Firm Heterogeneity and the Aggregate Labour Share

Richiardi, Matteo G. and Valenzuela, Luis

Institute for Social and Economic Research, University of Essex, Institute for New Economic Thinking, Oxford Martin School, University of Oxford, Nuffield College, University of Oxford, Collegio Carlo Alberto

4 June 2019

Firm Heterogeneity and the Aggregate Labour Share

Matteo G. Richiardi[1,2,3,4] and Luis Valenzuela †[2,3]

[1] *Institute for Social and Economic Research, University of Essex*
[2] *Institute for New Economic Thinking, Oxford Martin School, University of Oxford*
[3] *Nuffield College, University of Oxford*
[4] *Collegio Carlo Alberto*

4th June 2019

Abstract

Using a static model of firm behaviour with imperfect competition on the product and labour markets, we quantify the effect of firm heterogeneity in total factor productivity, market power, capital, wages and prices on the aggregate labour share. In particular, we suggest a new decomposition of the aggregate labour share in terms of the first moments of the joint distribution of these variables across firms, providing a bridge between the micro and the macro approach to functional distribution. We provide an application of our method to the UK manufacturing sector, using firm-level data for the period 1998-2014. The analysis confirms that heterogeneity matters: in an economy populated only by representative firms, the labour share would be 10 percentage points lower. Among all the dimensions studied, heterogeneity in total factor productivity and labour market power are the most relevant ones, whereas heterogeneity in product market power matters the least, with wages and prices in between. However, the observed fall in the aggregate labour share over the period is mostly explained by a widening of the disconnect between average productivity and real wages, with a smaller role for an increase in the average product and labour market power of firms after the Great Recession, while changes in the dispersion of these variables mostly offset each other.

KEYWORDS: labour share, firm heterogeneity, market power, firm level data.

JEL CLASSIFICATION: D33, E25, L10, D20, D42, D43.

†Corresponding author. Email: luis.valenzuelarivera@spi.ox.ac.uk.

1 Introduction

A marked decrease in the labour share over the recent decades has been documented in many countries. Updating the data collected by Karabarbounis and Neiman (2014), the IMF shows that in advanced economies the labour share decreased from around 75% in the first half of the 1970s to less than 40% in the first half of the 2010s (IMF, 2017). A downward trend, although of smaller magnitude, is also observed for European countries (Dimova, 2019). There is considerable debate about the causes underlying the documented decline in the labour share, ranging from capital-augmenting technological change; a decline in the price of capital relative to labour, capital accumulation, globalisation, deregulation of product and labour markets, an increase in firms' product and labour market power, financial deepening, monetary policy, the rise of "superstar firms", and even an increase in the cost of housing.

Most studies have adopted either a macro or a representative agent perspective, linking the aggregate labour share (at the national, regional or industry level) to aggregate values of those determinants, with a smaller number of studies looking at the determinants of the labour share at the level of the individual firm, and a few others focusing on compositional issues, that is explaining the decline in the aggregate labour share with an increase in the relative importance of firms with a lower than average labour share.

In this paper we propose a different statistical approach, based on a static model of firm behaviour with CES production functions and imperfect competition in the product and labour markets and geared towards empirical analysis, where all the determinants of the aggregate labour share can be jointly studied building up from the micro level, even in absence of an aggregate production function. This allows us to offer a full characterisation of the aggregate labour share in terms of the distribution of the individual determinants (in particular, we focus on wages, output prices, productivity, capital, and market power). The full characterisation can be approximated by a parsimonious characterisation in terms of the mean, variance and covariance of all those variables. Hence, we are able to generalise the three approaches described above, provide a quantification of the overall effect of heterogeneity, and look into the relative importance of the different sources of heterogeneity.

Our main theoretical result is that when the elasticity of substitution between capital and labour is below 1 —the empirically relevant case— an increase in the dispersion of productivity or monopsony power increases the aggregate labour share, while an increase in the dispersion of real wages or product market power decreases it. By contrast, in the Cobb-Douglas case, where the elasticity is equal to 1, only heterogeneity in market power has a direct effect on the labour share, while heterogeneity in prices, productivity, capital, and wages affects the labour share only if it is correlated with market power.

Empirically, we show on UK firm-level data that that firm heterogeneity increases the *level* of the aggregate labour share by roughly 10 percentage points, with respect to the labour share of a "representative" firm. This wedge however has remained fairly constant over time: the observed fall in the aggregate labour share is mainly explained in terms of the fall in the labour share of the representative firm, particularly due to an increased pay-productivity gap, and to a lesser extent to increased market power. The contribution of heterogeneity in explaining the *change* in the aggregate labour share is minor and mainly comes from an increased dispersion in TFP and labour market power.

The remaining of the paper is structured as follows. Section 2 presents a brief review of the literature. Section 3 presents a simple model of firm optimisation, where firms have a CES technology with constant returns to scale, with imperfect competition both in the product and the labour markets, and derives our theoretical results and main decomposition formula. Section 4 describes the data and our empirical strategy, while Section 5 describes our main findings. Section 6 summarises and concludes.

2 Literature

2.1 The determinants of the aggregate labour share

Abstracting from measurement issues, we can divide the existing empirical work on the labour share in three categories: (i) studies based on aggregate data, at the national, regional or industry level, where the outcome variable is the aggregate labour share; (ii) studies based on micro data, either at the firm or at the establishment level, where the outcome variable is the individual-level labour share, and (iii) studies where the aggregate labour share is analysed as an average of the individual-level labour shares, for instance by means of a shift-share analysis.[1]

Analyses of the aggregate labour share typically consider only aggregate variables —that is, totals or averages— as controls.[2] This might sound natural but as our contribution shows, the whole

[1]The imperfect measurement of capital and labour income is recognised as a potential confounder, although its importance is debated. Koh et al. (2018) suggest that the observed decline in the labour share is mostly explained by unaccounted intangible investments in R&D. This is however contrasting with Corrado et al. (2009), who show that a proper measurement of intangibles would point to a stronger increase in labour productivity, with a correspondingly stronger decline in the labour share. Elsby et al. (2013) refine the treatment of self-employment income and show that this slightly reduces the decline in the labour share, in the US. Karabarbounis and Neiman (2014) consider the case of using higher depreciation to account for less durable capital such as computers and software, but find similar trends in gross and net labour share, worldwide.

[2]A comprehensive review of the papers adopting the 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. Zeira (1998); Acemoglu (2003); Brynjolfsson and McAfee (2014) and Acemoglu and Restrepo (2018) point to (capital augmenting) technological change as a main driver, with Autor and Salomons (2018) and Eden and Gaggl (2018) focusing in particular on the role of automation. Piketty (2014); Piketty and Zucman (2014) and Glover and Short (2019) stress the role of capital accumulation. Harrison (2002); Bentolila and Saint-Paul (2003); Acemoglu (2003) and Karabarbounis and Neiman (2014) point to the decline in the price of capital relative to labour, while Hergovich and Merz (2018) and León-Ledesma and Satchi (2018) stress increased factor substitutability between capital and labour, and Grossman et al. (2018) bring the attention to a slowdown in productivity. González and Trivín (2017) point to increased asset prices, which lower investment —an explanation which is however at odds with the emphasis on capital deepening as a driver of the decline in the labour share (see above). Harrison (2002); Lee and Jayadev (2005); Guscina (2006); Daudey and García-Peñalosa (2007); Jayadev (2007); IMF (2007) and Elsby et al. (2013), among others, focus on globalisation and its implications in terms of the balance of power between capital and labour. Deregulation of product and labour markets, including privatisation policies, de-unionisation and the decline of employment-protection policies, is emphasised by Bassanini and Duval (2006); Annett (2006); Bental and Demougin (2010); Stiglitz (2012); Barkai (2016); Ciminelli et al. (2018); Dizon and Lim (2018); Dimova (2019) and Pak and Schwellnus (2019), among others. Blanchard and Giavazzi (2003), in an influential theoretical work, take into account the general equilibrium effects of deregulation policies and show that workers lose from product market deregulation but gain as consumers, and they eventually gain even from labour market deregulation, although only in long run, due to lower unemployment. Weil (2017) generically refer to the disempowerment of labour and the consequential reduction in the labour share, linked to practices such subcontracting, franchising, and a global supply chain, as *financialisation*. Furceri et al. (2018) point to financial globalisation and the liberalisation of international capital flows. Cantore et al. (2018) find an empirical relation between the decrease in the labour share and monetary policy *easing*, pointing to a new theoretical puzzle as this is inconsistent with a broad range of standard models. Rognlie (2015) and Gutierrez Gallardo (2017) point to the increase in the cost of housing and the

distribution of these variables matters. The aggregate approach is sometimes either explicitly or implicitly justified with reference to an aggregate production function at the country/industry level. Our criticism simply reflects the fact that aggregate production functions typically do not exist (see below). Thus, macro studies generally fall short of establishing causal relationships in the data. Only when a theoretical model of firm behaviour predicts an unambiguous association between variables at the micro level, irrespective of other firm's characteristics, this association can be safely tested using aggregate data. This is for instance the case of Azmat et al. (2012), where they test the prediction that privatization is associated with a lower labour share on cross-country industry-level OECD data.

Studies that analyse composition effects (e.g. Valentinyi and Herrendorf, 2008; Abdih and Danninger, 2017) can offer valuable insights about the dynamics of the aggregate labour share, but are often more descriptive in nature as they typically do not dig into what caused the shift in the composition of firms, nor they model firms' behaviour. For instance, Hopenhayn et al. (2018) point to the decline in population growth, which reduces firm entry rates and shifts the distribution of firms towards older firms with a lower labour share, in the US, but they do not offer an explanation why older firms have a lower labour share.[3]

Microeconometric studies are more causal in nature, but they generally do not derive implications for the aggregate labour share, or make the explicit or implicit assumption that what is relevant at the micro level is also relevant at the macro level, again a sort of aggregate production function type of argument. Studies that follow this approach include Siegenthaler and Stucki (2015), who study the determinants of the firm-level labour share on a panel of Swiss firms. They conclude that the most important factor in driving down the labour share is the diffusion of information and communication technologies (ICT). The aggregate labour share however remained fairly constant due to slow technological progress and sectoral reallocation towards industries with above-average labour share. Perugini et al. (2017) find a negative effect of internationalisation (in terms of export propensity, offshoring and foreign direct investment) on the firm-level labour share, using balance sheet data for six EU countries. De Loecker and Eeckhout (2018) document a rise in markups in the US from around 20% in 1980 to around 60% in the mid 2010s, well exceeding the rise in overhead costs. They link this to the decline in the labour share, mostly to the benefit of profits. An increase in product market power, coupled with a decline in rent sharing with employees, is also foud in the UK (Bell et al., 2018).

We also analyse the aggregate labour share as a weighted average of individual-level labour shares, but we characterise the behaviour of individual firms and map it directly into the aggregate outcome. Hence, in our analysis it is the joint distribution of firm-level characteristics that matters for the aggregate labour share, and not only average values. We can therefore explain the dynamics of the aggregate labour share in terms of changes in the moments of this joint distribution. Our approach

related increase in the value of capital and in real estate profits.

Our framework considers most of those determinants, in terms of their effects on firm-level variables. Technological change, given an elasticity of substitution between capital and labour smaller than 1, has a negative impact on the labour share in our model, as well as between-sectors and within-sectors shifts to relatively more capital-intensive technologies, changes in the relative price of capital with respect to labour (as brought about by a decreasing bargaining power of workers connected to globalisation and/or a weakening of labour market institutions, e.g. unions, collective bargaining and other industrial relations, minimum wages, employment protection and unemployment benefits), increased product and labour market power of firms. Only capital deepening *per se*, that is an increase in the K/L ratio at given technologies, has no effects on the labour share, in our model.

[3] Using macro data, Short and Glover (2017) point to a decreased ability of older workers to extract their marginal product of labour as a wage.

thus provides a bridge between the three perspective considered above.

Some theoretical models of firm behaviour take firm heterogeneity into account and have clearcut implications in terms of the aggregate labour share. This is the case of the theory of superstar firms (Autor et al., 2017b,a), where the driving force is an increasing "winner takes most" feature of cont markets, and of the model proposed by Aghion et al. (2019), where the driving force is a reduction in the cost of spanning multiple markets, leading to the selection of more productive firms characterised by a lower labour share, with an initial outburst of growth, followed by a low-innovation, low-growth regime. Consistently with the superstar firms narrative, Kehrig and Vincent (2017) find that the labour share has increased in most plants, but the reallocation of production towards hyper-productive, low labour share plants has caused the aggregate labour share to decline, in the US.

With respect to those papers, our model of firm behaviour is much simpler, as we follow a static, partial equilibrium approach. On the other hand, we are able to fully characterise and quantify the impact of the different dimensions of heterogeneity on the aggregate labour share, offering a comprehensive and novel decomposition method.

A paper closely related to our work is Mertens (2019). He develops a parsimonious theory of firm behaviour where three factors can affect the firm-level labour share: product market power, labour market power, and the output elasticity of labour, reflecting the importance of labour in production. Using German firm-level data, Mertens shows that his framework accounts for 94% of the observed variation in the labour share in manufacturing, between 1995 and 2014. Product and labour market power however account for only 30% of this explained change, leaving the remaining 70% to generic changes in production processes. Our theoretical framework is slightly more elaborated than his, allowing us to identify more determinants, at the cost of using a specific functional form for the production function, albeit quite general. In particular, we remain agnostic about the nature of imperfect competition in both the product and the labour market and characterise it following a reduced-form approach where a negatively sloped product demand curve and a positively sloped labour supply curve introduce a wedge between marginal costs and marginal revenues in the optimal firms' plans. This wedge is assumed to be constant irrespective of what other firms do.

2.2 Firm heterogeneity

Our focus is on between-firm heterogeneity, as opposed to within-firm heterogeneity. The literature has long recognised that some firms are more productive than others (e.g. Bernard et al., 2003; Foster et al., 2008; Hsieh and Klenow, 2009; Syverson, 2011; Aiello and Ricotta, 2015; Bartelsman and Wolf, 2017) and pay higher wages, for equally skilled workers (e.g. Dunne et al., 2004; Abowd et al., 1999; Goux and Maurin, 1999; Abowd et al., 2002; Gruetter and Lalive, 2009; Holzer et al., 2011) —see also the comprehensive review of the evidence in Van Reenen (2018).

More recently, a new generation of papers has shown that these between-firm wage differentials account for most of the overall wage inequality, and they have generally widened over time — see Barth et al. (2016) and Song et al. (2018) for the US; Faggio et al. (2010) for the UK; Card et al. (2013) for Germany; Håkanson et al. (2015) for Sweden; Card et al. (2016) for Portugal; and Elhanan Helpman and Oleg Itskhoki and Marc-Andreas Muendler and Stephen J. Redding (2017) and Alvarez et al. (2018) for Brazil.[4] Evidence across OECD countries show that the productivity gap between firms at the technology frontier and the rest has risen since the mid-2000s (Andrews et al., 2016), as well as the prevalence of and the resources sunk in "zombie" firms (McGowan et al., 2017); between-firm wage dispersion has also increased substantially, with most of the between-firm wage variance being driven by differences in pay across firms within sectors rather than by differences in average wages across sectors (Berlingieri et al., 2017). Also, Hartman-Glaser et al. (2019) make the point that as volatility of productivity has increased, the owners of the firm require an increased risk premium. Moreover, uncertainty about future productivity levels delays exit and increases the importance of mega-firms. Both factors lower the labour share.

The availability of firm-level data has allowed researchers to assess the dispersion of product market power —see among other De Loecker and Warzynski (2012) for the US, Tamminen and Chang (2013) for Finland, Forlani et al. (2016) for Belgium. The general agreement is that product market power has increased *and* has become more dispersed among firms (Epifani and Gancia, 2011; De Loecker and Eeckhout, 2018; De Loecker et al., 2018).[5] In addition to a marked increase in the average markup (see previous section), De Loecker and Eeckhout (2018) also document a substantial increase in its dispersion, with the median markup remaining roughly constant, and the 90th percentile increasing from 1.5 to 2.3. They relate the decrease in the aggregate labour share to the increase in average market power; however, they do not make any connection with its increasing dispersion.

A smaller number of papers look at labour market power, and they also find significant heterogeneity. Ransom and Oaxaca (2010) infer the elasticity of labour supply at the firm level – a measure of monopsonistic power – from the elasticity of the quit rates with respect to wages, and find for the US elasticities between 2.4 and 3 for men and between 1.5 and 2.5 for women. (Hirsch et al., 2010) for Germany and Weber (2015) for the US compute labour supply elasticities directly, using large linked employer–employee datasets, and also find considerable variation, between 1.9 and 3.7 for Germany and lognormally distributed with an average of 1.08 for the US.

Other papers look jointly at product and labour market power. Dobbelaere and Mairesse (2013) estimate production functions for different French manufacturing industries and compute firm specific price-cost markups and elasticities of labour supply as a wedge between the factor elasticities and their corresponding shares in revenues. They find considerable dispersion in both parameters. Félix and Portugal (2017) follow a similar approach for Portugal, while also decomposing the impact of the estimated labour supply elasticity on wages within an efficient bargaining setting. They estimate an average price-cost markup of 1.2, with a standard deviation of .3, and an average wage elasticity of labour supply of 3.3, with a standard deviation of 4.2. They also show that heterogeneity in monopsonistic power affects heterogeneity of wages across firms, with a one unit increase in a firm's labour supply elasticity being associated with an increase in earnings between 5 and 16 percent. Card et al. (2016) also link wage heterogeneity to labour market power, in terms of a random utility model of worker preferences which leads to firm-specific labour supply elasticities.[6]

[4]Both Card et al. (2013) and Song et al. (2018) show that the increase in between-firm wage heterogeneity is mostly due to increased worker sorting / assortative worker-firm matching (high-wage workers becoming increasingly likely to work in high-wage firms) and segregation / assortative worker-worker matching (high-wage workers becoming increasingly likely to work with each other), with little role for an increase of firm fixed effect.

[5]Fernández et al. (2015) show that in Spain heterogeneity in markups has increased significantly in some sectors (professional services, telecommunications, accommodation and food, utilities) after the Great Recession, while it has decreased in others (manufacturing).

[6]Another paper is Hornstein et al. (2011), which in the context of a search model obtain smaller wage dispersion.

Finally, the substantial heterogeneity in relative price variation, as measured typically by sectoral inflation and inflation persistence, is well documented (see, among others Blinder et al., 1998; Bils and Klenow, 2004; Lünnemann and Mathä, 2004; Bilke, 2005; Clark, 2006; Altissimo et al., 2009; Boivin et al., 2009; Wolman, 2011; Duarte and Restuccia, 2016; Kato and Okuda, 2017).

While remaining agnostic about the causes of between-firm heterogeneity, we look at the evidence of increased dispersion in wages, productivity, product and labour market power and relative inflation, and relate it to the observed changes in the labour share, for the UK manufacturing sector. As already anticipated, we find that between-firm heterogeneity is an important determinant of the aggregate labour share, particularly heterogeneity in total factor productivity and labour market power. However, its contribution has remained fairly constant over time, and therefore cannot explain the observed decline in the aggregate labour share.

3 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 (Green, 1964; Fisher, 1969; Zambelli, 2004; Felipe and McCombie, 2014). Thus, under firm heterogeneity the aggregate labour share cannot be computed 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 adding up labour costs and value added across firms. Here we use a simple neoclassical model of firm behaviour in order to characterise the relationship between the distribution of firms' characteristics and the aggregate labour share in the economy, in a partial equilibrium setting.

3.1 Setup

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

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

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.[7]

The aggregate labour share, defined as aggregate labour costs over aggregate value added, can then be expressed as a weighted average of λ_i:

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

[7] In practice, workers are heterogeneous (e.g. in terms of skills, type of contract, or hours worked). However, most datasets, including ours, only report the total number of employees. Therefore, because of necessity rather than desire, the theory assumes workers are homogeneous within the firm. In the empirical analysis, $w_i L_i$ is taken to be the reported total labour costs, which means w_i is defined as the average wage per worker.

where $\delta_i = \dfrac{p_i Y_i}{\sum_i p_i Y_i}$ corresponds to the share of aggregate value added produced by firm i.

Our aim is to characterise 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 *value added* production function (i.e. a mathematical relation between capital, labour and value added).[8] In particular, we assume a CES production function:

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

where $\sigma = \frac{1}{1-\rho}$ is the elasticity of substitution between capital and labour (hence: $\rho < 1$). Notice firms have the same technology in terms of elasticities (ρ and α), but they might have heterogeneous total factor productivity (TFP), A_i. A justification for the assumptions in equation (3) is presented later, once the main result is obtained.

We assume firms have a certain degree of *monopolistic power* in the pricing of the final good. Importantly, the degree of market power might be heterogeneous across firms. Formally, firms face an inverse demand function for their good given by$p_i(Y_i) = f(\eta_i^Y, \Theta_i^Y)$, where η_i^Y corresponds to the own-price elasticity of output demand, and Θ_i^Y refers to arbitrary characteristics of the product Y_i, idiosyncratic to firm i, which are valuable to consumers. Similarly, we assume firms have some degree of *monopsony power* in the labour market, which could also be heterogeneous across firms (for example, because of some non-pecuniary location effects valued by workers). Formally, firms face an inverse labour supply function given by $w_i(L_i) = g(\eta_i^L, \Theta_i^L)$, where η_i^L is the own-price elasticity of the labour supply, and Θ_i^L represents idiosyncratic firm characteristics, valuable for workers. The role of Θ_i^Y and Θ_i^L is to permit heterogeneous prices and wages even when firms have the same level of market power, or when they have no market power at all. The latter is not unknown to the literature, both in the case of firms with an homogeneous final good (Dahlby and West, 1986; Hosken and Reiffen, 2004) and homogeneous labour (Rosen, 1987; Hamermesh, 1999).

With the above assumptions in place, the profit function of the firm is $\Pi_i(L_i, K_i) = p_i(Y_i)Y_i - w_i(L_i)L_i - r_i K_i$. The first order 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 $\chi_i^Y = 1 + \frac{1}{\eta_i^Y}$. The term $\frac{\chi_i^L}{\chi_i^Y}$ represents the wedge between the real wage 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 maximisation requires $|\eta_i^Y| > 1$, so that χ_i^Y is always positive.

[8] 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. Unfortunately, because of the multiple non-linearities in our model, we were unable to test the elasticity of substitution of the gross output production function linked to our model (a nested CES). Regarding the price condition, we do not observe the price of intermediary inputs, and so cannot test this assumption either. For further details on the gross output production function associated with our model, see Appendix A.

From equation (4) we obtain the optimal $\frac{L_i}{Y_i}$ as a function of the firm characteristics:

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

This is then replaced into the formula for the firm level labour share (equation 1), leading to:

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

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). This property emanates from the fact that the CES function is homothetic. This means it has a linear expansion path, which is to say, 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., 2017b; Schwellnus et al., 2018). Additionally, in our framework, wages and prices do depend on L_i whenever there is imperfect competition.

Second, 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.[9]

Third, there is a close relationship between the pay-productivity disconnect (with productivity understood as TFP) and the labour share. In particular, the latter changes whenever a given increase in TPF does not translate into a similar increase in the real wage (i.e when $\frac{A_i p_i}{w_i}$ falls). Again, the final effect depends on ρ. Further analysis of the effect of individual firm level variables on the firm level and aggregate labour share is presented in Appendix C.

Finally, notice we do not model firms' choice of capital, as it is not needed in our framework. 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).

3.2 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

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

the aggregate labour share:

$$\lambda = \sum_i \left(\frac{\alpha\chi_i^Y}{\chi_i^L}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i p_i}{w_i}\right)^{\frac{\rho}{1-\rho}} \delta_i \tag{7}$$

We measure firm heterogeneity with respect to an hypothetical "average" firm. More specifically, for given relative weights $\{\omega_i\}$ we define $\bar{A} = \sum_i \omega_i A_i$, $\bar{w} = \sum_i \omega_i w_i$, $\bar{p} = \sum_i \omega_i p_i$, $\bar{\chi}^Y = \sum_i \omega_i \chi_i^Y$, $\bar{\chi}^L = \sum_i \omega_i \chi_i^L$. 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 a similar employment level (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 one often used in the literature to aggregate firms (e.g. De Loecker and Eeckhout, 2018; De Loecker et al., 2018, in the context of mark-ups). Notice however that the method itself is agnostic regarding the weights chosen. What is needed is that heterogeneity is quantified with respect to a given counterfactual, just as the variance is computed with respect to a mean. As long as there is heterogeneity in a given dimension (except capital alone, as Proposition 1 below states), such decomposition is **always possible**.

Having defined weighted averages for every variable we can then re-write the aggregate LS as:

$$\lambda = \lambda^{HOM} \sum_i \left(\frac{\chi_i^Y}{\bar{\chi}^Y}\right)^{\frac{1}{1-\rho}} \left(\frac{\bar{\chi}^L}{\chi_i^L}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i}{\bar{A}}\right)^{\frac{\rho}{1-\rho}} \left(\frac{\bar{w}}{w_i}\right)^{\frac{\rho}{1-\rho}} \left(\frac{p_i}{\bar{p}}\right)^{\frac{\rho}{1-\rho}} \delta_i \tag{8}$$

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

$$\lambda^{HOM} = \left(\frac{\bar{\alpha}\bar{\chi}^Y}{\bar{\chi}^L}\right)^{\frac{1}{1-\rho}} \left(\frac{\bar{A}\bar{p}}{\bar{w}}\right)^{\frac{\rho}{1-\rho}} \tag{9}$$

Equation (8) is our 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 6).[10] The proof can be trivially seen in equation (2), once we assume $\lambda_i = \lambda^{HOM}$.

The following proposition summarises 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 Cobb-Douglas). 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);

[10] 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.

(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 of the model. Yet, provided we can produce an estimate for each element in equation (8), this is sufficient for our purposes. This partial equilibrium approach assumes any observed dataset reflects a situation of general equilibrium. The drawback of this partial method 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 case, where the aggregate LS is:

$$\lambda = \frac{\alpha \bar{\chi}^Y}{\bar{\chi}^L} \sum_i \left(\frac{\chi_i^Y}{\bar{\chi}^Y} \right) \left(\frac{\bar{\chi}^L}{\chi_i^L} \right) \delta_i \tag{8'}$$

This highlights that for firm heterogeneity to affect the aggregate labour share if the technology is Cobb-Douglas, there must be *heterogeneous* imperfect competition. With perfect competition (where an exact aggregate production function exists), $\lambda = \alpha$, a well-known property of a Cobb-Douglas production function. The following proposition summarises 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 and 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 Cobb-Douglas 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.

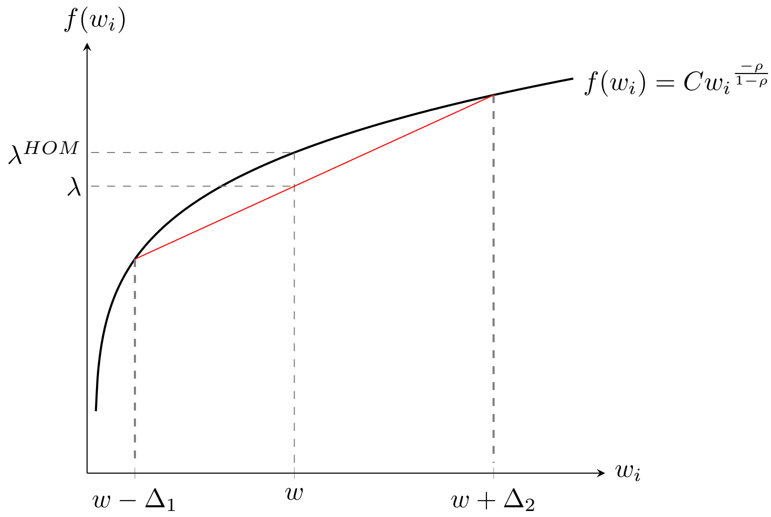
On the other hand, a CES enables a richer set of determinants for the labour share, reason why it is our preferred choice. However, it might seem odd that in our CES analysis we assume the production function to be homogeneous across firms (i.e. common parameters ρ and α). This is necessary as allowing heterogeneity in ρ impedes decomposition, and allowing heterogeneity in α greatly complicates estimation (see Appendix A for details).

Our last remark is on the choice of a value added production function. Using a gross output production function does not yield a decomposable formula for the aggregate labour share, except when the conditions suggested by Bruno (1978) are fulfilled. This is, when a value added production function exists, as assumed here. Appendix A provides further insights on these points.

3.3 Exercise: A mean-preserving increase in wage dispersion

To better illustrate the implications of equation (8), we now consider the case of a mean-preserving increase the dispersion of one variable only, namely wages.

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



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 the 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 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.[11]

In this setting, each firm's LS is (equation 8):

$$\lambda_i = Cw_i^{\frac{-\rho}{1-\rho}} \tag{10}$$

where $C = \left(\frac{\alpha\chi^Y}{\chi^L}\right)^{\frac{1}{1-\rho}} (Ap)^{\frac{\rho}{1-\rho}}$ (the part of the labour share which is identical across firms).

This function is monotonically increasing in wages and concave for $\rho < 0$, and monotonically decreasing in wages and convex 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 is identical to the ones existed before the change (hence the purpose of the mean-preserving spread), 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 Cobb-Douglas, where dispersion in wages **alone** does not change the aggregate LS, which is constant over w_i.

The reason why the aggregate LS falls in the example above is nothing else than Jensen's inequality, given the shape of the LS function. But why does the LS function depend on ρ? To understand this, let's first look at the first derivative, and explain why the LS is increasing in wages for $\rho < 0$,

[11] 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). Being our model a partial equilibrium one, we assume any resource constraints are fulfilled. In other words, prices represent an equilibrium.

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 Cobb-Douglas 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 $w \xrightarrow{\infty}$); 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.

The above example of wage heterogeneity also holds in the case of an unweighted mean-preserving spread of wages (i.e. where $\Delta_1 = \Delta_2 = \Delta$). The only difference is that the counterfactual wage that produces an equivalent level to that of the new (heterogeneous) aggregate LS is no longer w but $w - \Delta(\delta_2 - \delta_1)$. This level is lower (higher) than w for $\rho < 0$ ($\rho > 0$), only strengthening the result. Furthermore, it can be shown that the same conclusion arises for changes from an already heterogeneous economy, under plausible circumstances.

Finally, a similar analysis to that of wages could be made for other sources of heterogeneity. The ultimate behaviour of a given increase in heterogeneity rests on the exponent of the term in the firm level LS function. For instance, for $\rho < 0$, an increase in productivity dispersion leads to an increase in the aggregate LS.

3.4 Distributional characterisation

Proposition 1 is very general. In particular, it does not quantify how heterogeneity affects the aggregate labour share: the summation term in equation (8) 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 (8) 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. After dropping all interaction terms of order higher than two, equation (8) can be approximated by:[12]

$$\lambda \approx \lambda^{HOM} \sum_i \delta_i \Big[1 + \frac{1}{1-\rho}\left(\frac{\Delta\chi_i^Y}{\bar{\chi}^Y}\right) - \frac{1}{1-\rho}\left(\frac{\Delta\chi_i^L}{\bar{\chi}^L}\right) + \frac{\rho}{1-\rho}\left(\frac{\Delta A_i}{\bar{A}}\right) - \frac{\rho}{1-\rho}\left(\frac{\Delta w_i}{\bar{w}}\right) + \frac{\rho}{1-\rho}\left(\frac{\Delta p_i}{\bar{p}}\right)$$
$$+ \frac{\rho}{2(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\bar{\chi}^Y}\right)^2 + \frac{2-\rho}{2(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\bar{\chi}^L}\right)^2 + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\left(\frac{\Delta A_i}{\bar{A}}\right)^2 + \frac{\rho}{2(1-\rho)^2}\left(\frac{\Delta w_i}{\bar{w}}\right)^2 + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\left(\frac{\Delta p_i}{\bar{p}}\right)^2$$
$$- \frac{1}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\bar{\chi}^Y}\right)\left(\frac{\Delta\chi_i^L}{\bar{\chi}^L}\right) + \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\bar{\chi}^Y}\right)\left(\frac{\Delta A_i}{\bar{A}}\right) - \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\bar{\chi}^Y}\right)\left(\frac{\Delta w_i}{\bar{w}}\right)$$
$$+ \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^Y}{\bar{\chi}^Y}\right)\left(\frac{\Delta p_i}{\bar{p}}\right) - \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\bar{\chi}^L}\right)\left(\frac{\Delta A_i}{\bar{A}}\right) + \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\bar{\chi}^L}\right)\left(\frac{\Delta w_i}{\bar{w}}\right)$$
$$- \frac{\rho}{(1-\rho)^2}\left(\frac{\Delta\chi_i^L}{\bar{\chi}^L}\right)\left(\frac{\Delta p_i}{\bar{p}}\right) - \frac{\rho^2}{(1-\rho)^2}\left(\frac{\Delta A_i}{\bar{A}}\right)\left(\frac{\Delta w_i}{\bar{w}}\right) + \frac{\rho^2}{(1-\rho)^2}\left(\frac{\Delta A_i}{\bar{A}}\right)\left(\frac{\Delta p_i}{\bar{p}}\right)$$
$$- \frac{\rho^2}{(1-\rho)^2}\left(\frac{\Delta w_i}{\bar{w}}\right)\left(\frac{\Delta p_i}{\bar{p}}\right)\Big]$$

This can be simplified further. First, notice that when $\bar{z}$ 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), both of which are dimensionless and scale invariant:

$$\lambda \approx \lambda^{HOM} \sum_i \delta_i \Big[1$$
$$+ \frac{\rho}{2(1-\rho)^2}\mathrm{CV}^2(\chi_i^Y) + \frac{2-\rho}{2(1-\rho)^2}\mathrm{CV}^2(\chi_i^L) + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\mathrm{CV}^2(A_i) + \frac{\rho}{2(1-\rho)^2}\mathrm{CV}^2(w_i) + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\mathrm{CV}^2(p_i)$$
$$- \frac{1}{(1-\rho)^2}\mathrm{r}(\chi^Y,\chi^L)\mathrm{CV}(\chi^Y)\mathrm{CV}(\chi^L) + \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^Y,A)\mathrm{CV}(\chi^Y)\mathrm{CV}(A) - \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^Y,w)\mathrm{CV}(\chi^Y)\mathrm{CV}(w)$$
$$+ \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^Y,p)\mathrm{CV}(\chi^Y)\mathrm{CV}(p) - \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^L,A)\mathrm{CV}(\chi^L)\mathrm{CV}(A) + \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^L,w)\mathrm{CV}(\chi^L)\mathrm{CV}(w)$$
$$- \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^L,p)\mathrm{CV}(\chi^L)\mathrm{CV}(p) - \frac{\rho^2}{(1-\rho)^2}\mathrm{r}(A,w)\mathrm{CV}(A)\mathrm{CV}(w) + \frac{\rho^2}{(1-\rho)^2}\mathrm{r}(A,p)\mathrm{CV}(A)\mathrm{CV}(p)$$
$$- \frac{\rho^2}{(1-\rho)^2}\mathrm{r}(w,p)\mathrm{CV}(w)\mathrm{CV}(p)\Big] \tag{12}$$

This final equation reflects that it's the joint distribution of all the variables that affects the aggregate labour share. In particular, heterogeneity in each variable (defined in terms of the coefficient of variation), increases or decreases the aggregate labour share, depending on ρ. The only exception is monopsony power: an increase in the dispersion of monopsony power always increases the aggregate labour share (remember $\rho < 1$). Importantly, heterogeneity matters even if all variables are orthogonal to each other, i.e if all correlations are zero. As this is not likely to be the case however, the correlation structure does matter, at it pushes up or down the aggregate labour share again depending on ρ, for most variables.

The above result is summarised in the following proposition.

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

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 heterogeneous variables, and for most of the variables, on ρ. For the most 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. Other things being equal, 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 visualise, as any change in firm-level variables will typically trigger a change in the market shares. This in turns will cause a change in λ^{HOM}, which refers to the hypothetical labour share of a weighted average firm. Hence, the *ceteris paribus* clause typically won't hold in simple thought experiments. Returning to our example in section 3.3, in the case of nominal wage heterogeneity only equation (12) simplifies to:

$$\lambda \approx \lambda^{HOM}\left[1 + \frac{\rho}{2(1-\rho)^2}\mathrm{CV}^2(w)\right] \tag{12'}$$

In section 3.3 we showed the effects of a mean-preserving increase in wage dispersion, that is a case where λ^{HOM} remains constant while $\mathrm{CV}(w)$ increases. We also showed that the result still holds (and is indeed reinforced) if the increase in wage dispersion happens around the unweighted mean, rather than the weighted one. Equation $12'$ quantifies the effect. Recall that for keeping the *weighted* mean (hence, λ^{HOM}) constant, we need to have $\Delta_1 = \Delta_2\frac{\delta_2}{\delta_1}$: the increase in wages at firm 1 is bigger than the decrease at firm 2. To keep the *unweighted* mean constant we have to further lower wages at firm 2. This effectively lowers λ^{HOM}. Hence, the aggregate labour share falls because of both a reduction in λ^{HOM} and an increase in $\mathrm{CV}(w)$, when $\rho < 0$.

4 Data[13]

4.1 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 1998-2014 period.[14] We focus on the manufacturing sector because value added is very imperfectly measured in other sectors,

[13]Unfortunately, we are unable to provide the data underlying our results because this can only be accessed through the UK Data Service's secure lab. Nevertheless, we plan to make our code public, so anyone with access to the dataset can reproduce our results. Information about the dataset and how to access it can be found at http://doi.org/10.5255/UKDA-SN-7989-4

[14]Although our analysis is for Geat Britain only, for simplicity to the reader, we refer to our sample as the UK. The approximation might still be valid enough, given that during the sample period, manufacturing's GVA in Northern Ireland has been constantly below 3% of the UK-wide level, according to ONS data.

where intermediate inputs are less clearly identified – see for instance the discussion in Autor et al. (2017b).

We use data from 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.[15] 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 (for example, 97% in 2014). We focus only on manufacturing, where a production function at the firm level is more likely to be well defined.[16]

Several sample selection procedures were made. First, 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 (combination which might depend on the tax system). This might distort the computation of the labour share in ways unrelated to the theory.[17] Second, non-profit and other non-market oriented firms were excluded, as these are less likely to be characterised by profit maximising behaviour. Third, firms with missing information (e.g. no investment data, needed to compute capital stocks) were also dropped. Fourth, outliers in terms of top and bottom 0.5% percentiles, computed independently for different variables (including 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. In any case, all the analysis presented here is carried out using turnover-based sampling weights, in order to represent the whole sector as good as possible.

4.2 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 firm.

[15] 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 organisations), 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 ONS (2012).

[16] As suggested by Schwellnus et al. (2017), we drop subsector 19 in the SIC07 classification ("manufacturing of coke and refined petroleum"), because of the noise introduced by the volatility of oil prices.

[17] Two things must be mentioned here. First, regarding the actual threshold of 10 employees, 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. Second, 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 D). These issues and other information about sampling in ARD can be found in ONS (2012).

Prices

Our theory is build upon firm-level prices; however, no price information is available in ARD. Instead, we use the most disaggregated industry-level producer price index available (4 digits), provided by the Office for National Statistics.

There are, of course, differences between the industry price index and the firm price level, p_i. In practice however, using the former is not only our only option but it is, theoretically speaking, more informative. To see this, notice that in the theory, p_i refers to the price of a unit of physical output (or "real" value added), e.g. apples. This poses two difficulties. On the one hand, the unit of measurement of prices is arbitrary. For instance, if we were to measure p_i in terms of a kilo of apples instead of a unit of apples, the price level would change. On the other hand, comparing the price level of different goods (e.g. apples with cars) is uninformative. Because of these issues, price heterogeneity among heterogeneous goods can only be made sense if the price of goods are normalised to a common unit. This is precisely the goal of a price index.[18]

To see this more clearly, define as $p_{j,0}$ the price *level* in sector j in the base period of the index, and $\pi_j = p_j/p_{j,0}$ as the sectoral price index. We do not observe p_i but the π_j to which the firm belongs. Thus, in all the theoretical analysis presented so far, we could replace p_i with π_j. In other words, when we talk about price heterogeneity, we are referring to a *relative* price heterogeneity, in the sense that it is the *change* in the measure which is informative, and not the level. In effect, by construction, price heterogeneity is zero in the base year. Again, this is the only type of heterogeneity which is informative when dealing with heterogeneous goods.

Now, the replacement of p_i with π_j induces a second change. If we are dividing p_i by $p_{j,0}$, we must also multiply it somewhere else. As it turns out, it is the TFP term which captures the extra $p_{j,0}$ because, as shown later, TFP is estimated from a regression where value added is deflated by the same price index. Thus, when we replace p_i with π_j, we are also replacing A_i with $\tilde{A}_i$, where $\tilde{A}_i = A_i p_{j,0}$. This transformation is not a problem. In fact, TFP suffers from the same problem than prices. Notice that the units of measurement of TFP are directly related to those of output. Thus, if we are talking about heterogeneous goods like apples and cars, comparing A_i *levels* across firms is just as uninformative as comparing p_i across firms. Thus, TFP must also be normalised in order to be comparable. And $p_{j,0}$ is precisely such a normalisation.

In summary, by replacing firm level prices with that of an industry-level index (in terms of the equations, by replacing p_i with π_j and A_i with $\tilde{A}_i$), we achieve an empirical implementation of the general theoretical result, where the *interpretation* of price and TFP heterogeneity is informative. For notation economy though, we keep using p_i and A_i.

Estimation of TFP, α and ρ

A very important element of the model not available in the sample is a measure of firms' total factor productivity, A_{it}. Additionally, we do not observe the production function parameters α and ρ. The solution is to estimate the CES value added production function, like that in equation (3). As extensively noted in the literature (e.g. Olley and Pakes, 1996), it is necessary to account

[18] Notice that the two issues are independent. If firm-level prices were available, we could construct a price index based on firm level prices. As these are not available, we use industry level prices as a proxy.

for the potential endogeneity of employment which, being a variable factor, might respond to contemporary unobserved shocks to TFP. This is done following 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 D. Here we just highlight that the estimated elasticity of substitution (for the manufacturing sector as a whole) is 0.48 ($\hat{\rho} = -1.07$), significant at the 1% 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 D for further details.[20]

Finally, having estimated α and ρ, we can use the production function to compute $\hat{A}_{it}$ as a residual.[21] This can be done also for observations not used in the estimation of the production function (for example, because of missing data in a given year). This means that the final sample used for the decomposition is larger than the one used in the regression. For details, see Appendix D.

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. Importantly, this share is computed for each occupational group, after which a weighted average is produced for each firm. The aim of this occupation-adjusted share is to reflect different occupational composition of firms vis-a-vis that of the local labour market (e.g. a firm employing mostly high skill workers in a local labour market with mostly low skill workers has more market power than a firm mostly employing low skill workers in the same local labour

[19] The resulting model is highly non-linear (see equation (25) in Appendix D), and GMM does not converge on our data (in effect, most of the literature estimates Cobb-Douglas production functions, which are log-linear in the parameters). We therefore estimated 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 -1.07.

[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 Cobb-Douglas setting.

[21] As explained in Appendix D, it is impossible to identify the shock to value added. This is therefore included in the computation of $\hat{A}_{it}$. 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 footnote 34.

market).[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^L_{it}. 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^L_{it} 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

$$\eta^L_{it} = -c_1 \left(\frac{1}{ln\,(1 - s_{it})} \right)^{c_2}$$

This is flexible because it is possible to shape this function by changing the positive terms c_1 and c_2, holding the end points mentioned above fixed. These constants are chosen in order to match the scarce evidence available in the literature about η^L_{it} at the firm level. In effect, we match two empirical properties of η^L_{it}. First, Manning (2003) estimates an average firm level elasticity of supply for the UK of around 0.75. Second, Webber (2015) provides a characterisation 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).

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., 2018).[26]

4.3 Theoretical versus empirical decomposition

Before moving forward, an important issue must be dealt with. The decomposition formulas are built upon the optimisation behaviour of firms. Thus, they refer to the *predicted* labour share of

[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] The official documentation from the Office for National Statistics defines a TTWA as "[i]n concept, a self-contained labour market area is one in which all commuting occurs within the boundary of that area. In practice, it is not possible to divide the UK into entirely separate labour market areas as commuting patterns are too diffuse. TTWAs have been developed as approximations to self-contained labour markets reflecting areas where most people both live and work." 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 Table 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] Unfortunately, we can not implement the mark-up estimation method put forward by De Loecker and Warzynski (2012), where mark-ups are derived from the first order condition of the *gross output* production function with respect to a fully variable and competitive input (in our case, intermediate inputs, as labour is subject to imperfect competition). As already mentioned before, estimating a gross output function consistent with our framework (that is a nested CES, where the value added production function is one of the inputs, and materials is the other), proved impossible.

firms, as given by equation (6). However, the objective of the decomposition is to characterise *observed* labour shares (in terms of observed value added and labour costs). Naturally, there will be differences between these two. There are a multitude of reasons why predicted and observed values can be different. For a start, in terms of our theory, firms' optimisation process might be more complex (e.g. dynamic rather than static), or firms might face constraints that lead to misallocation of resources (i.e. firms are not efficient). There is substantial evidence in the literature for this. Even if the theory is a good enough approximation to reality, discrepancies might arise from imprecise or inconsistent estimates for the variables and parameters of the model (e.g. market power, or TFP). Additionally, variables in the data could be measured with errors. Last but not least, the stochastic nature of firms' production means the latter is subject to idiosyncratic shocks (e.g. productivity), which will always deviate the observed values from the predicted ones. This is an irreducible source of theory-data mismatch, at least at the firm level.

The discrepancy between predicted and observed labour share introduces an extra term into the decomposition.[27] To see this, let us define $\tau_{it} \equiv \dfrac{\lambda_{it}^{obs}}{\lambda_{it}}$, which captures the divergence between the observed and predicted labour share for firm i in period t, where the latter is given by equation (6). Now, consider the following identity regarding the aggregate labour share (for notation economy, time index is omitted throughout):

$$\lambda^{obs} \equiv \sum_i \lambda_i^{obs}\, \delta_i^{obs}$$

where $\delta_i^{obs} = \frac{p_i Y_i^{obs}}{\sum p_i Y_i^{obs}}$. This is the empirical counterpart to the labour share definition in equation (2). The connection between the model and data is done precisely through τ_i. This is, we can write:

$$\lambda^{obs} = \sum_i \lambda_i\, \tau_i\, \delta_i^{obs} \tag{13}$$

Replacing the predicted firm level labour share λ_i by its components (equation 6), and introducing the counterfactual λ^{HOM}, produces:

$$\lambda^{obs} = \lambda^{HOM} \sum_i \left(\frac{\chi_i^Y}{\bar{\chi}^Y}\right)^{\frac{1}{1-\rho}} \left(\frac{\bar{\chi}^L}{\chi_i^L}\right)^{\frac{1}{1-\rho}} \left(\frac{A_i}{\bar{A}}\right)^{\frac{\rho}{1-\rho}} \left(\frac{\bar{w}}{w_i}\right)^{\frac{\rho}{1-\rho}} \left(\frac{p_i}{\bar{p}}\right)^{\frac{\rho}{1-\rho}} \left(\frac{\tau_i}{\bar{\tau}}\right) \delta_i^{obs} \tag{14}$$

where

$$\lambda^{HOM} = \bar{\tau} \left(\frac{\alpha \bar{\chi}^Y}{\bar{\chi}^L}\right)^{\frac{1}{1-\rho}} \left(\frac{\bar{A}\bar{p}}{\bar{w}}\right)^{\frac{\rho}{1-\rho}} \tag{15}$$

and $\bar{\tau} \equiv \sum_i \omega_i \tau_i$, a weighted average of the discrepancy term. The introduction of $\bar{\tau}$ in the above equation is useful because $\dfrac{\tau_i}{\bar{\tau}}$ can be redefined as $1 + \frac{\Delta_{\tau_i}}{\bar{\tau}}$, which is similar in structure to the second order Taylor approximation of the other terms inside the summation. This ensures the final decomposition is defined in terms of correlations and coefficients of variation only, as was equation (12) before.

[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.

With the above modification, and looking at the ratio w/p to further simplify the notation (this has also the advantage to identify a real wage term in the expression for λ^{HOM}), we obtain the final "empirical" decomposition formula, used in the subsequent analysis:

$$\begin{aligned}
\lambda \approx \lambda^{HOM} \Bigg[1 &+ \frac{\rho}{2(1-\rho)^2}\mathrm{CV}^2(\chi^Y) + \frac{2-\rho}{2(1-\rho)^2}\mathrm{CV}^2(\chi^L) + \frac{\rho(2\rho-1)}{2(1-\rho)^2}\mathrm{CV}^2(A) + \frac{\rho}{2(1-\rho)^2}\mathrm{CV}^2\left(\frac{w}{p}\right) \\
&- \frac{1}{(1-\rho)^2}\mathrm{r}(\chi^Y,\chi^L)\mathrm{CV}(\chi^Y)\mathrm{CV}(\chi^L) - \frac{\rho^2}{(1-\rho)^2}\mathrm{r}(A,\frac{w}{p})\mathrm{CV}(A)\mathrm{CV}\left(\frac{w}{p}\right) \\
&+ \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^Y,A)\mathrm{CV}(\chi^Y)\mathrm{CV}(A) - \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^L,A)\mathrm{CV}(\chi^L)\mathrm{CV}(A) \\
&- \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^Y,\frac{w}{p})\mathrm{CV}(\chi^Y)\mathrm{CV}\left(\frac{w}{p}\right) + \frac{\rho}{(1-\rho)^2}\mathrm{r}(\chi^L,\frac{w}{p})\mathrm{CV}(\chi^L)\mathrm{CV}\left(\frac{w}{p}\right) \\
&+ \frac{1}{1-\rho}\mathrm{r}(\chi^Y,\tau)\mathrm{CV}(\chi^Y)\mathrm{CV}(\tau) - \frac{1}{1-\rho}\mathrm{r}(\chi^L,\tau)\mathrm{CV}(\chi^L)\mathrm{CV}(\tau) \\
&+ \frac{\rho}{1-\rho}\mathrm{r}(A,\tau)\mathrm{CV}(A)\mathrm{CV}(\tau) - \frac{\rho}{1-\rho}\mathrm{r}(\frac{w}{p},\tau)\mathrm{CV}\left(\frac{w}{p}\right)\mathrm{CV}(\tau)\Bigg] \qquad (16)
\end{aligned}$$

The only difference with the theoretical counterpart is the addition of the last four terms, and the introduction of $\bar{\tau}$ in λ^{HOM}. 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. $\mathrm{CV}(\tau) = 0$), there is no difference between the theoretical and empirical decomposition.

5 Results[28]

5.1 Descriptive analysis of the labour share

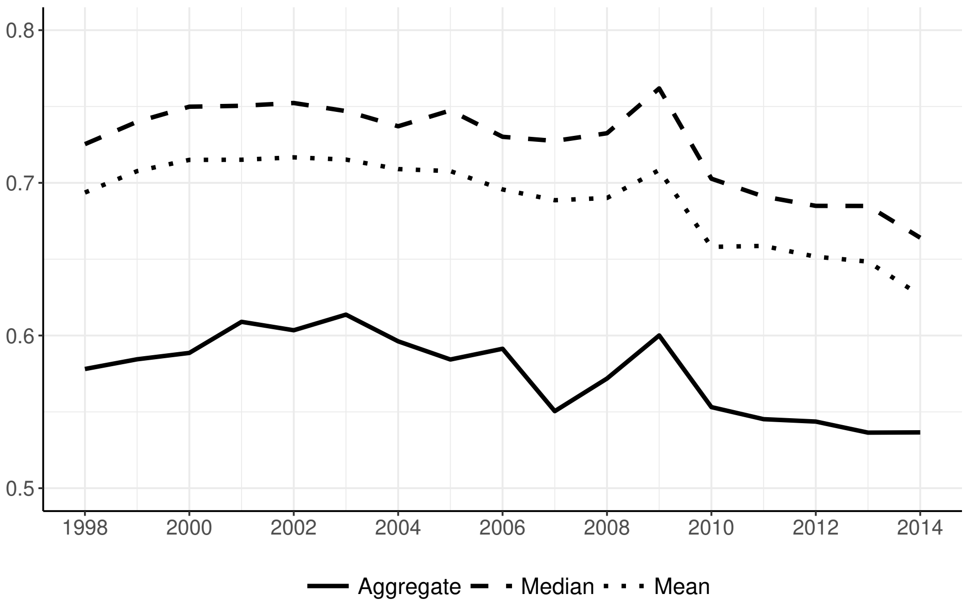
Before proceeding with the decomposition, it is useful to describe the labour share in our sample. Figure 2 presents different metrics for the latter, covering UK manufacturing, between 1998 and 2014. Starting with the aggregate labour share, we see a net fall over the period, from 0.58 in 1998 to 0.53 in 2014.[29] 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 also presents the mean and median labour share, which are above the aggregate labour share, highlighting that firms with higher value added (our measure of firm size) have a lower labour share. This is consistent with other findings in the literature (e.g. Autor et al., 2017b; Schwellnus et al., 2018).

Since the aggregate labour share is defined in terms of a weighted sum of firm level labour share (equation 2), changes in the labour share can be due to changes in the magnitude of the firm level labour shares, in the distribution of weights across different labour share levels, or both. Figure

[28] 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 1,000 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 and Tibshirani, 1986). This takes the point estimates as the center 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.

[29] The labour share in manufacturing, computed from national accounts, shows an increase in the labour share 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).

Figure 2: Aggregate, mean and median labour share



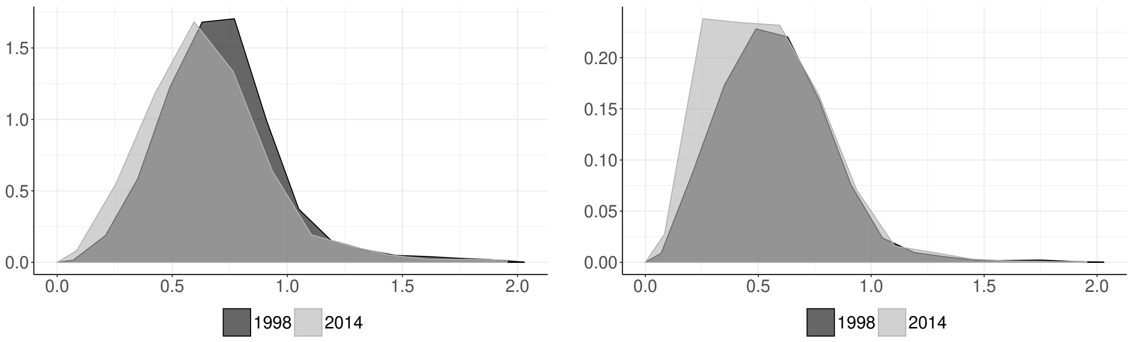
Source: our calculation based on ARD data.
Sample: UK manufacturing firms with 10 employees or more, ARD data.
Note: 95% confidence intervals are displayed as a shadowed area.

2 already shows a fall in the average labour share. Another perspective is seen in Panel (a) of Figure 3, which shows that the sample distribution of firms' labour share in 2014 has more mass at lower labour share levels than in 1998. Similarly, Panel (b) captures the distribution of value added across different labour share levels, showing that in 2014 more value added was produced by firms with lower labour share than in 1998.[30]

Changes in Panel (b) however might merely be reflecting changes in Panel (a), i.e. the fact that there are more firms at lower labour share levels. To identify the effect of changes in the composition of weights on the aggregate labour share, we compute the ratio between the average and aggregate labour share. This ratio is a measure of the covariance between λ_i and δ_i. In particular, when this ratio is one, these two variables are perfectly orthogonal; unweighted and weighted labour share are the same. Conversely, the higher the ratio, the more negative this covariance is.[31] As said earlier, smaller firms have higher labour share, so this covariance is negative. Figure 4 presents the evolution of this ratio. There is no obvious trend in this variable, 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. One way to confirm this is by a simple counterfactual exercise, where we either keep the distribution of the labour share (Panel (a) in Figure 3) or the distribution of value added (Panel (b) in same figure) to its value in 1998, and measure the counterfactual aggregate labour share in 2014. This exercise reveals that changes in the distribution of value added has a minor impact on the aggregate labour share; it's the change in the level of the labour share that matters the most.

[30]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).
[31]The mathematical characterisation of this ratio is presented in Appendix B.

Figure 3: Histograms related to changes in the labour share, 1998 and 2014

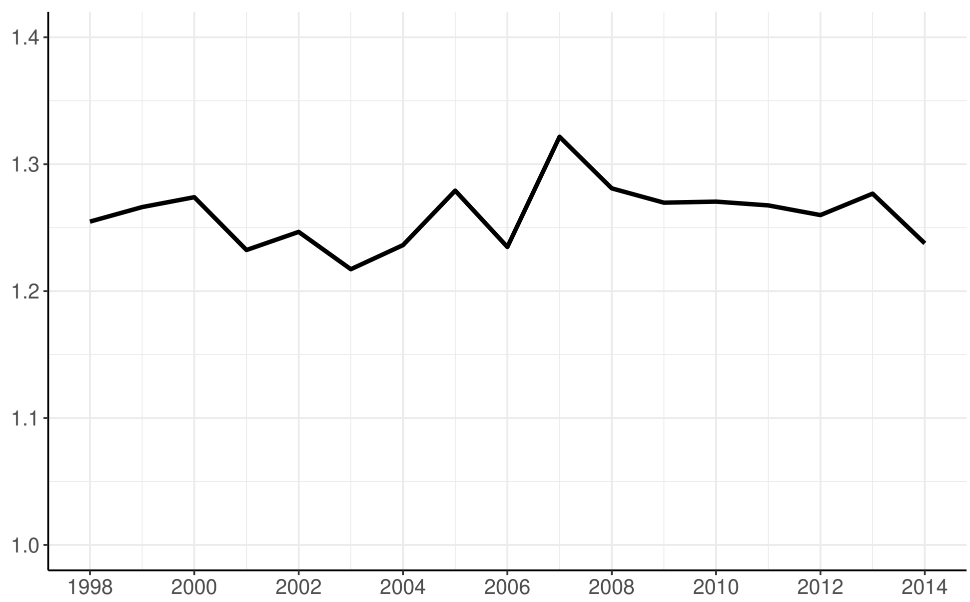


(a) Labour share distribution at the firm level (b) Total value added across labour share levels

Source: our calculation based on ARD data.

Sample: UK manufacturing firms with 10 employees or more.

Figure 4: Ratio between average and aggregate labour share

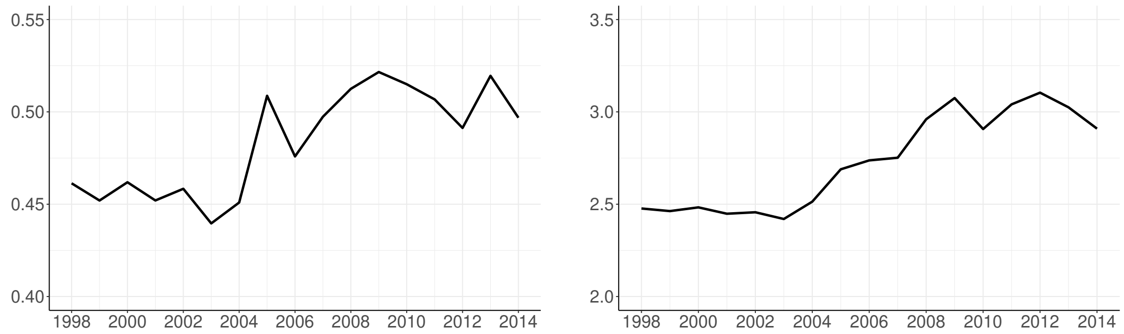


Source: our calculation based on ARD data.

Sample: UK manufacturing firms with 10 employees or more.

Note: 95% confidence intervals are displayed as a shadowed area.

Figure 5: Dispersion in the labour share



(a) Coefficient of variation

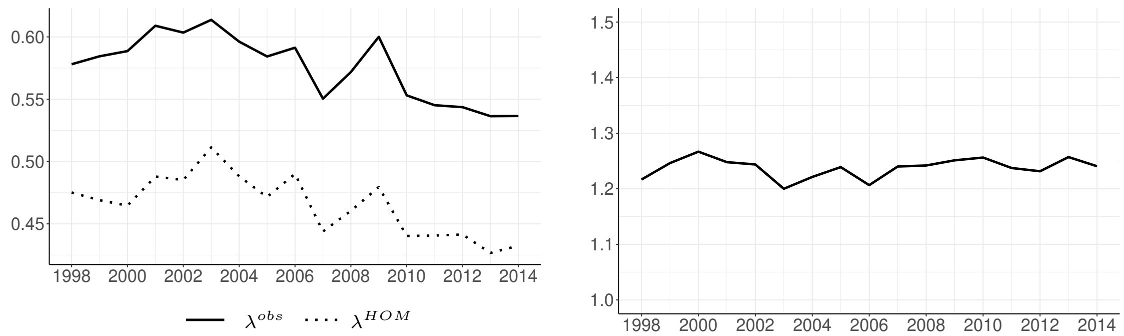
(b) P90/P10 ratio

Source: our calculation based on ARD data.

Sample: UK manufacturing firms with 10 employees or more.

Note: 95% confidence intervals are displayed as a shadowed area.

Figure 6: Decomposition of the aggregate labour share



(a) Aggregate labour share (λ^{obs}) and counterfactual labour share (λ^{HOM})

(b) Heterogeneity component, $\sum$

Source: our calculation based on ARD data.

Sample: UK manufacturing firms with 10 employees or more.

Note: 95% bootstrapped confidence intervals are displayed as a shadowed area.

Importantly, as Panel (a) in Figure 3 shows, the fall in the level of the labour share has not been an homogeneous phenomenon. In effect, the upper tail of the distribution barely changed between the two years. This reflects an increase in the dispersion of the labour share. Figure 5 documents this change, computed either as a coefficient of variation or a p90/p10 ratio. Dispersion changed particularly after 2003. There seems to be, actually, a relatively strong inverse relation between the aggregate labour share and the dispersion of firm level labour share (correlation of -0.70 or higher).

5.2 Decomposition of the aggregate labour share

Equation (14) 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' multidimensional dispersion with respect to that counterfactual firm). The decomposition for the manufacturing sector as a whole is depicted in Figure 6 (see Appendix E for an equivalent analysis at the subsector level).

Two things are important to notice. First, the level of the aggregate labour share is significantly different when compared with the counterfactual homogeneous firm. This is, if all firms were identical, the labour share would be significantly smaller. This result supports the theoretical result of the paper, and echoes the result in the example presented in Figure 1, namely that heterogeneity matters. Second, as Panel (b) shows, the role of heterogeneity has been relatively stable over the period. This is, 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 that of λ^{HOM} in movements of λ^{obs}, we carry out a simple growth accounting decomposition of the equation $\lambda^{obs} = \lambda^{HOM} \sum$. Such decomposition is given by

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

where g_Z stands for the growth rate of factor Z, over a given period. Table 1 presents the result of this exercise for two sub-periods, using 2003 (the year that the labour share reached it highest level) as threshold, as well as for the entire period. In all cases we see that the bulk of the change in the labour share has been due to λ^{HOM}. Meanwhile, $\sum$ has partly counteracted the effect of the former, by around 20%.

Table 1: Contribution to change in aggregate labour share (λ^{obs}), different periods

Period	λ^{obs}	λ^{HOM}	$\sum$	Interaction
1998-2003	6.16%	7.62%	-1.36%	-0.10%
2003-2014	-12.57%	-15.43%	3.38%	-0.52%
1998-2014	-7.18%	-8.98%	1.98%	-0.18%

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

5.3 Decomposition of the labour share of the representative firm

The homogeneous labour share, which formula is given by equation (15), can be further analysed by looking at its constituent elements. Figure 7 presents the evolution of the different variables for the whole manufacturing sector (again, see Appendix E for a subsector level analysis).[32] Recall that these represent weighted averages of the sector's firms. Panel (a) shows a fairly unstable but overall increase in TFP over the period (trend interrupted by a 2008-9 dip). Real wages (Panel b) show a stable pre-2008 growth, with a subsequent dip (particularly in 2009). Interestingly, such growth rate has 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.[33] Labour market power (Panel d) fell in the early years

[32] 2007 presents an unusual behaviour, with significantly more more missing observations in the original dataset, particularly for small firms. This selection means weighted averages are exaggerated. This can be seen in the graphs by particularly high values for productivity, real wages and labour market power. For instance, it looks like the Great Recession hit in 2008, whereas it did so towards the end of that year, and particularly in 2009 (visible in the graphs).

[33] 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

of the period, and subsequently increased post-2008 (recall lower χ^L means less less market power for the firm). The sharp rise in 2007 is artificial (see footnote 32). Lastly, the discrepancy term $\bar{\tau}$ (Panel e) is fairly stable, meaning this is unlikely to drive any of the results.[34]

Table 2 presents the growth rates of each variable in Figure 7, over the subperiods of interest. As equation (15) 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 (15). This decomposition is given by

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

Table 2: Change in λ^{HOM} components

Period	$g_{\bar{A}}$	$g_{\bar{w}/\bar{p}}$	$g_{\bar{\chi}^Y}$	$g_{\bar{\chi}^L}$	$g_{\bar{\tau}}$
1998-2003	8.69%	16.63%	1.45%	-4.26%	0.91%
2003-2014	35.30%	13.96%	-4.86%	7.86%	-1.81%
1998-2014	47.06%	32.91%	-3.48%	3.27%	-0.91%

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

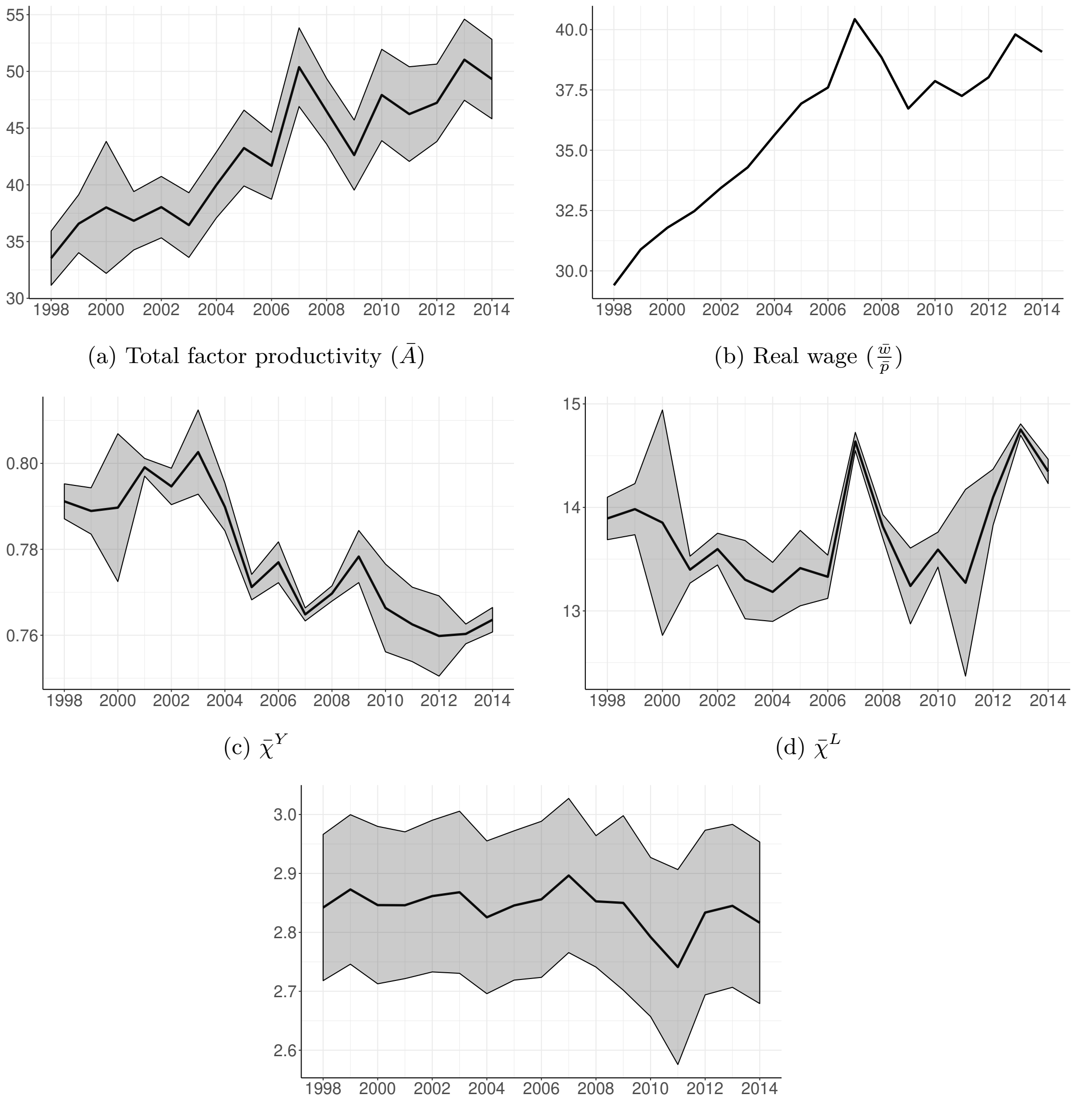
Table 3 shows the resulting contribution of each component of λ^{HOM}. It can be seen that real wages did not grow as fast as productivity did, and the gap between the two can explain most of the actual change in λ^{HOM}. In particular, we see that the pay-productivity disconnect is the key driver in both sub-periods. Given that the second sub-period is longer, and what was observed previously regarding the slower growth in real wages post recession, part of the blame is on the latter. In fact, if we impose the same annual growth rate of real wages observed between 1998 and 2003 (3.1%) for 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 market power both contributed to an overall fall in the labour share. As commented earlier in relation to Figure 7, there is a marked different behaviour of market power within the whole period. This is also seen in the Table. Between 1998 and 2003, both product and labour market power fell (reflected in higher and lower $\bar{\chi}^Y$ and $\bar{\chi}^L$, respectively). This reversed in the second sub-period. And the latter changes dominated. Thus, we can see an overall increase in

not use micro-data but balance-sheet data, covering sectors beyond manufacturing; they assume a Cobb-Douglas 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.

[34] At first, the level of $\bar{\tau}$ might appear to be relatively high. Recall this is computed as the ratio between the observed and predicted labour share across firms. Thus, $\bar{\tau}$ around two suggests predicted λ_i is around half of the observed labour share. This is however not necessarily true. As Appendix D shows, $\hat{A}_{it}$ contains both the shock to TFP and the shock to value added (terms ξ_{it} and ϵ_{it} in equations (23) and (24), respectively). While the latter has zero mean in terms of the *logarithm* of value added (again, see equation 23), it does not do so around value added itself. This bias is captured by the level of $\hat{A}_{it}$ (bias that should be constant as long as the variance of ϵ_{it} is constant). It can be shown that $E(\hat{A}_{it}|\Phi_t) = A_{it}e^{\frac{\sigma^2}{2}}$, 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 6) contains $\hat{A}_{it}$ to the power of $\frac{\rho}{1-\rho}$, and ρ is estimated to be -1.07, such bias is lowering the predicted labour share, which in turns raises τ_i and therefore $\bar{\tau}$. 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.

Figure 7: Evolution of different elements of λ^{HOM}



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

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

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

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

(e) $\bar{\tau}$

Source: our calculation based on ARD data.

Sample: UK manufacturing firms with 10 employees or more.

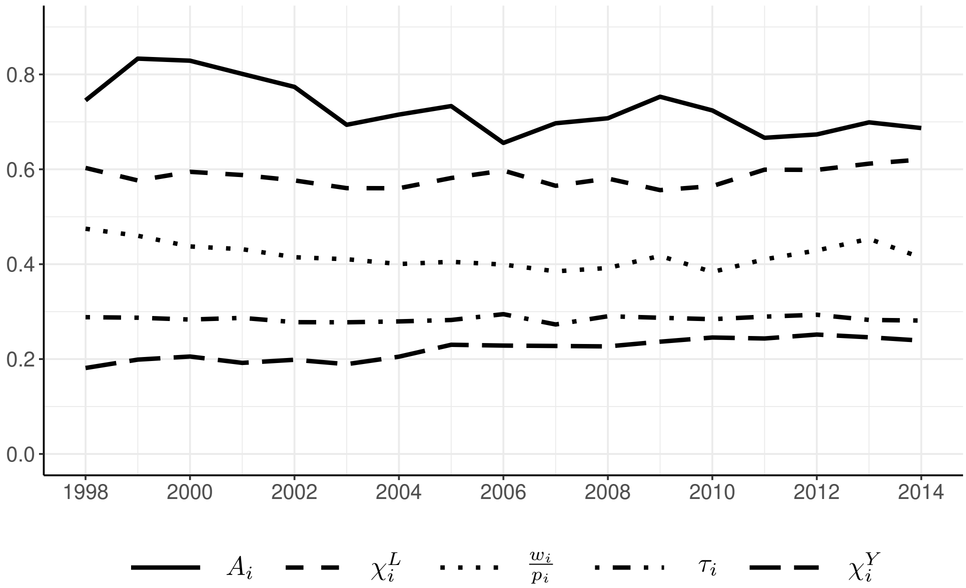
Note: 95% confidence intervals are displayed as a shadowed area (except for real wages, which are observed).

Table 3: Contribution to changes in λ^{HOM}

Period	λ^{HOM}	$\bar{A}$	$\bar{w}/\bar{p}$	$\bar{\chi}^Y$	$\bar{\chi}^L$	$\bar{\tau}$	Interaction
1998-2003	7.62%	-4.49%	8.59%	-0.70%	2.06%	0.91%	-0.15%
2003-2014	-15.43%	-18.25%	7.22%	-2.35%	-3.80%	-1.81%	3.55%
1998-2014	-8.98%	-24.33%	17.01%	-1.68%	-1.58%	-0.91%	2.51%

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

Figure 8: Coefficient of variation of each dimension of firm heterogeneity



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

product and labour market power of firms by 2014, jointly pushing for a 3.20% fall in the labour share. Yet, as said earlier, since the model offers a partial equilibrium approach, the total effects cannot be fully disentangled. In fact, one of the factors driving the slowdown in wage growth could well be the higher labour market power of firms. This, combined with higher rents enabled by more product market power, can help explain the the subsequent fall in the labour share.

5.4 Firms' heterogeneity

The last decomposition exercise focuses on the heterogeneity component, $\sum$. As equation (16) shows, this is a function of the coefficient of variation and the correlation among variables (this is, the structure of the joint distribution of heterogeneity). Before carrying out this decomposition, it is then interesting to evaluate these elements.

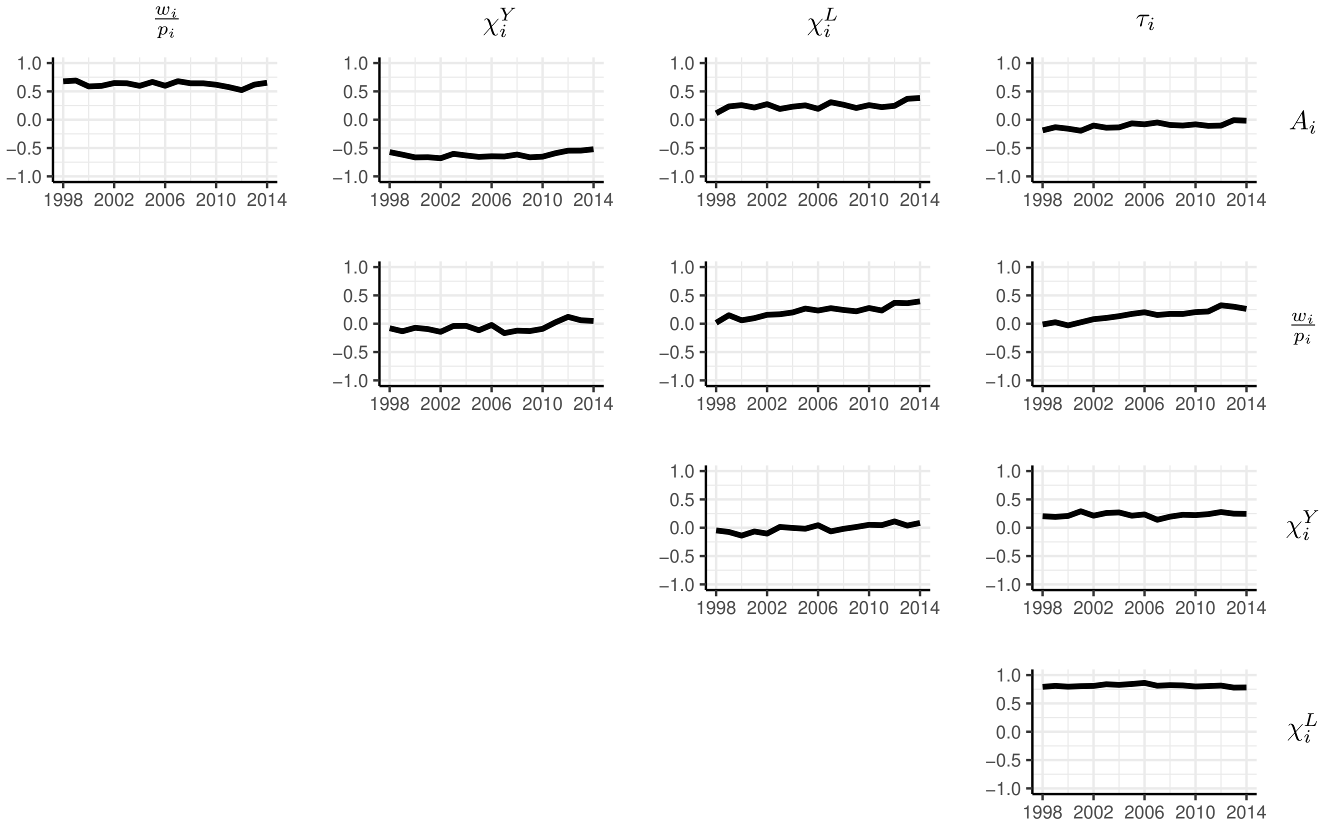
Figure 8 characterises firm heterogeneity in the five different dimensions under study, measured by their coefficient of variation —which, being dimensionless, can be compared across variables. In terms of levels, TFP and labour market power have the highest variation across firms. (Deflated) wages are somewhere in the middle, with product market power and the "discrepancy" component having the lowest variability across firms. It should not be surprising that TFP varies more than wages, given that the latter is a much more "structural" process, driven by market trends relatively common across firms, and regulated by legal contracts; conversely, TFP might be quite idiosyncratic to the firm's conditions, slow to reproduce elsewhere. What's more interesting is the relatively high variability of labour market power vis-a-vis that of product market power. The logic might be the same though. Product market power is computed in terms of mark-ups, which depend on the prices of final goods and variable inputs. Prices adjust easily and they move based on fairly common trends. Conversely, labour market power (calculated as local labour market shares) reflect the spatial heterogeneity of firms, with all the geographical idiosyncrasies involved. Spatial mobility of firms is a slow process. Finally, it is good news that the term describing the mismatch between the data and the theory has relatively low variability. In terms of changes of heterogeneity over time, there are no major movements, excepting TFP and real wages, both moving towards less heterogeneity.

The second important element of firm heterogeneity refers to the correlation among variables across firms over time, presented in Figure 9. Most of them are fairly close to zero. Noticeable exceptions are the positive correlation between TFP and real wages and the negative correlation between product market power and TFP (suggesting that firms with higher TFP have higher product market product). The other very high correlation is between labour market power and τ_i, perhaps suggesting measurement errors associated with the former. Interestingly, product market power is virtually uncorrelated with real wages, reflecting perhaps a low bargaining power of workers, as they are unable to capture rents by firms. As commented earlier, this might be part of the reason the labour share fell over the period. Regarding changes of these correlations over time, the most noticeable changes are those of labour market power with TFP and real wages, both increasing, albeit remaining below 0.40 in 2014. Most of the rest remain fairly stable.

Having provided some background evidence regarding the structure of the joint distribution of firms' characteristics, we can now proceed to the decomposition of $\sum$, which Table 1 above suggested had a minor role in the observed labour share. Two issues are of particular interest here. First, which are the most relevant sources of firm heterogeneity? Second, how have these sources changed over time? Figure 10 helps answering these two questions by presenting the evolution of the different components of $\sum$. The graph is a stacked area plot, meaning the sum of all terms (or equivalently, the difference between the positive and negative totals), yields $\sum$. For ease of visualisation, $\sum$ is centered around zero, whereas, as equation (16) shows, this moves around one.

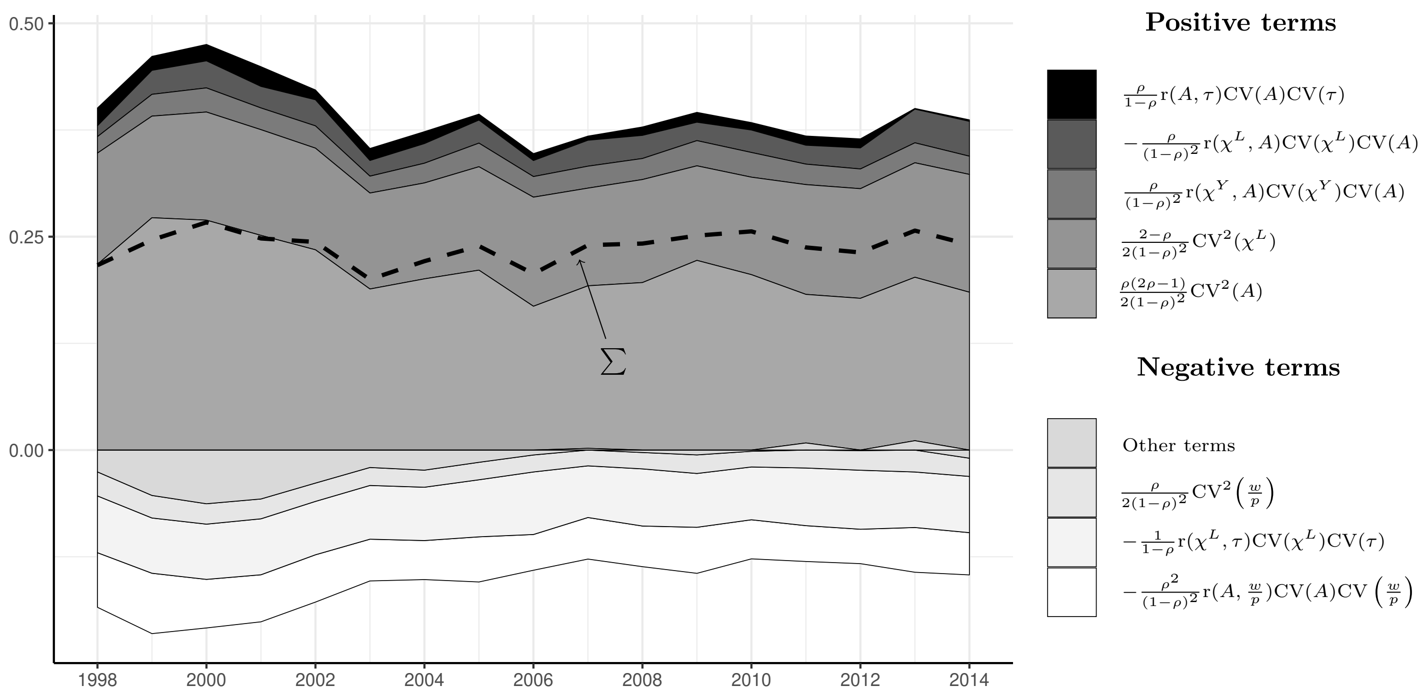
What this figure shows is that the bulk of the effect of firm heterogeneity on the aggregate labour share is due to two elements, namely TFP and labour market power. Figure 8 has already shown these are the dimensions with the highest variability. Figure 10 shows that they also have the biggest impact on the labour share, taking into account the effect of the 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 equation 16). In terms of correlations, the size of the effect again mimics those seen in Figure 9. The correlation of TFP with deflated wages, and labour market power with τ_i are the most significant, followed

Figure 9: Evolution of correlation among different dimensions of firm heterogeneity



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

Figure 10: Stacked area plot showing the decomposition of $\sum$



Notes: As equation (16) shows, $\sum$ is centered around one. For ease of visualisation, here we center it around zero. Positive (negative) terms are those above (below) zero, thereby increasing (decreasing) $\sum$. "Other terms" encompasses all terms in equation (16) 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.

by that of TFP with both product and labour market power. Recall the latter correlation is not significantly high, but its combination with CV(χ^L) and CV(A) (both high), pushed the effect up.

It is worth point it 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 equation (16), and therefore, acts as an empirical test for the validity of the decomposition. It is therefore revealing to see that our approximation is sufficient for capturing the bulk of the changes in $\sum$. Naturally, it is impossible to extrapolate this conclusion to every empirical application of the method, but it is our suspicion the method is in general good enough for its purpose.

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). Naturally, TFP and labour market power are grounded in idiosyncratic processes (e.g. organisational knowledge specific to the firm, or geographical amenities, respectively), more difficult to arbitrate across firms, and therefore with higher and more persistent heterogeneity. Conversely, prices are particularly mobile, meaning less variation and persistence of such variation across firms. Luckily for our analysis, the term capturing the discrepancy between the data and our theory is a highly irrelevant driver of the results.

6 Conclusions

This paper presented a novel approach to study the aggregate labour share. The method is based on a simple, yet powerful 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, real wages and product and labour market power). The method characterises the aggregate economy by means of a weighted average firm (also called a "counterfactual" 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 in ways that are 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 apply 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% over the period, something that seems related mostly to a generalised fall in the firm level labour share across the firm size spectrum. Albeit the distribution of the labour share moved towards the left (hence the overall fall), the upper tail remained stable, implying an increase in the dispersion of the labour share.

The main decomposition exercise produces two results. First, firm heterogeneity contributes significantly to the aggregate labour share. In effect, the weighted average labour share is around ten points lower than the aggregate labour share. Second, most of the 7% fall in the aggregate labour share can be accounted for changes in the weighted average labour share. In other words, the effect of firm heterogeneity on the aggregate labour share is roughly constant over the period.

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 (and particularly after 2003), which alone can explain most of the change in the weighted average labour share. 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 organisational 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, we need to move towards a general equilibrium setting, which we expect to do in future research.

References

Abdih, Y. and 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, International Monetary Fund.

Abowd, J. M., Creecy, R., and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. Technical Report TP2002-06, U.S. Census Bureau, LEHD Program.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. (2003). Labor- And Capital-Augmenting Technical Change. *Journal of the European Economic Association*, 1(1):1–37.

Acemoglu, D. and Restrepo, P. (2018). The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment. *American Economic Review*, forthcoming.

Ackerberg, D., Lanier Benkard, C., Berry, S., and Pakes, A. (2007). Econometric tools for analyzing market outcomes. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6A, chapter 63. Elsevier, 1 edition.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P., and Li, H. (2019). A Theory of Falling Growth and Rising Rents. mimeo.

Aiello, F. and Ricotta, F. (2015). Firm heterogeneity in productivity across Europe: evidence from multilevel models. *Economics of Innovation and New Technology*, 25(1):57–89.

Altissimo, F., Mojon, B., and Za§aroni, P. (2009). Can Aggregation Explain the Persistence of Inflation? *Journal of Monetary Economics*, 56(2):231–241.

Alvarez, J., Benguria, F., Engbom, N., and Moser, C. (2018). Firms and the Decline in Earnings Inequality in Brazil. *American Economic Journal: Macroeconomics*, 10(1):149–189.

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

Annett, A. (2006). Reform in Europe: What Went Right? *Finance and Development*, 43(3).

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Reenen, J. V. (2017a). Concentrating on the Fall of the Labor Share. *American Economic Review: Papers & Proceedings*, 107(5):180=185.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Reenen, J. V. (2017b). The Fall of the Labor Share and the Rise of Superstar Firms. Working Paper 23396, National Bureau of Economic Research.

Autor, D. and Salomons, A. (2018). Is Automation Labor-Dispacing? Productivity Growth, Employment, and the Labor Share. NBER Working Paper Series 24871, National Bureau of Economic Research.

Azmat, G., Manning, A., and Reenen, J. V. (2012). Privatization and the Decline of Labour's Share: International Evidence from Network Industries . *Economica*, 79(315):470–492.

Barkai, S. (2016). Declining Labor and Capital Shares. Job Market Paper.

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

Bartelsman, E. J. and Wolf, Z. (2017). Measuring Productivity Dispersion. Tinbergen Institute Discussion Paper TI 2017-033/VI, Tinbergen Institute.

Barth, E., Bryson, A., Davis, J. C., and Freeman, R. (2016). It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States. *Journal of Labor Economics*, 34(S2):S67–S97.

Bassanini, A. and Duval, R. (2006). Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions. OECD Economics Department Working Papers 486, OECD Publishing.

Bell, B., Bukowski, P., and Machin, S. (2018). Rent Sharing and Inclusive Growth. Discussion Paper Series 12060, IZA Institute of Labor Economics.

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

Bentolila, S. and Saint-Paul, G. (2003). Explaining Movements in the Labor Share. *The B.E. Journal of Macroeconomics*, 3(1):1–33.

Berlingieri, G., Blanchenay, P., and Criscuolo, C. (2017). The Great Divergence(s). OECD Science, Technology and Innovation Policy Papers 39, OECD.

Bernard, A. B., Eaton, J., Jensen, J. B., and Kortum, S. (2003). Plants and Productivity in International Trade. *American Economic Review*, 93(4):1268–1290.

Bilke, L. (2005). Break in the mean and persistence of inflation: A sectoral analysis of French CPI. ECB Working Paper 179, European Central Bank.

Bils, M. and Klenow, P. J. (2004). Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, 112(5):947–985.

Blanchard, O. and Giavazzi, F. (2003). Macroeconomic Effects of Regulation and Deregulation in Goods and Labor Markets. *The Quarterly Journal of Economics*, 118(3):879–907.

Blinder, A. S., Canetti, E. R. D., Lebow, D. E., and Rudd, J. B. (1998). *Asking About Prices. A New Approach to Understanding Price Stickiness*. Russel Sage Foundation.

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

Boivin, J., Giannoni, M., and Mihov, I. (2009). Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data. *American Economic Review*, 99(1):350–384.

Branston, J. R., Cowling, K. G., and 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 at http://www.peterbrummund.com/docs/pwb_monopsony.pdf.

Bruno, M. (1978). Duality, Intermediate Inputs and Value Added. In Fuss, M. and McFadden, D., editors, *Production Economics: A Dual Approach to Theory and Applications*, volume 2. North Holland.

Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.

Cantore, C., Ferroni, F., and Leó-Ledesma, M. A. (2018). The Missing Link: Monetary Policy and The Labor Share. CEPR Discussion Paper DP13551, Center for Economic Policy Research.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics*, 131(2):633–686.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3):967.

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

Ciminelli, G., Duval, R., and Furceri, D. (2018). Employment Protection Deregulation and Labor Shares in Advanced Economies. IMF Working Paper WP/18/186, International Monetary Fund.

Clark, T. (2006). Disaggregate Evidence on the Persistence of Consumer Price Inflation. *Journal of Applied Econometrics*, 21(5):563–587.

Collard-Wexler, A. and De Loecker, J. (2016). Production Function Estimation with Measurement Error in Inputs. NBER Working Paper 22437, National Bureau of Economic Research.

Corrado, C., Hulten, C., and Sichel, D. (2009). Intangible Capital and U.S. Economic Growth. *The Review of Income and Wealth*, (55):661–685.

Dahlby, B. and West, D. (1986). Price Dispersion in an Automobile Insurance Market. *Journal of Political Economy*, 94(2):418–38.

Daudey, E. and García-Peñalosa, C. (2007). The personal and the factor distributions of income in a cross-section of countries. *The Journal of Development Studies*, 43(5):812–829.

De Loecker, J. and Eeckhout, J. (2018). Global Market Power. NBER Working Papers 24768, National Bureau of Economic Research, Inc.

De Loecker, J., Eeckhout, J., and Unger, G. (2018). The Rise of Market Power and the Macroeconomic Implications. Available at http://www.janeeckhout.com/wp-content/uploads/RMP.pdf.

De Loecker, J. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–71.

Dimova, D. (2019). The Structural Determinants of the Labour Share in Europe. IMF Working Paper WP/19/67, International Monetary Fund.

Dizon, R. and Lim, G. C. (2018). Labor's Share, the Firm's Market Power, and Total Factor Productivity. *Economic Inquiry*, 56(4):2058–2076.

Dobbelaere, S. and Mairesse, J. (2013). Panel Data Estimates of the Production Function and Product and Labor Market Imperfections. *Journal of Applied Econometrics*, 28:1–46.

Duarte, M. and Restuccia, D. (2016). Relative Prices and Sectoral Productivity. Working Papers 555, University of Toronto, Department of Economics.

Dunne, T., Foster, L., Haltiwanger, J., and Troske, K. R. (2004). Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment . *Journal of Labor Economics*, 22(2):397–429.

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

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

Elhanan Helpman and Oleg Itskhoki and Marc-Andreas Muendler and Stephen J. Redding (2017). Trade and inequality: From theory to estimation. *Review of Economic Studies*, 84(1):357–405.

Elsby, M., Hobijn, B., and Sahin, A. (2013). The Decline of the U.S. Labor Share. *Brookings Papers on Economic Activity*, 44(2 (Fall)):1–63.

Epifani, P. and Gancia, G. (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics*, 83(1):1 – 13.

Faggio, G., Salvanes, K., and Van Reenen, J. (2010). The evolution of inequality in productivity and wages: panel data evidence. *Industrial and Corporate Change*, 19(6):1919–1951.

Felipe, J. and McCombie, J. S. (2014). The Aggregate Production Function: 'Not Even Wrong'. *Review of Political Economy*, 26(1):60–84.

Félix, S. and Portugal, P. (2017). Labor Market Imperfections and the Firm's Wage Setting Policy. Working Papers 2017/4, Banco de Portugal.

Fernández, C., Lacuesta, A., Montero, J. M., and Urtasun, A. (2015). Heterogeneity of markups at the firm level and changes during the Great Recession: The case of Spain. Documentos de Trabajo 1536, Banco de España.

Fisher, F. M. (1969). The Existence of Aggregate Production Functions. *Econometrica*, 37(4):553–577.

Forlani, E., Martin, R., Mion, G., and Muûls, M. (2016). Unraveling Firms: Demand, Productivity and Markups Heterogeneity. CEP Discussion Paper 1402, Centre for Economic Performance.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425.

Furceri, D., Loungani, P., and Ostry, J. D. (2018). The Aggregate and Distributional Effects of Financial Globalization: Evidence from Macro and Sectoral Data. IMF Working Paper WP/18/83, International Monetary Fund.

Glover, A. and Short, J. (2019). Can Capital Deepening Explain the Global Decline in Labor's Share? Staff Working Paper 2019-3, Bank of Canada.

González, I. and Trivín, P. (2017). The Global rise of Asset Prices and the Decline of the Labor Share. Available at ssrn: http://dx.doi.org/10.2139/ssrn.2964329.

Goux, D. and Maurin, E. (1999). Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data. *Journal of Labor Economics*, 17(3):492–533.

Green, H. J. (1964). *Aggregation in Economic Analysis*. Princeton University Press.

Grossman, G. M., Helpman, E., Oberfield, E., and Sampson, T. (2018). The Productivity Slowdown and the Declining Labor Share: A Neoclassical Exploration. NBER Working Paper Series 23853, National Bureau of Economic Research.

Gruetter, M. and Lalive, R. (2009). The importance of firms in wage determination. *Labour Economics*, 16(2):149–160.

Guscina, A. (2006). Effects of Globalization on Labor's Share in National Income. Working Paper 06/294, International Monetary Fund.

Gutierrez Gallardo, G. (2017). Investigating Global Labor and Profit Shares. Available at ssrn: https://ssrn.com/abstract=3040853 or http://dx.doi.org/10.2139/ssrn.3040853.

Haldane, A., Aquilante, T., Chowla, S., Dacic, N., Masolo, R., Schneider, P., Seneca, M., and Tatomir, S. (2018). Market Power and Monetary Policy. Speech, Bank of England. Available at https://www.bankofengland.co.uk/speech/2018/andy-haldane-speech-at-the-economic-policy-symposium-panel-jackson-hole.

Hamermesh, D. S. (1999). Changing Inequality in Markets for Workplace Amenities. *The Quarterly Journal of Economics*, 114(4):1085–1123.

Harrison, A. E. (2002). Has Globalization Eroded Labor's Share? Some Cross-Country Evidence. Working paper, University of California at Berkeley and NBER, Berkeley.

Hartman-Glaser, B., Lustig, H., and Xiaolan, M. Z. (2019). Capital Share Dynamics When Firms Insure Workers. *The Journal of Finance*, (forthcoming).

Hergovich, P. and Merz, M. (2018). The Price of Capital, Factor Substitutability, and Corporate Profits. Discussion Paper Series 11791, IZA Institute of Labor Economics.

Hirsch, B., Schank, T., and Schnabel, C. (2010). Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany. *Journal of Labor Economics*, 28(2):291–330.

Holzer, H. J., Lane, J. I., Rosenblum, D. B., and Andersson, F. (2011). *Where Are All the Good Jobs Going?: What National and Local Job Quality and Dynamics Mean for U.S. Workers*. Russell Sage Foundation.

Hopenhayn, H., Neira, J., and Singhania, R. (2018). Rent Sharing and Inclusive Growth. NBER Working Paper Series 25382, National Bureau of Economic Research.

Hornstein, A., Krusell, P., and Violante, G. L. (2011). Frictional Wage Dispersion in Search Models: A Quantitative Assessment. *American Economic Review*, 101(7):2873–98.

Hosken, D. and Reiffen, D. (2004). Patterns of Retail Price Variation. *RAND Journal of Economics*, 35(1):128–146.

Håkanson, C., Lindqvist, E., and Vlachos, J. (2015). Firms and skills: the evolution of worker sorting. Research Papers in Economics 2015:4, Stockholm University, Department of Economics.

Hsieh, C.-T. and Klenow, P. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

IMF (2007). World Economic Outlook, October 2007: Globalization and Inequality. Technical report, International Monetary Fund.

IMF (2017). World Economic Outlook: Gaining Momentum? Technical report, International Monetary Fund.

Jayadev, A. (2007). Capital account openness and the labour share of income. *Cambridge Journal of Economics*, 31(3):423–443.

Karabarbounis, L. and Neiman, B. (2014). The Global Decline of the Labor Share. *The Quarterly Journal of Economics*, 129(1):61–103.

Kato, R. and Okuda, T. (2017). Market Concentration and Sectoral Inflation under Imperfect Common Knowledge. IMES Discussion Paper Series 17-E-11, Institute for Monetary and Economic Studies, Bank of Japan.

Kehrig, M. and Vincent, N. (2017). Growing Productivity without Growing Wages: The Micro-Level Anatomy of the Aggregate Labor Share Decline. CESifo Working Papers 6454, CESifo.

Koh, D., Santaeulalia-Llopis, R., and Zheng, Y. (2018). Labor Share Decline and Intellectual Property Products Capital. Working Paper 873, School of Economics and Finance, Queen Mary University.

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

Lee, K. and Jayadev, A. (2005). Capital Account Openness and its Effects on Growth and Distribution: A Review of the Cross Country Evidence. In Epstein, G. A., editor, *Capital Flight and Capital Controls in Developing Countries*, chapter 2, pages 15–57. Edward Elgar.

León-Ledesma, M. A. and Satchi, M. (2018). Appropriate Technology and Balanced Growth. *Review of Economic Studies*, 86:807–835.

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

Lünnemann, P. and Mathä, T. Y. (2004). How persistent is disaggregate inflation? An analysis across EU15 countries and HICP Sub-indexes. ECB Working Paper 415, European Central Bank.

Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

McGowan, M. A., Andrews, D., and Millot, V. (2017). The Walking Dead?: Zombie Firms and Productivity Performance in OECD Countries. OECD Economics Department Working Papers 1372, OECD.

Mertens, M. (2019). Micro-mechanisms behind declining labour shares: market power, production processes, and global competition. IWH-CompNet Discussion Papers 3/2019, Halle Institute for Economic Research (IWH).

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.* Office for National Statistics, [original data producer(s)]. Office for National Statistics. SN 7989, http://doi.org/10.5255/UKDA-SN-7989-4.

Olley, G. S. and 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. Technical report, Office for National Statistics.

Pak, M. and Schwellnus, C. (2019). Labour share developments over the past two decades: The role of public policies. Economics Departmetn Working Papers 1541, OECD.

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

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T. and Zucman, G. (2014). Capital is Back: Wealth-Income Ratios in Rich Countries 1700-2010. *The Quarterly Journal of Economics*, 129(3):1255–1310.

Ransom, M. R. and Oaxaca, R. L. (2010). New Market Power Models and Sex Differences in Pay. *Journal of Labor Economics*, 28(2):267–289.

Rognlie, M. (2015). Deciphering the Fall and Rise in the Net Capital Share. *Brookings Papers on Economic Activity*, 46(1 (Spring)):1–69.

Rosen, S. (1987). The theory of equalizing differences. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics*, volume 1, chapter 12, pages 641–692. Elsevier, 1 edition.

Schwellnus, C., Kappeler, A., and Pionnier, P.-A. (2017). Decoupling wages from productivity: Macro-level facts. OECD Economics Department Working Papers 1373, OECD.

Schwellnus, C., Pak, M., Pionnier, P.-A., and 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, OECD.

Short, J. and Glover, A. (2017). The Age-Distribution of Earnings and the Decline in Labor's Share. 2017 Meeting Papers 1369, Society for Economics Dynamics.

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

Song, J., Price, D. J., Guvenen, F., Bloom, N., and von Wachter, T. (2018). Firming Up Inequality. Working Paper 750, Federal Reserve Bank of Minneapolis.

Stiglitz, J. (2012). *The Price of Inequality: How Today's Divided Society Endangers Our Future*. W. W. Norton.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, 49(2):326–365.

Tamminen, S. and Chang, H.-H. (2013). Firm and sectoral heterogeneity in markup variability. *The Journal of International Trade & Economic Development*, 22(1):157–178.

Valentinyi, A. and Herrendorf, B. (2008). Measuring Factor Income Shares at the Sector Level. *Review of Economic Dynamics*, 11(4):820–835.

Van Reenen, J. (2018). Increasing Differences Between Firms: Market Power and the Macro-Economy. CEP Discussion Paper 1576, Centre for Economic Performance.

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

Weber, D. A. (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*. Harvard University Press.

Wolman, A. L. (2011). The Optimal Rate of Inflation with Trending Relative Prices. *Journal of Money, Credit and Banking*, 43(2-3):355–384.

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

Zeira, J. (1998). Workers, Machines, and Economic Growth. *The Quarterly Journal of Economics*, 113(4):1091–1117.

A Alternative production functions and the aggregate labour share

The assumption of a constant returns to scale CES value added production function with homogeneous parameters is not arbitrary, but chosen for necessity and parsimony. To see this, consider the consequences for the decomposition formula of alternative assumptions. First, let us recall our main assumptions. The production function is

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

This implies the firm level labour share is

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

and the aggregate labour share is

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

In the case of a Cobb-Douglas production function (i.e. when $\rho = 0$), the latter is given by

$$\lambda = \sum_i \left(\frac{\alpha \chi_i^Y}{\chi_i^L}\right) \delta_i$$

Thus, there is no explicit role for productivity and real wages (and thus for the pay productivity disconnect) in the aggregate labour share; only market power affects the latter. This result holds even if there are non constant returns to scale.

Regarding the assumption of constant returns to scale (CRS) for the general CES, this is necessary for a decomposition to be possible. In effect, for a CES like the following:

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

it can be shown that the firm level labour share is

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

This is, the labour share depends on the actual *level* of employment (except, of course, under CRS). To be consistent with our framework, where we replace value added and employment by their optimal values in terms of TFP, wages, prices and market power, we need to replace L_i above with its solution in terms of these same variables. This solution is a highly non-linear function, obtained from the combination of the FOC of profits with respect to labour and capital. The final expression for the firm's labour share is then not a neat, multiplicative function of each variable, and thus it is not possible to decompose using our method.

What if technologies are heterogeneous? If α varies across firms, it is still possible to achieve a clear decomposition. In one sense, α is like any other variable inside λ. It would be possible to compute a counterfactual $\bar{\alpha}$ and add it to λ^{HOM}. The problem is empirical. The model is already too complicated for it to be estimated, and adding non-linear firm specific parameters would not improve things (see equation 25 in Appendix D). Meanwhile, heterogeneity in ρ denies the possibility to arrive at a decomposition formula altogether. More precisely, there is no way to write terms like $\frac{A_i}{A}$ and separate λ into a counterfactual firm and dispersion with respect to it, as in equation (8).

One might also question the use of a value added production function as a starting point, rather than, for instance, a CES gross output production function. To see this, consider such a function:

$$Q_i = B_i \left(\alpha L_i^\gamma + \beta M_i^\gamma + (1-\alpha-\beta)K_i^\gamma\right)^{\frac{1}{\gamma}} \tag{17}$$

where Q_i is gross output and M_i is intermediary inputs. The first order condition with respect to L_i is:

$$\frac{\partial Q_i}{\partial L_i} \equiv \alpha B_i^\gamma (Q_i)^{1-\gamma}(L_i)^{\gamma-1} = \left(\frac{w_i}{p_i}\right)\frac{\chi_i^L}{\chi_i^Y}$$

From here, we obtain a formula for optimal $\frac{Q_i}{L_i}$, given by:

$$\frac{Q_i}{L_i} = \left(\frac{w_i \chi_i^L}{\alpha B_i^\gamma p_i \chi_i^Y}\right)^{\frac{1}{1-\gamma}}$$

Repeating for intermediary inputs, we obtain:

$$\frac{Q_i}{M_i} = \left(\frac{p_i^M}{\beta B_i^\gamma p_i \chi_i^Y}\right)^{\frac{1}{1-\gamma}}$$

where p_i^M is the price of intermediary inputs. Now, the labour share for the firm is given by:

$$\lambda_i \equiv \frac{w_i L_i}{p_i Q_i - p_i^M M_i} = \frac{w_i}{p_i \frac{Q_i}{L_i} - p_i^M \frac{M_i}{L_i}}$$

where nominal value added is defined as $p_i Q_i - p_i^M M_i$. Combining the above results, it is possible to show that λ_i is:

$$\lambda_i = \left(\frac{\alpha}{\chi_i^L}\right)^{\frac{1}{1-\gamma}} \frac{w_i^{\frac{\gamma}{1-\gamma}}}{(\beta p_i)^{\frac{\gamma}{\gamma-1}}\left(\chi_i^Y\right)^{\frac{1}{\gamma-1}} - \beta^{\frac{1}{1-\gamma}}\left(p_i^M\right)^{\frac{\gamma}{\gamma-1}}}$$

It is evident that the above is unhelpful in achieving a decomposition of the aggregate labour share, not even in the special cases of a Cobb-Douglas or Leontief gross output production function ($\gamma = 0$ or $\gamma = -\infty$, respectively). An alternative is to specify the production function in equation (17) as a nested CES, as follows:

$$Q_i = B_i \left(b{Y_i}^\gamma + (1-b){M_i}^\gamma\right)^{\frac{1}{\gamma}} \tag{18}$$

where $Y_i = A_i \left(\alpha L_i^\rho + (1-\alpha) K_i^\rho\right)^{\frac{1}{\rho}}$, this is, a CES of labour and capital. Bruno (1978) shows that, if this gross output production function is either linear or Leontief (i.e. $\gamma = -\infty$ or $\gamma = 1$), Y_i corresponds to real value added (where both output and inputs are deflated). In other words, $Y_i = f(L_i, K_i)$ is a value added production function. Such is the implicit assumption in our model, one which unfortunately cannot be tested, due to the complexity of applying the dynamic panel method (or the control function method) to the production function in equation (18).

It must be noted here that, albeit it is tempting to estimate equation (18) using value added directly as an input in equation (18) (rather than the nested CES with capital, labour and intermediary inputs) this is spurious. Value added is *defined* as gross output minor intermediary inputs (properly deflated). Estimating such equation will produce (and does produce, in our data) $\hat{B}_i \approx 1$, $\hat{b} \approx 0.5$ and $\hat{\gamma} \approx 1$. The latter might suggest the conditions for the existence of a value added production function set out in Bruno (1978) are met. Yet, the problem with this approach is that when using value added as an input, the existence of such production function is actually imposed in the equation. In effect, replacing the estimates given above into equation (18) leads to $Q_i \approx Y_i - M_i$, which is the identity for the definition of real value added. In other words, we are merely estimating an identity.

B Statistical decomposition

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

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

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

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

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

$$\lambda^{obs} = \hat{\mathrm{E}}(\lambda) + N\hat{\mathrm{Cov}}(\delta, \lambda) \tag{19}$$

where $\hat{\mathrm{E}}(\lambda)$ is the observed unweighted average labour share. Therefore, the unweighted over the weighted average labour share is:

$$\frac{\hat{\mathrm{E}}(\lambda)}{\lambda^{obs}} = 1 - N\frac{\hat{\mathrm{Cov}}(\delta, \lambda)}{\lambda^{obs}} \tag{20}$$

This ratio is greater the more negative the covariance between firm size and labour share is, *ceteris paribus*.

C Effects of changes at one specific firm

C.1 Effects on the firm level labour share

How the firm level labour share reacts to changes in prices, wages, technology, market power and capital depends on ρ. By taking the partial derivative of (6) with respect to each variable, it is immediate to identify the effects of the different (firm level) variables, which are summarised in Tables A1 and A2, for $\rho < 0$ and $\rho > 0$ respectively. For simplicity, the two forms of market power are combined in one term, with $\chi_i = \frac{\chi_i^L}{\chi_i^Y}$.

Table A1: Determinants of the firm level labour share, $\rho < 0$.

		Sign of the effect
$\lambda_i'(A_i)$		$-$
$\lambda_i'(K_i)$		0
$\lambda_i'(w_i)$	originated by a change in Θ_i^L	$+$
$\lambda_i'(p_i)$	originated by a change in Θ_i^Y	$-$
$\lambda_i'(\chi_i)$	direct effect	$-$
$\lambda_i'(\chi_i)$	indirect effect through prices	$-$
$\lambda_i'(\chi_i)$	indirect effect through wages	$-$
$\lambda_i'(\chi_i)$	overall effect	$-$

Table A2: Determinants of the firm level labour share, $\rho > 0$.

		Sign of the effect
$\lambda_i'(A_i)$		$+$
$\lambda_i'(K_i)$		0
$\lambda_i'(w_i)$	originated by a change in Θ_i^L	$-$
$\lambda_i'(p_i)$	originated by a change in Θ_i^Y	$+$
$\lambda_i'(\chi_i)$	direct effect	$-$
$\lambda_i'(\chi_i)$	indirect effect through prices	$+$
$\lambda_i'(\chi_i)$	indirect effect through wages	$+$
$\lambda_i'(\chi_i)$	overall effect	?

Note that when analysing the *ceteris paribus* effects of a change in wages or output price on λ_i, we are assuming that they were originated by a change in the idiosyncratic preference parameters Θ_i^L and Θ_i^Y, and not in a change in market power χ_i. Market power, on the other hand, has two effects: a *direct* effect, and an *indirect* effect through prices and wages. The direct effect of market power on the labour share is negative (see the first term of equation 6). Said differently, a lower $|\eta_i^Y|$ or η_i^L, which are equivalent to a higher market power χ_i for the firm, translate into a lower labour share. When $\rho < 0$, the indirect effect goes in the same direction, so that the overall effect of market power is unambiguously negative. Note also that the amount of capital employed, given the CRS assumption, does not affect λ_i.

Table A1 or variants obtained with different theoretical assumptions informed most of the micro-econometric investigations on the determinants of the firm-level LS.

C.2 Effects on the aggregate labour share

The effects of a variation in one firm level variable on the aggregate labour share, keeping all other firms unchanged, depend among other things on whether the firm considered has an above-average or below-average labour share. Recall that the aggregate labour share is a weighted average of the individual labour shares (equation 2), and write it as

$$\lambda = \lambda_J(1-\delta_i) + \lambda_i\delta_i \tag{2'}$$

where $\lambda_J = \sum_{j\neq i} \lambda_j \frac{1-\delta_j}{1-\delta_i}$ is the average labour share excluding firm i.

The change in the aggregate labour share following from changes in the production plans at firm i is therefore given by:

$$\lambda'(\cdot) = (\lambda_i - \lambda_J)\delta_i'(\cdot) + \lambda_i'(\cdot)\delta_i \tag{21}$$

The direction of change in λ depends therefore on (i) how the labour share changes at firm i, (ii) how the relative weight of firm i changes, and (iii) whether the labour share at firm i is above or below the average.

Note that equation (21) does not depend on firm i having the same production function (e.g the same α and ρ) of any other firm j. What originates the change in λ is simply a change in λ_i for a given firm i, and a change in the weight of that individual firm in the whole economy: no changes occur to any other firm, so how the labour share is determined in those other firms does not matter. These simple results provide the basics for understanding the effects of simultaneous changes taking places at different firms. Each change in firm level variables pulls in a direction as identified by Table A1 and A2, and the overall effect is nothing else than a simple composition of all these individual effects.

We know from Tables A1 and A2 how λ_i reacts to changes in firm level variables. As for what concerns the weights, these depend, keeping all other firms fixed, on the value added of firm i. From equation (3) we know that value added Y_i grows with A_i, L_i and K_i, while it is easy to show from equation (4) that employment at firm i expands as productivity A_i, prices p_i (keeping market power constant) and capital K_i increase, and contracts as wages w_i (keeping market power constant), and market power χ_i (keeping wages fixed) increase. Putting all this together, by the envelop theorem we get that the relative weight of firm i in the economy increases with A_i, K_i and p_i, and decreases with w_i, keeping market power constant.

The latter has an ambiguous effect on market share. When $\rho < 0$, the partial derivative of Ω_i with respect to χ_i is negative: the direct effect of market power is thus to reduce market share. However, an increase in market power originates either from a decrease in $|\eta_i^Y|$ (output is more rigid) or from a decrease in η_i^L (labour supply is more rigid), translating either into a higher price or into lower wages. Both increase production and hence market share, counteracting the negative direct effect.

The overall effects of changes in the characteristics of one firm on the aggregate labour share are reported in Tables A3 and A4, for $\rho < 0$ and $\rho > 0$ respectively.

Table A3: Effects of changes in the characteristics of one firm on the aggregate labour share, $\rho < 0$.

		δ_i'	λ_i'	λ'	
				$\lambda_i > \lambda_J$	$\lambda_i < \lambda_J$
A_i		$+$	$-$	?	$-$
K_i		$+$	0	$+$	$-$
w_i	originated by a change in Θ_i^L	$-$	$+$	?	$+$
p_i	originated by a change in Θ_i^Y	$+$	$-$	?	$-$
χ_i	direct + indirect effect	?	?	?	?

Table A4: Effects of changes in the characteristics of one firm on the aggregate labour share, $\rho > 0$.

		δ_i'	λ_i'	λ'	
				$\lambda_i > \lambda_J$	$\lambda_i < \lambda_J$
A_i		$+$	$+$	$+$	?
K_i		$+$	0	$+$	$-$
w_i	originated by a change in Θ_i^L	$-$	$-$	$-$	?
p_i	originated by a change in Θ_i^Y	$+$	$+$	$+$	?
χ_i	direct + indirect effect	?	?	?	?

D 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 firms. 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 that might exist when estimation productivity using micro data.[35] Here we follow the dynamic panel approach (e.g. Blundell and 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, expanding the sample size.[36] In short, the method works as follows. We start

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

[36] 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. ARD has a companion dataset with the universe of firms for all years, including minimal data like employment (dataset built from administrative tax registry data).

with our production function, extended to an econometric notation:

$$Y_{it} = e^{\omega_{it}} \left(\alpha L_{it}^{\rho} + (1-\alpha) K_{it}^{\rho}\right)^{\frac{1}{\rho}} e^{\epsilon_{it}} \tag{22}$$

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, $\omega_{it} = E\left[\omega_{it}|\omega_{it-1}\right] + \xi_{it}$, where ξ_{it} is an idiosyncratic *iid* shock to productivity, known to the firm. Taking logs of (22), we get:

$$y_{it} = \frac{1}{\rho} \ln\left(\alpha L_{it}^{\rho} + (1-\alpha) K_{it}^{\rho}\right) + \omega_{it} + \epsilon_{it} \tag{23}$$

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 (23). 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):

$$\omega_{it} = \theta \omega_{it-1} + \xi_{it} \tag{24}$$

If we combine equations (22) and (24) (i.e. if we "θ-differentiate" the production function), we get:

$$y_{it} = \frac{1}{\rho} \ln\left(\alpha L_{it}^{\rho} + (1-\alpha) K_{it}^{\rho}\right) + \theta \left[y_{it-1} - \frac{1}{\rho} \ln\left(\alpha L_{it-1}^{\rho} + (1-\alpha) K_{it-1}^{\rho}\right)\right] + \xi_{it} + (\epsilon_{it} - \theta \epsilon_{it-1}) \tag{25}$$

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 Cobb-Douglas production functions, which are log-linear in parameters. Unfortunately, in our data GMM does not converge. Hence, in practice, we estimate a translog production function, which is an approximation of the CES around an elasticity of substitution equal to 1.[37] This production function is:

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

Similarly, using the perpetual inventory method, we can compute capital in $t-1$ for every firm sampled in t. Therefore, we are able to extend the sample with past values of the key instruments, thereby expanding the sample size. 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 parameter outside the theoretical domain).

[37] As commented earlier, Monte Carlo 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 (see end of appendix).

The final model estimated by GMM is obtained by "θ-differencing" the equation above, with instruments $\{\ln(K_{it}), \ln^2(K_{it}), \ln(K_{it-1}), \ln^2(K_{it-1}), \ln(L_{it-1}), \ln^2(L_{it-1}), \ln(L_{it-1})\ln(K_{it-1})\}$. Estimation produces the following values, all significant at the 1%:[38] $\hat{\alpha} = 0.50$, $\hat{\rho} = -1.07$, $\hat{\theta} = 0.92$, and a constant of 3.8.

Finally, having estimated α and ρ, we can use equation (22) to compute $\hat{A}_{it}$ 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}), $\hat{A}_{it}$ also includes the realised shock to value added, $\hat{\epsilon}_{it}$. In effect, from equation (22) we can see that $\hat{A}_{it} = e^{c+\hat{\omega}_{it}+\hat{\epsilon}_{it}}$ (where c is the constant term in the regression). This means that $\hat{A}_{it}$ 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.[39] Notice equation (22) allows us to compute the "realised" value of A even for observations not part of the regression sample (for example, because of missing data in a given year). We follow this approach, and "extrapolate" $\hat{A}_{it}$ whenever possible. Around 50% of final observations used in the analysis are extrapolated.

Capital stock

Having outlined the estimation procedure for the production function, we should mention that 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% of firms), the level of capital may be greatly underestimated with such method.

Instead, we follow the strategy proposed by Chen and Plotnikova (2014), which estimates 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).[40] This relationship is given by the following formula:

$$K_{it} = \left(\frac{L_{it}}{L_t}\right)^a \left(\frac{M_{it}}{M_t}\right)^{1-a} K_t \tag{26}$$

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.

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

[39]See footnote 34 for further details.

[40]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 patters.

In practice, a is unknown. Furthermore, this cannot be estimated from equation (26), since K_{it} is also unknown. The solution is to combine equation (26) with that of capital accumulation, namely $K_{it} = (1-\delta)K_{it-1} + I_{it}$, 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:

$$I_{it} = \left(\frac{L_{it}}{L_t}\right)^a \left(\frac{M_{it}}{M_t}\right)^{1-a} K_t - (1-\delta)\left(\frac{L_{it-1}}{L_{t-1}}\right)^a \left(\frac{M_{it-1}}{M_{t-1}}\right)^{1-a} K_{t-1} \tag{27}$$

The above can be estimated using GMM. Results for the whole manufacturing sector yield $\hat{a} = 0.42$, significant at 1%.[41] With this value is then possible to impute capital at the firm level using equation (26). 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 one year. The extrapolation is valid as long as the "structural relationship" does not depend on properties of the sample selection (for instance, firm size).

E Decomposition results for manufacturing subsectors

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

Unfortunately, only 13 out of 23 subsectors produced meaningful results (in terms of parameters within the theoretical boundaries), suggesting not every subsector might be represented by a CES/translog production function.[42] Overall, the 13 subsectors cover 62% of the total observations (firm-years) available the manufacturing sector, and used in the main text.

Table A5 presents the decomposition for each subsector's labour share. Additionally, the table presents an extra row ("combined subsectors") with the decomposition of an aggregate series of λ^{obs}, computed from a weighted average of subsectors' λ^{obs}, using value added as weights. For comparison, another row is added wth the results for the whole manufacturing sector presented in the main text. Finally, to give a sense of the importance of different subsectors, the table includes an extra column with the 2014's share of value added of each subsector with respect to all manufacturing.

The overall picture is the same as in our results for the whole manufacturing sector, namely that firm heterogeneity has not been a major driver of the labour share. This is true both for subsectors individually and for their combination. The latter decomposition is also quite similar to the results for manufacturing as a whole. Still, some disparity is observed in $\sum$ across subsectors, 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 subsectors, albeit fell in most of them.

[41] The depreciation rate is assumed to be 4.58%, the average for the 1998-2014 period, according to Office for National Statistics data for the UK.

[42] In particular, 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).

Table A5: Contribution to change in subsector labour share (λ^{obs}), 1998-2014

Subsector	λ^{obs}	λ^{HOM}	$\sum$	Interaction	Subsector's share of value added (2014)
13 (Manufacture of textiles)	-9.19%	-9.56%	0.40%	-0.04%	1.4%
14 (Manufacture of wearing apparel)	-25.66%	-34.00%	12.65%	-4.30%	0.5%
16 (Manufacture of wood and products of wood and cork, excl. furniture)	-10.28%	-15.61%	6.32%	-0.99%	1.7%
17 (Manufacture of paper and paper products)	-8.89%	-9.56%	0.74%	-0.07%	2.3%
18 (Printing and reproduction of recorded media)	-6.15%	-7.06%	0.98%	-0.07%	3.1%
23 (Manufacture of other non-metallic mineral products)	-10.45%	-7.08%	-3.63%	0.26%	3.5%
25 (Manufacture of fabricated metal products, excl. machinery and equip.)	-12.70%	-14.66%	2.29%	-0.34%	10.8%
26 (Manufacture of computer, electronic and optical products)	6.38%	12.49%	-5.42%	-0.68%	5.3%
27 (Manufacture of electrical equipment)	1.78%	8.73%	-6.39%	-0.56%	3.0%
28 (Manufacture of machinery and equipment n.e.c.)	-6.60%	-7.54%	1.02%	-0.08%	8.1%
29 (Manufacture of motor vehicles, trailers and semi-trailers)	-19.77%	-21.55%	2.27%	-0.49%	10.2%
30 (Manufacture of other transport equipment)	-11.46%	-19.66%	10.21%	-2.01%	6.4%
33 (Repair and installation of machinery and equipment)	10.25%	10.90%	-0.59%	-0.06%	3.9%
Combined subsectors	-8.75%	-9.39%	0.69%	-0.05%	60.4%
All manufacturing	-7.18%	-8.98%	1.98%	-0.18%	100%

Source: our calculation based on ARD data.
Sample: UK manufacturing firms with 10 employees or more. Subsectors represent 2 digit (division) SIC07 codes. Subsectors 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"). Subsector 19 ("manufacture of coke and refined petroleum") is omitted from main analysis and thus also omitted here.

Finally, Table A6 shows the decomposition of λ^{HOM} across subsectors (similar to Table 3).[43] In line with results at the aggregate level, the key driver of the homogeneous labour share (and thus of the subsector labour shares) is the disconnect between pay and productivity. In most subsectors, productivity grew faster than real wages. Exceptions are subsectors 26 and 27 ("manufacture of computer, electronic and optical products" and "manufacture of electrical equipment", respectively), where wages grew faster than productivity, and subsector 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.

[43] Unlike in Table A5, no decomposition is shown for the "combined" subsectors because this requires an estimate of ρ, which was only estimated at the subsector levels.

Whereas in the latter both product and market power had an equally minor role in λ^{HOM}, in most subsectors the contribution of labour market power is significantly greater than that of product market power. In fact, in some subsectors the change in labour market power is high enough to make a significant difference to the subsector's λ^{HOM}. For instance, in subsector 26, the pay-productivity disconnect changed very little over the period; it is $\bar{\chi}^L$ which defines the bulk of the change. In particular, the labour market power of firms in this subsector fell importantly over the period.[44]

Overall, subsector and industry-wide results differ only where the latter masks heterogeneity in the former. As Table A6 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 subsectors but one (subsector 33).

Table A6: Contribution to changes in subsectoral λ^{HOM}, 1998-2014

Subsector	λ^{HOM}	$\bar{A}$	$\bar{w}/\bar{p}$	$\bar{\chi}^Y$	$\bar{\chi}^L$	$\bar{\tau}$	Interaction
13	-9.56%	-32.94%	24.37%	-2.20%	-3.36%	1.33%	3.22%
14	-34.00%	-136.42%	64.55%	-3.87%	26.76%	-34.73%	49.71%
16	-15.61%	-22.44%	7.68%	-2.21%	4.85%	-6.97%	3.49%
17	-9.56%	-28.03%	13.97%	-0.62%	1.29%	-5.03%	8.86%
18	-7.06%	-21.37%	11.88%	-1.66%	-3.24%	8.30%	-0.97%
23	-7.08%	-13.25%	9.68%	0.84%	-0.12%	-5.19%	0.96%
25	-14.66%	-20.15%	9.05%	-3.83%	-6.83%	2.98%	4.11%
26	12.49%	-79.48%	81.38%	-1.49%	10.61%	-6.51%	7.97%
27	8.73%	-7.90%	17.46%	0.18%	10.29%	-11.64%	0.35%
28	-7.54 %	-34.12%	23.59%	-0.89%	1.48%	-3.43%	5.84%
29	-21.55%	-87.24%	55.68%	0.41%	-15.14%	2.49%	22.25%
30	-19.66%	-14.08%	0.52%	-1.76%	-0.63%	-7.58%	3.87%
33	10.90%	4.94%	-0.39%	0.08%	17.46%	-15.09%	3.90%

Note: the extreme behaviour of subsector 14 is due to a significant reduction in the sample size available, from 340 firms in 1998 to 56 firms in 2018.
Source: our calculation based on ARD data.
Sample: see Table A5.

[44] Recall from the growth accounting decomposition that $g_{\bar{\chi}^L}$ is multiplied by $-\left(\frac{1}{1-\rho}\right)$. The estimated ρ for this subsector is negative, which, combined with a fall in $\bar{\chi}^L$ (i.e. a fall in firms' labour market power) yields the positive contribution of this variable to λ^{HOM}, as Table A6 shows.