Accepted: 18 January 2024

DOI: 10.1111/labr.12265

ORIGINAL ARTICLE

Firm heterogeneity and the aggregate labour share

Matteo G. Richiardi[1] | **Luis Valenzuela**[2,3]

[1]Centre for Microsimulation and Policy Analysis (CeMPA), Institute for Social and Economic Research, University of Essex, Colchester, UK

[2]Center of Economics for Sustainable Development (CEDES), Faculty of Economics and Government, Universidad San Sebastián, Concepción, Chile

[3]Millennium Institute on the Evolution of Work (MNEW), Santiago, Chile

Correspondence
Matteo G. Richiardi, Centre for Microsimulation and Policy Analysis (CeMPA), Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK.
Email: matteo.richiardi@essex.ac.uk

Funding information
Agencia Nacional de Investigación y Desarrollo Millennium Science Initiative program, Grant/Award Number: NCS2021-033; Agencia Nacional de Investigación y Desarrollo (Fondecyt Iniciación), Grant/Award Number: 11221344

Abstract

We propose a model-based decomposition method for the aggregate labour share in terms of the first moments of the joint distribution of total factor productivity, market power, wages and prices, and apply it to UK manufacturing using firm-level data for 1998–2014. Contrary to a narrative focussing on increasing disparities between firms, the observed decline in the aggregate labour share over the period is driven entirely by the decline in the labour share of the representative firm, mostly due to an increasing disconnect between average productivity and real wages. Changes in the dispersion of firm-level variables have contributed to slightly contain this decline.

JEL CLASSIFICATION

D33, E25, L10, D20, D42, D43

1 | INTRODUCTION

In this paper, we propose and test a model-based decomposition approach that characterizes the aggregate labour share in term of the first moments (mean, variance and covariance) of key firm-level characteristics, namely total factor productivity, product and labour market power, wages and output prices. By doing so, we contribute to the literature investigating the role of firm heterogeneity in explaining aggregate labour market outcomes.

A marked decrease in the aggregate labour share over the recent decades has been documented in many countries (Dimova, 2019; IMF, 2017; Karabarbounis & Neiman, 2014). A rich debate emerged about the factors driving this decline. Abstracting from measurement issues (e.g. Koh et al., 2020), many studies have analysed the aggregate labour share (at the national, regional or industry level) from a macro perspective.[1] Others have looked at the drivers of the firm-level labour share (e.g. Bell et al., 2018; De Loecker & Eeckhout, 2018; Mertens, 2022; Perugini et al., 2017; Siegenthaler & Stucki, 2015). A third group has produced *statistical* decompositions of the aggregate labour share, either looking at changes in the relative importance of firms with a higher/lower than average labour share (Abdih & Danninger, 2017; Bauer & Boussard, 2020; Valentinyi & Herrendorf, 2008), or comparing across-firm changes in wages and labour productivity, the two factors that *define* the labour share (Bockerman & Maliranta, 2012; Kehrig & Vincent, 2021).[2] These studies however fall short from offering an understanding of the drivers behind such changes. One notable exception is Autor et al. (2020) and their theory of superstar firms, where they offer a link between firm-level mark-ups and the aggregate labour share. Still, in their model there is no other determinant of the aggregate labour share but mark-ups, a by-product of assuming a linear production function where labour is the only factor of production (hired in a competitive market).[3]

Instead, in this paper we propose a novel approach that explicitly offers a microfoundation of the aggregate labour share in term of multiple firms' characteristics. More precisely, starting from a static model of firm behaviour with CES production functions and imperfect competition in the product and labour markets, we offer a *model-based* decomposition formula for the aggregate labour share that allows us to quantify the overall effect of firm heterogeneity on the aggregate labour share, looking into the relative importance of the different sources. This method builds up from the micro level, without requiring the existence of an aggregate production function.

In particular, our decomposition formula shows that when the elasticity of substitution between capital and labour is below 1 —the empirically relevant case— a ceteris paribus increase in the dispersion of productivity or monopsony power increases the aggregate labour share, while a ceteris paribus increase in the dispersion of real wages or product market power decreases it. Another important theoretical lesson is derived for the case of a Cobb–Douglas (CD) production function —where the elasticity is equal to 1, and the most common empirical assumption. We show that, unless models assume *heterogeneity* in product or labour market power, heterogeneity in other variables does not affect the labour share in a CD world. As such, models assuming CD production functions might severely constrain the potential determinants of the labour share. Relatedly, since an aggregate production function cannot be derived from the micro level production functions under imperfect competition (regardless of its shape), firm heterogeneity is by definition invisible to a model based on an aggregate production function.

We then apply this decomposition method to firm-level data covering UK manufacturing between 1998 and 2014. We show that the labour share of a 'representative', average firm, is roughly 10 percentage points lower than the one actually observed in the data. In other words, firm heterogeneity increases the *level* of the aggregate labour share, in our data. Empirically as well as theoretically therefore, 'heterogeneity matters'. Second, we find that this wedge between the aggregate labour share and the average labour share is due largely to two dimensions of heterogeneity, namely TFP and, to a lower extent, labour market power. Heterogeneity in wages and product market power has little effect on this wedge. This is likely to reflect the fact that TFP and labour market power are more difficult to arbitrate across firms (e.g., due to organizational knowledge specific to the firm or geographical amenities, respectively). Third, we find that the aforementioned wedge between average and aggregate labour share remained fairly constant over time. The fall in the aggregate labour share observed over the period is therefore attributable by and large to changes in the characteristics of the average firm: in particular, to an increased pay-productivity gap, and to a lesser extent to increased market power. Interestingly, and differently from a common narrative highlighting the role of 'superstar firms', we find that changes in firm heterogeneity contributed to slightly *reduce* the fall in the aggregate labour share.

In the remaining of the paper, Section 2 presents a simple model of firm optimisation where firms have a constant returns to scale CES technology with imperfect competition in the product and labour markets, and derives our theoretical results; Section 3 describes the data and our empirical strategy, while Section 4 reports our main findings. Section 5 summarizes and concludes.

2 | MODEL

A well-studied though often neglected result from the neoclassical theory of production is that when input and output prices and quantities are heterogeneous across firms, or when firms differ in terms of fundamental factors like total factor productivity, aggregation of firms' technologies into a single production function is not possible (Felipe & McCombie, 2014; Fisher, 1969; Green, 1964; Zambelli, 2004). Thus, under firm heterogeneity the aggregate labour share cannot be computed solely with reference to an optimal production plan of a 'representative firm', using aggregates of input and output prices and factors. Instead, it must be computed from the bottom up, adding up labour costs and value added across firms. Here, we use a simple neoclassical model of firm behaviour in order to characterize the relationship between the distribution of firms' characteristics and the aggregate labour share in the economy, in a partial equilibrium setting.

2.1 | Setup

First, let us define the firm level labour share, upon which all the analysis is built:

$$\lambda_i \equiv \frac{w_i L_i}{p_i Y_i}, \tag{1}$$

where w_i are wages, L_i is the level of employment, p_i is output price, and Y_i is real value added, for a given firm i. The aggregate labour share, defined as aggregate labour costs over aggregate value added, can then be expressed as a weighted average of λ_i:

$$\lambda \equiv \frac{\sum_i w_i L_i}{\sum_i p_i Y_i} = \sum_i \lambda_i \delta_i, \tag{2}$$

where $\delta_i = \frac{p_i Y_i}{\sum_j p_j Y_j}$ corresponds to the share of aggregate value added produced by firm i. Our aim is to characterize λ in terms of firms' choices. Since the latter depends on λ_i, which in turns depends on $\frac{Y_i}{L_i}$, we need assumptions about technology, market structure and firm's behaviour which enables us to find the optimal $\frac{Y_i}{L_i}$ ratio for firms. Our starting point is a CES *value added* production function (i.e., a mathematical relation between capital, labour and value added)[4]:

$$Y_i = A_i \left(\alpha L_i^{\rho} + (1-\alpha) K_i^{\rho}\right)^{\frac{1}{\rho}}, \tag{3}$$

where $\sigma = \frac{1}{1-\rho}$ is the elasticity of substitution between capital and labour. Notice firms have the same technology in terms of elasticities (ρ and α), but they might have heterogeneous total factor productivity (TFP), A_i. We assume firms have a certain degree of *monopolistic power* in the pricing of the final good, and some degree of *monopsony power* in the labour market. Importantly, the degrees of market power might be heterogeneous across firms. Accordingly, the inverse product demand function firms face is denoted with $p_i = p_i(Y_i)$, while the inverse labour supply function is $w_i = w_i(L_i)$.

Given profits $\Pi_i(L_i, K_i) = p_i(Y_i)Y_i - w_i(L_i)L_i - r_iK_i$, the first order maximizing condition with respect to labour is given by:

$$\frac{\partial Y_i}{\partial L_i} \equiv \alpha A_i^{\rho}(Y_i)^{1-\rho}(L_i)^{\rho-1} = \left(\frac{w_i}{p_i}\right)\frac{\chi_i^L}{\chi_i^Y}, \tag{4}$$

where $\chi_i^L = 1 + \frac{1}{\eta_i^L}$ and $\chi_i^Y = 1 + \frac{1}{\eta_i^Y}$, and where η_i^Y corresponds to the own-price elasticity of output demand and η_i^L is the own-price elasticity of the labour supply. Importantly, the term $\frac{\chi_i^L}{\chi_i^Y}$ represents the wedge between wages and the marginal product of labour when markets are not perfectly competitive. The higher labour and/or product market power are, the higher this ratio is. Conversely, in the case of perfectly competitive product and labour markets (i.e., $\eta_i^L = \infty$ and $\eta_i^Y = -\infty$), $\frac{\chi_i^L}{\chi_i^Y} = 1$. Note that profit maximization requires $|\eta_i^Y| > 1$, so that χ_i^Y is always positive. From Equation (4), we obtain the optimal $\frac{L_i}{Y_i}$ as a function of the firm characteristics, which is then replaced into the formula for the firm level labour share (Equation 1), leading to:

$$\lambda_i = \left(\frac{\alpha \chi_i^Y}{\chi_i^L}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i p_i}{w_i}\right)^{\frac{\rho}{1-\rho}}. \tag{5}$$

A few insights are worth pointing out here. First, λ_i does not explicitly depend on the size of the firm (either in terms of K_i or L_i), a property emanating from the homotheticity assumption of a linear expansion path (optimal K_i/L_i and L_i/Y_i ratios are constant). However, a correlation between λ_i and firm size might be observed in practice, provided the other determinants of λ_i (TFP, market power, wages or prices) do depend on the size of the firm. In effect, there is evidence of such correlation, not the least because bigger firms tend to be more productive and have more market power (e.g., Autor et al., 2020; Schwellnus et al., 2018). Additionally, in our framework wages and prices do depend on L_i whenever there is imperfect competition. Second, the direction of the effect on the labour share of all parameters but market power depends on the sign of ρ. For instance, a ceteris paribus increase in TFP increases (decreases) λ_i if ρ is positive (negative). Meanwhile, both higher monopoly power (i.e., a decrease in χ_i^Y) and higher monopsony power (i.e., an increase in χ_i^L) lower λ_i. In the limiting case of $\rho = 0$ (Cobb–Douglas), only market power affects λ_i.[5] Third, there is a close relationship between the pay-productivity disconnect (with productivity understood as TFP) and the labour share, at the firm level. In particular, TFP changes unmatched by wage changes affect the labour share. Again, the direction of the effect depends on ρ.

Finally, while we do not model firms' choice of capital, this does not mean capital is necessarily fixed. Rather, we remain agnostic about the precise capital accumulation mechanism (for instance, in addition to the first order optimality condition for capital, firms might take into account adjustment costs to the capital stock).

2.2 | Heterogeneity and the aggregate labour share

Ultimately, we are interested in the effects of firm heterogeneity on the aggregate labour share. Replacing the individual firm labour share λ_i into Equation (2) leads to the following expression for the aggregate labour share:

$$\lambda = \sum_i \left(\frac{\alpha \chi_i^Y}{\chi_i^L}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i p_i}{w_i}\right)^{\frac{\rho}{1-\rho}} \delta_i. \tag{6}$$

We measure firm heterogeneity with respect to a hypothetical 'average' firm. More specifically, for given relative weights $\{\psi_i\}$ we define $\overline{A} = \sum_i \psi_i A_i$, $\overline{w} = \sum_i \psi_i w_i$, $\overline{p} = \sum_i \psi_i p_i$, $\overline{\chi}^Y = \sum_i \psi_i \chi_i^Y$, $\overline{\chi}^L = \sum_i \psi_i \chi_i^L$. This is, we compute a weighted average of all heterogeneous parameters in the model, which then define the parameters of the benchmark firm.

It is natural to weight variables by some measure of firm size. Whilst employment might seem a reasonable option, there is often significant capital-labour variability at similar employment levels (something which is true in our data too). Since a given level of value added can be achieved with different capital and labour combinations, we consider value added a more suited weighting variable. In effect, value added (or sales) is also often used in the literature to aggregate firms (e.g., De Loecker & Eeckhout, 2018; De Loecker et al., 2020, in the context of mark-ups). Notice however that the method itself is agnostic regarding the weights chosen. What we need is a benchmark against which to quantify heterogeneity, just as the variance is computed with respect to a mean. Having defined weighted averages for every variable we can then re-write the aggregate LS as:

$$\lambda = \lambda^{\text{HOM}} \sum_i \left(\frac{\chi_i^Y}{\overline{\chi}^Y}\right)^{\frac{1}{1-\rho}} \left(\frac{\overline{\chi}^L}{\chi_i^L}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i}{\overline{A}}\right)^{\frac{\rho}{1-\rho}} \left(\frac{\overline{w}}{w_i}\right)^{\frac{\rho}{1-\rho}} \left(\frac{p_i}{\overline{p}}\right)^{\frac{\rho}{1-\rho}} \delta_i, \tag{7}$$

where λ^{HOM} is the labour share of the counterfactual firm, and defined as:

$$\lambda^{\text{HOM}} = \left(\frac{\alpha \overline{\chi}^Y}{\overline{\chi}^L}\right)^{\frac{1}{1-\rho}} \left(\frac{\overline{A}\,\overline{p}}{\overline{w}}\right)^{\frac{\rho}{1-\rho}}. \tag{8}$$

Note that λ^{HOM} is not the average labour share, but the labour share of an optimizing average firm.[6]

Equation (7) is our first decomposition formula, which shows that *any* form of heterogeneity affects the aggregate labour share, with the exception of capital *alone*. If firms differ only with respect to capital, their labour shares are identical (see Equation 5).[7] The proof can be trivially seen in Equation (2), once we assume $\lambda_i = \lambda^{\text{HOM}}$.

The following proposition summarizes the CES result:

> **Proposition 1.** *Assume firms have identical CES technologies* (i.e., α and ρ are the same across firms), *and* $\rho \neq 0$ (i.e., technology is not CD). *Then, it is true that* (i) *heterogeneity in wages, price dynamics, TFP or market power affects the aggregate labour share* (directly and through δ_i); (ii) *heterogeneity in capital affects the aggregate labour share* (through δ_i) *only if other forms of heterogeneity are also present.*

Notice the decomposition formula is purely descriptive of the optimal production plans of the different firms, reflecting the partial equilibrium nature of the model. Yet, provided we can produce an estimate for each element in Equation (7), this is sufficient for our purposes. The drawback of this partial equilibrium approach is, of course, that we cannot provide a deeper understanding of why heterogeneity in wages and prices occurs in the first place.

This result can be contrasted with the Cobb–Douglas (CD) case, where the aggregate LS is:

$$\lambda = \frac{\alpha \overline{\chi}^{Y}}{\overline{\chi}^{L}} \sum_{i} \left(\frac{\chi_i^{Y}}{\overline{\chi}^{Y}} \right) \left(\frac{\overline{\chi}^{L}}{\chi_i^{L}} \right) \delta_i . \tag{9}$$

This highlights that for firm heterogeneity to affect the aggregate labour share under CD technology, there must be *heterogeneous* imperfect competition. With perfect competition (where an exact aggregate production function exists), $\lambda = \alpha$, a well-known property of a CD production function. The following corollary summarizes the result:

Corollary 1. *Assume firms have identical Cobb–Douglas technologies* (i.e., α is the same). *If market power is homogeneous across firms* (including the limit case of perfect competition), *then firm heterogeneity is irrelevant for the aggregate labour share: the labour share is identical across firms* (with perfect competition, it is equal to α). *On the other hand, with heterogeneous market power, firm heterogeneity of any dimension affects the aggregate labour share. In particular, heterogeneity in capital, wages, prices and TFP affect the labour share indirectly through* δ_i.

The above result is very simple but makes an important point, given the extensive use of CD production functions with perfect competition in the literature: even when firms are heterogeneous along many dimensions (including TFP), and an aggregate production function hence does not exist, in competitive markets the aggregate labour share only depends on technology. Conversely, a CES enables a richer set of determinants for the labour share, reason why it was chosen here.[8]

2.3 | Exercise: A mean-preserving increase in the dispersion of one variable

To better illustrate the implications of Equation (7), we now consider a simple scenario where only one dimension of heterogeneity is present. First, we focus on wages. Then, we extend the result to other forms of heterogeneity.

For simplicity, we consider only two (types of) firms $i = \{1,2\}$. We start from a situation where the two firms are identical, with wage w. Since the LS does not depend on firm's size, both firms (and the aggregate economy) have the same labour share, λ. Now, consider an exogenous *value added-weighted* mean-preserving spread in wages. This is a change in wages such that their weighted average (using value added as weights) yields the same original average, w. Mathematically, for new wages $w_1 = w + \Delta_1$ and $w_2 = w - \Delta_2$, this is true if $\Delta_1 = \Delta_2 \frac{\delta_2}{\delta_1}$, where δ_i represents the firm's share of value added in the economy with this new set of wages.[9] In this setting, each firm's LS is (Equation 7):

$$\lambda_i = C w_i^{\frac{-\rho}{1-\rho}}, \tag{10}$$

where $C = \left(\frac{\alpha\chi^Y}{\chi^L}\right)^{\frac{1}{1-\rho}}(Ap)^{\frac{\rho}{1-\rho}}$ (identical across firms). This function is strictly concave and increasing in wages for $\rho < 0$, and strictly convex and decreasing in wages for $\rho > 0$. The case of $\rho < 0$ is depicted in Figure 1. The aggregate LS is a weighted mean of the individual LS, with weights equal to δ_i (Equation 2). Jensen's inequality ensures that the aggregate LS is lower the bigger the dispersion in wages, Δ. In other words, starting from a situation of firm homogeneity, an increase in the dispersion of wages such that the counterfactual firm does not change leads to a fall in the aggregate LS, if the elasticity of substitution between capital and labour is lower than one. Again, notice the limiting case of the CD, where dispersion in wages *alone* does not change the aggregate LS, which is constant over w_i.[10]

A similar analysis holds for other firm-level variables. In the end, all depends on the shape of the equivalent of function $f(\cdot)$ in Figure 1. For most dimensions, this shape depends on the value of ρ. In particular, product market power exhibits the same behaviour than wages. Namely, for $\rho < 0$ ($\rho > 0$), an increase in product market power heterogeneity lowers (raises) the aggregate LS. The relationship is the opposite for TF: for $\rho < 0$ ($\rho > 0$), an increase in the dispersion of productivity leads to an increase (decrease) in the aggregate LS.[11] This contrasting relationship for wages and TFP makes sense. Recall that the firm level LS depends on the pay-productivity disconnect. So, if wages and productivity change in tandem, the effect on the firm LS is muted. This must be reflected also in the aggregate LS. Note the relationship with the theory of superstar firms (Autor et al., 2020). That theory predicts that because some firms become much more productive thanks to new technologies, the aggregate labour share falls. Hence, the theory says both that (i) the average productivity goes up and (ii) the dispersion in productivity goes up. We show that these two things have different implications with respect to the aggregate labour share. An increase in the average productivity pushes—ceteris paribus—the aggregate labour share down (see Equation 5), while the corresponding increase in its dispersion reduces this effect.

Finally, labour market power is an exceptional case because the function $f\left(\chi_i^L\right)$ is strictly convex (and decreasing) for every ρ (even for the CD case of $\rho = 0$; see endnote 5). This means that an increase in the dispersion of labour market power *always* raises the aggregate

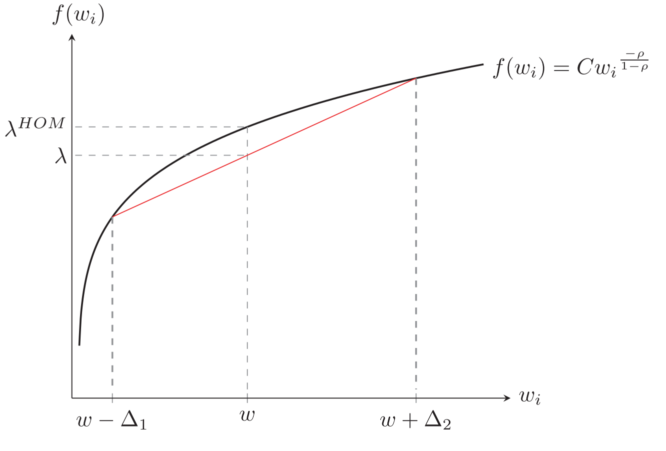


FIGURE 1 A VA-weighted mean-preserving spread of wages, $\rho < 0$.

LS. The intuition of this particular case is also evident. Recall that higher χ_i^L means more monopsonistic power by the firm. Since the labour share represents the proportion of firm's value accruing to workers, it is reasonable to expect that this proportion falls with χ_i^L, regardless of the degree of complementarity between capital and labour. This is why $f\left(\chi_i^L\right)$ is strictly convex and decreasing for every ρ. Conversely, product market power affects directly the 'size of the pie' (value added), and thus its final effect on the labour share does depend on the degree of complementarity between capital and labour.

2.4 | Distributional characterization

Proposition 1 is very general. In particular, it does not quantify how heterogeneity affects the aggregate labour share: the summation term in Equation (7) is obscure enough for this to be seen. In order to shed more light on the issue, we approximate each of the fractions inside the summation term in Equation (7) by means of a second-order Taylor expansion around the respective weighted average. For each $z=\{\chi^Y,\chi^L,A,w,p\}$, this approximation is:

$$\left(\frac{z_i}{\bar{z}}\right)^{\phi} \approx 1+\phi\left(\frac{\Delta_{z,i}}{\bar{z}}\right)+\frac{\phi(\phi-1)}{2}\left(\frac{\Delta_{z,i}}{\bar{z}}\right)^2, \tag{11}$$

where $\bar{z}$ is the weighted mean of the respective variable, and $\Delta z_i = z_i - \bar{z}$ is the deviation from that mean. As shown in Appendix A, after dropping all interaction terms of order higher than two,[12] Equation (7) can be approximated by

$$\lambda \approx \lambda^{HOM} \sum, \tag{12}$$

where $\sum$ is a unidimensional expression that can be summarized in the following matrix multiplication:

$$\sum = \frac{\rho}{2(1-\rho)^2}\begin{bmatrix} \mathrm{CV}(\chi^Y) \\ \mathrm{CV}(\chi^L) \\ \mathrm{CV}(A) \\ \mathrm{CV}(w) \\ \mathrm{CV}(p) \end{bmatrix}^T \begin{bmatrix} 1 & -\frac{r(\chi^Y,\chi^L)}{\rho} & r(\chi^Y,A) & -r(\chi^Y,w) & r(\chi^Y,p) \\ -\frac{r(\chi^Y,\chi^L)}{\rho} & \frac{2-\rho}{\rho} & -r(\chi^L,A) & r(\chi^L,w) & -r(\chi^L,p) \\ r(\chi^Y,A) & -r(\chi^L,A) & 2\rho-1 & -\rho r(A,w) & \rho r(A,p) \\ -r(\chi^Y,w) & r(\chi^L,w) & -\rho r(A,w) & 1 & \rho r(w,p) \\ r(\chi^Y,p) & -r(\chi^L,p) & \rho\rho r(A,p) & \rho r(w,p) & 2\rho-1 \end{bmatrix} \begin{bmatrix} \mathrm{CV}(\chi^Y) \\ \mathrm{CV}(\chi^L) \\ \mathrm{CV}(A) \\ \mathrm{CV}(w) \\ \mathrm{CV}(p) \end{bmatrix}, \tag{13}$$

where $\mathrm{CV}(\cdot)$ is the coefficient of variation, and $r(\cdot)$ is the correlation coefficient, both of which are dimensionless and scale invariant.[13] This term reveals two components of heterogeneity. First, the variability (coefficient of variation) of each dimension (market power, wages, etc), given by the outer vectors. Trivially, if there is no variability in a given dimension, $\mathrm{CV}(\cdot)=0$, and thus no contribution to heterogeneity. Second, the (symmetric) correlation matrix across dimensions, weighted by a function of ρ. The different signs map the sign of their correlation in the individual labour share formula, sometimes depending on the value of ρ (see Equation 6).

Notice that if two dimensions are not correlated, they only affect the heterogeneity term independently (through their CV). If two dimension are correlated, then heterogeneity gets amplified (or diminished) depending on the sign of their correlation and ρ. Importantly, heterogeneity matters even if all variables are orthogonal to each other, i.e., if all correlations are zero.

Our result is presented in the following proposition.

Proposition 2. *If firms have a CES technology with identical α and ρ and constant returns to scale, the aggregate labour share is approximately given by Equation* (12). *Hence, the effect of firm heterogeneity on the labour share depends on the joint distribution of all firm-level variables, and for most of the variables, on ρ. The total effect can be separated in two components, a direct effect unaffected by the correlation structure among variables, and an interaction effect that depends on the correlation structure. The direct effect is such that, for the empirically relevant case* (i.e., when the elasticity of substitution between capital and labour is smaller than 1, that is to say, $\rho < 0$), *and other things being equal, an increase in the dispersion of productivity or monopsony power increases the aggregate labour share, while an increase in the dispersion of wages or product market power decreases it. The interaction effect holds that ceteris paribus changes in the correlation structure affect the labour share. In particular, for $\rho < 0$, an increase in the correlation between labour market power and TFP, or between product market power and wages increases the aggregate labour share, whereas an increase in the correlation between product market power and labour market power, product market power and TFP, labour market power and wages, or TFP and wages decreases it.*

The implications of Proposition 2 are somewhat difficult to visualize, as any change in firm level variables will typically trigger a change in the market shares, δ_i. This in turns will cause a change in λ^{HOM}, which refers to the hypothetical labour share of a weighted average firm. Moreover, coefficients of variation and correlations will change too (since they are weighted by δ_i). Hence, the ceteris paribus clause typically will not hold in simple thought experiments. However, the above approximation allows us to generalize and quantify the unidimensional heterogeneity exercise presented in Section 2.3 beyond two (types of) firms, to the case of many heterogeneous firms. For instance, in the case of nominal wage heterogeneity only, Equation (12) simplifies to:

$$\lambda \approx \lambda^{\text{HOM}} \left[1 + \frac{\rho}{2(1-\rho)^2} \text{CV}^2(w) \right]. \tag{14}$$

We see here that an increase in firm heterogeneity (defined in terms of the coefficient of variation) lowers the aggregate labour share when $\rho < 0$, as predicted in Section 2.3. Similar parallels exist for the other firm dimensions.

3 | DATA

3.1 | Sample

Equation (12) provides a model-based decomposition of the aggregate labour share in terms of firm heterogeneity vis-a-vis a counterfactual firm. We apply this decomposition to the manufacturing sector in Great Britain (UK without Northern Ireland), for the period 1998–2014.[14,15] We focus on the manufacturing sector because value added is very imperfectly measured in other sectors, where intermediate inputs are less clearly identified (e.g., see the discussion in Autor et al., 2020).

We use data from the 3rd edition of the Annual Respondent Database (ARD), which contains a census of all enterprises with at least 250 employees, plus a sample of all those firms with less than 250 employees.[16] The dataset has information both at the plant and 'reporting unit' level. The latter is the smallest unit that contains detailed financial information needed for the analysis (like labour costs, investment, and so on), and so it is our working definition of firm. Still, most of firms only have one plant (e.g., 97 per cent in 2014).

Several sample selection procedures were made. First, as suggested by Schwellnus et al. (2017), sub-sector 19 in the SIC07 classification ('manufacturing of coke and refined petroleum') was dropped, because of the noise introduced by the volatility of oil prices. Second, firms with less than 10 employees were dropped. This is because for small firms (and particularly for firms with 1 or 2 employees, the bulk of those dropped) the level of wages might not so much be associated with market mechanism, as both capital income and labour income can be used to reward the firm's owners (for instance for fiscal reasons). This might distort the computation of the labour share in ways unrelated to the theory.[17] Third, non-profit and other non-market oriented firms were excluded, as these are less likely to be characterized by profit maximizing behaviour. Fourth, firms with missing information (e.g., no investment data, needed to compute capital stocks) were also dropped. Fifth, outliers in terms of top and bottom 0.5 per cent percentiles, computed independently for different variables (including the firm level labour share, Y/L and L/K), were discarded. The final sample used contains 115,150 observations, covering around 38,000 unique firms. The analysis is then carried out using turnover-based sampling weights, in order to represent the whole sector as good as possible.

3.2 | Variables not in ARD

Although ARD is a very rich dataset, in terms of our needs it only contains information on number of employees, total labour costs (including pension funds contributions) and value added (the latter either directly available, or computed using gross output and intermediaries, when missing). Therefore, we need to either add or produce our own estimates for the remaining terms, namely firm-level prices, TFP, production function parameters (α and ρ), and product and factor market power. We also need to impute the capital stock of the firms.

3.2.1 | Prices

Our theory is build upon firm-level prices; however, no price information is available in ARD. Simple algebra (see Appendix B) shows that it is the volatility in the relative price changes, not in the level of prices, that matters for our decomposition. Consequently, we use the most disaggregated (4 digits) industry-level producer price *index* available, as provided by the Office for National Statistics.[18] The use of an index has the advantage of making firm-level prices comparable. The same issue of comparability arises with TFP, whose units of measurement are directly related to those of output. It can be shown that using a price index also makes TFP comparable, because the base term used in the price index is captured by the predicted TFP, obtained from a regression where value added is deflated by the same price index.

3.2.2 | Estimation of TFP, α and ρ

As extensively noted in the literature (e.g., Olley & Pakes, 1996), when estimating production functions it is necessary to account for the potential endogeneity of employment which, being a variable factor, might respond to contemporary unobserved shocks to TFP. In order to estimate our CES value added production function (Equation 3) we follow the dynamic panel method proposed by Blundell and Bond (2000). In this method, unobserved TFP is assumed to follow an AR(1) with parameter θ, and the model is then θ-differentiated, and estimated with GMM.[19] This dynamic panel approach is preferred to the, also common, control function method, because the latter is more demanding on the data, reducing the sample size.

Full details of the estimation method are presented in Appendix C. Here we just highlight that the estimated elasticity of substitution (for the manufacturing sector as a whole) is 0.46 ($\widehat{\rho} = -1.18$), significant at the 1 per cent confidence level. This elasticity implies capital and labour are gross substitutes, a result that is generally consistent with other firm-level evidence (an example using UK data is Barnes et al., 2008). Importantly, the firm-level capital stock is not available in the data, and yet it is required for estimating the production function. We therefore impute capital using a combination of the perpetual inventory method and information from the capital stock for the whole sector, obtained from the Office for National Statistics. See Appendix C for further details.[20] Finally, having estimated α and ρ, we can use the production function to compute $\widehat{A}_{it}$ as a residual.[21] This can be done also for observations not used in the estimation of the production function (e.g., because of missing data in a given year). This means that the final sample used for the decomposition is larger than the one used for estimation. For details, see again Appendix C.

3.2.3 | Market power

Labour and product market power are defined in terms of labour supply and output demand elasticities, respectively. As these are not directly observable, we calibrate χ_i^L and χ_i^Y using proxies. For labour market power, we start by measuring the employment share of each firm in the local labour market they are situated: this share is computed for each occupational group, after which a weighted average is produced for each firm.[22] The local labour market is understood to be a 'travel to work area' (TTWA).[23] The final measure of hiring concentration ranges between 0 and 1. We then need to map the measure of monopsony power derived above (which we denote as s_{it}) into the labour supply elasticity faced by the firm, η_{it}^L. The method we use is relatively simple. Notice that the elasticity of supply is a number that goes between 0 and ∞. Therefore, any relationship between s_{it} and η_{it}^L must be such that, in competitive markets, $s_{it} \approx 0$, whereas in complete monopsony power, $s_{it} \approx 1$. Albeit there are several functional forms producing such relationship, a flexible one is

$$\eta_{it}^L = -c_1 \left(\frac{1}{ln(1 - s_{it})} \right)^{c_2}. \tag{15}$$

The parameters c_1 and c_2 are chosen in order to match the scarce evidence available in the literature about η_{it}^{L} at the firm level. First, Manning (2003) estimates an average firm level elasticity of supply for the UK of around 0.75. Second, Webber (2015) provides a characterization of the distribution of this elasticity for the US, from where it is possible to calibrate with decent fit a log-normal distribution of this elasticity.[24] For lack of a better alternative, we assume the UK also follows this distribution, but scaled to match the UK average estimated by Manning (2003).[25] This enable us to compute c_1 and c_2 (found to be 0.01 and 0.37, respectively).[26]

Regarding product market power, our theory defines this in terms of the firm's elasticity of demand. Albeit this is also unobserved, there is a direct relationship between this elasticity and the mark-up (price over marginal cost). In particular, under monopolistic competition, if the mark-up of a firm is μ, with $\mu \geq 1$, its elasticity of demand is $\eta^{L} = -\frac{\mu}{\mu - 1}$. We compute the firm-level mark-up as the sales to total variable costs' ratio, which approximates marginal costs with average (e.g., Branston et al., 2014; De Loecker et al., 2020).

3.3 | Theoretical versus empirical decomposition

The decomposition formulas are built upon the optimisation behaviour of firms. Thus, they refer to the *predicted* labour share of firms, as given by Equation (5). However, the objective of the decomposition is to characterize *observed* labour shares (in terms of observed value added and labour costs). Naturally, there will be differences between these two, either because of theoretical problems (e.g., specification errors), empirical problems (e.g., biased estimation of parameters), measurement errors, or the very stochastic nature of production. This discrepancy between predicted and observed labour share introduces an extra term into the decomposition.[27] To see this, let us define $\tau_{it} \equiv \frac{\lambda_{it}^{obs}}{\lambda_{it}}$, which captures the divergence between the observed and predicted labour share for firm i in period t, where the latter is given by Equation (5). The connection between the model and data is done through τ_i. This is, we can write (the time index is omitted):

$$\lambda^{\text{obs}} \equiv \sum_i \lambda_i^{obs} \delta_i^{obs} = \sum_i \lambda_i \tau_i \delta_i^{obs}, \tag{16}$$

Replacing the predicted firm level labour share λ_i by its components (equation 5), and introducing the counterfactual λ^{HOM}, produces:

$$\lambda^{\text{obs}} = \lambda^{\text{HOM}} \sum_i \left(\frac{\chi_i^{Y}}{\overline{\chi}^{Y}}\right)^{\frac{1}{1-\rho}} \left(\frac{\overline{\chi}^{L}}{\chi_i^{L}}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i}{\overline{A}}\right)^{\frac{\rho}{1-\rho}} \left(\frac{\overline{w}}{w_i}\right)^{\frac{\rho}{1-\rho}} \left(\frac{p_i}{\overline{p}}\right)^{\frac{\rho}{1-\rho}} \left(\frac{\tau_i}{\overline{\tau}}\right) \delta_i^{obs}, \tag{17}$$

where

$$\lambda^{\text{HOM}} = \overline{\tau} \left(\frac{\alpha \overline{\chi}^{Y}}{\overline{\chi}^{L}}\right)^{\frac{1}{1-\rho}} \left(\frac{\overline{A}\overline{p}}{\overline{w}}\right)^{\frac{\rho}{1-\rho}}, \tag{18}$$

and $\overline{\tau} \equiv \sum_i \psi_i \tau_i$, a weighted average of the discrepancy term. We then get to an approximation of the decomposition similar to that of Equation (12), but including four extra terms representing the correlation of τ with productivity, labour and product market power and real wages, multiplied by the respective coefficients of variations.[28] It is this 'discrepancy-adjusted' equation (12) that we take to the data. Notice that heterogeneity in τ itself does not affect the labour share, unless it is correlated with other factors. Also, if the discrepancy is constant across firms (i.e., $\mathrm{CV}(\tau) = 0$), the only difference between the theoretical and empirical decomposition is that λ^{HOM} is multiplied by the average discrepancy term, as shown in Equation (18).

4 | RESULTS

4.1 | Descriptive analysis of the labour share

In our sample, we observe a net fall of the labour share over the period, from 0.58 in 1998 to 0.53 in 2014 (see Figure 3 below).[29],[30] It's interesting to notice an initial period of increase in the labour share (peaking at 0.61 in 2003), and a subsequent fall, with a minor interruption during the financial crisis. Figure 2 compares the distribution of the unweighted and (value added) weighted labour share at the beginning and at the end of our period. Panel (A) shows that the sample distribution of firms' labour share in 2014 has more mass at lower levels than in 1998. Similarly, Panel (B) shows that in 2014 more value added was produced by firms with lower labour share than in 1998.[31] Importantly, the fall in the level of the labour share has not been a homogeneous phenomenon. In effect, the upper tail of the distribution barely changed between the 2 years. This reflects an increase in the dispersion of the labour share.

Before presenting the results for our model-based decomposition, it is informative to discuss those from a simple statistical decomposition. We compute the ratio between the weighted (aggregate) and the unweighted (average) labour share, which is a measure of the correlation between λ_i and δ_i (details in Appendix E). If this correlation is positive (negative), the ratio will be above (below) one. If the two variables are uncorrelated, the ratio is one: unweighted and weighted labour share are the same. As said earlier, smaller firms have higher labour share, so this correlation is negative. We find no trend in this variable (results not shown), suggesting that most of the change in the aggregate labour share is due to the fall in the level of the labour share across the firm size spectrum. The statistical decomposition, however, has little to say about the drivers of such change. Not only our decomposition approach can shed light into the

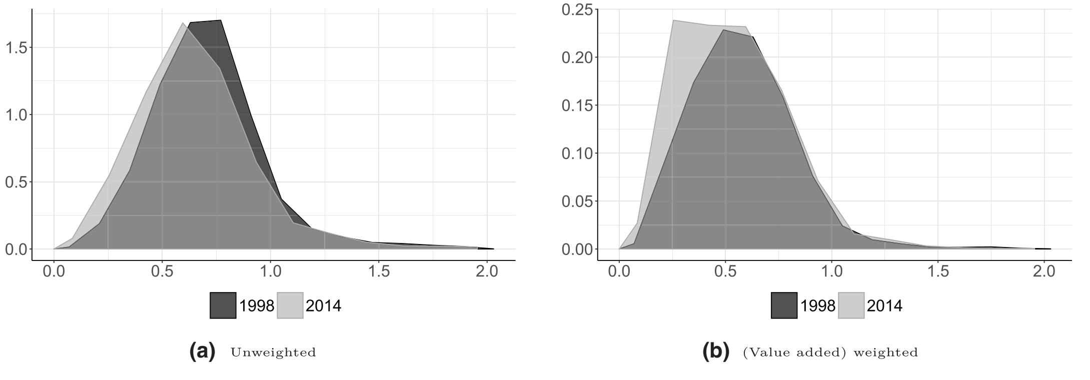


FIGURE 2 (a) Unweighted and (b) (value added) weighted distribution of the labour share, 1998 and 2014. *Source*: Our calculation based on ARD data. *Sample*: UK manufacturing firms with 10 employees or more.

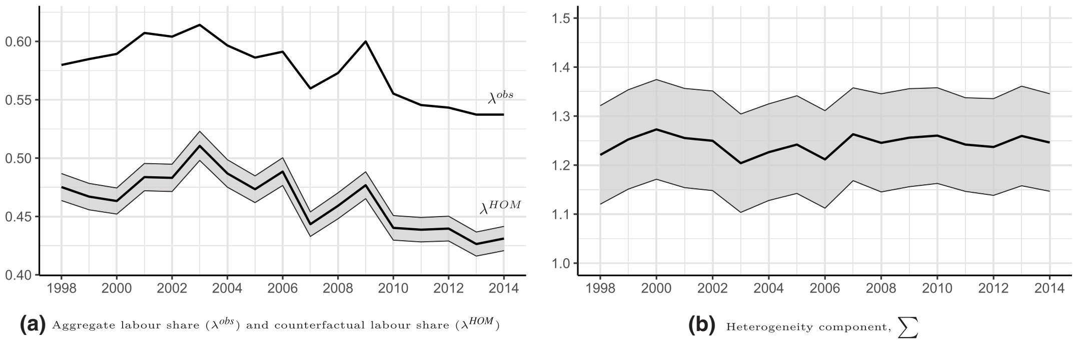


FIGURE 3 Decomposition of the aggregate labour share. (a) Aggregate labour share (λ^{obs}) and counterfactual labour share (λ^{HOM}). (b) Heterogeneity component, Σ. The graphs display weighted firm averages. 95 per cent confidence intervals are displayed as a shadowed area (except for observed aggregate labour share). *Source*: Our calculation based on ARD data. *Sample*: UK manufacturing firms with 10 employees or more.

evolution of the labour share of the representative firm, but we can also dig further into the flat trend for the impact of heterogeneity. As we will see, this is due to conflicting trends in the underlying determinants, which offset each other.

4.2 | Decomposition of the aggregate labour share

Equation (17) decomposes the aggregate labour share in terms of λ^{HOM} (i.e., the labour share of a counterfactual 'representative' firm) and $\sum$ (i.e., a quantification of firms' multi-dimensional dispersion with respect to that counterfactual firm). The decomposition for the manufacturing sector as a whole is depicted in Figure 3 (see Appendix F for an equivalent analysis at the sub-sectoral level). In all the analysis that follows we assume value-added weights.

Two things are important to notice. First, the level of the aggregate labour share is significantly different when compared with the level of the counterfactual, average firm. This is, if all firms were identical to the average firm, the labour share would be significantly smaller. This result supports the theoretical result of the paper, namely that heterogeneity matters. Second, as Panel (B) shows, the role of heterogeneity has been relatively stable over the period: changes in firm heterogeneity has not been a major driver of the movements of the aggregate labour share observed over the period. Importantly, this does not mean firm heterogeneity has not changed. As shown later, changes have partly offsetted each other.

To quantify the role of firm heterogeneity vis-a-vis λ^{HOM} on changes in λ^{obs}, we carry out a simple growth accounting decomposition of $\lambda^{\text{obs}} = \lambda^{\text{HOM}} \sum$:

$$g_{\lambda^{\text{obs}}} = g_{\lambda^{\text{HOM}}} + g_{\Sigma} + \text{interaction effect},$$

where g_z stands for the growth rate of factor Z, over a given period. Table 1 presents the result of this exercise for the entire period and for two sub-periods, pre- and post-2003 (the year that the labour share reached its highest level). In all cases, we see that the bulk of the change in the

TABLE 1 Disaggregation of the growth rate of the aggregate labour share (λ^{obs}).

Period	$g_{\lambda^{\text{obs}}}$ %	$g_{\lambda^{\text{HOM}}}$	$g_{\sum}$	Interaction
1998–2003	6.04	7.43 [7.41, 7.45]	−1.29 [−0.41, −1.17]	−0.10 [−0.21, 0.01]
2003–2014	−12.63	−15.58 [−15.60, −15.56]	3.50 [3.38, 3.62]	−0.54 [−0.65, −0.43]
1998–2014	−7.36	−9.31 [−9.33, −9.29]	2.16 [2.08, 2.28]	−0.20 [−0.31, −0.09]

Note: $g_{\lambda^{\text{obs}}} = g_{\lambda^{\text{HOM}}} + g_{\sum} +$ interaction effect. 95 per cent confidence intervals in parenthesis.
Source: Our calculation based on ARD data.
Sample: UK manufacturing firms with 10 employees or more.

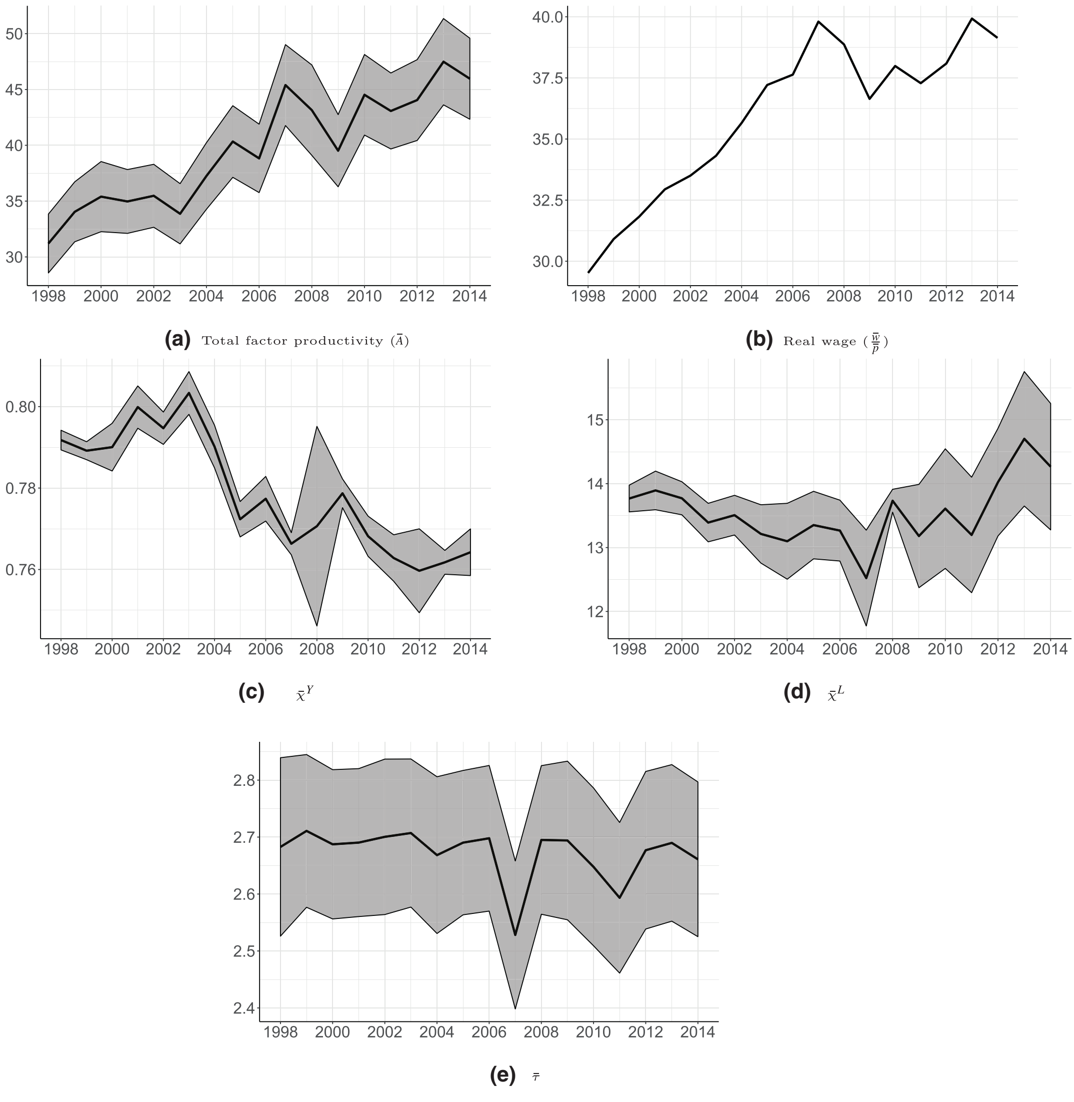
labour share has been due to λ^{HOM}. Meanwhile, $\sum$ has partly counteracted the effect of the former, from 9 per cent to 7 per cent.

4.3 | Decomposition of the labour share of the representative firm

The homogeneous labour share (Equation 18) can be further analysed by looking at its constituent elements.[32] Figure 4 presents the evolution of the different variables for the whole manufacturing sector (again, see Appendix F for a sub-sectoral analysis).[33] Panel (A) shows a fairly unstable but overall increase in TFP over the period (trend interrupted by a 2008–2009 dip). Real wages (Panel B) show a stable pre-2008 growth, with a subsequent dip (particularly in 2009). Interestingly, such growth rate slowed down post-2008, a trend consistent with ONS aggregate data. Product market power (Panel C) has increased over the period (recall lower χ^Y means more product market power), albeit also not in a steady fashion.[34] Labour market power (Panel D) fell in the early years of the period, and subsequently increased post-2008 (recall lower χ^L means less less market power for the firm). Lastly, the discrepancy term $\overline{\tau}$ (Panel E) is fairly stable (except for 2007; see endnote 31), meaning this is unlikely to drive any of the results.[35] Table 2 presents the growth rates of each variable in Figure 4, over the subperiods of interest. As Equation (18) indicates, the effect of these variables on λ^{HOM} is mediated by ρ. In order to see the final effect of each of these variables on λ^{HOM}, we carry out a growth accounting decomposition of Equation (18). This decomposition is given by

$$g_{\lambda^{\text{HOM}}} = \left(\frac{\rho}{1-\rho}\right) g_{\overline{A}} - \left(\frac{\rho}{1-\rho}\right) g_{\overline{w}/\overline{p}} + \left(\frac{1}{1-\rho}\right) g_{\overline{\chi}^Y} - \left(\frac{1}{1-\rho}\right) g_{\overline{\chi}^L} + g_{\overline{\tau}}. \quad (19)$$
$$+\text{interaction effect}$$

Table 3 shows the resulting contribution of each component of λ^{HOM}. It can be seen that real wages and productivity have different growth rates over the period, with an overall increase in the disconnect. This gap explains most of the actual change in λ^{HOM}: in fact, if we impose the same annual growth rate of real wages observed between 1998 and 2003 (3.1 per cent) for



(a) Total factor productivity ($\bar{A}$)

(b) Real wage ($\frac{\bar{w}}{\bar{p}}$)

(c) $\bar{\chi}^Y$

(d) $\bar{\chi}^L$

(e) $\bar{\tau}$

FIGURE 4 Evolution of the components of λ^{HOM} (a–e: total factor productivity, real wages, product market power, labour market power and the discrepancy term, respectively). The graphs display weighted firm averages. 95 per cent confidence intervals are displayed as a shadowed area (except for real wages, which are observed). *Source*: Our calculation based on ARD data. *Sample*: UK manufacturing firms with 10 employees or more.

TABLE 2 Growth rates of the components of λ^{HOM}.

Period	$g_{\bar{A}}$	$g_{\bar{w}/\bar{p}}$	$g_{\bar{\chi}^Y}$	$g_{\bar{\chi}^L}$	$g_{\bar{\tau}}$
	%				
1998–2003	8.54	16.24	1.46	−4.02	0.91
2003–2014	35.72	14.05	−4.87	7.96	−1.71
1998–2014	47.32	32.57	−3.48	3.62	−0.81

Source: Our calculation based on ARD data.
Sample: UK manufacturing firms with 10 employees or more.

TABLE 3 Determinants of the growth rate of λ^{HOM}.

Period	$\boldsymbol{g}_{\lambda^{HOM}}$	$\overline{\boldsymbol{A}}$	$\overline{\boldsymbol{w}}/\overline{\boldsymbol{p}}$	$\overline{\chi}^{\boldsymbol{Y}}$	$\overline{\chi}^{\boldsymbol{L}}$	$\overline{\tau}$	**Interaction**
	%						
1998–2003	7.43	−4.62	8.79	0.67	1.85	0.91	−0.17
	[7.41, 7.45]	[−4.66, −4.54]	[8.67, 8.91]	[0.66, 0.68]	[1.82, 1.88]		[−2.53, 2.19]
2003–2014	−15.58	−19.34	7.61	−2.23	−3.65	−1.71	3.74
	[−15.60, −15.56]	[−19.61, −19.07]	[7.50, 7.72]	[−2.25, −2.21]	[−3.71, −3.59]		[1.80, 5.68]
1998–2014	−9.31	−25.61	17.63	−1.60	−1.66	−0.81	2.73
	[−9.33, −9.29]	[−25.95, −25.25]	[17.35, 17.84]	[−1.61, −1.59]	[−1.69, −1.63]		[0.67, 4.80]

Note: Effects are attributed as per Equation (19). 95 per cent confidence intervals in parenthesis.
Source: Our calculation based on ARD data.
Sample: UK manufacturing firms with 10 employees or more.

the post-crisis period, no pay-productivity disconnect would have emerged over the period, virtually muting any change in λ^{HOM}, ceteris paribus.

Table 3 also indicates that (product and labour) market power contributed to the fall in the labour share. However, as commented earlier in relation to Figure 4, there is a marked difference in the effects of market power within the whole period. Between 1998 and 2003, both product and labour market power fell, albeit only slightly (reflected in higher and lower $\overline{\chi}^{Y}$ and $\overline{\chi}^{L}$, respectively). The second sub-period is characterized by a marked reversal of this initial timid trend. By 2014, both measures of market power are significantly higher than in 1998, jointly pushing for a 3.09 per cent fall in the aggregate labour share.[36]

4.4 | Firms' heterogeneity

The last decomposition exercise focuses on the heterogeneity component, $\sum$. As Equation (12) shows, this is a function of the coefficient of variation and the correlation among variables. Before carrying out this decomposition, it is then interesting to evaluate these elements.

Figure 5 characterizes firm heterogeneity in the five different dimensions under study, and shows no drastic changes over the period. TFP and real wages moved towards less heterogeneity (with some oscillation over the period), whereas product market power heterogeneity increased over the period; labour market power remained relatively stable, except for the artificial jump in 2007 mentioned earlier.

The second important element of firm heterogeneity refers to the correlation among variables across firms over time. Most of them are fairly close to zero, with little variation over time. Noticeable exceptions are (i) a positive correlation between TFP and real wages; (ii) a negative correlation between product market power and TFP; (iii) a positive correlation between labour market power and real wages. The latter increased from almost zero in 1998 to 0.40 in 2014. The link between these two variables seems counter-intuitive, but it reflects the fact that larger firms tend to have both higher wages and labour market power.

Having provided some background evidence regarding the structure of the joint distribution of firms' characteristics, we proceed to the decomposition of $\sum$.

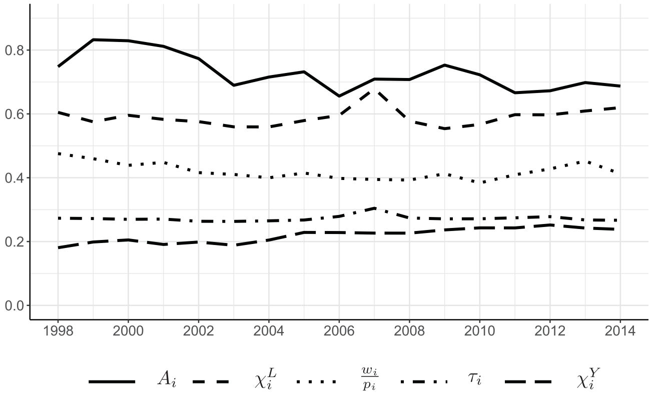


FIGURE 5 Evolution of the dispersion of firm-level variables. The graph displays the coefficient of variation of the variables. *Source*: Our calculation based on ARD data. *Sample*: UK manufacturing firms with 10 employees or more, ARD data.

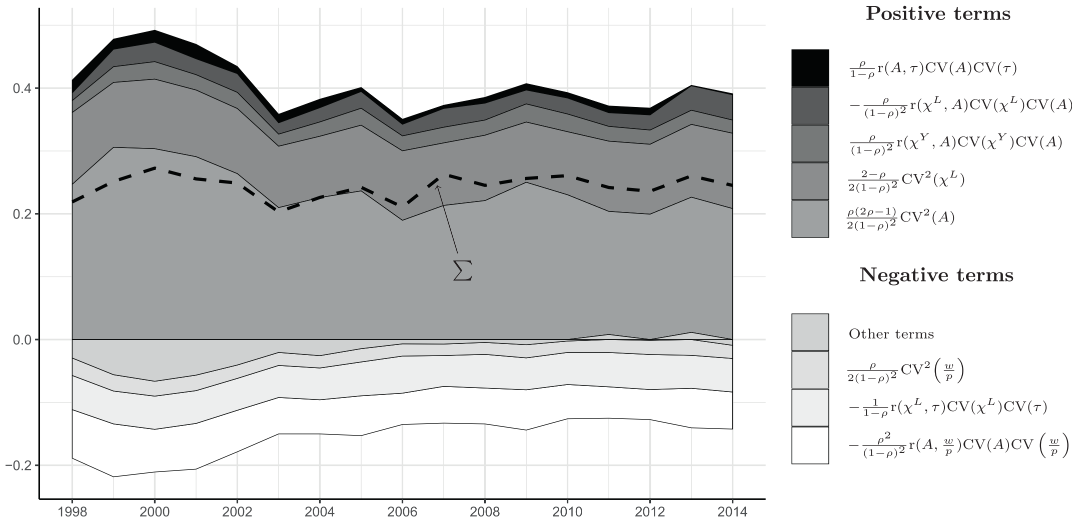


FIGURE 6 Stacked area plot for the decomposition of $\sum$. As Equation (12) shows, $\sum$ is centred around one. For ease of visualization, here we centre it around zero. Positive (negative) terms are those above (below) zero, thereby increasing (decreasing) $\sum$. 'Other terms' encompasses all terms in the discrepancy-adjusted Equation (12) inside $\sum$ not listed in the plot. It also considers all higher order terms not part of the approximation. Adding up all positive and negative terms yields $\sum$. *Source*: Our calculation based on ARD data. *Sample*: UK manufacturing firms with 10 employees or more.

Figure 6 presents the evolution of the different components of $\sum$.[37] What this figure shows is that the bulk of the effect of firm heterogeneity on the aggregate labour share is due to two elements, namely TFP and labour market power. Figure 5 has already shown these are the dimensions with the highest variability. Figure 6 shows that they also have the biggest impact on the labour share, taking into account the effect of the elasticity of substitution parameter ρ. Variability in the real wage is of second order of importance (and its effect goes in the other direction), whereas variability in product market power is completely irrelevant (its value averages -0.006 over the period), as it is that of τ_i, which in itself does not affect λ (see the discussion in Section 3.3).

It is worth pointing out that the terms not explicitly mentioned in the decomposition (included in 'Other terms') are mostly irrelevant for the labour share. Crucially, this component includes every other term excluded from the approximation in the discrepancy-adjusted Equation (12). It is therefore revealing to see that our approximation is sufficient for capturing the bulk of the changes in $\sum$, at least in this empirical application.

Another relevant result from the decomposition presented in Figure 6 refers to what in our theoretical section was defined as the direct versus interaction effect of heterogeneity on the labour share. Results suggest that the direct effects (the terms containing coefficients of variation alone) are much more relevant than the interaction effects (the terms containing a correlation term). This is at least true for TFP and labour market power, with large direct effects that go in the same direction, and smaller interaction effects that offset each other. For real wages, direct and interaction effects are roughly similar in size, both operating in a negative direction. Product market power is the exception, in that the interaction effect is much more relevant than the direct effect. This is due to the interesting combination of a relatively low coefficient of variation and a very high correlation between product market power and TFP (which itself has a high coefficient of variation).

To conclude this section, we can see what type of heterogeneity matters the most for the aggregate labour share (TFP and χ_i^L), what matters the least (χ_i^Y and τ_i), and what matters in between (wages and prices).

5 | CONCLUSIONS

This paper presented a novel approach to study the aggregate labour share, without relying on an aggregate production function. The method is based on a simple, yet insightful enough model of firm behaviour, which allows for a detailed decomposition of the aggregate labour share in terms of different dimensions of firm heterogeneity (TFP, product and labour market power, wages and output prices). The method characterizes the aggregate economy by means of a weighted average firm, and quantifies heterogeneity with respect to such average. The main theoretical result presents the conditions under which firm heterogeneity affects the labour share. The role of the joint distribution of firm-level variables is captured in the decomposition formula in terms of the coefficient of variation for each variable and the correlation among variables. Importantly, the paper shows that firm heterogeneity matters, and this remains invisible when using models based on an aggregate production function. In this sense, our model provides a bridge between the micro and the macro approach to the analysis of the labour share.

To prove the value of the method, we applied the decomposition to a firm level dataset from the UK manufacturing sector, covering the 1998–2014 period. Descriptively speaking, the data indicates that the aggregate labour shares fell around 7 per cent over the period, something that seems related mostly to a generalized fall in the firm level labour share across the firm size spectrum. Albeit the distribution of the labour share moved towards the left, the upper tail remained stable, implying an increase in the dispersion of the labour share.

Analysis showed that —contrary to a narrative focussing on increasing disparities between firms— the observed decline in the aggregate labour share over the period is driven almost entirely by the decline in the labour share of the representative firm, mostly due to an increasing disconnect between average productivity and real wages. Changes in the dispersion of firm-level variables have contributed to slightly contain this decline.

More specifically, the decomposition exercise produced two results. First, firm heterogeneity has a significant impact on the aggregate labour share: the weighted average labour share is around 10 points lower than the aggregate labour share. Second, the fall in the aggregate labour share (7.3 per cent over the period) is mostly accounted for by changes in the weighted average labour share. Indeed, the fall in the weighted average labour share is even bigger (9.3 per cent), indicating that the change in the dispersion of the firm-level determinants of the labour share has softened the downward trend.

Then, we provide further insights on the drivers of the observed fall in the weighted average labour share. We show that the pay-productivity gap widened over the period (particularly after 2003), which alone can explain 90 per cent of the change in the weighted average labour share (8.3 out of 9.3 percentage points). Firm market power (in the product and labour market) grew somewhat over the period too (particularly after the Great Recession), also contributing to the lower labour share.

Lastly, we look deeper into the factors that produce the wedge between the weighted average labour share and the aggregate labour share. This is, we look at what type of heterogeneity matters. The analysis reveals that TFP and labour market power are the two key sources of heterogeneity driving the wedge. The least relevant dimension is product market power heterogeneity (which is fairly low), with wages and price dispersion somewhere in between. This result seems intuitive enough. TFP and labour market power reflect phenomena which are much more difficult to arbitrate across space and time (e.g., because of some organizational knowledge specific to the firm, or the reduced mobility of workers across space). Conversely, product market power and real wages are rooted in prices, which by definition can adjust much quicker across space and time. Different degrees of persistence matter.

Some issues remain to be solved. In particular, even though our analysis benefits from relatively low degrees of (bi-variate) correlation across variables, our approach is still that of partial equilibrium. To get a more fundamental grasp of the deep drivers of our results, a general equilibrium analysis would be needed, something we leave for future research.

ENDNOTES

[1] A review of this macro approach is outside the scope of this work. It is, however, interesting to consider what this literature has identified as the main determinants of the fall in the aggregate labour share, from (capital augmenting) technological change and automation (e.g., Acemoglu & Restrepo, 2018; Autor & Salomons, 2018; Bergholt et al., 2022; Eden & Gaggl, 2018) to the decline in the price of capital relative to labour (e.g., Karabarbounis & Neiman, 2014), increased factor substitutability between capital and labour (e.g., León-Ledesma & Satchi, 2018), a productivity slowdown (e.g., Grossman et al., 2017), increased asset prices that lower investment (e.g., González & Trivín, 2019), deregulation of product and labour markets, including privatization policies, de-unionization and the decline of employment-protection policies (e.g., Bental & Demougin, 2010; Weil, 2017), the increase in the cost of housing and the related increase in the value of capital and in real estate profits (e.g., Rognlie, 2015), and population ageing (Hopenhayn et al., 2022).

[2] The statistical decomposition approach includes the extension of the Olley-Pakes decomposition of aggregate productivity to the labour share (see Autor et al., 2020).

[3] Another related paper is Aghion et al. (2022), which offers a computable general equilibrium model with two types of firms. This framework however does not focus explicitly on the role of firm heterogeneity on the aggregate labour share, but how certain model parameters affect the latter.

[4] The existence of a value-added production function hinges on some assumption about the underlying gross output production function (which relates capital, labour and intermediate inputs to gross output), as Bruno (1978) demonstrated. In particular, the elasticity of substitution between intermediate inputs and the rest of inputs (in our case, capital and labour) must be either zero (i.e., a Leontief) or infinity (i.e., a linear production function). Alternatively, a value-added production function is well defined when the relative price of intermediate inputs to output is constant.

[5] In particular, the labour share is equal to $\frac{\alpha\chi_i^Y}{\chi_i^L}$. Perfect competition yields the familiar result that $\lambda_i = \alpha$.

[6] Looking at the labour share of an 'average' firm is consistent with a model-based decomposition approach, while looking at the average labour share as in the Olley-Pakes decomposition is consistent with a statistical decomposition approach.

[7] Incidentally, this is exactly the case where an aggregate production function exists, namely when firms only differ in their size. Because they have identical K/L ratios, it is possible to mechanically redistribute factor of productions among them without altering factor prices (abstracting from competition considerations). Equivalently, it is possible to combine all firms into one big firm; the production function of this firm 'becomes' the aggregate production function of the economy.

[8] It is interesting to note that—according to our results—the most important source of heterogeneity in explaining aggregate movements of the labour share is precisely TFP, which in a CD setting has no effects.

[9] One might suggest here that the aggregate demand for labour in the two scenarios has not been restricted to be the same. However, the labour supply has not been restricted either (in fact, nothing has been said about

the source of the change in wages). In our partial equilibrium setting, we assume any resource constraints are fulfilled, and wages represent an equilibrium.

[10] To understand why the LS function depends on ρ, let's first look at the first derivative, and explain why the LS is increasing in wages for $\rho < 0$, and decreasing for $\rho > 0$. Consider first the case of $\rho < 0$, where there is relatively low degree of substitution between capital and labour. Starting from a given wage w, an increase in such wage by Δ produces a fall in employment and in value added. Yet, because of low substitution between K and L, such fall in output is relatively significant. Thus, L/Y falls (because of CRS), but not so much. In fact, precisely because of this low substitution, the firm labour share actually increases (recall the labour share is $\frac{w}{p}\frac{L}{Y}$). This is, the 'price effect' outweighs the 'quantity effect'. Conversely, if $\rho > 0$ (high substitution), L/Y falls considerably more, in which case the quantity effect dominates and the labour share falls. In the CD case, these two effects cancel out. Let's now look at the second derivative, and explain why the LS is concave in wages for $\rho < 0$, and convex for $\rho > 0$. Consider again the case of $\rho < 0$. As we said, an increase in the wage from w by Δ lowers L/Y by relatively little. As we further increase wages by Δ, L/Y falls again, but because of decreasing marginal product of labour, the overall change in Y gets smaller, and therefore L/Y falls (again because of CRS) in an increasing fashion, as employment just cannot raise output fast enough. In turn, the price effect of higher w, which always outweighs the quantity effect for $\rho < 0$, is less capable of rising the labour share. This effect plateaus in the limit (i.e., as $w\infty \rightarrow$); hence, its concavity. The argument is the same for the case of $\rho > 0$. Recall that when $\rho > 0$ the LS is decreasing with wages, as the quantity effect outweighs the price effect. Yet, because of decreasing marginal product of labour, such outweighing looses force with w_i and it plateaus in the limit; hence, its convexity.

[11] This result runs counter to Gouin-Bonenfant (2022), who however points to a different mechanism whereas productivity dispersion effectively shields high-productivity firms from wage competition.

[12] For instance, terms like $\frac{\Delta A_i}{\overline{A}} \frac{\Delta \chi_i^Y}{\overline{\chi}^Y} \frac{\Delta \chi_i^L}{\overline{\chi}^L}$ and $\frac{\Delta A_i}{\overline{A}} \left(\frac{\Delta \chi_i^Y}{\overline{\chi}^Y}\right)^2$ are dropped. In our empirical analysis, this omitted residual is never above 5 per cent of the total value.

[13] Both are defined as value added weighted measures.

[14] The data used in this paper can only be accessed through the UK Data Service's secure lab, reason why we are unable to provide it with the paper. Information about access can be found at http://doi.org/10.5255/UKDA-SN-7989-4. We can share the code so users with access to the data can reproduce our results.

[15] The GVA of manufacturing in Northern Ireland has been constantly below 3 per cent of the UK-wide level in our period of analysis, according to ONS data.

[16] ARD covers the Non-Financial Business Economy of Great Britain, between 1998 and 2014. In terms of SIC07 codes, all sectors are included except O (Public administration, defence and compulsory social security), T (mainly activities of households as employers of domestic personnel), U (activities of extraterritorial organizations), sections 01.1 to 01.5 (inclusive) of Agriculture, section 65.3 of Financial and Insurance activities, any educational activity carried out by the public sector in P, section 86.2 (medical and dental practice activities) and any other public provision of human health and social work activities in Q. The coverage is around two-thirds of the GB gross value added. The sample does not cover self-employees (formally called sole proprietors or traders), unless they are registered with the UK tax authority, HMRC (which is not necessary for businesses below a given income threshold). For further details, see Office for National Statistics (2017).

[17] ARD is based on stratified sampling, using industry, region and employment size as strata. The latter uses 0–9 employees band as one cell for sampling. Hence, it is natural to exclude the whole band together. Moreover, firms with less than 10 employees tend not to be sampled in consecutive years. This means their capital stock cannot be imputed, nor be used in the production function estimation (see Appendix C). These issues and other information about sampling in ARD can be found in ONS (2012).

[18] Replacing individual prices with an aggregate price index amounts to introducing measurement error in the outcome variable (value added), resulting in increased uncertainty around the estimates. However, the level of aggregation used in our analysis is relatively low (4-digit industry classification). Montecarlo simulations show that the proportion of the overall uncertainty attributable to measurement error in prices is minor.

[19] The resulting model is highly non-linear (see Equation C5 in Appendix C), and GMM does not converge in our data (in effect, most of the literature estimates CD production functions, which are log-linear in the parameters). We therefore consider a translog production function, which is a non-linear approximation of

the CES around an elasticity of substitution equal to 1. According to Monte Carlo simulations in Lagomarsino (2017), the bias of a second order (i.e., non-linear) Taylor approximation of a two-input CES is neglectible for $|\rho| < 1$, and it is still relatively small at $\rho = -2$. Our main estimates situate ρ around -1.18.

[20] Collard-Wexler and De Loecker (2016) show that measurement errors in the capital stock introduce a downward bias in the estimates of the production function parameters. To deal with the problem, they suggest a hybrid IV-Control function approach that instruments capital with lagged investment. However, the method relies on log-linearity and is therefore not directly applicable outside a CD setting.

[21] As explained in Appendix C, it is impossible to identify the shock to value added. This is therefore included in the computation of $\widehat{A}_{it}$. This introduces a bias in the latter, which is constant as long as the variance of the shock to value added is also constant. For further details, see also endnote 33.

[22] Unfortunately, ARD does not contain information on the skill level of the workers employed. These are instead imputed from the Annual Survey of Hours and Earnings (ASHE). In particular, we compute the share of workers in each of the nine occupation groups (SOC2010 major groups), in a given industry (SIC07 division), and year. Then, we assign this share to firms in ARD in that given industry-year cluster. Total employment for each occupation group in the local labour market is also computed from ASHE.

[23] More details about the definition and methodology for computing the TTWA can be found in https://ons.maps.arcgis.com/home/item.html?id=379c0cdb374f4f1e94209e908e9a21d9.

[24] In particular, we fit a log-normal distribution using the percentiles presented in tab. 6 in Webber (2015).

[25] Notice Manning (2003) derives an elasticity for the whole economy. In consequence, we apply this method *before* removing other sectors and firms from our sample. This is, we use the maximum sample available in ARD.

[26] Sensitivity analysis performed on c_1 and c_2 (Appendix D) shows that the higher dispersion there is in firm level labour supply elasticities, the higher the importance of heterogeneity in explaining the level of the aggregate labour share. Only for unreasonably low levels of dispersion in these elasticities (inconsistent with the literature mentioned here) our main result that heterogeneity increases the aggregate labour share is reversed. In terms of changes in the aggregate labour share over time (Tables 1 and 3 below), c_1 and c_2 have no meaningful effect.

[27] An approach where this term would not show up is when one of the variables of the model is *not* computed using an optimality condition, but as a residual. For example, we could measure labour market power implicitly, as the value that makes the rest of the measured variables fit that equation (e.g., as in Brummund, 2012). This however confounds any “true” discrepancy with the measure of labour market power.

[28] More precisely, the four terms are $\frac{1}{1-\rho}\mathrm{r}(\chi^Y,\tau)\mathrm{CV}(\chi^Y)\mathrm{CV}(\tau)$, $-\frac{1}{1-\rho}\mathrm{r}(\chi^L,\tau)\mathrm{CV}(\chi^L)\mathrm{CV}(\tau)$, $+\frac{\rho}{1-\rho}\mathrm{r}(A,\tau)\mathrm{CV}(A)\mathrm{CV}(\tau)$ and $-\frac{\rho}{1-\rho}\mathrm{r}\left(\frac{w}{p},\tau\right)\mathrm{CV}\left(\frac{w}{p}\right)\mathrm{CV}(\tau)$.

[29] In order to produce standard errors for the estimated variables (e.g., TFP), we bootstrap the whole estimation procedure (i.e., the imputation of capital, the estimation of the production function, and the decomposition), with 1000 repetitions. Bootstrap is actually needed in order to compute the correct standard errors for the parameters of the production function, given that capital is a generated regressor. To compute the confidence intervals presented in this section we use the *percentile method* (e.g., see Efron & Tibshirani, 1986). This takes the point estimates as the centre of the interval, rather than the bootstrap average. Because of the non-linearities involved in the imputation process, a bias might emerge when adding normally distributed variability to the estimations via bootstrap. In practice, the two means have a correlation above 0.98, for every variable. The major discrepancy arises with the mean of TFP, which is 14% higher in the bootstrap case. Trends are however the same.

[30] The labour share in manufacturing, computed from national accounts, shows an increase between 1998 and 2009, and a fall thereafter, with the 2014 level being roughly the same as that in 1998. The level is also around 0.10 points higher in the national accounts. There is however no reason why they should be the same. For instance, the sample used here focuses only on firms with more than 10 employees (with smaller firms tending to have a higher labour share).

[31] Notice the labour share is always positive, because the (few) observations with negative value added are removed from the sample (as they cannot be used in the estimation of the production function).

[32] For simplicity, we look at the joint effect of w_i/p_i in the analysis below. This has also the advantage of identifying a real wage term in the expression of λ^{HOM}.

[33] Year 2007 presents an unusual behaviour, with significantly more missing observations in the original dataset, particularly for small firms (this issue is to be resolved in the 4th edition of the dataset, unavailable at the moment of producing this paper).

[34] De Loecker and Eeckhout (2018) also document a mild increase in mark-ups for the UK, although with a different timing than the one described here. However, the difference between their method and ours are major. They do not use micro-data but balance-sheet data, covering sectors beyond manufacturing; they assume a CD gross output production function; and they use sales rather than value added to compute national level averages. A similar method and data to the latter is applied by Haldane et al. (2018), who report an increase in mark-ups in UK manufacturing, starting around 2005.

[35] At first, the level of $\overline{\tau}$ might appear to be relatively high. Recall this is computed as the ratio between the observed and predicted labour share across firms: $\overline{\tau}$ around two then suggests the predicted λ_i is around half of the observed labour share. This is however not necessarily true. As Appendix E shows, $\widehat{A}_{it}$ contains both the shock to TFP and the shock to value added (terms ξ_{it} and ϵ_{it} in Equations C3 and C4, respectively). While the latter has zero mean in terms of the *logarithm* of value added (again, see Equation C3), it does not do so around value added itself. This bias is captured by the level of $\widehat{A}_{it}$ (bias that should be constant as long as the variance of ϵ_{it} is constant). It can be shown that $E\left(\widehat{A}_{it}|\Phi_t\right)=A_{it}e^{\frac{\sigma^2}{2}}$, where σ^2 is the variance of ϵ_{it}. The magnitude of such bias is unknown because the two shocks cannot be empirically identified, and thus σ^2 cannot be estimated. The sign however is evidently positive; TFP is overestimated. Furthermore, since the predicted labour share (Equation 5) contains $\widehat{A}_{it}$ to the power of $\frac{\rho}{1-\rho}$, and ρ is estimated to be -1.18, such bias is lowering the predicted labour share, which in turns raises τ_i and therefore $\overline{\tau}$. Again, as long as σ^2 is constant over time, such bias is only a level effect, without affecting trends and therefore the decomposition exercise.

[36] The documented increase in product market power contributing to a lower labour share seems consistent with the 'winner-take-most' literature. To relate more to that literature, we compute changes in market shares and market concentration (which are not necessarily related to mark-ups, our measure of product market power). In particular, we compute the gross output-based Herfindahl–Hirschman Index (HHI) for the 218 4-digit SIC07 sub-industries available in the data. We find that this index rose in two thirds of these sub-industries, between 1998 and 2014. If we aggregate the subindustry HHI up to the division level (2-digit SIC07, 22 divisions in total), properly weighted, we observe that 52 per cent of divisions had a higher HHI index in 2014 than in 1998. Finally, the aggregated manufacturing-level HHI index also went up over the period, from 0.10 to 0.12. So, even if concentration rose in most sectors, the change is relatively small. Moreover, the level itself is relatively low, according to traditional interpretations of the HHI index. This evidence seem broadly consistent with results from other studies like Bell and Tomlinson (2018) and Valletti et al. (2017) (minding the differences in terms of sample, indicator and period). Overall, the 'winner-take-most' phenomenon is only weakly present in UK manufacturing, if present at all.

[37] For ease of visualization, $\sum$ is centred around zero, whereas, as Equation (12) shows, this moves around one.

[38] For a survey, see Ackerberg et al. (2007).

[39] Another common approach is the control function method, based on Olley and Pakes (1996) and Levinsohn and Petrin (2003). This semi-parametric method is based on stronger assumptions and requires greater data availability than the dynamic panel approach. The latter is the case because the control function method relies on past values of investment (in the case of Olley and Pakes, 1996), or intermediary inputs (in the case of Levinsohn and Petrin, 2003) as instruments, which in our sample are only available when a firm is selected for the survey (say, in period t). Conversely, we can implement the dynamic panel approach using past values of employment and capital as instruments, without need for investment or intermediary inputs. In any case, for comparative purposes we also estimated the production function with the control function approach, using intermediate inputs as proxy. Unfortunately, it yielded invalid results (in terms of parameters outside the theoretical domain). We also estimated a model without accounting for endogeneity (that is, assuming $\theta=0$), which also yielded bad behaved parameters (which we think is evidence of presence of endogeneity). Finally, we also estimated a Cobb–Douglas model, but as stated in our theoretical section, this function does not allow us to capture all sources of heterogeneity. Moreover, the CES estimations indicate an elasticity of substitution significantly different to 1, hence ruling out a Cobb–Douglas functional form.

[40] As commented in the main text, Montecarlo simulations in Lagomarsino (2017) show that the non-linear translog used here is a very good approximation of the underlying CES, for ρ close to or below 1, as it is our case.

[41] Since capital is a generated regressors (see next subsection), standard errors are based on bootstrap estimates, with 1000 replications.

[42] See footnote 35 for further details.

[43] In principle, this relationship is not testable, since we lack capital data at the firm level. However, the correlation between capital estimated with this method and the perpetual inventory method, for firms observed to be born in the sample period, is 0.56. This is a significantly high correlation, considering that new firms are likely to be significantly different that established firms, e.g., in terms of their investment patterns.

[44] The depreciation rate is assumed to be 4.58 per cent, the average for the 1998–2014 period, according to Office for National Statistics data for the UK.

[45] This is the equivalent for the labour share of the Olley-Pakes decomposition formula for productivity (Olley and Pakes, 1996).

[46] Specifically, a meaningful result is one where ρ is not greater than 1 (for which the elasticity of substitution is properly defined), and where α is between 0 and 1 (otherwise, one factor of production would have negative marginal product).

[47] Unlike in Table A1, no decomposition is shown for the 'combined' sub-sectors because this requires an estimate of ρ, which was only estimated at the sub-sector levels.

[48] Recall from the growth accounting decomposition that $g_{\overline{\chi}^L}$ is multiplied by $-\left(\frac{1}{1-\rho}\right)$. The estimated ρ for this sub-sector is negative, which, combined with a fall in $\overline{\chi}^L$ (i.e., a fall in firms' labour market power) yields the positive contribution of this variable to λ^{HOM}, as Table A2 shows.

REFERENCES

Abdih, Y. & Danninger, S. (2017) *What explains the decline of the U.S. labor share of income? An analysis of state and industry level data*, IMF Working Paper WP/17/167. Washington D.C.: International Monetary Fund.

Acemoglu, D. & Restrepo, P. (2018) The race between machine and man: implications of technology for growth, factor shares and employment. *American Economic Review*, 108(6), 1488–1542.

Ackerberg, D., Lanier Benkard, C., Berry, S. & Pakes, A. (2007) Econometric tools for analyzing market outcomes. In: Heckman, J. & Leamer, E. (Eds.) *Handbook of Econometrics*, Vol. 6A. Amsterdam: Elsevier.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P.J. & Li, H. (2022) *A theory of falling growth and rising rents*. Federal Reserve Bank of San Francisco Working Paper 2019-11. San Francisco, CA: Federal Reserve Bank of San Francisco.

Autor, D., Dorn, D., Katz, L.F., Patterson, C. & Reenen, J.V. (2020) The fall of the labor share and the rise of superstar firms. *Quarterly Journal of Economics*, 135(2), 645–709.

Autor, D. & Salomons, A. (2018) *Is automation labor-Dispacing? Productivity growth, employment, and the labor share*, NBER Working Paper Series 24871. Cambridge, MA: National Bureau of Economic Research.

Bockerman, P. & Maliranta, M. (2012) Globalization, creative destruction, and labour share change: evidence on the determinants and mechanisms from longitudinal plant-level data. *Oxford Economic Papers*, 64, 259–280.

Barnes, S., Price, S. & SebastiÃ¡-Barriel, M. (2008) *The elasticity of substitution: evidence from a UK firm-level data set*, Working Paper 348. London, UK: Bank of England.

Bauer, A. & Boussard, J. (2020) Market power and labor share. *Economie et Statistique/Economics and Statistics*, 520-521, 125–146.

Bell, B., Bukowski, P. & Machin, S. (2018) *Rent sharing and inclusive growth*, Discussion paper series 12060. Bonn: IZA Institute of Labor Economics.

Bell, T. & Tomlinson, D. (2018) *Is everybody concentrating? Recent trends in product and labour market concentration in the UK*. London, UK: Briefing, Resolution Foundation. Available from: https://www.resolutionfoundation.org/publications/is-everybody-concentrating-recent-trends-in-product-and-labour-market-concentration-in-the-uk/

Bental, B. & Demougin, D. (2010) Declining labor shares and bargaining power: an institutional explanation. *Journal of Macroeconomics*, 32(1), 443–456.

Bergholt, D., Furlanetto, F. & Maffei-Faccioli, N. (2022) The decline of the labor share: new empirical evidence. *American Economic Journal: Macroeconomics*, 14(3), 163–198.

Blundell, R. & Bond, S. (2000) Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), 321–340.

Branston, J.R., Cowling, K.G. & Tomlinson, P.R. (2014) Profiteering and the degree of monopoly in the great recession: recent evidence from the United States and the United Kingdom. *Journal of Post Keynesian Economics*, 37(1), 135–162.

Brummund, P. (2012) Variations in monopsonistic behavior across establishments: Evidence from the indonesian labor market. Available from: http://www.peterbrummund.com/docs/pwbmonopsony.pdf

Bruno, M. (1978) Duality, intermediate inputs and value added. In: Fuss, M. & McFadden, D. (Eds.) *Production economics: a dual approach to theory and applications*, Vol. 2. Amsterdam: North Holland.

Chen, X. & Plotnikova, T. (2014) *Retrieving initial capital distributions from panel data*, MPRA Paper 61154. Germany: University Library of Munich.

Collard-Wexler, A. & De Loecker, J. (2016) *Production function estimation with measurement error in inputs*, NBER Working Paper 22437. Cambridge, MA: National Bureau of Economic Research.

De Loecker, J. & Eeckhout, J. (2018) *Global Market Power*, NBER Working Papers 24768. Cambridge, MA: National Bureau of Economic Research.

De Loecker, J., Eeckhout, J. & Unger, G. (2020) The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2), 561–644.

Dimova, D. (2019) *The structural determinants of the labour share in Europe*, IMF Working Paper WP/19/67. Washington D.C: International Monetary Fund.

Eden, M. & Gaggl, P. (2018) On the welfare implications of automation. *Review of Economic Dynamics*, 29, 15–43.

Efron, B. & Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1), 54–75.

Felipe, J. & McCombie, J.S. (2014) The aggregate production function: Not even wrong. *Review of Political Economy*, 26(1), 60–84.

Fisher, F.M. (1969) The existence of aggregate production functions. *Econometrica*, 37(4), 553–577.

González, I. & Trivín, P. (2019) The global rise of asset prices and the decline of the labor share. MPRA Paper No. 94587.

Gouin-Bonenfant, É. (2022) Productivity dispersion, between-firm competition, and the labor share. *Econometrica*, 90(6), 2755–2793.

Green, H.J. (1964) *Aggregation in economic analysis*. Princeton, NY: Princeton University Press.

Grossman, G.M., Helpman, E., Oberfield, E. & Sampson, T. (2017) *The productivity slowdown and the declining labor share: a neoclassical exploration*, NBER Working Paper Series 23853. Cambridge, MA: National Bureau of Economic Research.

Haldane, A., Aquilante, T., Chowla, S., Dacic, N., Masolo, R., Schneider, P. et al. (2018) *Market power and monetary policy, Speech*. London, UK: Bank of England. Available from: https://www.bankofengland.co.uk/speech/2018/andy-haldane-speech-at-the-economic-policy-symposium-panel-jackson-hole

Hopenhayn, H., Neira, J. & Singhania, R. (2022) From population growth to firm demographics: implications for concentration, entrepreneurship and the labor share. *Econometrica*, 90(4), 1879–1914.

IMF. (2017) *World economic outlook: gaining momentum?*, Technical report. Washington D.C: International Monetary Fund.

Karabarbounis, L. & Neiman, B. (2014) The global decline of the labor share. *The Quarterly Journal of Economics*, 129(1), 61–103.

Kehrig, M. & Vincent, N. (2021) The micro-level anatomy of the labor share decline. *The Quarterly Journal of Economics*, 136(2), 1031–1087.

Koh, D., Santaeulàlia-Llopis, R. & Zheng, Y. (2020) Labor share decline and intellectual property products capital. *Econometrica*, 88(6), 2609–2628.

Lagomarsino, E. (2017) *A study of the approximation and estimation of CES production functions*, PhD thesis. Edinburgh, UK: Heriot-Watt University. Available from: https://www.ros.hw.ac.uk/handle/10399/3972

León-Ledesma, M.A. & Satchi, M. (2018) Appropriate technology and balanced growth. *Review of Economic Studies*, 86(2), 807–835.

Levinsohn, J. & Petrin, A. (2003) Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2), 317–341.

Manning, A. (2003) *Monopsony in motion: imperfect competition in labor markets*. Princeton, NJ: Princeton University Press.

Mertens, M. (2022) Micro-mechanisms behind declining labor shares: rising market power and changing modes of production. *International Journal of Industrial Organization*, 81, 102808.

Office for National Statistics, Virtual Microdata Laboratory (VML), University of West England, Bristol. (2017) *Annual Respondent Database X, 1998-2014: Secure Access. [data collection]*, 4th edition. London: Office for National Statistics, [original data producer(s)]. SN 7989. Available from: http://doi.org/10.5255/UKDA-SN-7989-4

Olley, G.S. & Pakes, A. (1996) The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6), 1263–1297.

ONS. (2012) *Annual business survey (ABS)*, Technical Report. London: Office for National Statistics. Available from: https://www.ons.gov.uk/ons/guide-method/method-quality/specific/business-and-energy/annual-businesssurvey/quality-and-methods/abs-technical-report.pdf

Perugini, C., Vecchi, M. & Venturini, F. (2017) Globalisation and the decline of the labour share: a microecomic perspective. *Economic Systems*, 41, 524–536.

Rognlie, M. (2015) Deciphering the fall and rise in the net capital share. *Brookings Papers on Economic Activity*, 46(1), 1–69.

Schwellnus, C., Kappeler, A. & Pionnier, P.-A. (2017) *Decoupling wages from productivity: macro-level facts*, OECD economics department working papers 1373. Paris: OECD.

Schwellnus, C., Pak, M., Pionnier, P.-A. & Crivellaro, E. (2018) *Labour share developments over the past two decades: the role of technological progress, globalisation and "winner-take-most" dynamics*, OECD economics department working papers 1503. Paris: OECD.

Siegenthaler, M. & Stucki, T. (2015) Dividing the pie: firm-level determinants of the labor share. *Industrial and Labor Relations Review*, 68(5), 1157–1194.

Valentinyi, A. & Herrendorf, B. (2008) Measuring factor income shares at the sector level. *Review of Economic Dynamics*, 11(4), 820–835.

Valletti, T., Koltay, G., Lorincz, S. & Zenger, H. (2017) *Concentration trends in Europe*, Presentation. Brussels: European Commission. Available from: https://ecp.crai.com/wp-content/uploads/2017/12/Valleti-Concentration_Trends_TV_CRA-002.pdf

Webber, D. (2015) Firm market power and the earnings distribution. *Labour Economics*, 35(C), 123–134.

Weil, D. (2017) *The fissured workplace: why work became so bad for so many and what can be done to improve it*. Cambridge, MA: Harvard University Press.

Zambelli, S. (2004) The 40% neoclassical aggregate theory of production. *Cambridge Journal of Economics*, 28(1), 99–120.

How to cite this article: Richiardi, M.G. & Valenzuela, L. (2024) Firm heterogeneity and the aggregate labour share. *LABOUR*, 1–36. Available from: https://doi.org/10.1111/labr.12265

APPENDIX

A | EXPRESSION FOR $\sum$

For a given, $z = \{\chi^Y, \chi^L, A, w, p\}$, the second-order Taylor expansion around the respective weighted average is:

$$\left(\frac{z_i}{\overline{z}}\right)^{\phi} \approx 1 + \phi\left(\frac{\Delta_{z,i}}{\overline{z}}\right) + \frac{\phi(\phi-1)}{2}\left(\frac{\Delta_{z,i}}{\overline{z}}\right)^2, \tag{A1}$$

where $\overline{z}$ is the weighted mean of the respective variable, and $\Delta z_i = z_i - \overline{z}$ is the deviation from that mean. After dropping all interaction terms of order higher than two, Equation (7) can be approximated by:

$$
\begin{aligned}
\lambda \approx \lambda^{\mathrm{HOM}} \sum_i \delta_i \Bigg[& 1 + \frac{1}{1-\rho}\left(\frac{\Delta\chi_i^Y}{\overline{\chi}^Y}\right) - \frac{1}{1-\rho}\left(\frac{\Delta\chi_i^L}{\overline{\chi}^L}\right) + \frac{\rho}{1-\rho}\left(\frac{\Delta A_i}{\overline{A}}\right) - \frac{\rho}{1-\rho}\left(\frac{\Delta w_i}{\overline{w}}\right) + \frac{\rho}{1-\rho}\left(\frac{\Delta p_i}{\overline{p}}\right) \\
& + \frac{\rho}{2(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\overline{\chi}^Y}\right)^2 + \frac{2-\rho}{2(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\overline{\chi}^L}\right)^2 + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\left(\frac{\Delta A_i}{\overline{A}}\right)^2 + \frac{\rho}{2(1-\rho)^2}\left(\frac{\Delta w_i}{\overline{w}}\right)^2 + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\left(\frac{\Delta p_i}{\overline{p}}\right)^2 \\
& - \frac{1}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\overline{\chi}^Y}\right)\left(\frac{\Delta\chi_i^L}{\overline{\chi}^L}\right) + \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\overline{\chi}^Y}\right)\left(\frac{\Delta A_i}{\overline{A}}\right) - \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\overline{\chi}^Y}\right)\left(\frac{\Delta w_i}{\overline{w}}\right) \\
& + \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\overline{\chi}^Y}\right)\left(\frac{\Delta p_i}{\overline{p}}\right) - \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\overline{\chi}^L}\right)\left(\frac{\Delta A_i}{\overline{A}}\right) + \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\overline{\chi}^L}\right)\left(\frac{\Delta w_i}{\overline{w}}\right) \\
& - \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\overline{\chi}^L}\right)\left(\frac{\Delta p_i}{\overline{p}}\right) - \frac{\rho^2}{(1-\rho)^2}\left(\frac{\Delta A_i}{\overline{A}}\right)\left(\frac{\Delta w_i}{\overline{w}}\right) + \frac{\rho^2}{(1-\rho)^2}\left(\frac{\Delta A_i}{\overline{A}}\right)\left(\frac{\Delta p_i}{\overline{p}}\right) \\
& - \frac{\rho^2}{(1-\rho)^2}\left(\frac{\Delta w_i}{\overline{w}}\right)\left(\frac{\Delta p_i}{\overline{p}}\right) \Bigg].
\end{aligned}
$$

This can be simplified further. First, notice that when $\overline{z}$ is defined using value added as weights, $\sum_i \delta_i \Delta z_i = 0$. Thus, the first four terms in the parenthesis above (representing the weighted sum of all deviations from the weighted average) are zero. Second, notice that $\sum_i \delta_i (\Delta z_i)^2 = \mathrm{Var}(z)$ and $\sum_i \delta_i \Delta x_i \Delta z_i = \mathrm{Cov}(x,z)$, with both defined as value added weighted measures, and not in the standard, unweighted fashion. Then, we can restate our decomposition formula solely in terms of variances and covariances or, equivalently, in terms of correlations (r) and coefficient of variations (CV):

$$
\begin{aligned}
\lambda \approx \lambda^{\mathrm{HOM}} \Bigg[& 1 \\
& + \frac{\rho}{2(1-\rho)^2}\mathrm{CV}^2(\chi^Y) + \frac{2-\rho}{2(1-\rho)^2}\mathrm{CV}^2(\chi^L) + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\mathrm{CV}^2(A) + \frac{\rho}{2(1-\rho)^2}\mathrm{CV}^2(w) + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\mathrm{CV}^2(p) \\
& - \frac{1}{(1-\rho)^2} r(\chi^Y,\chi^L)\mathrm{CV}(\chi^Y)\mathrm{CV}(\chi^L) + \frac{\rho}{(1-\rho)^2} r(\chi^Y,A)\mathrm{CV}(\chi^Y)\mathrm{CV}(A) - \frac{\rho}{(1-\rho)^2} r(\chi^Y,w)\mathrm{CV}(\chi^Y)\mathrm{CV}(w) \\
& + \frac{\rho}{(1-\rho)^2} r(\chi^Y,p)\mathrm{CV}(\chi^Y)\mathrm{CV}(p) - \frac{\rho}{(1-\rho)^2} r(\chi^L,A)\mathrm{CV}(\chi^L)\mathrm{CV}(A) + \frac{\rho}{(1-\rho)^2} r(\chi^L,w)\mathrm{CV}(\chi^L)\mathrm{CV}(w) \\
& - \frac{\rho}{(1-\rho)^2} r(\chi^L,p)\mathrm{CV}(\chi^L)\mathrm{CV}(p) - \frac{\rho^2}{(1-\rho)^2} r(A,w)\mathrm{CV}(A)\mathrm{CV}(w) + \frac{\rho^2}{(1-\rho)^2} r(A,p)\mathrm{CV}(A)\mathrm{CV}(p) \\
& - \frac{\rho^2}{(1-\rho)^2} r(w,p)\mathrm{CV}(w)\mathrm{CV}(p) \Bigg].
\end{aligned}
$$

The latter can be transformed into a matrix multiplication through isolating the vector $[\mathrm{CV}(\chi^Y)\ \mathrm{CV}(\chi^L)\ \mathrm{CV}(A)\ \mathrm{CV}(w)\ \mathrm{CV}(p)]$, resulting in Equation (13) in the main text.

B | ROLE OF PRICE HETEROGENEITY

From theory, the firm level LS is given by Equation (5), which for simplicity we can rewrite as:

$$
\lambda_i = \left(\frac{\alpha}{\chi_i}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i p_i}{w_i}\right)^{\frac{\rho}{1-\rho}},
$$

where $\chi_i = \chi_i^L / \chi_i^Y$. This can be further rewritten as

$$\lambda_i = \left(\frac{\alpha}{\chi_i}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i p_{j,0} \pi_j}{w_i}\right)^{\frac{\rho}{1-\rho}},$$

where $p_{j,0}$ is the price *level* in sector j in the base period, and $\pi_j = p_j / p_{j,0}$ is the sectoral price index. The initial price level can then be included as a multiplicative factor of productivity, by defining $\tilde{A}_i = A_i p_{j,0}$. This leads to

$$\lambda_i = \left(\frac{\alpha}{\chi_i}\right)^{\frac{1}{1-\rho}} \left(\frac{\tilde{A}_i \pi_j}{w_i}\right)^{\frac{\rho}{1-\rho}}. \tag{B1}$$

Finally, defining $\pi = \sum \delta_j \pi_j$ (aggregate price index), we get a modified expression for the aggregate LS as:

$$\lambda = \lambda^{\mathrm{HOM}} \sum_i \left(\frac{\chi}{\chi_i}\right)^{\frac{1}{1-\rho}} \left(\frac{\tilde{A}_i}{\tilde{A}}\right)^{\frac{\rho}{1-\rho}} \left(\frac{\pi_j}{\pi}\right)^{\frac{\rho}{1-\rho}} \left(\frac{w}{w_i}\right)^{\frac{\rho}{1-\rho}} \delta_i, \tag{B2}$$

where

$$\lambda^{\mathrm{HOM}} = \left(\frac{\alpha}{\chi}\right)^{\frac{1}{1-\rho}} \left(\frac{\tilde{A}\pi}{w}\right)^{\frac{\rho}{1-\rho}}. \tag{B3}$$

$\frac{\pi_j}{\pi}$ is a measure of how prices in sector j are growing relative to the average price growth. More heterogeneity in $\frac{\pi_j}{\pi}$ has thus a straightforward interpretation in terms of *relative price changes*.

$\frac{\tilde{A}_i}{\tilde{A}}$ has a less clear interpretation. However, *changes* in this ratio are only related to changes in TFP, as the base prices do not change.

C | ESTIMATION OF VALUE-ADDED PRODUCTION FUNCTION

As said in the main text, TFP and the production function parameters are not observed in the data. We draw from the abundant literature on estimating production functions in order to compute these missing terms.

The point of departure in this analysis is the fact that A_{it} is not observed by the econometrician, but might be observed by the firm. Thus, if firms choose inputs based on their productivity level, a simple estimation of TFP using least squares would suffer from endogeneity. Since the seminal paper by Olley and Pakes (1996), several techniques have been put forward to address the endogeneity problem when estimating productivity using micro data.[38] Here, we follow the dynamic panel approach (e.g., Blundell & Bond, 2000), where endogeneity is eliminated by assuming TFP follows an AR(1) process with parameter θ, and then the main model is θ-differentiated. The dynamic panel approach has some advantages with respect to other methods, including less stringent data requirement, which allows to increase the sample size.[39] In short, the method works as follows. We start with our production function, extended to an econometric notation:

$$Y_{it} = e^{\omega_{it}}(\alpha L_{it}^{\rho} + (1-\alpha)K_{it}^{\rho})^{\frac{1}{\rho}} e^{\varepsilon_{it}} \tag{C1}$$

where for convenience we have defined $A_{it} \equiv e^{\omega_{it}}$, and where ϵ_{it} is a idiosyncratic *iid* shock to output. As the rest of the literature, we assume ω_{it} follows a first-order Markov process. This is, $\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it}$, where ξ_{it} is an idiosyncratic *iid* shock to productivity, known to the firm. Taking logs of (C1), we get:

$$y_{\text{it}} = \frac{1}{\rho}\ln\left(\alpha L_{\text{it}}^{\rho} + (1-\alpha)K_{\text{it}}^{\rho}\right) + \omega_{\text{it}} + \epsilon_{\text{it}}. \tag{C2}$$

The common assumption in the literature about the informational setting is that capital is a state variable (in the sense that it is chosen in period $t-1$), whereas labour is a flexible factor (in the sense that it can be chosen in period t). This informational structure is relevant because under the assumption that a firm knows its shock to productivity (ξ_{it}), labour is correlated with the unobserved (by the econometrician) error term, and hence endogenous in Equation (C2). Non-linear least squares would then yield inconsistent results.

To move further, we align with the literature by assuming ω_{it} follows an AR(1):

$$\omega_{it} = \theta\omega_{it-1} + \xi_{it}. \tag{C3}$$

If we combine Equations (C2) and (C3) (i.e., if we 'θ-differentiate' the production function), we get:

$$\begin{aligned} y_{\text{it}} = &\frac{1}{\rho}\ln\left(\alpha L_{\text{it}}^{\rho} + (1-\alpha)K_{\text{it}}^{\rho}\right) + \theta\left[y_{\text{it}-1} - \frac{1}{\rho}\ln\left(\alpha L_{\text{it}-1}^{\rho} + (1-\alpha)K_{\text{it}-1}^{\rho}\right)\right] \\ &+\xi_{\text{it}} + (\epsilon_{\text{it}} - \theta\epsilon_{\text{it}-1}) \end{aligned}. \tag{C4}$$

Thus, 'θ-differencing' the model eliminates unobserved productivity from the equation. The above can then be estimated using GMM.

It is important to notice here that the above model is highly non-linear. In effect, most of the literature estimates CD production functions, which are log-linear in parameters. To a void convergence issues due to non-linearity, we estimate a translog production function, which is an approximation of the CES around an elasticity of substitution equal to 1.[40] This production function is:

$$\begin{aligned} y_{\text{it}} \approx &\ln(A_{\text{it}}) + \alpha\ln(L_{\text{it}}) + (1-\alpha)\ln(K_{\text{it}}) + \frac{\rho\alpha(1-\alpha)}{2}\ln^2(L_{\text{it}}) \\ &-\rho\alpha(1-\alpha)\ln(L_{\text{it}})\ln(K_{\text{it}}) + \frac{\rho\alpha(1-\alpha)}{2}\ln^2(K_{\text{it}}) + \mu_{\text{it}} \end{aligned}. \tag{C5}$$

The final model estimated by GMM is obtained by 'θ-differencing' the equation above, with instruments $\{\ln(K_{\text{it}}),\ \ln^2(K_{\text{it}}),\ \ln(K_{\text{it}-1}),\ \ln^2(K_{\text{it}-1}),\ \ln(L_{\text{it}-1}),\ \ln^2(L_{\text{it}-1}),\ \ln(L_{\text{it}-1})\ln(K_{\text{it}-1})\}$. Estimation produces the following values, all significant at the 1 per cent[41]: $\widehat{\alpha} = 0.38$, $\widehat{\rho} = -1.18$, $\widehat{\theta} = 0.92$, and a constant of 3.8.

Finally, having estimated α and ρ, we can use Equation (C1) to compute $\widehat{A}_{it}$ as a residual, for every firm and period. Importantly, since the productivity shock (ξ_{it}) cannot be identified separately from the idiosyncratic shock to value added (ϵ_{it}), $\widehat{A}_{it}$ also includes the realized shock to value added, $\widehat{\epsilon}_{it}$. In effect, from Equation (C1) we can see that $\widehat{A}_{it} = e^{c+\widehat{\omega}_{it}+\widehat{\epsilon}_{it}}$ (where c is the constant term in the regression). This means that $\widehat{A}_{it}$ is a biased predictor of A_{it}. Nevertheless, as long as the variance of ϵ_{it} is constant over time, such bias is constant too, not affecting the decomposition, which focuses on changes over time.[42] Notice Equation (C1) allows us to compute the 'realized' value of A even for observations not part of the regression sample (e.g., because of missing data in a given year). We follow this approach, and 'extrapolate' $\widehat{A}_{it}$ whenever possible. Around 50 per cent of final observations used in the analysis are extrapolated.

Capital stock

Firm-level capital stock is not available in the dataset. Nonetheless, firms report information on their capital expenditures (investment) for a variety of assets like buildings, vehicles, and so on. One method often used to compute capital at the firm level is the perpetual inventory method. Whilst this is a good approximation for firms that are *observed* to be born during the sample period (i.e., for those which are sampled during their first year of existence), for firms that do not (in our sample, 99.99 per cent of firms), the level of capital may be greatly underestimated with such method.

Instead, we follow the strategy proposed by Chen and Plotnikova (2014), who estimate capital at the firm level using the aggregate level of capital stocks in the manufacturing sector (obtained from the Office for National Statistics). First, we select a few 'proxy' variables, which are likely to be positively correlated with unobserved firm-level capital, and are observed both at the firm and at the aggregate level. We use intermediate inputs and employment. Then, we estimate the 'structural relationship' between these proxies and capital (based on an assumed stability of their joint distribution).[43] This relationship is given by the following formula:

$$K_{\mathrm{it}} = \left(\frac{L_{\mathrm{it}}}{L_t}\right)^a \left(\frac{M_{\mathrm{it}}}{M_t}\right)^{1-a} K_t, \tag{C6}$$

where L_t, M_t and K_t represent the observed values of employment, intermediate inputs and capital in the whole of manufacturing sector in year t; parameter a accounts for the relative importance of each proxy in the structural relationship. This parameter is assumed constant over time.

In practice, a is unknown. Furthermore, this cannot be estimated from Equation (C6), since K_{it} is also unknown. The solution is to combine Equation (C6) with that of capital accumulation, namely, where I_{it} is firm level investment (available in the dataset), and d is the depreciation rate of the capital stock in manufacturing. This leads to the following equation:

$$I_{\mathrm{it}} = \left(\frac{L_{\mathrm{it}}}{L_t}\right)^a \left(\frac{M_{\mathrm{it}}}{M_t}\right)^{1-a} K_t - (1-\delta)\left(\frac{L_{\mathrm{it}-1}}{L_{t-1}}\right)^a \left(\frac{M_{\mathrm{it}-1}}{M_{t-1}}\right)^{1-a} K_{t-1}. \tag{C7}$$

The above can be estimated using GMM. Results for the whole manufacturing sector yield $\widehat{a} = 0.42$, significant at 1 per cent.[44] With this value is then possible to impute capital at the firm level using Equation (C6). Notice this imputation allows for extrapolation from the estimation sample to firms which are not observed in consecutive years (condition required by the regression), or which are sampled only in 1 year. The extrapolation is valid as long as the 'structural relationship' does not depend on properties of the sample selection (for instance, firm size).

D | MEASUREMENT ERROR IN LABOUR MARKET POWER

In our specfication for labour market power, we adopted the functional form Equation 15 because it produces a theoretically valid mapping between the domain of s_{it} (the share of local employment of firm i at time t) and its elasticity of labour supply η^L_{it}, whilst being quite flexible (c_1 affects the level of the function and c_2 affects its curvature). The two parameters are set to match the two-parameter (mean and variance) lognormal distribution of elasticities estimated for the US (Webber, 2015), shifted to match the UK average as calculated in Manning (2003). The solution is unique.

Within this framework, one way to assess the impact of measurement error in χ_L is to acknowledge uncertainty around the two key parameters c_1 and c_2. Unfortunately, Webber (2015) does not report confidence intervals for the mean and variance of his estimated lognormal distribution. Our measure of monopsony power s_{it} comes from the data and is therefore not stochastic either. Hence, for lack of a better alternative, we test some extreme values for c_1 and c_2. Notice this is altering the unique optimal solution from the calibration exercise. In order for the test not to be totally arbitrary, we still match the average elasticity reported in Manning (2003), but allow changes in its variance. We adopt as extreme values for c_1 values respectively 50 per cent lower and 50 per cent higher than the calibrated values: 0.005 and 0.015, centred at 0.01. We then compute c_2 in these two extreme scenarios to match the average elasticity reported by Manning (2003). The two values for c_2 are 0.4255 and 0.3373, respectively. Finally, we use these new pairs of parameters (0.005, 0.4255) and (0.015, 0.3373) to recalculate χ^L, and then λ^{HOM} (Equation 18) and Σ (Equation 17). Figure A1 shows the original estimate for Σ (black line), together with the new estimates corresponding to the extreme values of c_1 and c_2.

The red line corresponds to a log-normal distribution with higher variance than the optimally calibrated one (black line), whereas the blue line corresponds to a distribution with lower variance. More specifically, the variance in the red scenario is 0.61 and the variance in the blue scenario is 0.33, compared with an observed (for the US) variance of 0.41 (black line). Even with 'extreme' values of c_1 and c_2, therefore, heterogeneity matters, increasing the level of the aggregate labour share.

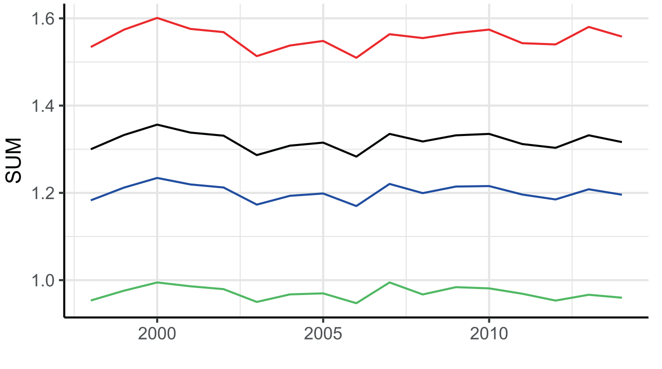


FIGURE A1 Evolution of the heterogeneity component of the labour share. *Source*: our calculation based on ARD data. *Sample*: UK manufacturing firms with 10 employees or more.

As shown in the text, the higher the variance in labour market power, the more the heterogeneity component Σ matters. Consequently, there is a sufficiently low level of heterogeneity in labour market power for which Σ is equal to 1 ($c_1 = 0.04$ and $c_2 = 0.256$), but this is very far off the evidence in the referred literature (the implied variance is 0.17, less than half of the observed one, 0.41). We see no reason to prefer any of these estimates to the 'optimal' one that matches externally calibrated data.

Another important point from this exercise is that, as it is clear from the graph and can be shown mathematically, the trends in Σ and λ^{HOM} are parallel over time. This means Tables 2 and 3 are essentially invariant to c_1 and c_2.

In summary, the importance of heterogeneity (the level of Σ) does depend on χ^L, but the sensitivity is not excessive within 'reasonable' alternative levels of the variance; its change over time, on the other hand, does not depend on χ^L.

E | STATISTICAL DECOMPOSITION

The aggregate labour share is defined as a weighted average of firms' labour share:

$$\lambda^{\text{obs}} = \sum_i \delta_i \lambda_i,$$

where δ_i is the total economy's share of value added of firm i. As the sample size grows, sample moments converge to population moments (ultimately, if we were to have a census of all firms, the two would be the same, provided no other issues like measurement errors exist). One such moment is $E(\delta\lambda)$, for which the Law of Large Numbers states that

$$\lim_{N\to\infty} \frac{\lambda^{\text{obs}}}{N} = \mathrm{E}(\delta\lambda).$$

Using the formulas from the covariance, and replacing population moments with sample equivalent, it is trivial to show that

$$\lambda^{\text{obs}} = \widehat{\mathrm{E}}(\lambda) + N\widehat{\mathrm{Cov}}(\delta,\lambda), \tag{E1}$$

where $\widehat{\mathrm{E}}(\lambda)$ is the observed unweighted average labour share.[45] Since $\widehat{\mathrm{E}}(\delta) = \frac{1}{N}$, it follows that

$$\lambda^{\text{obs}} = \widehat{\mathrm{E}}(\lambda) + \widehat{\mathrm{E}}(\lambda)\widehat{\mathrm{Corr}}(\delta,\lambda). \tag{E2}$$

Therefore, the weighted over the unweighted average labour share is:

$$\frac{\lambda^{\text{obs}}}{\widehat{E}(\lambda)} = 1 + \widehat{\mathrm{Corr}}(\delta,\lambda). \tag{E3}$$

This ratio is smaller the more negative the correlation between firm size (in terms of value added) and labour share is, ceteris paribus.

F | DECOMPOSITION RESULTS FOR MANUFACTURING SUB-SECTORS

In the main text we presented the decomposition analysis for the whole manufacturing sector. Here, we repeat the exercise for manufacturing sub-sectors, defined as 2 digit SIC07 (divisions). Instead of assuming a common production function across sub-sectors, we estimate the (translog) production function for each division separately.

Thirteen out of 23 sub-sectors produced meaningful results (in terms of parameters within the theoretical boundaries), suggesting not every sub-sector might be represented by a CES/translog production function.[46] Overall, the 13 sub-sectors cover 62 per cent of the total observations (firm-years) available across the manufacturing sector, and used in the main text.

Table A1 presents the decomposition for each sub-sector's labour share. Additionally, the table presents an extra row ('combined sub-sectors') with the decomposition of an aggregate series of λ^{obs}, computed from a weighted average of sub-sectors' λ^{obs}, using value added as weights. For comparison, another row is added with the results for the whole manufacturing sector presented in the main text. Finally, to give a sense of the importance of different

TABLE A1 Contribution to changes in the sub-sectoral labour share (λ^{obs}), 1998–2014.

Sub-sector	λ^{obs}	λ^{HOM}	$\sum$	Interaction	δ
	%				
13 Manufacture of textiles	−9.21	−9.38	0.19	−0.02	1.4
14 Manufacture of wearing apparel	−25.02	−33.28	12.38	−4.12	0.4
16 Manufacture of wood and products of wood and cork, excl. furniture	−10.28	−15.61	6.32	−0.99	1.8
17 Manufacture of paper and paper products	−8.67	−9.40	0.81	−0.08	2.4
18 Printing and reproduction of recorded media	−5.59	−6.09	0.54	−0.03	3.2
23 Manufacture of other non-metallic mineral products	−10.45	−7.08	−3.63	0.26	3.6
25 Manufacture of fabricated metal products, excl. machinery and equip.	−11.47	−13.84	2.75	−0.38	11.1
26 Manufacture of computer, electronic and optical products	4.16	7.35	−2.97	−0.22	5.4
27 Manufacture of electrical equipment	1.84	8.81	−6.40	−0.56	3.1
28 Manufacture of machinery and equipment n.e.c.	−6.46	−7.39	1.00	−0.07	8.4
29 Manufacture of motor vehicles, trailers and semi-trailers	−19.49	−21.27	2.27	−0.48	10.5
30 Manufacture of other transport equipment	−11.22	−19.61	10.43	−2.04	6.6
33 Repair and installation of machinery and equipment	10.44	11.12	−0.61	−0.07	3.9
Combined sub-sectors	−8.75	−9.39	0.69	−0.05	61.9
All manufacturing	−7.36	−9.31	2.16	−0.20	100

Note: $g_{\lambda^{obs}} = g_{\lambda^{HOM}} + g_{\sum} +$ interaction effect. δ is the sub-sector's share of the value added in 2014.
Source: Our calculation based on ARD data.
Sample: UK manufacturing firms with 10 employees or more. Sub-sectors represent 2 digit (division) SIC07 codes. Sub-sectors which estimates were spurious and thus omitted are 10 ('manufacture of food products'), 11 ('manufacture of beverages'), 12 ('manufacture of tobacco products'), 15 ('manufacture of leather and related products'), 20 ('manufacture of chemicals and chemical products'), 21 ('manufacture of pharmaceutical products'), 22 ('manufacture of rubber and plastic products'), 31 ('manufacture of furniture') and 32 ('other manufacturing'). Sub-sector 19 ('manufacture of coke and refined petroleum') is omitted from main analysis and thus also omitted here.

TABLE A2 Contribution to changes in sub-sectoral λ^{HOM}, 1998–2014.

Sub-sector	λ^{HOM} %	$\overline{A}$	$\overline{w}/\overline{p}$	$\overline{\chi}^{Y}$	$\overline{\chi}^{L}$	$\overline{\tau}$	Interaction
13	−9.38	−33.27	24.61	−2.24	−2.33	0.54	3.31
14	−33.28	−135.16	65.08	−3.77	26.50	−34.21	48.28
16	−15.61	−22.44	7.68	−2.21	4.85	−6.97	3.49
17	−9.40	−28.04	14.11	−0.61	1.26	−4.88	8.76
18	−6.09	−19.68	11.57	−1.62	−3.38	8.87	−1.84
23	−7.08	−13.25	9.68	0.84	−0.12	−5.19	0.96
25	−13.84	−19.79	9.11	−3.73	−7.29	3.90	3.96
26	7.35	170.52	−148.98	−10.04	81.55	−7.60	−78.10
27	8.81	−7.83	17.43	0.20	10.21	−11.53	0.33
28	−7.39	−34.00	23.65	−0.88	1.43	−3.29	5.70
29	−21.27	−86.78	55.69	0.45	−15.63	3.04	21.97
30	−19.61	−14.03	0.49	−1.75	−0.61	−7.56	3.87
33	11.12	4.93	−0.30	0.12	17.37	−14.92	3.90

Note: Effects are attributed as per Equation (19). The extreme behaviour of sub-sector 14 is due to a significant reduction in the sample size available, from 340 firms in 1998 to 56 firms in 2018. Meanwhile, large numbers in sub-sector 26 reflect the rapid fall in this sector's output prices, between 1998 and 2005.
Source: Our calculation based on ARD data.
Sample: See Table A1.

sub-sectors, the table includes an extra column with the 2014's share of value added of each sub-sector with respect to all manufacturing.

The overall picture is the same as in our results for the whole manufacturing sector, namely that changes in firm heterogeneity have not been a major driver of the changes in the labour share. This is true both for sub-sectors individually and for their combination (exception is sub-sector 27). The latter decomposition is also quite similar to the results for manufacturing as a whole. Still, some disparity is observed in $\sum$ across sub-sectors, both in terms of direction of change and magnitude, with most of the effect of heterogeneity going against the observed change in the labour share (just like in the main results). Notice also that the labour share went up in some sub-sectors, albeit fell in most of them.

Finally, Table A2 shows the decomposition of λ^{HOM} across sub-sectors (similar to Table 3).[47] In line with results at the aggregate level, the key driver of the homogeneous labour share (and thus of the sub-sector labour shares) is the disconnect between pay and productivity. In most sub-sectors, productivity grew faster than real wages. Exceptions are sub-sectors 26 and 27 (‘manufacture of computer, electronic and optical products’ and ‘manufacture of electrical equipment’, respectively), where wages grew faster than productivity, and sub-sector 33 (‘repair and installation of machinery and equipment’), where both productivity and wages shrank over the period, the former more than the latter.

Regarding market power, there are differences with respect to the results for the whole sector. Whereas in the latter both product and market power had an equally minor role in λ^{HOM}, in most sub-sectors the contribution of labour market power is significantly greater than that of product market power. In fact, in some sub-sectors the change in labour market power is high enough to make a significant difference to the sub-sector's λ^{HOM}. For instance, in sub-sector 26, the pay-productivity disconnect changed very little over the period; it is $\overline{\chi}^{L}$ which defines the bulk of the change. In particular, the labour market power of firms in this sub-sector fell importantly over the period.[48]

Overall, sub-sector and industry-wide results differ only where the latter masks heterogeneity in the former. As Table A2 reveals, this is particularly relevant for labour market power, which contribution contains both large positive and negative values. Conversely, variables like TFP and real wages have the same sign in all sub-sectors but one (sub-sector 33).