BANK OF ENGLAND

Staff Working Paper No. 740

Decomposing differences in productivity distributions

Patrick Schneider

July 2018

BANK OF ENGLAND

Staff Working Paper No. 740

Decomposing differences in productivity distributions

Patrick Schneider[1]

Abstract

I analyse the post-crisis slowdown in UK productivity growth using a novel decomposition framework, applied to firm-level data. The framework tracks flexibly defined distributions over time, and links changes in the shape of these distributions to aggregate movements. It encompasses many existing methods, which typically track firms over time, and also provides opportunities for various new types of analysis, particularly where firms are not repeatedly observed in survey data. In my application, I show that the slowdown in productivity growth is driven entirely by post-crisis reallocations of workers to firms with less-productive characteristics, rather than changes in the productivity associated with these characteristics (which have actually supported growth since the crisis). I further show that the puzzle is located in the top tail of the distribution, as is the negative contribution from these allocation effects.

Key words: Labour productivity, productivity decomposition, productivity distribution, UK productivity puzzle.

JEL classification: C14, C21, O47, L11.

(1) Bank of England. Email: patrick.schneider@bankofengland.co.uk

The views expressed in this paper are those of the author, and not necessarily those of the Bank of England or its committees. I am grateful to Will Abel, Tommaso Aquilante, Pawel Adrjan, Nikola Dacic, Rebecca Freeman, Joanna Konings, Marko Melolinna, Steve Millard, Patrick Moran, Nick Oulton, Oren Schneorson and Angelos Theodorakopoulos for their comments on an earlier version. Any remaining errors are mine.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Publications and Design Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

UK productivity growth has been puzzlingly slow since the 2008-09 global financial crisis. After averaging 2% p.a. over the pre-crisis decade, growth in labour productivity (output per hour worked) slowed to an average of only 0.5% since the crisis. Extensive research and commentary on the productivity puzzle has suggested myriad causes for the malaise—including 'zombie' firms hoarding resources, sluggish investment in the face of uncertainty, mismeasurement and more (e.g. Barnett et al., 2014; Goodridge et al., 2013; Haskel et al., 2015)—and have dismissed others that no longer seem plausible, such as temporary labour hoarding.

One of the live questions is whether the slowdown is attributable to particular groups of firms (e.g. in particular sectors, as in Tenreyro (2018) and Riley et al. (2018)). A strand of this research emphasises the role the weakest firms play in keeping aggregate productivity down—observing that a long tail of unproductive firms drags down on the aggregate (Haldane, 2017) and that a diverging top end of 'frontier firms' signifies stalled technology diffusion, the cause of flagging growth (Andrews et al., 2015; Andrews et al., 2016). The common thread here is that different sections of the distribution, or firms with particular features within it, could be driving aggregate results. But these analyses often lack a mechanism that links distribution level results to the aggregate, and so it can be hard to identify appropriate policy conclusions.

I propose a decomposition framework that allows us to link distributional observations to aggregate productivity directly. This is complementary to existing 'bottom-up' decompositions (Balk, 2016), with which researchers and policymakers describe changes in aggregate productivity measured with corporate micro-data (e.g. Barnett et al., 2014; Andrews et al., 2015; Riley and Bondibene, 2016; Borio et al., 2016; Decker et al., 2017). Such decompositions are typically achieved with one of two approaches.

1. **Panel** decompositions track firms over time and attribute changes in the aggregate to three contribution terms—the 'within' effect of continuing firms' productivity changing, the 'between' effect of labour moving between continuing firms, altering contributions to the average, and the 'net entry' effect of firms coming into and out of existence (e.g. Griliches and Regev, 1995; Foster et al., 2001; Baily et al., 2001; Diewert and Fox, 2005).

2. **Cross-sectional** decompositions attribute changes in productivity to changes in two contribution terms—the 'average' effect of a change in average productivity across firms and the 'allocative efficiency' effect of a change in a covariance term, relating firm productivity and employment (Olley and Pakes, 1996, SOP (Static Olley–Pakes)). This can be further augmented with a net entry effect, termed Dynamic Olley-Pakes (DOP) by Melitz and Polanec (2015).

In general, these methods require very high-quality data. Except for the SOP decomposition, they all track firms over time. As a result, unless the firm-level sample is a balanced panel, they must either be applied to a restricted set of repeatedly observed firms or imputed data for unobserved firms. They also offer limited insights. As discussed, for example, one cannot apply them to observations about the distribution with much flexibility.

In this paper, I show that panel methods are a special case of difference-in-mean decompositions, which are themselves a sub-class of methods for analysing changes in distributional statistics, outlined in Fortin et al. (2011) (FLF). Placed within the FLF framework, changes in aggregate productivity are equivalent to changes in the mean of the unconditional distribution of productivity across workers; and these changes in the unconditional mean are driven by changes in firm 'structure' (the conditional distribution of firm-productivity, given firm characteristics) and in the 'allocation' of workers (the distribution of workers across firm characteristics).

Suppose, for example, that a firm's export status is the only characteristic that affects its productivity (that exporting firms are more productive than others). In this case, aggregate productivity depends on two things—how much more productive exporting firms are (structure) and the proportion of workers employed by exporting firms (allocation). In this set-up, changes in aggregate productivity are driven by changes in either the export premium or the relative size of exporters' workforces, or both. This is a basic description of the Oaxaca (1973) and Blinder (1973) (OB) decomposition of the mean with respect to a set of characteristics.

The framework I outline encompasses many existing methods. Indeed, the panel methods described earlier *are* a OB decomposition, but with the characteristics set boiled down to a single, special dimension—a vector of firm identity dummies[1]. But placing productivity analysis within this framework adds many new, complementary methods to the researcher's toolkit, with three general benefits:

1. Relaxing a data quality restriction. By tracking firm characteristics, rather than identies, we rid ourselves of the need for balanced panels or imputation because the distributions, rather than the firms, are our objects of interest.

2. Allowing for insights in new dimensions. By thinking of aggregate productivity in terms of these distributions, we can look to the influence of economic structure and reallocations of activity to describe changes, potentially opening up new opportunities to test theoretical results.

3. Opening up the target statistics we can analyse. The framework applies to any distributional statistic. As well as being able to describe changes in aggregate productivity (the mean), it can be used to address other, increasingly distributional (Syverson, 2011), questions.

The paper is structured as follows. In section 2, I recast aggregate productivity as a statistic of the productivity distribution across workers, where the latter is conditional on the distribution of firm characteristics. This places our question squarely within the FLF framework, which I sketch. I then implement this framework, in section 3, with decompositions of a mean from two angles—an exact application using a OB decomposition and an approximate application, averaging over centiles, themselves decomposed following Chernozhukov et al. (2013). In section 4, both of these methods are applied to UK data to explain the change in aggregate UK labour-productivity over different periods between 2002 and 2014, with a focus on the puzzle. Section 5 concludes.

2 Theory

Aggregate productivity can be defined in terms of the distributions of firm structure and the allocations of workers across firms. Doing so allows us to use a general decomposition framework to analyse changes in productivity in these terms. In the following, I outline the two steps necessary to analyse productivity thus—first, I show that aggregate productivity is a statistic (the mean) of the unconditional distribution of productivity across workers, and that this distribution can be expressed as the integral of a conditional distribution (structure) with respect to the distribution of conditioning variables (allocation), the form of the general framework in FLF; second, I sketch the FLF framework for decomposing changes in generic distributional statistics.

[1] As shown in Section 3.1.

2.1 Aggregate productivity is a distributional statistic

Aggregate labour-productivity (Π) is some measure of total output (say value-added VA) per some measure of total labour-input (say number of workers L). This can be rearranged into a labour-weighted average of firm-level productivity[2] (π_i), where firms are indexed by i and weights are s_i.

$$\Pi = \frac{VA}{L} = \frac{\sum_i VA_i}{\sum_i L_i} = \sum_i \frac{L_i}{\sum_i L_i}\pi_i = \sum_i s_i \pi_i \tag{1}$$

This is the sample estimate of a population statistic—the mean of the productivity distribution across workers. For ease of notation, let Y denote worker productivity, a random variable with the unconditional distribution F_Y. Being an unbiased estimator, Π will equal the mean of Y, in expectation.

$$E[\Pi] = E[Y] \equiv \int y \, \mathrm{d}F_Y(y) \tag{2}$$

From equation (2), we can see that differences in F_Y must drive any differences in mean between groups, or over time. We can expand F_Y to include the influence of a set of characteristics describing a worker's employer (X) as conditioning variables.

$$F_Y(y) = \int F_{Y|X}(y|x) \, \mathrm{d}F_X(x) \tag{3}$$

So the distribution of productivity is determined by the 'structure' of the economy ($F_{Y|X}$), which relates the distribution of firms' productivity to their characteristics, and by the 'allocation' of workers (F_X), which marks the prevalence of firm characteristics across workers. Because the level is determined by structure and allocation, changes are also attributable to differences in these two objects.

2.2 Decomposing distributional statistics

I have shown that aggregate productivity is the sample estimate of the worker productivity distribution's mean, and that this distribution combines the effects of firm structure and worker allocations. Now I outline FLF's general framework for decomposing changes in distributions, and therefore their statistics, into contributions from differences in the distributions of structure or allocation[3].

In general, suppose we have two unconditional productivity distributions, describing two mutually exclusive groups of firms (e.g. two different time periods, or London based and not).

$$F_Y(y) = \int F_{Y|X}(y|x) \, \mathrm{d}F_X(x) \qquad \text{and} \qquad F'_Y(y) = \int F'_{Y|X}(y|x) \, \mathrm{d}F'_X(x) \tag{4}$$

And that we wish to describe the difference in these distributions ($\Delta F_Y = F_Y - F'_Y$) in terms of contributions from the difference in structure ($\Delta F_{Y|X}$) and in allocation (ΔF_X). These contributions can be constructed in two steps. The first step is to generate a counterfactual distribution by substituting F_X for F'_X in F_Y and leaving the other element fixed such that

$$F^C_Y = \int F_{Y|X}(y|x) \, \mathrm{d}F'_X(x) \tag{5}$$

or vice versa $F^C_Y = \int F'_{Y|X}(y|x) \, \mathrm{d}F_X(x)$. In terms of the example in the introduction, this counterfactual

[2] Although I work with labour-productivity here, the methods described are applicable whenever the aggregate is defined as an index which is a weighted average of lower level observations.

[3] FLF use the term 'characteristics' for what I am calling 'allocation'.

tells us what the distribution of productivity would be if either the export premium were fixed and workers re-allocated, or alternatively if workers stayed put but the export premium varied. It's important to recognise that these counterfactuals are not equivalent. They represent distinct experiments and either (or some combination of the two) may be appropriate depending on the question at hand.

Having constructed F_Y^C, the second step is then to add and subtract it to ΔF_Y and rearrange so that the contributions are identified[4].

$$\underbrace{\Delta F_Y(y)}_{\text{Difference}} = \underbrace{\int F'_{Y|X}(y|x)\,\mathrm{d}\Delta F_X(x)}_{\text{Allocation}} + \underbrace{\int \Delta F_{Y|X}(y|x)\,\mathrm{d}F_X(x)}_{\text{Structure}} \tag{6}$$

FLF show that the same logic applies to any distribution functional $v(F_Y)$—for example the mean, variance, other moments or any quantile—as long as three assumptions hold

1. Simple counterfactual: there are no general equilibrium effects in the calculation of the counterfactual distribution;
2. Overlapping support: both groups must be definable by the same types of characteristics, though their likelihood may vary; and
3. Ignorability: any unobserved features are orthogonal to the variable distinguishing the groups, when conditioning on observed features[5].

Under these assumptions, overall differences in any distribution functional (Δv_O) can be attributed to contributions from a change in structure (Δv_S) and from a change in allocation (Δv_X).

$$\underbrace{\Delta v_O}_{\text{Difference}} = \underbrace{\Delta v_X}_{\text{Allocation}} + \underbrace{\Delta v_S}_{\text{Structure}} \tag{7}$$

Finally, because Δv_O is observed, we need only calculate one of the right hand side terms; the other will be the residual[6].

There are a plethora of ways to actually apply this framework that differ in (a) the statistic of interest $v(\cdot)$, and (b) how the counterfactual is calculated. For example, OB can be used where $v(\cdot)$ is the mean and we assume the structure is linear; and Nopo (2008) provides a non-parametric alternative when F_X and F'_X have different supports. Various papers have also dealt with OB equivalents for non-linear models with specific functional forms, e.g. Fairlie (2005); Bauer and Sinning (2008). DiNardo et al. (1996) implement the decomposition for various $v(\cdot)$ by reweighting $\mathrm{d}F_X$, avoiding assumptions about the functional form of $F_{Y|X}$. And Machado and Mata (2005) and Chernozhukov et al. (2013) both provide

[4] This can be achieved in a few ways which are equal in sum but have different mid-points, representing the different experiments they impose on the counterfactual. Mechanically, the difference is in how the double-Δ term in the second line below is divided between the existing terms. The below roughly sketches the required algebra.

$$\begin{aligned}\Delta xy &= x_1y_1 - x_0y_0\\ &= \Delta xy_0 + x_0\Delta y + \Delta x\Delta y\\ &= \Delta xy_0 + x_1\Delta y\\ &= \Delta xy_1 + x_0\Delta y\\ &= \Delta x\bar{y} + \bar{x}\Delta y\end{aligned}$$

The difference is analogous to the distinction between Laspeyres and Paasche indices (Diewert and Fox, 2005).

[5] This last assumption is weaker than the exogeneity assumption in a classical linear regression model; in that setting, ignorability equates to assuming that if the linear estimator is biased, that it is biased *in the same way* between the two comparison groups and thus the bias cancels out in the differencing.

[6] This seems to be where ignorability does its work—if ignorability is violated, then the residual after allocation effects are calculated will include both true structure effects plus any allocation effects due to uncontrolled-for characteristics.

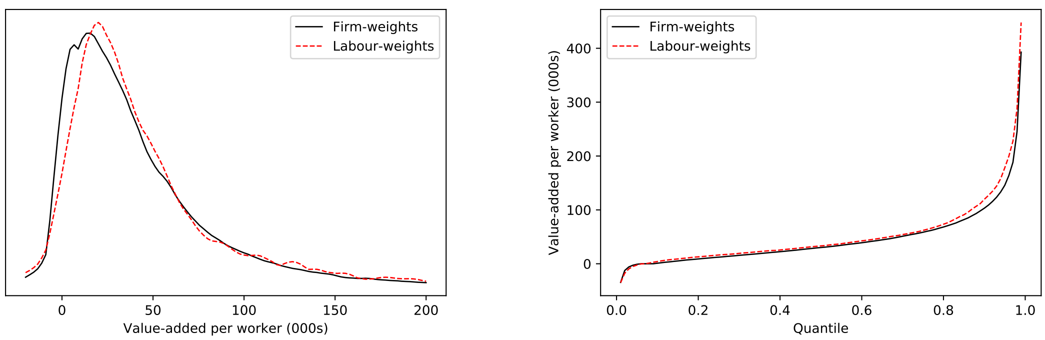
methods to decompose differences in whole distributions, differing primarily in whether F'_X or $F'_{Y|X}$ is used to generate the counterfactual. These, and many others, others are surveyed in FLF[7].

3 Empirical strategy

I have shown that aggregate productivity can be thought of as a distributional statistic, and that changes in such statistics can be decomposed into contributions from changes in the underlying structure of firms and the allocation of workers in the economy. In the next section I will apply the framework to the question: *What drove the change in aggregate productivity over different periods from 2002 to 2014?* using two implementations of the framework that I outline below. Given aggregate productivity is the target in both questions, our distribution functional for both implementations is the mean (i.e. $v(F_Y) = E[Y]$), though recall that it need not be.

Note that the expectations operator ($E[Y]$) describes the mean of the productivity distribution *across workers* (see equations (1) and (2)). Given that we usually measure productivity at the firm level, we can't just calculate the simple average or other distribution statistics from our data—they need to be labour weighted. This weighting can matter to varying degrees—more productive firms tend to be larger, so the (unweighted) firm distribution has more mass at lower productivity levels than the (weighted) worker distribution does (figure 1). The difference in average worker-productivity and average firm-productivity is the allocative efficiency term in the SOP decomposition—equal to the covariance of deviations in firm employment shares and productivity from their averages across firms—which varies over time (e.g. Decker et al., 2017).

Figure 1: Firm– and labour–productivity distributions in 2014



(a) Density functions (b) Quantile functions

In the following I outline the two different implementations of the framework, two ways to calculate the mean and then decompose contributions to differences, both of which will be used to answer our research question. The first, a OB decomposition, is exact and the second, averaging over changes in equally spaced centiles that are themselves decomposed as in equation (7), is an approximation. The former is a straight forward application. The latter, although an approximation, allows us to identify the sections of the distribution most responsible for the change in the mean; this is useful even without attributing such changes to underlying structure and allocation. The following outlines the high level theory behind each method.

[7]Implementations are also readily available for statistics packages such as Stata. For example, the 'oaxaca' and 'nldecompose' commands in Stata implements the OB decomposition for linear and non-linear models and 'cdeco' implements the Chernozhukov et al. (2013) quantile decomposition.

3.1 The Oaxaca–Blinder decomposition

Any mean is the expected value of a conditional expectation function, by the law of iterated expectations

$$E[Y] = E\left[E[Y|X]\right] = E[m(X)] \tag{8}$$

and if this function is linear, then[8] $E[Y] = E[X]\beta$. The OB decomposition, equation (9), estimates two linear regressions, one for each comparison group, then creates one of two counterfactuals—$E'[x]\beta$ or $E[x]\beta'$—and applies the algebra outlined in footnote 2.2 to recover the contributions[9] from differences in allocations ($\Delta E[x]$) or structure ($\Delta\beta$).

$$\Delta E[Y] = \underbrace{\Delta E[X]\beta'}_{\text{Allocation}} + \underbrace{E[X]\Delta\beta}_{\text{Structure}} \tag{9}$$

In the context of productivity analysis, the familiar 'within' and 'between' contributions of panel methods described earlier are a special case of a OB decomposition: where a difference in mean is decomposed with respect to the identities of firms. The 'between' firm contribution, from reallocations of labour between surviving firms, is equivalent to the contribution from a difference in allocations ($\Delta E[X]\beta'$); the 'within' firm contribution, from changes in surviving-firm productivity, is equivalent to the contribution from a change in structure ($E[X]\Delta\beta$); and the problem posed by the entry and exit of firms between periods is equivalent to the distributions of characteristics having different supports between groups, as in Nopo (2008).

To see this equivalence for the set of surviving firms, for example, one could expand the dataset so there are repeated observations of firms (one for each worker), add some random noise to the productivity variable (to eliminate perfect collinearity between workers at the same firm) and then perform the OB decomposition of productivity conditioned on firm fixed-effects.

3.2 The quantile approximation and decomposition

The mean of a distribution F_Y is equal to the integral of the distribution's quantile function $q(i|F_Y)$, with respect to a standard uniform distribution $F(i)$[10]. Furthermore, it can be approximated (11) by taking the simple average over a number (Q) of equally spaced quantiles.

$$E[Y] = \int_0^1 q(i|F_Y)\,\mathrm{d}F(i) \tag{10}$$

$$\approx \frac{1}{Q}\sum_{i=1}^{Q} q(i|F_Y) \tag{11}$$

[8] The proof of this is as follows. Substituting (3)→(2) and rearranging

$$E[Y] = \int_y y\,\mathrm{d}\int_x F_{Y|X}(y|x)\,\mathrm{d}F_X(x) = \int_x\int_y y\,\mathrm{d}F_{Y|X}(y|x)\,\mathrm{d}F_X(x) = \int_x m(x)\,\mathrm{d}F_X(x)$$

Where $m(X) = E[Y|X]$ is the conditional expectation function. Now suppose $m(X)$ linear, i.e. $m(X) = X\beta$, then

$$= \int_x x\beta\,\mathrm{d}F_X(x) = \int_x x\,\mathrm{d}F_X(x)\beta = E[X]\beta$$

[9] More generally referred to as composition and coefficient effects in OB decompositions.

[10] The proof of this is as follows, where the first step is to apply a probability integral transform to F_Y.

$$E[Y] \equiv \int_y y\;\mathrm{d}F_Y(y) = \int_y y\;\mathrm{d}\left[\int_0^1 F_{Y|i}(y|i)\;\mathrm{d}F(i)\right] = \int_0^1\left[\int_y y\;\mathrm{d}F_{Y|i}(y|i)\right]\;\mathrm{d}F(i) = \int_0^1 q(i|F_Y)\;\mathrm{d}F(i)$$

The approximation is not exact and will be biased if there is skew in the distribution of Y (in the opposite direction of the skew), but it becomes better and less biased as Q grows, such that $\lim_{Q\to\infty} \frac{1}{Q}\sum_{i=1}^{Q} q(i|F_Y) = E[Y]$, as in equation (10).

This approximation offers its own decomposition. Changes in aggregate productivity can be measured as the average of the difference between quantile functions (12) and so we can identify the sections of the distribution driving a change in the mean. Even absent contributions from allocations and structure, we can use this approximation to locate changes over time, cross-country differences and many other comparisons, at different parts of the distribution.

Futhermore, each quantile is itself the product of underlying structure and allocation distributions. As such, we can decompose the quantile-by-quantile differences into contributions from changes in these distributions, as in equation (7). There are various available methods for affecting such a decomposition; I follow Chernozhukov et al. (2013) in the following. This method estimates the full distribution function, conditioning on characteristics, and integrates the function over these characteristics to arrive at the unconditional quantile function. The counterfactual is constructed by integrating the base group's conditional distribution function over the comparison group's characteristics (i.e. $F^C_Y = \int F'_{Y|X}(y|x)\mathrm{d}F_X(x)$).

$$\Delta E[Y] \approx \frac{1}{Q}\sum_{i=1}^{Q} q(i|F_Y) - q(i|F'_Y) \tag{12}$$

$$= \underbrace{\frac{1}{Q}\sum_{i=1}^{Q}\left(q(i|F^C_Y) - q(i|F'_Y)\right)}_{\text{Allocation}} + \underbrace{\frac{1}{Q}\sum_{i=1}^{Q}\left(q(i|F_Y) - q(i|F^C_Y)\right)}_{\text{Structure}} \tag{13}$$

The two methods outlined above are both novel ways of decomposing a difference in aggregate productivity over time or groups, two examples of the many opportunities made possible by placing the research question within the distribution decomposition framework.

4 Application

In this section, I apply these two implementations of the framework to the question: *What drove the change in aggregate productivity over different periods from 2002 to 2014?*. I first introduce the dataset, then present results and finally discuss limitations of the applications I've chosen.

4.1 Data

I use micro-data from the ONS's (2016) Annual Respondents Database X from 2002 to 2014 to understand changes in aggregate productivity over this period. The dataset combines the Annual Business Inquiry (to 2008) and Annual Business Survey (from 2008) datasets, which cover the population of reporting units of firms with over 250 employees in the UK (excluding Northern Ireland) and samples remaining firms—I have used sample weights in the following to ensure appropriate aggregation.

There are 35,000–47,000 observations per year. The surveys cover the non-finanical business sector and all observations in the dataset are included with a few exceptions. The Finance and Insurance Activities (SIC07 64–66), Agriculture, Forestry and Fishing (SIC07 01–03) and Public Administration and Defence (SIC07 84) industries are dropped due to low coverage. Also, some industries are only surveyed after 2008—Mining and Quarrying (SIC07 05–09), Retail Trade (SIC07 47), and Accommodation and Food Services Activities (SIC07 55–56)—and are excluded, to ensure consistency. Any aggregate figures constructed from this dataset therefore represent the UK economy, except tfor these sectors.

Productivity is measured as the ratio of value-added at market prices (deflated using 2-digit SIC07

Figure 2: Aggregate productivity over time

industry deflators) to total employment. Chart 2 shows the time–series of annual, aggregate productivity in the dataset. The crisis (slump from 2007-09) and productivity puzzle (stagnant growth from 2010 onward) are both clearly present, even without including financial sectors[11]. I analyse the change over our whole sample, and then break the sample into distinct periods—pre-crisis (2002-07), crisis (2007-09) and post-crisis (2009-14)—with a focus on comparing post- and pre-crisis rates of change to analyse the UK's productivity puzzle.

Finally, the task at hand is to explain these differences in terms of contributions from allocations of workers across firm characteristics and the structure relating these characteristics to firm productivity. We therefore need a set of characteristics. I have opted for a very simple set—a reporting unit's SIC07 'division'[12], its size class (defined by employment in bins of {1, 2–9, 10–24, 25–99, 100–249, 250–999, >1,000} workers), its region and whether it has a foreign owner or not—to ensure as many observations can be included as possible (trade exposure variables, for example, are only available in the second half of the data-set). This limits the inferences one can make about the decomposition contributions, as I discuss in section 4.3, but allows for a good demonstration of the framework.

4.2 Results

4.2.1 Growth from 2002 to 2014

Let's start by analysing the change in the aggregate over our whole sample period. The results for both decomposition methods are presented in table 1, with the OB outputs recorded in the first row of the table, labelled 'Mean'—the three columns report the measured absolute difference in productivity over this period, and the contributions to this difference from changes in the allocation of workers and the estimated structure of firms[13], respectively. The average worker in 2014 produced over £12k worth of value-added more than they did in 2002 (under the Δ symbol). The OB decomposition estimates suggest that this increase in productivity is almost entirely due to changes in structure; that reallocations of workers across firm characteristics over this period supported productivity growth only mildly, if we held the structure fixed at 2002 level[14].

The next row in table 1, labelled 'Quantile approx.', reports the difference in aggregate productivity

[11] Which are important for understanding the whole economy puzzle (Tenreyro, 2018)

[12] A little more detailed than the 1–digit sectors.

[13] The counterfactual used here is to change the allocation in the base (earlier) year first, and then the structure. There are alternatives to this, as outlined in footnote 2.2, which will deliver different results. I've chosen this particular one to ensure consistency between the OB decomposition and the quantile approximation.

[14] Note that this does not imply all changes in the allocation of workers across characteristics supported productivity. Rather, that the changes were a net-positive if we hold the 2002 productivity-returns to characteristics fixed.

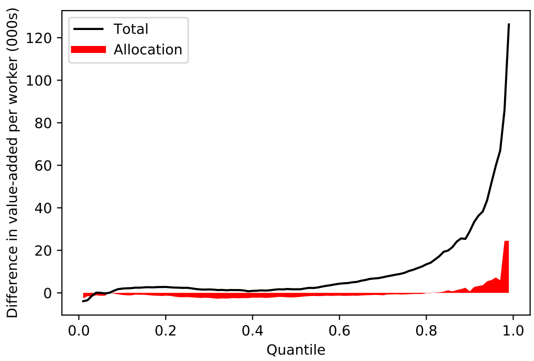
Table 1: Summary of results over time

	2011 £000s CVM		
	Δ	Allocations	Structure
Mean	12.76	0.22	12.53
Quantile approx.	10.22	-0.31	10.54
q1–q50	*0.69*	*-0.91*	*1.61*
q51–q75	*1.32*	*-0.31*	*1.63*
q76–q99	*8.21*	*0.91*	*7.30*

between the 2002 and 2014 as estimated by averaging over the 99 centiles[15]. This difference is close to, but a bit less than the exact mean difference (£10.22k compared to the actual £12.76k), reflecting the bias originating from the skew in the distributions, and the missed observations above the 99th centile. This approximation of the difference is then decomposed into the average contributions from changes in allocations and structure, estimated following Chernozhukov et al. (2013)[16]. The contribution from changing allocation here has a different sign to the OB decomposition—the growth in productivity over this period is estimated to have occurred despite disadvantagoues reallocation of labour across firm characteristics, though this latter effect is small relative to the total change.

The final three rows apportion the total quantile approximation numbers to three sections of the distribution—the bottom half '*q1-q50*', and splitting the remainder into two quartiles, '*q51-q75*' and '*q76-q99*'. These are binned figures, averages over the labelled sections of the distribution, but we can see the whole range of results plotted in figure 3. These results show that the bulk (£8.21k of £10.22k) of the change in the aggregate productivity is driven by the top quartile (the most productive workers are more productive still), and figure 3 shows that even the top quartile results are themselves concentrated in the upper centiles. The allocation contributions do not affect the distribution uniformly—changes in worker allocation across firm characteristics appear to have dragged on the bulk of the distribution, but supported growth at the very top. As with the total difference, the positive allocation contributions are stronger the higher the quantile. The bias from missing the top 1% of workers, therefore, could also affect the estimated allocations contribution in this decomposition and be driving the difference in sign between the OB and quantile approx. decompositions.

Figure 3: Quantile decomposition of the change from 2002–14



The dominance of the top tail for aggregate growth is a natural result of the extreme skew in the distribution (see figure 1)—the top tail has a very strong influence on the level of aggregate productivity in any given year, just as large outliers will push up on any average, and also on changes in this level. This latter observation is similar to Andrews et al. (2016)'s result that the top tail of 'frontier' firms

[15] 999 mintiles or any other set of equally spaced quantiles would do as well, with varying degrees of accuracy.
[16] The conditional distributions are approximated by 5,000 logit models over the support of productivity.

is diverging from the rest. The latter paper speculates that this divergence could be the cause of the aggregate growth slow-down; that it signifies the failure of firms in the rest of the distribution to keep up with innovations at the frontier, thus holding back their growth, and that of the aggregate.

The quantile approximation allows us to go a step further and measure the implication of the divergence for aggregate productivity. And it appears that the divergence of the top tail is the *source* of most growth in the aggregate. There are a number of differences between the present analysis and that in Andrews et al. (2016)[17], the main one being these results do not account for the changing composition of firms at the top end of the distribution. But the present analysis does directly measure the relationship between the frontier *workers* (whichever firms employ them) and the aggregate and, as I show below, this relationship is crucial for understanding the UK's productivity puzzle.

4.2.2 The productivity puzzle

Looking at total growth from 2002–14 elides the pre- and post-crisis eras, so we miss most of the interesting changes within. If we instead break the the sample into two five-year periods—pre-crisis from 2002–07 and post-crisis from 2009-14—we can use the framework to analyse the UK's productivity growth puzzle; that is, the slow-down in aggregate growth after the crisis[18], or the difference-in-changes between these two periods.

Table 2: The productivity puzzle (2011 £000s CVM)

	Pre-crisis (2002–07)			Post-crisis (2009–14)			Puzzle		
	Δ	Alloc.	Struc.	Δ	Alloc.	Struc.	$\Delta\Delta$	ΔAlloc.	ΔStruc.
Mean	2.06	0.21	1.84	1.90	-0.12	2.02	-0.16	-0.34	0.18
Quantile approx.	1.85	0.13	1.72	1.59	-0.16	1.75	-0.26	-0.29	0.02
q1–q50	*0.10*	*-0.07*	*0.17*	*0.34*	*-0.05*	*0.39*	*0.25*	*0.02*	*0.23*
q51–q75	*0.34*	*-0.00*	*0.35*	*0.30*	*-0.04*	*0.34*	*-0.04*	*-0.03*	*-0.01*
q76–q99	*1.42*	*0.20*	*1.21*	*0.95*	*-0.07*	*1.02*	*-0.47*	*-0.27*	*-0.19*

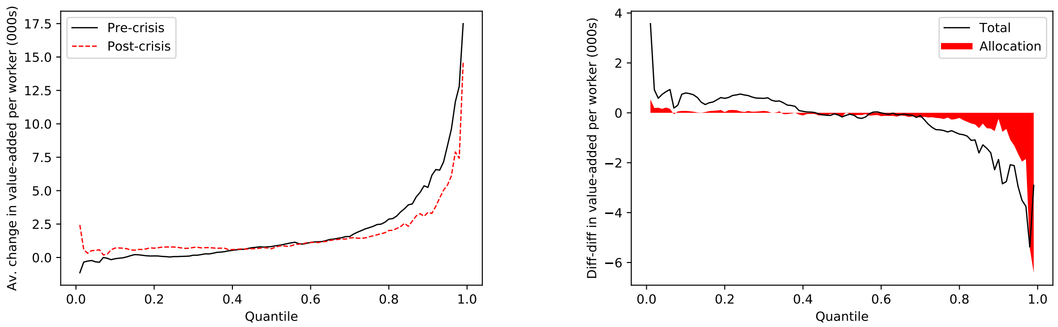
Table 2 shows the results for each of the pre- and post-crisis periods, in per-annum terms. The final three columns show the difference between the post- and pre-crisis periods to give a sense of the puzzle in our data. The puzzle measured in these data amounts about nearly a 10% slow-down in the change in aggregate productivity after the crisis (£2.06k p.a. before the crisis down to £1.9k p.a. after). Both the OB estimates of contributions and the quantile approximation find similar figures, and both decompositions attribute the slow-down entirely to negative relative contributions from reallocations of workers across firm characteristics to growth: reallocations contributed poisitively to growth before the crisis, but drag on it afterward. By contrast, the difference in changing structure after the crisis is net-positive; if there had been no change in these structure contributions after the crisis, the puzzle would be even deeper.

The quantile approximation allows us to see where the puzzle is located in the distribution. Table 2 shows that the slowdown in growth is almost entirely in the top quartile of the distribution; indeed, the lower section of the distribution grew faster, post-crisis, than it did before. We can see these differences

[17]The former is a within-industry analysis of firm-level total-factor productivity, and describes the firm-weighted distribution, whereas mine is a cross-industry analysis of labour productivity, and weights by labour. I have re-run this analysis at the industry level, and find similar results to the aggregate ones presented above.

[18]The UK productivity puzzle usually refers to the deviation of labour productivity from the exponential trend set before the crisis. There are two parts to this deviation—the 'level' and 'growth' puzzles. The level puzzle is that productivity did not quickly return to trend, as it has after other post-war recessions. But it's not actually so puzzling in the broader sweep of history. This is because recessions following financial crises tend to be deeper and more prolonged (Jorda et al., 2013; Cerra and Saxena, 2008) and are associated with permanent output losses within range of the UK's actual experience (Oulton and Sebastia-Barriel, 2017; Duval et al., 2017; Basel Committee, 2010). Hence, the UK's level puzzle is, to some extent, typical. The 'growth' puzzle is that productivity did not return to pre-crisis growth rates, even locking in the level-hit during the crisis, and that this has persisted for nearly a decade since the crisis. Such a long-run effect on growth following even a financial crisis is much more puzzling, and so is the focus of most current analysis.

Figure 4: The puzzle across the distribution



(a) Average annual change in quantiles (b) Decomposition contributions to the difference

more in figure 4. The left panel in this figure plots the average, annual change in productivity by quantile before and after the crisis; the gap between these two lines is the puzzle. The right panel plots this gap, as well as the difference between the estimated allocation contributions to each of the lines in the left panel. Both panels show that the puzzle is isolated to the top end of the distribution: the slowdown in growth after the crisis is isolated to the top quartile, whereas the third quartile grew at about the same rate as it did pre-crisis, and the quantiles below the median tended to grow more than before.

Turning now to the decomposition of the puzzle, table 2 shows that the aggregate contributions are not uniform across the distribution. Overall, we attributed the puzzle to the negative pull of allocations contributions, and found that changes in structure contributions have actually supported growth since the crisis (and so their absence would deepen the puzzle). But the allocation contributions are concentrated in the top end of the distribution. Prior to the crisis, reallocations supported growth in the top quartile of the distribution, and pulled down on the rest. Since then, reallocations are estimated to drag on the whole distribution. Hence, the loss of this support for growth in the top quartile explains the allocations contribution to the aggregate puzzle.

The story is very different for the estimated structure contributions, which have little overall influence on the puzzle in the quantile approximation. As table 2 shows, changes in structure are estimated to support growth in the whole distribution, both before and after the crisis. This support is estimated to be stronger after the crisis for quantiles below the median, and weaker after the crisis in the top quartile. The net contribution between these opposing forces on different points in the distribution is about zero. Hence, there is little overall structure effect in the quantile approximation to the puzzle, although these changes have affected the shape of the distribution by slowing its expansion rate.

4.3 Limitations

There are a number of limits to the interpretation of the results presented. First, because the results all come from models and statistics which include measurement error, proper inference requires confidence intervals—these can all be bootstrapped, and many of the packages I've employed here provide them. Second, because the results are all constructed from deflated nominal productivity statistics, we cannot interpret them as describing quantities unless we presume prices are consistent within 2-digit industries, which is unlikely. Third, each decomposition method can be applied to the same data in at least two ways by swapping the base and comparison groups and the contributions will change as a result. In the examples above, the signs and magnitudes of effects are actually quite stable across different specifications, but we should nonetheless be careful to interpret results in light of the specific counterfactual that was used.

Perhaps most importantly, we should distinguish features in the data from those that result from these modelling choices. In the above applications, the measured differences, and their attribution to different parts of the productivity distributions, are features of the data. As such, we do not require any assumptions to conclude that the bulk of the observed differences in productivity are driven by the top tails of the distributions, and that this is where the productivity puzzle can be found. By contrast, their attributions to allocation and structure rest on three assumptions—simple counterfactual, overlapping support and ignorability. The last of these is likely to pose problems that limit identification in my applications. For example, firms in 2014 may be more productive than those in 2002 because of an uncontrolled-for characteristic on which they also differ (for example trade-exposure). In this case, ignorability will be violated and the attributions are only partially identified—the allocation contribution *of the controlled-for characteristics* is identified, but the remainder is a mix of the remaining difference in allocations and structure effects, rather than just the latter.

5 Conclusion

I have used a novel decomposition framework to analyse the UK productivity puzzle. I have shown that the puzzling slow-down since the financial crisis is attributable to reallocations of labour into firms with less productive characteristics. By contrast, the growth in productivity associated with this simple set of characteristics has actually improved since the crisis, and so would have supported growth if the allocation of labour were fixed in 2009. Furthermore, the slowdown is entirely located in the top end of the distribution—workers at the most productive firms are not improving on their predecessors as quickly as they did prior to the crisis—and the negative pull from worker reallocations is also concentrated here.

These results are based on two implementations of the distribution-decomposition framework surveyed in FLF, which I apply to the analysis of productivity. This consists of viewing firms as bundles of characteristics and attributing changes in the productivity distribution to contributions from changes in (a) the structure distribution, which describes firm productivity conditional on characteristics, and (b) the allocation distribution, which describes the spread of workers across these characteristics.

This framework is very general. It encompasses many existing decomposition methods and can also be used in tandem with them. One could, for example, amend the quantile approximation to describe continuing firms only and add a net-entry term. And it is also extremely flexible. The two implementations in this paper demonstrate its utility for a familiar question—describing changes in aggregate productivity over time—but the framework is just as applicable to other moments of the distribution as it is to the mean, as well as to other comparisons and to richer firm characterics controls. The ability to analyse distributional questions is particularly useful, given the increasing focus on firm (and worker) heterogeneity.

References

Andrews, D., C. Criscuolo, P. Gal, et al. (2015). Frontier firms, technology diffusion and public policy: Micro evidence from OECD countries. Technical report, OECD Publishing.

Andrews, D., C. Criscuolo, and P. N. Gal (2016, December). The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy. OECD Productivity Working Papers 5, OECD Publishing.

Baily, M. N., E. J. Bartelsman, and J. Haltiwanger (2001). Labor productivity: structural change and cyclical dynamics. *The Review of Economics and Statistics 83*(3), 420–433.

Balk, B. M. (2016). The dynamics of productivity change: A review of the bottom-up approach. In *Productivity and Efficiency Analysis*, pp. 15–49. Springer.

Barnett, A., A. Chiu, J. Franklin, and M. Sebastiá-Barriel (2014). The productivity puzzle: a firm-level investigation into employment behaviour and resource allocation over the crisis. *Bank of England Working Paper* (495).

Basel Committee (2010). *An assessment of the long-term economic impact of stronger capital and liquidity requirements.* Bank for International Settlements.

Bauer, T. K. and M. Sinning (2008). An extension of the Blinder–Oaxaca decomposition to nonlinear models. *AStA Advances in Statistical Analysis 92*(2), 197–206.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436–455.

Borio, C. E., E. Kharroubi, C. Upper, and F. Zampolli (2016). Labour reallocation and productivity dynamics: financial causes, real consequences. *BIS Workping papers* (534).

Cerra, V. and S. C. Saxena (2008, March). Growth dynamics: The myth of economic recovery. *American Economic Review 98*(1), 439–57.

Chernozhukov, V., I. Fernndez-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2017, May). Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown. *American Economic Review 107*(5), 322–326.

Diewert, W. E. and K. A. Fox (2005). On measuring the contribution of entering and exiting firms to aggregate productivity growth. *Price and productivity measurement 6.*

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044.

Duval, M. R. A., M. G. H. Hong, and Y. Timmer (2017). *Financial frictions and the great productivity slowdown.* International Monetary Fund Working Paper.

Fairlie, R. W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of economic and social measurement 30*(4), 305–316.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. *Handbook of Labor Economics 4*, 1 – 102.

Foster, L., J. C. Haltiwanger, and C. J. Krizan (2001). Aggregate Productivity Growth: Lessons from Microeconomic Evidence. In *New Developments in Productivity Analysis*, NBER Chapters, pp. 303–372. National Bureau of Economic Research, Inc.

Goodridge, P., J. Haskel, and G. Wallis (2013). Can intangible investment explain the UK productivity puzzle? *National Institute Economic Review 224*(1), R48–R58.

Griliches, Z. and H. Regev (1995). Firm productivity in israeli industry 1979-1988. *Journal of Econometrics 65*(1), 175–203.

Haldane, A. (2017, March). Productivity puzzles. `https://www.bankofengland.co.uk/speech/2017/productivity-puzzles`. Speech at London School of Economics.

Haskel, J., P. Goodridge, and G. Wallis (2015). Accounting for the UK productivity puzzle: a decomposition and predictions.

Jorda, O., M. Schularick, and A. M. Taylor (2013). When credit bites back. *Journal of Money, Credit and Banking 45*(s2), 3–28.

Machado, J. A. F. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics 20*(4), 445–465.

Melitz, M. J. and S. Polanec (2015, 06). Dynamic Olley-Pakes productivity decomposition with entry and exit. *RAND Journal of Economics 46*(2), 362–375.

Nopo, H. (2008). Matching as a tool to decompose wage gaps. *The Review of Economics and Statistics 90*(2), 290–299.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*(3), 693–709.

Office for National Statistics. Virtual Microdata Laboratory (VML), University of the West of England, B. (2016). Annual respondents database X, 1998-2014: Secure access. http://doi.org/10.5255/UKDA-SN-7989-3.

Olley, G. S. and A. Pakes (1996, November). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64*(6), 1263–1297.

Oulton, N. and M. Sebastia-Barriel (2017). Effects of financial crises on productivity, capital and employment. *The Review of Income and Wealth 63*(1).

Riley, R. and C. R. Bondibene (2016). Sources of labour productivity growth at sector level in britain, after 2007: a firm level analysis. *NESTA Working Paper* (16/01).

Riley, R., A. Rincon-Aznar, and L. Samek (2018). Below the aggregate: A sectoral account of the UK productivity puzzle. *ESCoE Discussion Paper* (06).

Syverson, C. (2011, June). What determines productivity? *Journal of Economic Literature 49*(2), 326–65.

Tenreyro, S. (2018, January). The fall in productivity growth: causes and implications. https://www.bankofengland.co.uk/speech/2018/silvana-tenreyro-2018-peston-lecture. Speech at Peston Lecture Theatre, Queen Mary University of London.