow, Robert M. (1956). "A contribution to the theory of economic growth". *Quarterly Journal of Economics*, vol. 70(1), pp. 65-94.

Solow, Robert M. (1957) "Technical change and the aggregate production function". *Review of Economics and Statistics*, vol. 39(3), pp. 312-320.

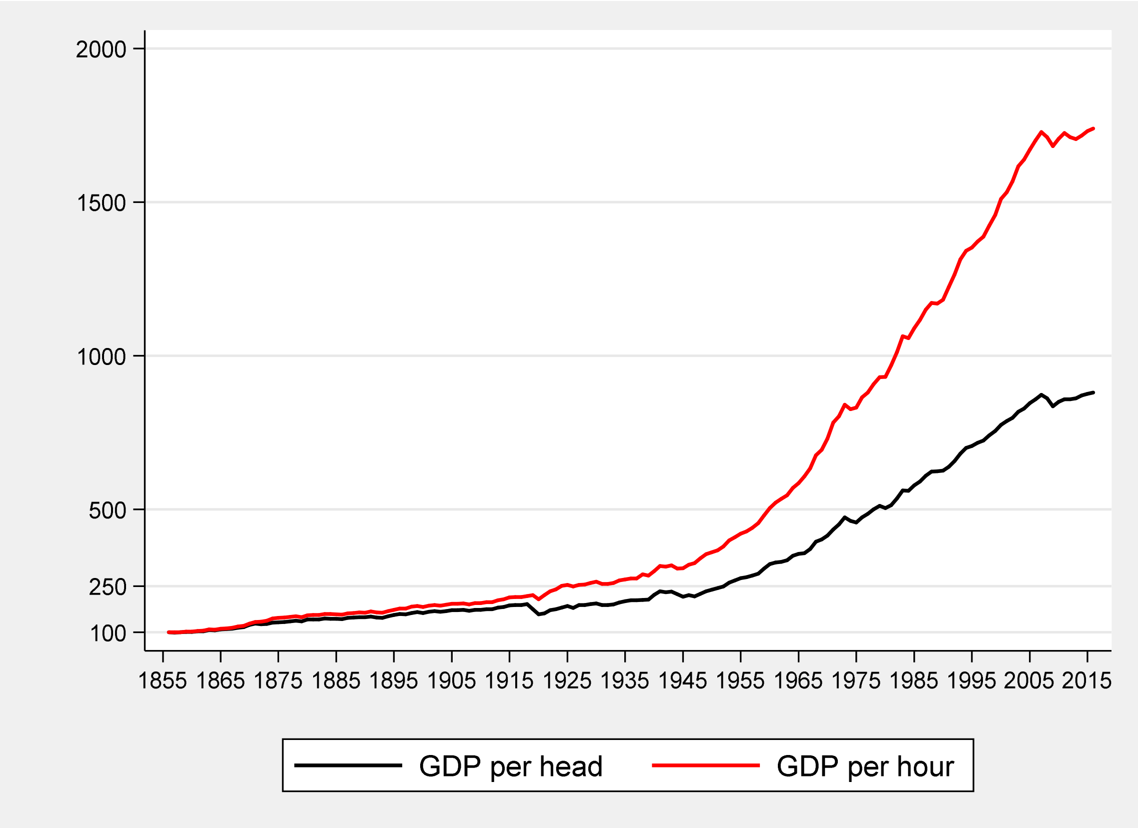
Syverson, Chad (2011). "What determines productivity?" *Journal of Economic Literature*, 49(2), pp. 326-65.

World Bank (2005). *ICP 2003-2006 Handbook*. http://go.worldbank.org/ MW520NNFK0.

World Bank Group (2015). *Purchasing Power Parities and the Real Size of World Economies: A Comprehensive Report of the 2011 International Comparison Program.* http://issuu.com/world.bank.publications/docs/9781464803291. The summary tables used here are at http://pubdocs.worldbank.org/en/108821487172249299/2011-International-Comparison-Program-results.xlsx.

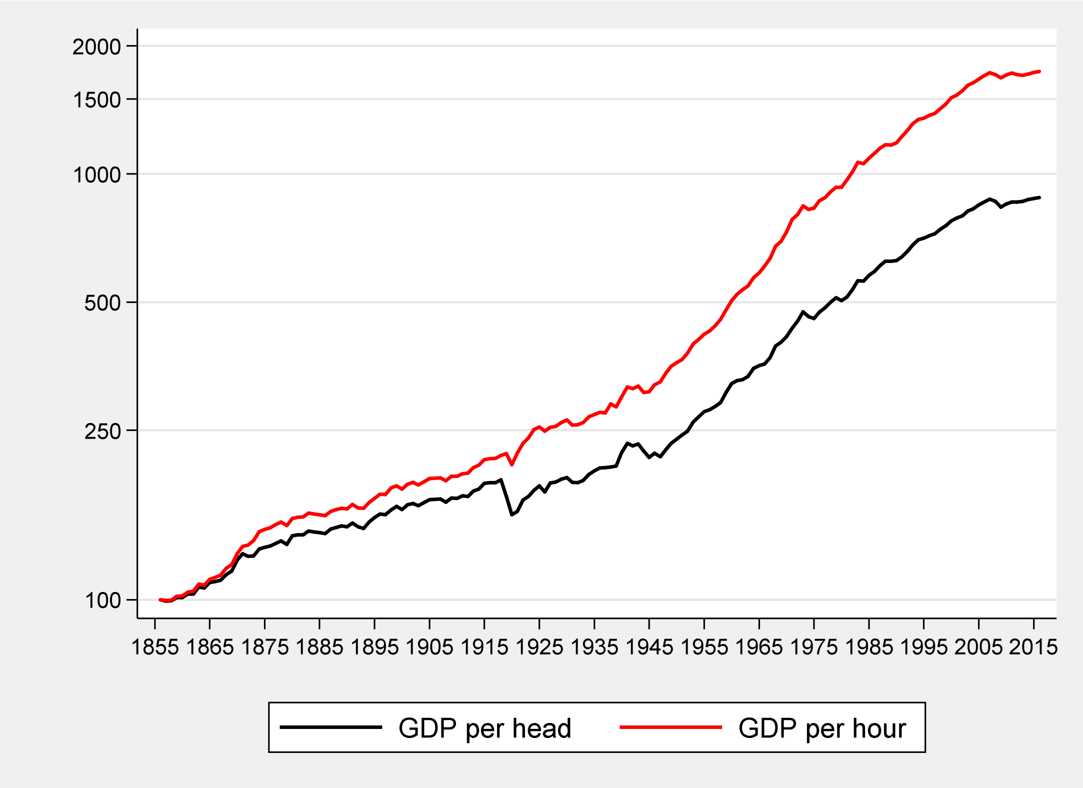
CHARTS

Chart 1 IGNORE THIS CHART!
GDP per hour and GDP per head in Britain, 1856-2016, 1856 = 100



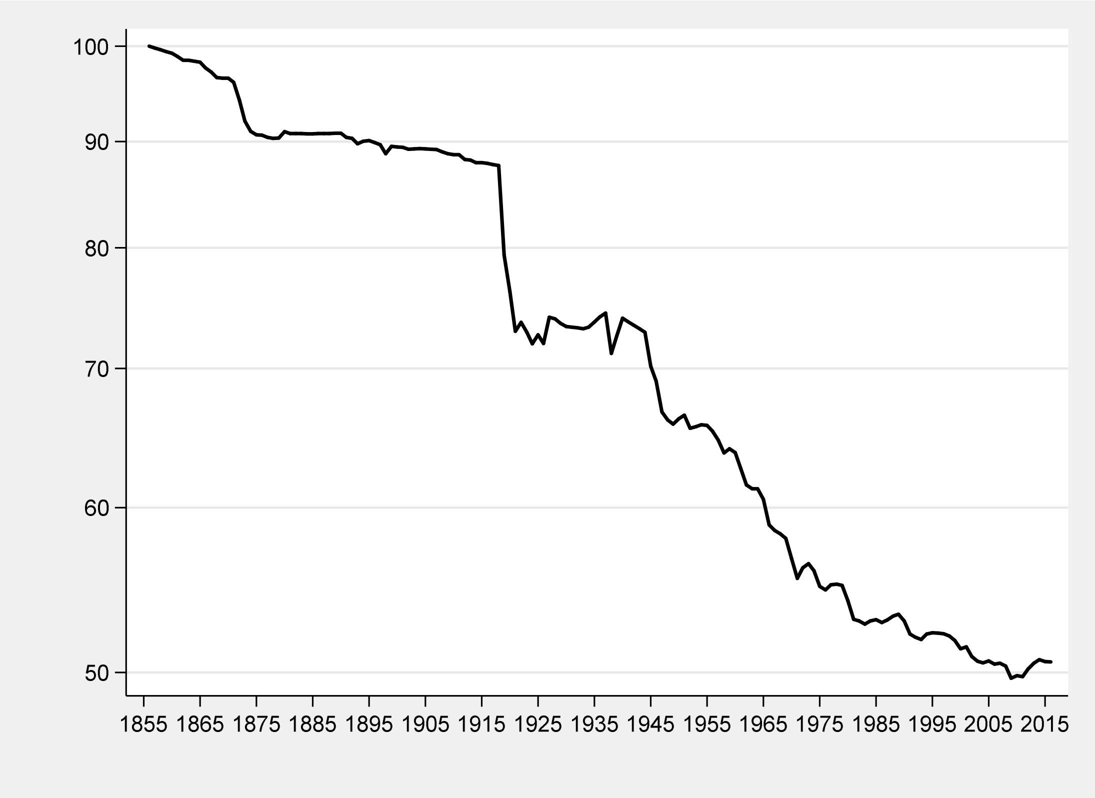
Source Bank of England (2017). Methods and sources explained in Chadha et al. (2019). Underlying data in spreadsheet for this paper.

Chart 2
GDP per hour and GDP per head in Britain, 1856-2016, 1856 = 100, LOG SCALE



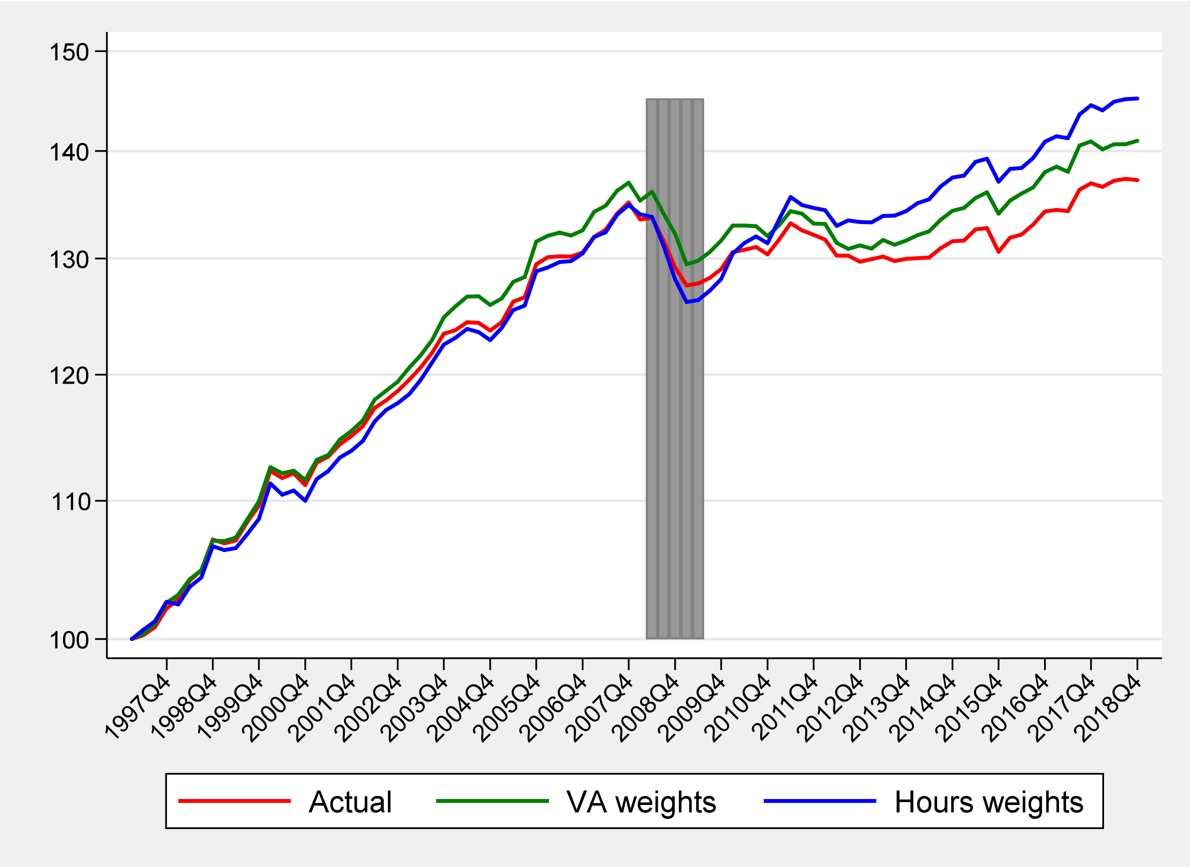
Source Bank of England (2017). Methods and sources explained in Chadha et al. (2019). Underlying data in spreadsheet for this chapter.

Chart 3
Hours worked per head, 1856-2016, 1856=100, log scale



Source Bank of England (2017). Methods and sources explained in Chadha et al. (2019). Hours worked per head calculated as GDP per head divided by GDP per hour. Underlying data in spreadsheet for this chapter. The very sharp fall just after the first world war was due to legislation shortening the working week.

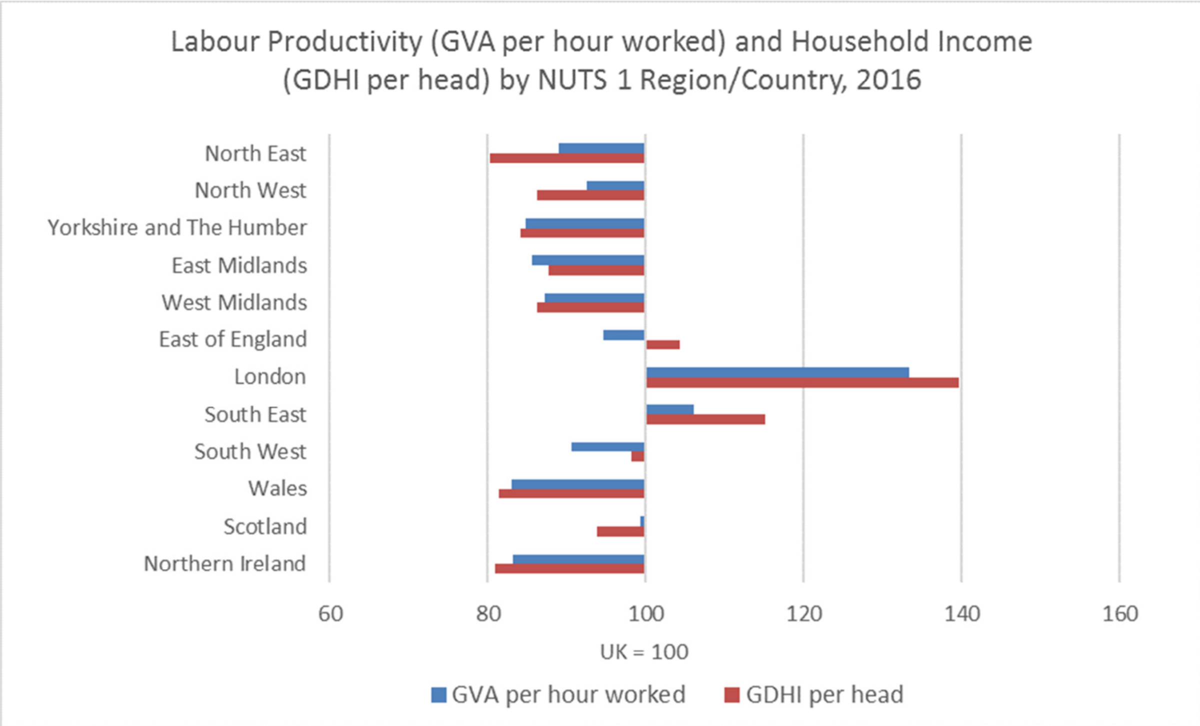
Chart 4
Labour productivity in the market sector, 1997Q1-2018Q4
1997Q1=100, log scale



Source ONS spreadsheet "prodconts.xls" containing data underlying the *Labour Productivity Bulletin* (ONS 2019d), available at https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/datasets/annualbreakdownofcontributionswholeeconomyandsectors

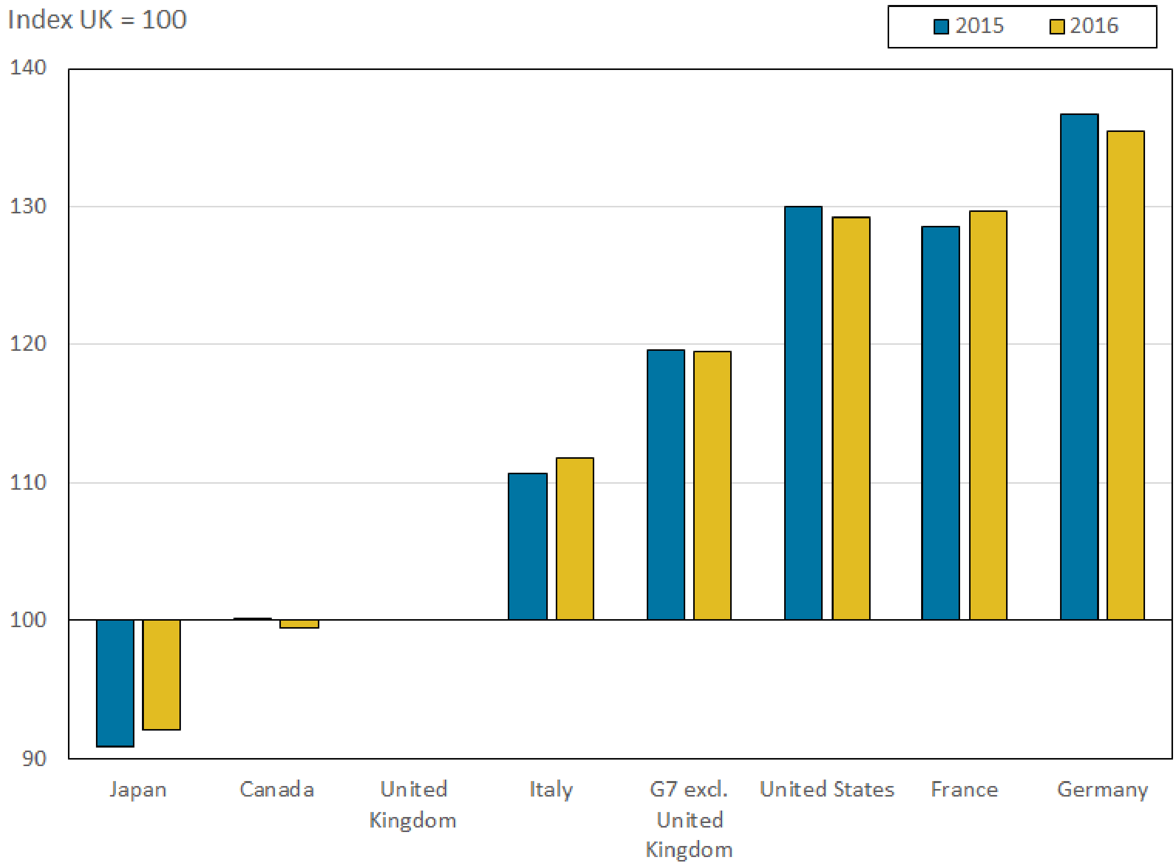
Note Grey bar marks Great Recession. Market sector defined as whole economy less sections L, O, P and Q.

Chart 5



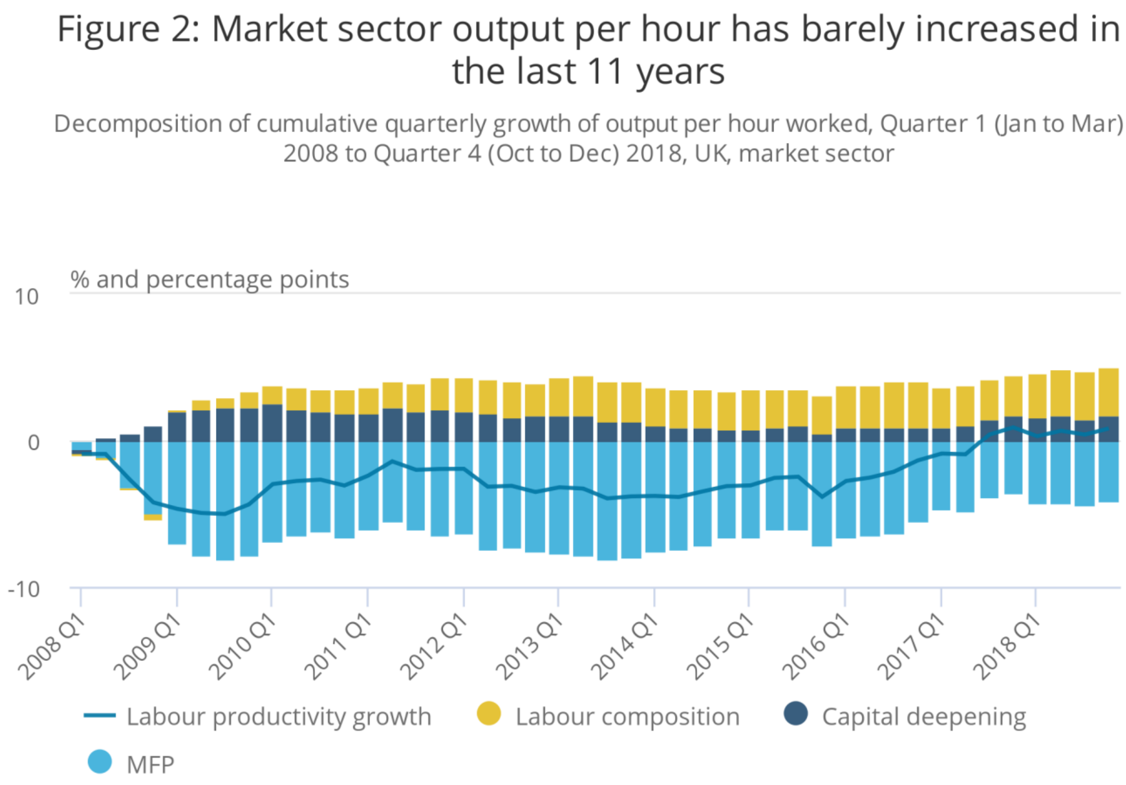
Source Prothero (2018)

Chart 6
Labour productivity in the G7 countries
GDP at PPP market prices per hour worked)



Source: ONS (2018a). Underlying data in spreadsheet for this chapter (Data for Oulton (2020).xlsx).

Chart 7
Output per hour, labour composition, capital deepenening, and MFP in the market sector, 2008Q1=2018Q4



Source ONS (2019b).

TABLES

Table 1
Labour productivity and livings standards in Britain, 1856-2016, 1856=100

	GDP per hour	*GDP per head*	*Hours per head*
1856	100	100	100
2016	1740	880	51

Source Bank of England (2017). Sources and methods discussed in detail in Chadha et al. (2019). Underlying data in spreadsheet for this paper (Data for chapter6.xlsx).

Table 2
Average annual growth rates of labour productivity and living standards, % per year

Period	*GDP per hour*	*GDP per head*	*Hours per head*
1856-1874	2.05	1.53	-0.52
1874-1914	0.90	0.81	-0.09
1914-1918	1.33	1.25	-0.09
1918-1939	1.25	0.36	-0.89
1939-1945	1.35	0.78	-0.56
1945-1973	3.59	2.81	-0.78
1973-2007	2.12	1.80	-0.32
2007-2016	0.07	0.09	0.02
1856-2016	1.79	1.36	-0.43

Source Bank of England (2017). Sources and methods discussed in detail in Chadha et al. (2019).
Note Growth rates are exponential (log differences divided by length of time interval). Underlying data in spreadsheet for this paper (Data for Oulton (2020).xlsx).

Table 3
The structure of the British economy in 2018: sections of the 2007 SIC

	Section	Value added, % of total	Hours worked, % of total	Labour productivity, % of average
		$(100\times v_{it})$	$(100\times w_{it})$	$(100\times(v_{it}/w_{it}))$
1	A: Agriculture	0.64	1.51	42.22
2	B: Mining and Quarrying	1.39	0.25	557.42
3	C. Manufacturing	8.53	7.95	107.26
4	D: Energy	1.75	0.53	332.01
5	E: Water Supply	0.99	0.69	144.25
6	F: Construction	6.06	8.12	74.68
7	G: Wholesale and Retail	10.61	14.08	75.31
8	H: Transport and Storage	4.26	5.18	82.30
9	I: Hotels and Catering	3.02	5.72	52.81
10	J: Information and Communication	6.62	5.13	129.10
11	K: Finance and Insurance	6.87	3.67	187.08
12	L: Real Estate	13.32	1.62	822.77
13	M: Business Services	7.85	8.56	91.63
14	N: Administrative and Support Services	4.92	8.36	58.86
15	O,P,Q: Public Services	17.53	22.63	77.44
16	R: Recreation and Culture	1.51	2.41	62.95
17	S,T,U: Other Services	2.64	2.58	102.55
18	Whole Economy	100.00	100.00	100.00

Source ONS (2019d) and underlying data in spreadsheet prodconts.xls available at https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/datasets/productivityjobsproductivityhoursmarketsectorworkersmarketsectorhours

Note Value added in section L, Real Estate, includes the imputed rent of owner occupiers, to which no hours worked are assigned. Hence the very high relative productivity in this sector should be disregarded. Underlying data in spreadsheet accompanying this paper (Data for Oulton (2020).xlsx).

Table 4
Growth of labour productivity in the market sector: within and between effects, per cent per year

	Within	Between	Total
During the boom (1997Q1-2008Q1)	2.70	-0.08	2.61
After the boom (2008Q1-2018Q4)	0.28	-0.03	0.25

Source Own calculations using data from ONS (2019d); underlying data in spreadsheet prodconts.xls available at https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/datasets/productivityjobsproductivityhoursmarketsectorworkersmarketsectorhours

Note Calculations use Törnqvist decomposition: see equation (8). Underlying data in spreadsheet accompanying this paper (Data for Oulton (2020).xlsx).

Table 5
Contributions of each section in the market sector to the growth of labour productivity, during and after the boom, and to the slowdown:
percentage points per year

Section	During the boom (1997Q1-2008Q1)	After the boom (2008Q1-2018Q4)	Slowdown	% of total slowdown
A: Agriculture	0.1055	0.0028	0.1027	4.34
B: Mining and Quarrying	-0.1350	-0.1121	-0.0229	-0.97
C. Manufacturing	0.6826	0.0465	0.6361	26.88
D: Energy	0.0429	-0.0542	0.0971	4.10
E: Water Supply	0.0370	-0.0081	0.0451	1.91
F: Construction	0.0405	0.0339	0.0066	0.28
G: Wholesale and Retail	0.2623	0.3169	-0.0546	-2.31
H: Transport and Storage	0.1656	-0.0419	0.2076	8.77
I: Hotels and Catering	0.0470	-0.0351	0.0821	3.47
J: Information and Communication	0.5170	0.1064	0.4106	17.35
K: Finance and Insurance	0.4792	-0.1632	0.6424	27.15
M: Business Services	0.3129	0.1437	0.1691	7.15
N: Administrative and Support Services	0.0651	0.0289	0.0361	1.53
R: Recreation and Culture	-0.0064	-0.0445	0.0381	1.61
S,T,U: Other Services	-0.0022	0.0279	-0.0301	-1.27
Market Sector	2.61	0.25	2.37	100.00

Source As Table 4.

Note Calculations use the Törnqvist decomposition: see equation (8). Figures for the market sector are the sum of the figures for the sections. Underlying data in spreadsheet accompanying this paper (Data for Oulton (2020).xlsx).

Table 6
The 30 largest economies in the world in 2011, based on PPPs

Country	*GDP at PPP US$, billions*	*Rank*	*GDP per head at PPP US$*	*Rank*	*AIC per head at PPP US$*	*Rank*	*Population Millions*
United States	15,533.80	1	49,782	12	37,390	2	312.04
China	13,495.91	2	10,057	100	4,331	123	1,341.98
India	5,757.53	3	4,735	128	3,023	136	1,215.96
Japan	4,379.75	4	34,262	33	24,447	23	127.83
Germany	3,352.10	5	40,990	24	28,478	9	81.78
Russian Federation	3,216.93	6	22,502	55	15,175	52	142.96
Brazil	2,816.32	7	14,639	80	9,906	80	192.38
France	2,369.59	8	36,391	30	26,486	15	65.11
United Kingdom	2,201.44	9	35,091	32	26,146	18	62.74
Indonesia	2,058.13	10	8,539	108	4,805	120	241.04
Italy	2,056.69	11	33,870	34	23,875	25	60.72
Mexico	1,894.55	12	16,377	72	11,844	67	115.68
Spain	1,483.22	13	32,156	36	21,484	31	46.13
Korea, Republic of	1,445.33	14	29,035	41	17,481	43	49.78
Canada	1,416.17	15	41,069	23	27,434	11	34.48
Saudi Arabia	1,366.70	16	48,163	14	17,797	42	28.38
Turkey	1,314.90	17	17,781	66	13,732	59	73.95
Iran, Islamic Rep.	1,314.24	18	17,488	67	8,576	85	75.15
Australia	955.98	19	42,000	20	27,089	12	22.76
Taiwan, China	907.14	20	39,059	27	25,129	21	23.22
Thailand	898.96	21	13,299	84	8,477	88	67.60
Egypt, Arab Rep.	843.83	22	10,599	98	8,529	87	79.62
Poland	838.05	23	21,753	57	16,307	47	38.53
Pakistan	788.13	24	4,450	130	3,926	128	177.11
Netherlands	720.27	25	43,150	16	25,983	19	16.69
South Africa	611.14	26	12,111	87	8,280	92	50.46
Malaysia	606.10	27	20,926	58	11,082	72	28.96
Philippines	543.66	28	5,772	126	4,490	122	94.19
Colombia	534.99	29	11,360	91	7,836	95	47.09
Nigeria	511.13	30	3,146	143	2,075	149	162.47

Source Spreadsheet containing the results of the 2011 ICP, accessed 26 April 2019: http://siteresources.worldbank.org/ICPINT/Resources/270056-1183395201801/2011-International-Comparison-Program-results.xlsx. Underlying data in spreadsheet for this paper (Data for Oulton (2020).xlsx).

Table 7
Asset types included in the VICS

1	Buildings other than dwellings
2	Other structures *(e.g. chemical works, motorways)*
3	Land improvements
4	Transport equipment
5	ICT equipment (excluding telecoms)
6	ICT equipment (telecoms equipment)
7	Other machinery and equipment
8	Cultivated biological resources *(e.g. cows)*
9	Research and development
10	Mineral exploration and evaluation
11	Computer software and databases (Own-Account)
12	Computer software and databases (Purchased)
13	Entertainment, literary or artistic originals

Source: ONS (2019c).