

Productivity Dispersion, Between-firm Competition and the Labor Share

Preliminary and incomplete version, please do not cite.

Émilien Gouin-Bonenfant*

August 2017

Abstract: In this paper, I study how the pass-through of productivity to wages depends on the distribution of productivity across firms. Using administrative data covering the universe of Canadian corporations, I document high concentration of value added within highly productive, low-labor-share firms. Importantly, these large firms do not have a higher capital-output ratio and achieve a low labor share *despite* paying above average salaries. To interpret these findings, I develop a tractable firm dynamics model (à la [Hopenhayn \[1992\]](#)) with search frictions and wage posting in the labor market (à la [Burdett and Mortensen \[1998\]](#)). In the model, more productive firms offer higher wages in order to increase their market share by poaching workers from lower paying firms. As in the data, most firms have a high labor share, routinely above one, yet the aggregate labor share is low due to the disproportionate effect of a small fraction of large, extremely productive “superstar firms”. The model predicts that the pass-through of aggregate labor productivity to average wages is lower when productivity dispersion across firm is high, meaning that all else equal, an increase in productivity dispersion decreases the aggregate labor share. The mechanism is that an increase in the productivity differential between high and low productivity firms increases profit margins at high productivity firms, who become effectively shielded from wage competition. I test the model’s prediction and mechanism using cross-country data and find support, thus suggesting that the measured rise in productivity dispersion has contributed to the decline of the global labor share.

Keywords: Labor share; productivity; firm dynamics; search frictions; monopsony

JEL Codes: E2, J4, L1, O4, D2.

*Department of Economics, University of California San Diego. Email: egouinbo@ucsd.edu

1 Introduction

The stability of the labor share has been enshrined as one of the stylized facts of economic growth. One implication of a constant labor share is that, at the macro level, productivity gains translate in proportional wage gains. However, evidence points to a decoupling of wage and productivity growth in the U.S. and across the world starting around the 1980s, leading to a decline of the global labor share.¹ But behind national measures of productivity and wages lies a lot of heterogeneity. Indeed, the dispersion in productivity and average wages across firms is large and has been increasing over time.² An emerging narrative emphasizes the importance of firm heterogeneity in understanding movements in the labor share. Autor et al. [2017] and Kehrig and Vincent [2017] present evidence that the decline of the U.S. labor share is primarily driven by a reallocation of value-added towards firms with a low labor share (i.e. *between-firm component* of the decline) as opposed to a broad-based decline of the labor share amongst incumbent firms (i.e. the *within-firm component*), thus challenging theories which imply a uniform effect across firms. In addition, Barkai [2016] shows that industries where sales concentration rose the most saw the largest declines in the labor share. Together, these facts suggest that the configuration of production amongst heterogeneous firms within an industry is an important determinant of the labor share.

In this paper, I study how the pass-through of aggregate productivity to average wages depends on the dispersion of productivity across firms. I argue that an increase in productivity dispersion leads to a decline the labor share. From a growth perspective, the implication will be that broad-based productivity gains increase wages one-for-one while gains achieved by a small set of firms will have a limited impact on wages. In order to formalize the argument, I propose a tractable model which links the distribution of labor productivity across firms to the distribution of employment and wages. The model is novel in that it embeds search frictions and wage-posting in the labor market (à la Burdett and Mortensen [1998]) in a firm dynamics model (à la Hopenhayn [1992]). At the heart of the model is a rent-sharing problem between the firm and its workers and between-firm competition, combined with on-the-job search, will play a crucial role in raising wages above the value of unemployment (as in Cahuc et al. [2006]). The presence of productivity dispersion weakens the *intensity* of competition, thus allowing high productivity firms to offer wages significantly below labor productivity while still being able to attract and retain workers. Effectively, productivity dispersion increases the *monopsony power* of firms at the top of the productivity distribution.

I discipline the model using administrative data covering the universe of Canadian corporations over the 2000-2015 period. The estimated model rationalizes several features of the data such as the negative

¹See Karabarbounis and Neiman [2014] and Dao et al. [2017] for a cross-country analysis and Elsby et al. [2013].

²See Syverson [2011] for a review of the evidence on productivity dispersion and Barth et al. [2014], Song et al. [2015] for evidence of rising inequality in average wage across firms. Berlingieri et al. [2017b] document an upward trend in productivity dispersion in OECD countries while Kehrig [2015] and Haltiwanger et al. [2016] find a similar pattern in the U.S. manufacturing sector.

correlation between labor share and a firm's size (Autor et al. [2017]), the firm-size and firm-productivity wage premium (Oi and Idson [1999], Berlingieri et al. [2018]) as well as the positive correlation between concentration and the labor share (Barkai [2016]). In addition, the model fits the Canadian microdata along many non-targeted dimensions despite being very parsimonious. Quantitatively, the model implies an aggregate elasticity of roughly -0.2 between productivity dispersion (measured as the 90/10 percentile ratio of labor productivity across firms) and the labor share. I then use cross-country data on the labor share, wage dispersion and productivity dispersion across firms to test the model's prediction. I find support for the mechanism and the estimated elasticities are in the same range of the model-implied elasticity.

Understanding the factors which caused the decline of the global labor share³ is important given that wages and salaries are the predominant source of income for most households while capital income is extremely concentrated. We would thus expect a decline of the labor share to lead to an increase in income inequality. The factor which I emphasize - *productivity dispersion* - is novel and particularly relevant in a context of rising productivity dispersion in advanced economies. Berlingieri et al. [2017b] present evidence that, over the last decades, productivity dispersion has increased significantly amongst advanced economies. The exact causes of the increase are unknown but evidence suggests that globalization and technological change played a role. For instance, Dunne et al. [2004] find that, in the U.S. manufacturing sector, an important fraction of the increase in productivity dispersion is accounted for by changes in the distribution of computer investment across plants. Compared to theories based on labor-capital substitution, the model predicts comovements in response to a productivity dispersion shock consistent with the *superstar firm hypothesis* fleshed out in Autor et al. [2017] as well as the facts documented in Kehrig and Vincent [2017] regarding the U.S. manufacturing sector. In particular, in response to an increase in productivity dispersion (1) output concentration increases, (2) the aggregate labor share decreases while (3) the unweighted labor share across firms increases.

The contribution of the paper is twofold. First, I provide microeconomic evidence on the determinants of firm-level labor shares. The data used is particularly suited for this exercise given that (1) it covers the universe of corporations in all sectors of the Canadian economy (2) it contains the variables necessary to compute gross operating surplus, wages paid and the value of the capital stock at the corporation level, thus allowing me to use the *System of National Accounts* definition of corporate-sector value-added and (3) it provides information on the ownership structure of firms and establishment. I document high concentration of value added within highly productive, low-labor-share firms. Importantly, these large firms do

³Although many studies present evidence that we are currently experiencing a decline of the global labor share (see for example Karabarbounis and Neiman [2014], Rognlie [2016], Dao et al. [2017]), it is worth mentioning that researchers disagree on the extent to which the global labor share has declined. This partly reflects the fact that measuring the labor share within an economy poses challenges with respect to the treatment of taxes and capital depreciation (Rognlie [2016], Bridgman [2017]) as well as intangible capital (Koh et al. [2016]) and self-employment (Gollin [2002]).

not have a higher capital-output ratio and achieve a low labor share *despite* paying above average salaries. Moreover, I show that the wage and labor share size relationships are driven by labor productivity alone, not employment. In particular, I show that after controlling for labor productivity, employment does not predict the labor share or wages paid at the firm.

The second contribution is theoretical. I introduce proper firm dynamics (i.e. endogenous entry and exit, productivity shocks, and firm life-cycle) in the canonical wage posting model presented in [Burdett and Mortensen \[1998\]](#) (from now on *BM*). Although traditional search models can generate realistic predictions regarding the link between firm-level productivity and wages (see [Cahuc et al. \[2006\]](#) and [Bagger et al. \[2014\]](#)), the addition of firm dynamics is necessary for the purpose of this paper due to two reasons. First, introducing productivity shocks generate churn in the productivity distribution which allows the model to rationalize the large share of firms with a labor share above one (i.e. negative profits) found in the data. With time-invariant firm productivity (as in the *BM* model), only profitable firms select into entry and never exit. Quantitatively, the problem is that the high profit margins at very productive firms are not compensated by negative profit margins at the bottom of the distribution. Second, the introduction of productivity shocks and endogenous entry and exit breaks the one-to-one relationship between employment, productivity and wages implied by models with time-invariant productivity. The introduction of a firm life-cycle “fixes” a counterfactual prediction of the *BM* model which is that wages are strictly increasing in firm size so that job-to-job flows are unidirectional from small firms towards large firms. [Haltiwanger et al. \[2017\]](#) present evidence that a significant portion of job-to-job flows occurs from large low-wage firms towards small high-wage firm. In their words: “One of the reasons firm size may be a much less good proxy for productivity of a firm is the role of firm age [...] At least some small, young firms are highly productive and are in the process of growing to become large firms.” This allows the model to match the fat upper tail of the firm size distribution through the well-known mechanism of *random growth* (see [Luttmer \[2007\]](#)). This is of crucial importance for the purpose of this paper which highlights the relationship between concentration and the labor share.

In terms of observables, the stationary equilibrium of the model consists of a distribution of firm level wages w , employment N , capital stock K and value-added y . Despite its parsimony, the model generates a distribution of observables which is quantitatively consistent with the data along key dimensions. First, it rationalizes the relationship between labor productivity, growth, and wages. The mechanism is that high productivity firms pay high wages in order to grow faster. Second, the endogenous firm size distribution conforms to Zipf’s law, meaning that it has a Pareto upper tail with Pareto exponent slightly above one.⁴ Random growth models relying on i.i.d. growth drawn from a distribution with high variance can generate Zipf’s law (for instance in [Gabaix \[1999\]](#)), the model I propose generates employment concentration as

⁴A random variable X is said to have a Pareto upper tail with exponent ζ if $\lim_{x \rightarrow \infty} \mathbb{P}(X \leq x) = cX^{-\zeta}$ for some constant $c > 0$.

a result of a small share of firms experiencing very high growth rates for sustained periods of time. The behavior of these “gazelle” firms have been extensively studied and is thought to be an engine of growth. As discussed in [Luttmer \[2011\]](#), persistent growth rate differentials across firms are needed to account for the relatively young average age of very large firms. In the estimated model, the mechanics of firm growth and wage-posting give rise to a realistic allocation of income between wages and profits. As in the data, most firms have a high labor shares (often above one) while the largest firms tend to have low labor shares, leading to an aggregate labor share of roughly two thirds.

Related literature. This paper relates to several strands of literature. First, my model is based on the influential work of [Burdett and Mortensen \[1998\]](#) and [Bontemps et al. \[2000\]](#) in that it features random matching, on-the-job search, take-it-or-leave-it wage offers and firm heterogeneity. [Postel-Vinay and Robin \[2002\]](#) and [Cahuc et al. \[2006\]](#) augment the standard model by introducing counter offers. Wage-posting models have been used to study informal labor markets [Meghir et al. \[2015\]](#), inequality [Moser et al. \[2016\]](#), and business cycle dynamics [Moscarini and Postel-Vinay \[2016\]](#) and more sophisticated models involving two-sided heterogeneity and sorting have been proposed ([Bagger et al. \[2014\]](#), [Bagger and Lentz \[2014\]](#), [Borovicková \[2016\]](#), [Lise and Robin \[2017\]](#)). This paper differs in that it features firm dynamics as in [Coles and Mortensen \[2016\]](#) and [Kaas and Kircher \[2015\]](#). Another strand of search theory - *directed search* - allows workers to direct their search effort towards specific sets of firms (see [Wright et al. \[2017\]](#) for a thorough literature review). This paper also contributes to the literature on the causes and consequences of the decline of the labor share ([Elsby et al. \[2013\]](#), [Karabarbounis and Neiman \[2014\]](#), [Piketty and Zucman \[2014\]](#), [Rognlie \[2016\]](#), [Barkai \[2016\]](#), [Kehrig \[2015\]](#), [Koh et al. \[2016\]](#), [Autor et al. \[2017\]](#), [Dao et al. \[2017\]](#), [De Loecker and Eeckhout \[2017\]](#), [Hartman-Glaser et al. \[2017\]](#)).

The remainder of the paper is organized as follows. Section 2 presents the dataset and provides microeconomic evidence on the determinants of the labor in the cross-section of firms. Section 3 presents the model. In Section 4, I estimate the model and validate the model. Finally, in Section 5, I use the estimated model to conduct counterfactual analysis and test the predictions using cross-country data.

2 A preliminary look at the data

The main dataset used is the *National Accounts Longitudinal Microdata File* (NALMF) produced by Statistics Canada which is obtained by merging administrative data from different sources. The NALMF contains de-identified data covering the universe of private sector employers in Canada over the 2000-2015 period. The unit of observation is an enterprise-year and I make use of two distinct sources of primary information: financial statements from the *Corporation Income Tax Return* (T2) and employment and earnings data from the *Statement of Remuneration Paid* (T4) and *Payroll Deductions and Remittances* (PD7). Corporations are

required by the *Canadian Revenue Agency* to file a T2 at the end of the fiscal year even if no taxes are owed. The tax form includes a Balance Sheet (schedule 100) and an Income Statement (schedule 125). In addition, employers must file a T4 at the end of the fiscal year for every worker employed at some point throughout the year. The T4 contains employment income as well as various contributions to programs such as Employment Insurance. Finally, employers must send a monthly PD7 form detailing payroll deductions.⁵ The monthly nature of this third program allows for an accurate assessment of the average number of employees at the firm throughout the year.

The main sample is restricted to the private corporate sector excluding Agriculture and Mining (NAICS two-digit sectors 11,21) due to data limitations as well as Education and Health (61 and 62) due to the fact that these sectors are dominated by public entities in Canada. Moreover, I restrict the analysis to firms with at least 20 employees. Finally, I remove firm-year observations which either have (1) negative value-added, (2) average salary above 500,000\$, and (3) missing values in payroll, employment, value-added, tangible assets or industry code. The resulting extract (from now on the *main sample*) covers roughly approximately 45% of private sector employment over the 2000-2015 period. Over the period considered, wage and price inflation was very limited. On average year-to-year CPI inflation was 1.5% and nominal wages in the main sample were pretty flat, with an increase of only 8% from 2000 to 2015. For most of descriptive part of the analysis, I will pool all years together to create a large cross-section. I use the total CPI to convert prices into 2002 Canadian dollars.

I now present facts related to the distribution of value-added Y , employment N , average wage w and stock of capital K across firms (see Appendix A.1 for a description of the variables construction). First, I pool all firm-year observations in a large cross-section and sort by value-added within two-digit North American Industry Classification System (NAICS) industries. For each industry, I bin firms into 4 size groups which contribute equally to sectoral value-added (i.e. each size group contribute a quarter of value-added).⁶ Then, I construct economy-wide size groups by pooling all industries together. The advantage of sorting firms in this manner is that (1) each group is equally important in terms of contribution to GDP, (2) each group has the same industrial composition as the aggregate economy, and (3) even if some industries are very concentrated, confidentiality is preserved.

⁵Employers with less than 25000CAD in average monthly withholding can submit a PD7 form on a quarterly basis.

⁶The size thresholds for industry j are compute according to the following formula.

$$q_{\tau,j} = \min \left\{ q : \sum_{\substack{i \in \mathcal{J} \\ Y_i \leq q}} Y_i = \tau Y_j \right\} \quad \tau \in \left\{ \frac{1}{4}, \frac{1}{2}, \frac{3}{4} \right\}$$

Where Y_j is the total value-added in industry j while $\mathcal{J}$ denotes the set if firm indices i which are part of industry j

2.1 Firm size and concentration

Value-added in the sample is highly concentrated, the largest 0.3% of firms account for a quarter of total value-added (see Table 1). The average firm in the sample produces 8.1 million Canadian dollars (CAD) of value-added per year whereas firms in the fourth group produce on average 736 million CAD which is roughly 350 times more than the average firm in the first group. These large differences in value-added across firms can partially be explained by differences in employment. While the average firm has 119 employees, firms in the fourth group employ on average 7008 employees which is roughly 130 times more than firms in the first group (see Table 1). The fact that the value-added differential across groups are larger than employment differentials reflects a critical feature of the data: high value-added firms not only employ more workers, they also achieve higher levels of value-added per worker (i.e. they have higher labor productivity).

Table 1: Share of firms, average value-added (M\$) and employment

Group	%	Y	N
1	83	2	54
2	14	14	212
3	2.7	75	941
4	0.3	736	7008
Total	100	8.1	119

2.2 Labor share size-effect

How does the allocation of value-added between workers and shareholders differ by size group? I find that high value-added firms tend to have a significantly lower labor share than small firms. As a whole, Group 1 has a labor share of 80% while Group 4 has a labor share of 45% (see Table 2). The labor share in the full sample is 62% and can be obtained as a simple average of the labor share in the four groups. High value-added firms thus contribute disproportionately to profits while low value-added firms contribute more than their share in wages. These findings are consistent with [Edmond et al. \[2015\]](#) and [Autor et al. \[2017\]](#) which show that the labor share is decreasing in market share.

Capital output ratio explanation. One candidate explanation for the fact that large firms have a low labor share is that those firms spent more on capital expenditures and that the high measured profit margins are in fact compensation for past investments. If that was the case, the capital output ratio (i.e. the ratio of the value of the firm-level capital stock to its value-added) should be higher at large firms. Surprisingly, the data provides no evidence in that regard. The capital output ratio is essentially constant across groups

Table 2: Aggregate contributions and labor share by size groups

Group	LS	K/Y
1	0.80	1.05
2	0.65	1.07
3	0.60	1.10
4	0.45	1.04
Total	0.62	1.06

around its aggregate value of 1.06 (see Table 2). Of course, the value of capital assets is measured at book and perhaps some firms paid less for assets of equal market value but accounting data does not allow for such a distinction.

What do we learn from a constant capital to output ratio across firms? In the neoclassical investment model with labor and capital as factors of production, perfect capital markets and *constant elasticity of substitution* technology,⁷ the capital output ratio is given by $\frac{K}{Y} = (\frac{\alpha}{R})^\rho z^{1-\rho}$ where ρ is the elasticity of substitution, z is total factor productivity, α is the share parameter and R is the user cost of capital. The special case where the elasticity of substitution between inputs is one $\rho = 1$ (which represent the *Cobb-Douglas* case) implies that the capital output ratio is constant across firms independently of their total factor productivity.

Capital rental versus ownership explanation. Another candidate explanation for the negative size and labor share correlation in the data is based on the differential accounting treatment of *current expenses* versus *capital expenses*.⁸ Consider the following example. Firms A and B operate in the retail industry and each have 100\$ of monthly sales, 25\$ of cost of goods sold and a 50\$ of payroll. Firm A rents its store for 10\$ a month while firm B owns the store. Although both firms have the same sales and payroll, firm A has generates $100\$ - 25\$ - 10\$ = 65\$$ of value-added while firm B generates $100\$ - 25\$ = 75\$$. The reason why value-added is lower at firm A is the the rental cost is expensed while for firm B, the cost of acquiring the building was capitalized at the moment of acquisition. In addition to generating value-added through retailing services, Firm B generates value-added by by creating “leasing services” from its asset. As a result, the labor share at firm A is $50/65 \approx 77\%$ while it is $50/75 \approx 66\%$ at firm B. If large firms are more likely to own their capital assets, perhaps the labor-share-size correlation in the data is the result of accounting conventions. I find that it is not the case as an alternative measure of the labor share which is invariant

⁷The constant return to scale CES production function with j inputs takes the form $Y = z \left(\sum_j \alpha_j x_j^\rho \right)^{\frac{1}{\rho}}$ where the *factor shares* $\{\alpha_j\}_{j=1}^J$ are non-negative and sum to one.

⁸As per the *Canadian Revenue Agency*, a *current expense* is one that generally reoccurs after a short period while a *capital expense* generally gives a lasting benefit or advantage. Current expenses directly affect the net income reported on the Income Statement while capital expenses increase the book value of assets reported on the Balance Sheet.

to the treatment of capital expenditures - the payroll to sales ratio - exhibits the same pattern as the labor share.⁹ The economics behind the choice between rental and ownership of capital has been extensively studied (see [Miller and Upton \[1976\]](#), [Eisfeldt and Rampini \[2008\]](#))

2.3 Wage size-effect

The labor share at a firm can be expressed as the average salary as a ratio of the labor productivity (i.e. $LS = \frac{wN}{Y} = \frac{w}{Y/N}$). A low labor share can thus be the result of low salaries and/or high labor productivity. I find that the low labor share is the result of high labor productivity, not low wages. In fact, average salaries are increasing in firm size (see Table 3), consistently with a long empirical literature document *firm-size wage premia* (see [Oi and Idson \[1999\]](#)). The magnitudes I measure are large: firms in group 4 have average salaries of 68,000 CAD per year which is more than twice the average salary of 33,500 CAD in group one. The lower labor share in group four is thus the result of extremely high labor productivity: firms in group four are nearly ten times more productive than firms in group one. These measurements are consistent with a rent-sharing view of wages where more productive firms share some of the firm-level rent with their workers (see [Card et al. \[2018\]](#) for a review of the empirical evidence).

Table 3: Average wage and labor productivity by size group

Group	w	LP	k
1	33.5	41.3	31.1
2	45.0	70.1	48.4
3	48.8	111.3	108.7
4	68.0	406.0	351.0
Total	35.6	48.3	36.5

Sorting explanation. An alternative explanation would be that high-value added firms tend to employ more productive workers and pay them higher salaries. Although I can not rule out the existence of sorting, in Appendix A.3 I provide evidence that the firm-size wage premium in the Canadian economy over the 2000-2015 period is not the result of sorting on observable characteristics (education, occupation, experience, etc) using microdata underlying the *Labor Force Survey*. Can differences in *unobserved heterogeneity* explain the firm-level correlation between labor productivity and wages? [Bagger et al. \[2014\]](#) find that worker heterogeneity (as measured by the worker fixed-effect in an employer-employee two-way fixed effect regression) is not an important determinant of firm-level labor productivity. They use matched employer-employee data from Denmark and find that, although worker heterogeneity accounts for most of the variance in log wages, it is not an important source of cross-sectional differences in labor productivity across

⁹[Autor et al. \[2017\]](#) use sales as a measure of size and find a negative correlation with the labor share using U.S. data.

firms.

Another feature of the data worth emphasizing is that the high labor productivity at large firms is achieved by having a much higher stock of capital per worker. The differentials are large with firms in group four having on average ten times more capital per worker than firms in group one. This finding seems a priori inconsistent with the fact that the capital to output ratio is constant across firm groups - it is not. To fix ideas, consider a constant return to scale production function for firm i 's output $Y_i = z_i K_i^\alpha N_i^{1-\alpha}$ where z_i is a measure of total factor productivity and $\alpha \in (0, 1)$. The production function implies that the capital-output ratio and capital intensity are related through

$$z_i = \left(\frac{K_i}{N_i} \right)^{1-\alpha} \frac{Y_i}{K_i} \quad (1)$$

which implies that a the observation that Y_i/K_i is constant across groups while K_i/N_i is not is consistent with larger firms having higher TFP (e.g. firms in Group 4 have higher TFP than firms in group 1). We thus have that large firms achieve high labor productivity through a combination of (1) high total factor productivity z and (2) high capital intensity $\frac{K}{N}$. If labor productivity differentials were driven purely by capital intensity, then the capital output ratio would have to be higher at large firms.

3 Model

3.1 Environment

Time is continuous and a period $[t, t + 1)$ represents a year. The economy is populated by an endogenous measure F of heterogeneous firms and a unit measure of identical workers. Firms are risk neutral, discount the future at exogenous rate $r > 0$, compete for workers by posting wage contracts w , and rent capital in a centralized market at rate R . Firms can change the wage policy at any time and therefore do not precommit to future wages. I assume that, since workers are equally productive, the firm is required to pay all of its workers the same wage. Workers can be either employed or unemployed and search for jobs in both states. The flow value of unemployment is exogenously given by $b > 0$. When a worker receives a job offer from a firm, she must form expectations regarding the future path of wages and compare it to the continuation value in her current state (either employed or unemployed). Firms produce an homogeneous good and differ only in their *total factor productivity* (TFP) level $z > 0$. I assume a Cobb-Douglas functional form with constant returns to scale for the production function:

$$y = zK^\alpha N^{1-\alpha} \quad (2)$$

where N denotes the measure of workers and K the stock of capital.

Firm turnover. At rate $\chi_1 > 0$, a potential entrant draws a productivity level z from a distribution $\Gamma_0(z)$ with support $[z_0, \bar{z}]$ and decides whether or not to enter. Similarly, active firms draw a new productivity level at rate $\chi_2 \geq 0$ from that same distribution Γ_0 . In equilibrium, the optimal strategy for the firm will be to enter whenever the draw of productivity is above an endogenous threshold $z_l \geq z_0$ and to exit whenever the new level of productivity falls below z_l . The optimal behavior of the firm will be characterized later but the intuition is that firms with $z < z_l$ are not productive enough to afford to pay workers their outside option of b and therefore cannot operate. To simplify notation, define the truncated distribution $\Gamma(z)$, the rate of firm exit $\chi_x(z_l)$, the flow of entrants $\chi_e(z_l)$ and the rate at which a firm's productivity changes without leading to an exit $\chi_s(z_l)$ by

$$\Gamma(z) = \frac{\Gamma_0(z) - \Gamma_0(z_l)}{1 - \Gamma_0(z_l)} \text{ for } z \in [z_l, \infty) \quad (3)$$

$$\chi_x(z_l) = \chi_2 \Gamma_0(z_l), \quad \chi_e(z_l) = \chi_1 (1 - \Gamma_0(z_l)), \quad \chi_s(z_l) = \chi_2 (1 - \Gamma_0(z_l)). \quad (4)$$

In stationary equilibrium, the measure of active firms is given by the ratio of rates of entry and exit $F = \chi_e(z_l) / \chi_x(z_l)$. For notational simplicity, I suppress the dependence of the rates χ_e, χ_s, χ_x on the endogenous threshold z_l .

Matching technology. In the *BM* model, a firm meets workers at rate λ , which implies that, as it grows, it becomes increasingly difficult to meet new workers. Following the insight of [Burdett and Vishwanath \[1988\]](#), who argue that the assumption of a size-independent meeting rate has the unrealistic implication that the value of a firm is higher if it splits itself in two,¹⁰ I assume balanced matching, which means that the rate at which a firm meets workers grow linearly with its size. Moreover, I assume random matching, meaning that the identity of the workers a firm meets is drawn uniformly from the population of workers independently of her current status (e.g. employed or unemployed). For example, a firm with a mass N of employees meets a new worker at rate λN . I also assume that entering firms meet a mass λ_0 of workers, that workers at exiting firms are sent into unemployment and that jobs at existing firms are exogenously destroyed at rate $\delta \geq 0$. Using the fact that existing as well as entering firms offer wages above the value of unemployment, the law of motion for the unemployment rate is given by:

$$\dot{u} = \underbrace{(1 - u)(\delta + \chi_x)}_{\text{Unemployment inflows}} - \underbrace{u\lambda(1 - u) - u\lambda_0\chi_e}_{\text{Unemployment outflows}} \quad (5)$$

¹⁰Typically, in job search models it is assumed that a worker is equally likely to contact any firm in the market. Thus, when there are n firms in the market, the probability that a worker contacts a given firm is $\frac{1}{n}$. This will be termed "random matching." This simple specification, routinely used with little justification, is subject to criticism. For example, an unfortunate consequence of random matching is that a firm, by splitting itself into two, can increase its number of potential employees since $\frac{1}{n}$ is less than $\frac{2}{n+1}$ and, thus, possibly increase its profits"

where $\dot{u} \equiv du/dt$. Notice that the right-hand side of Equation 5 is quadratic in u . In Appendix B.1, I show that the unique stationary solution (i.e. $\dot{u} = 0$) is given by the following formula:

$$u = \frac{\delta + \chi_x + \lambda_0 \chi_e + \lambda - \sqrt{(\delta + \chi_x + \lambda_0 \chi_e + \lambda)^2 - 4\lambda(\delta + \chi_x)}}{2\lambda} \quad (6)$$

The matching technology can be recast in terms of a matching function $M(S_F, S_W)$ where the rate at which a match occurs depends on aggregate search effort from firms S_F and aggregate search effort from workers S_W . The balanced matching assumptions can be interpreted as saying that firms generate one unit of search effort per employee. Since there is a measure $1 - u$ of workers, the aggregate firm search effort is $S_F = 1 - u$. Now assume that each worker (employed or unemployed) generates one unit of search effort (i.e. $S_W = 1$). The matching function consistent with the balanced random matching assumption is thus

$$M(S_F, S_W) = \lambda S_F^\theta S_W^{1-\theta} \text{ with } \theta = 1 \quad (7)$$

Which implies that the rate at which a worker meets a firm is $M(S_F, S_W)/S_W = \lambda(1 - u)$ while the rate at which a firm of size N meets a worker is $NM(S_F, S_W)/S_F = \lambda N$

Beliefs. The dynamics of the unemployment rate was easy to characterize due to the fact that unemployed workers accept any job offer above the value of unemployment. For employed workers receiving a competing job offer, it is more complicated as they need to form expectation regarding the path of future wages at both the current employer and the competing firm and accept the job offer only if the continuation value of quitting is higher than the continuation value of staying. I assume that the wage policy w at a firm as well as its TFP z are public information. I assume that workers form expectation regarding the wage of a continuing firm at time $t + s$ by conditioning only on the firm's wage at time t . The expectation $\mathbb{E}(w_{t+s}|w_t = w)$ is given by a belief function $\hat{w}(w, s)$. Moreover, I assume that when meeting an entering firm, which does not yet have a wage policy, workers form expectation regarding the wage at time $t + s$ by conditioning only on the firm's productivity at time t . The expectation $\mathbb{E}(w_{t+s}|z_t = z)$ is given by a belief function $\hat{w}_0(z, s)$.

Assumption 1. *The belief functions $\hat{w}$ and $\hat{w}_0$ satisfy the following monotonicity condition*

$$w' > w \implies \hat{w}(w', s) > \hat{w}(w, s) \quad \forall s > 0$$

$$z' > z \implies \hat{w}_0(z', s) > \hat{w}_0(z, s) \quad \forall s > 0$$

Risk-neutrality of workers ensures that the expected path of wages $\{w_{t+s}\}_{s>0}$ is a sufficient statistic to compute the continuation value of a job. The solution concept which I consider is a pure strategy bayesian equilibrium and will require that the belief functions $\hat{w}$ and $\hat{w}_0$ are indeed correct in equilibrium. For a thorough discussion of this solution concept and the monotone beliefs assumption, I refer the reader to [Coles and Mortensen \[2016\]](#). It follows from the assumption, workers' optimal decision rule is trivial: they accept any job offer which leads to a pay increase. The reason for this simple job acceptance rule is that

there is not trade-off, moving to a higher paying firm (or more productive in the case of entering firms) does not restrict the job offers one will get in the future.

Employment growth. I now characterize the link between firm-level pay policy and employment growth. The change in employment $\dot{N} \equiv \frac{dN}{dt}$ at a firm offering wage rate $w \geq b$ and with current employment N is given by

$$\dot{N} = \tilde{g}(w)N \quad (8)$$

where $\tilde{g}(w)$ is an endogenous object which I will refer to as the employment growth function. The growth rate depends only on the current wage due to the fact that (1) meeting rates are linear in the firm size (balanced matching assumption) and (2) workers job acceptance rule depends only the current wage (monotone belief assumption). A simple expression can be obtained for the employment growth function $\tilde{g}$.

$$\tilde{g}(w) \equiv \underbrace{\left[\lambda u + \lambda(1-u)\tilde{P}(w) \right]}_{\text{hiring rate}} - \underbrace{\left[\delta + \lambda(1-u)(1-\tilde{P}(w)) + \chi_e \lambda_0(1-\Gamma(w_0^{-1}(w,0))) \right]}_{\text{separation rate}} \quad \forall w \geq b \quad (9)$$

The function $\tilde{P}(w)$ denotes the endogenous wage distribution (CDF) in the population of workers. The distribution of wages at entering firms implied by the workers beliefs is $\Gamma(w_0^{-1}(w,0))$ where $w_0^{-1}(w,0)$ is the unique inverse function defined by $w_0(w_0^{-1}(w,0),0) = w$ for all $w \geq b$.¹¹ The hiring rate for firms paying w is given by the sum of the rate at which a firm with $N = 1$ meets an unemployed worker which is λu and the rate at which it meets a worker earning less than w which is $\lambda(1-u)P(w)$. By increasing its wage, a firm increases its rate of hiring from employment since it is able to poach workers higher in the wage distribution $P(w)$. Similarly, the separation rate is given by the sum of the exogenous job destruction rate δ , the rate of quits to existing firms $\lambda(1-u)(1-P(w))$ and the rate of quits to entering firms $\chi_e \lambda_0(1-\Gamma(w_0^{-1}(w,0)))$. Offering higher wages thus serves a dual purpose: attracting more workers and retaining them more effectively.

Firm's problem. The problem of an existing firm is to chose a sequence of wage rates and levels of capital to rent $\{w_s, K_s\}_{s=0}^T$ as well as a stopping time T in order to maximize the expected flow of profits discounted at rate $r > 0$. Firms rent capital in a centralized market at rate $R = r + \tau$ which reflects the user cost of capital (opportunity cost of funds r + depreciation rate τ). Given that firms are risk neutral, the value function $v(z, N)$ is given by the expected discounted sum of profits.

$$v(z_0, N_0) = \max_{T, \{w_s, K_s\}_{s=0}^T} \mathbb{E}_0 \int_0^T e^{-rs} \left(z_s K_s^\alpha N_s^{1-\alpha} - w_s N_s - R K_s \right) ds \quad (10)$$

¹¹Uniqueness of the inverse is ensured by the fact that the function $w_0(\cdot, 0)$ is strictly monotone under Assumption 1.

$$w_s \geq b \quad (11)$$

$$\dot{N}(w_t) = \tilde{g}(w_t)N_t \quad (12)$$

$$dz_s = dJ_s(z'_s - z_s) \quad (13)$$

where J_s is a Poisson process with intensity χ_2 and $\{z'_s\}_{s>0}$ is a sequence of i.i.d. productivity draws from Γ_0 . To ensure that the value function is well-defined, I make the following assumption.

Assumption 2. *The meeting rate λ , rate of arrival of productivity shocks χ_2 and exogenous job destruction rate δ satisfy*

$$\chi_2 + \delta - \lambda > 0$$

This assumption ensures that for any $r > 0$, that the value function is finite (i.e. that a firm cannot come to dominate the economy). Notice that the firm's problem exhibits constant return to scale both in terms of production technology and meeting technology which implies an important property for the value function.

Lemma 1. *The value function is homogeneous of degree one in N , which means that $v(z, N) = v(z, 1)N$, and the optimal stopping time is to exit whenever productivity falls below a threshold z_l .*

Proof. See Appendix B.2 □

Denote the value to a firm of a measure one of employees by $v(z) \equiv v(z, 1)$ and the stock of capital per worker by $k \equiv K/N$. Using Lemma 1, the value function can now be characterized by an HJB equation (Equation 14) combined with a boundary condition (Equation 15).

$$(r + \chi_x)v(z) = \max_{\substack{w \geq b \\ k \geq 0}} \left\{ \underbrace{zk^\alpha - w - Rk}_{\text{profits}} + \underbrace{v(z)\tilde{g}(w)}_{\text{growth}} \right\} + \underbrace{\chi_s \left(\int v(u)\Gamma(du) - v(z) \right)}_{\text{productivity shocks}} \quad (14)$$

$$z_l = \min \left\{ z : v(z) \geq 0 \right\} \quad (15)$$

Recall that the rates χ_s and χ_x depend on the endogenous threshold z_l . A critical implication of the homogeneity result (Lemma 1) is that the wage function $w(z)$ and the optimal the stock of capital per worker $k(z)$ are firm-size independent. The first order conditions for optimality, when the constraint $w \geq b$ does not bind, are given by:

$$\underbrace{z\alpha k(z)^{\alpha-1}}_{\text{marginal product of capital}} = \underbrace{R}_{\text{user cost of capital}} \quad (16)$$

$$\underbrace{1}_{\text{marginal increase in wage bill}} = \underbrace{v(z)\tilde{g}'(w(z))}_{\text{marginal increase in value of the firm}}. \quad (17)$$

The first-order condition for the wage captures the core insight that high wages are an investment in future growth. When a firm offers a more generous pay policy, it incurs higher labor costs but attracts and retains more workers more effectively. Define the equilibrium growth rate of employment by $g(z) \equiv \tilde{g}(w(z))$. To solve for the wage schedule $w(z)$, first notice that $g'(z) = \tilde{g}'(w(z))w'(z)$. From Equation 16, we have that $w'(z) = v(z)g'(z)$, which is an initial value problem with the general solution

$$w(z) = w(z_l) + \int_{z_l}^z v(u)g'(u)du \quad (18)$$

The initial condition $w(z_l)$ is obtained by noticing that the marginal entering/exiting firm is constrained by the worker's participation constraint $w \geq b$ which means that $w(z_l) = b$. To complete the proof, the second-order condition for optimality of the candidate solution is

$$v(z)\tilde{g}''(w(z)) < 0 \quad \forall z > z_l \quad (19)$$

which, in Appendix B.3, is shown to hold in equilibrium.

Lemma 2. *Capital intensity and the wage schedule are increasing in z and are given by*

$$k(z) = \left(\frac{z\alpha}{R}\right)^{\frac{1}{1-\alpha}} \quad (20)$$

$$w(z) = b + \int_{z_l}^z v(u)g'(u)du \quad (21)$$

Equation 21 highlights the fact that in equilibrium, workers earn a wage above their outside option of $b > 0$ and that this “premium” is higher at more productive firms. In contrast to models with competitive labor markets, the marginal product of labor MPN is not equalized across firms. In fact, the MPN is increasing in firm productivity.

$$MPN(z) = (1 - \alpha) \left(\frac{\alpha}{R}\right)^{\frac{\alpha}{1-\alpha}} z^{\frac{1}{1-\alpha}} \quad (22)$$

The dispersion in marginal (and average) labor productivity across firms is thus driven by the equilibrium distribution of TFP - a feature of the model which I explore in Section 5.

3.2 Equilibrium

Definition. A Stationary Equilibrium consists of a value function $v(z)$, policy functions $w(z), k(z), g(z)$, belief functions for the workers $\hat{w}(w, s), \hat{w}_0(z, s)$, a productivity threshold z_l , an unemployment rate u and wage distribution $\tilde{P}(w)$ which satisfy the following conditions:

- (i) Workers beliefs satisfy the monotonicity assumption (Assumption 1)
- (ii) $(v(z), w(z), k(z), z_l)$ solve the firm's problem (Equations 14, 15, 20, 21)
- (iii) The unemployment rate is in steady-state (Equation 6)
- (iv) The wage distribution $\tilde{P}$ is in steady-state
- (v) Workers beliefs are correct

Proposition 1. *Under Assumption 2 there exists a unique Stationary Equilibrium*

Proof. Fix $\hat{w}_0(z, 0) = w(z)$ and $\hat{w}(w(z), s) = \hat{w}_0(z, s)$. It remains to be shown that (1) there exists a unique $(v(z), z_l)$ which solves the HJB equation and (2) there exists a stationary wage distribution $\tilde{P}$, and (3) that the workers beliefs are correct. $\square$

A corollary of Lemma 2 is that the equilibrium exhibits *rank preservation*, meaning that the ranking of wages offered across firms is the same as the ranking of productivity. This result allows for a closed-form solution for the employment-weighted productivity distribution $P(z) \equiv \tilde{P}(w(z))$. The measure of workers working at firm with productivity less than z is thus given by $P(z)$ which has support $[z_l, +\infty)$. The *Kolmogorov Forward Equation*, which determines the evolution through time of the function $P(z)$, is given by:

$$\begin{aligned} \dot{P}(z)(1-u) = & \underbrace{(1-u)^2 \lambda P(z)(P(z)-1)}_{\text{Job offers to employed (by existing firm)}} + \underbrace{(1-u)\chi_e \lambda_0 P(z)(\Gamma(z)-1)}_{\text{Job offers to employed (by entrant)}} \\ & + \underbrace{u(1-u)\lambda P(z)}_{\text{Job creation (by existing firms)}} + \underbrace{u\chi_e \lambda_0 \Gamma(z)}_{\text{Job creation (by entrants)}} - \underbrace{(1-u)(\delta + \chi_x)P(z)}_{\text{Job destruction (by existing firm)}} + \underbrace{(1-u)\chi_s(\Gamma(z)-P(z))}_{\text{productivity shocks}}. \end{aligned} \quad (23)$$

Notice that the right-hand size of Equation 23 is quadratic in $P(z)$ which means that if a solution exists (i.e. $P(z)$ such that $\dot{P}(z) = 0$ for all $z \geq z_l$), it can be obtained in closed-form.

Proposition 2. *Under Assumption 2, there exists a unique stationary distribution $P(z)$ which is given by*

$$\begin{aligned} P(z) = & \frac{\lambda(1-2u) + \chi_e \lambda_0 \bar{\Gamma}(z) + \delta + \chi_s + \chi_x}{2\lambda(1-u)} \\ & - \frac{\sqrt{(\lambda(1-2u) + \chi_e \lambda_0 \bar{\Gamma}(z) + \delta + \chi_s + \chi_x)^2 - 4\lambda(1-u)(\frac{u}{1-u}\chi_e \lambda_0 + \chi_s)\Gamma(z)}}{2\lambda(1-u)} \end{aligned} \quad (24)$$

over the range $[z_l, +\infty)$

Proof. In Appendix B.4, I provide a derivation of the *Kolmogorov Forward Equation* and prove that the solution is unique and given by $P(z)$. $\square$

As a corollary, the equilibrium growth function $g(z)$ is obtained in directly from the definition of $\tilde{g}(w)$ (Equation 9) and the fact that $\hat{w}^{-1}(w(z))(z) = z$ in equilibrium.

$$g(z) = \lambda u + \lambda(1 - u)P(z) - \delta - \lambda(1 - u)\bar{P}(z) - \chi_e \lambda_0 \bar{\Gamma}(z) \quad (25)$$

Where $\bar{P}(z)$ and $\bar{\Gamma}(z)$ denote the complementary CDFs (i.e. $\bar{P}(z) \equiv 1 - P(z)$ and $\bar{\Gamma}(z) \equiv 1 - \Gamma(z)$). Notice that the growth rate is non-decreasing in z

$$g'(z) = 2\lambda(1 - u)P'(z) + \chi_e \lambda_0 \Gamma'(z) \geq 0 \quad (26)$$

Equipped with a solutions for the equilibrium employment growth function, the wage function $w(z)$ can be easily computed over a grid through value-function iteration and numerical integration (see Appendix C for a detailed description of the solution method).

Firm-level labor share. In order to map the model to the data, I construct firm-level measures of value-added and its components in the model by following the *National Accounts* definitions. The individual state variable for the firm is its TFP level z so the value-added components will be indexed by z . As established in Lemma 1, firm-level employment N only scales the firm but does not affect its policies (i.e. wage and capital intensity). In Table 4, I present the definition of the main GDP components and I normalize N to be one.

Table 4: National Accounts definitions in the model

Component	Value
Wages and salaries	$w(z)$
Net operating surplus	$zk(z)^\alpha - w(z) - \tau k(z)$
Consumption of fixed capital	$\tau k(z)$
Value added	$zk(z)^\alpha$
Net value-added	$zk(z)^\alpha - \tau k(z)$

To be consistent with the data, I will use net-value added to compute the firm-level labor share (share of wages and salaries in net value-added) and labor productivity (net value-added per worker). Figure 1a presents the wage policy and labor productivity as a function of TFP in the estimated model (I discuss the estimation procedure in Section 4.1). The range of TFP values in the graph represent the bottom 99.9% of firms operating in equilibrium (it excludes the top 0.1% most productive firms). First, notice that the fact that capital intensity $k(z)$ is increasing in TFP leads to an endogenous amplification of productivity differentials (the labor productivity curve has a slope above one). Second, notice that the wage schedule lies *above* labor productivity for the least productive firms. The productivity distribution is highly skewed so

that an important fraction of firms are in that range. Finally, notice that the gap between the wage and labor productivity increases with TFP. These features of the equilibrium policies lead to an labor share schedule which start above one and declines with which productivity (see Figure 1b).

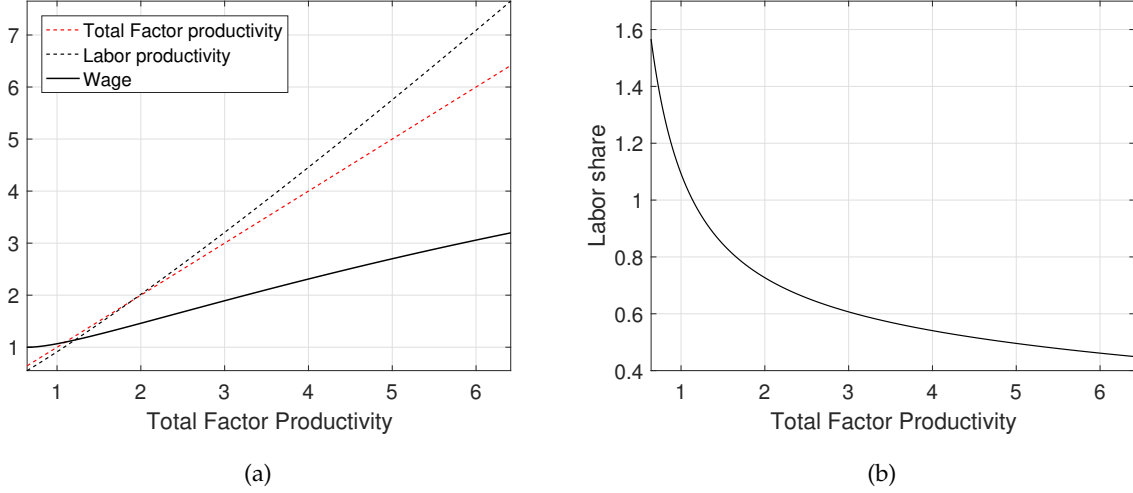


Figure 1: Wage and labor share (estimated model)

Wages at highly productive firms. Why do wages seem to disconnect from productivity in the upper tail of the distribution? Recall that more productive firms increase wages in order to grow faster. The first-order condition (Equation 17) implies that $w'(z) = v(z)g'(z)$ so that a decrease in the sensitivity of wages to productivity can only be explained by the fact that the present-value of an employee $v(z)$ grows slower than the growth effect $g'(z)$, as a function of z . Asymptotically,¹² it can be shown that the value of an employee grows at rate $z^{\frac{1}{1-\alpha}}$ while the growth effect grows at rate $z^{-(1+\eta)}$. We thus have the following asymptotic equivalence

$$w'(z) \sim cz^{-(1+\eta-\frac{1}{1-\alpha})} \quad (27)$$

for some $c > 0$. As firms become more productive, their wage policy becomes increasingly disconnected from productivity and in the limit we have that $w'(z) \rightarrow 0$ (since $\alpha \in (0, 1)$ and $\eta > 0$). The reason is that, for very productive firms, increasing the wage schedule does not increase employment growth as they are already the highest paying firms. Every time they meet a worker, they are able to hire her and their growth is only constrained by labor market frictions, something which is unrelated to their pay policy. In the language of monopsony theory, as the wage increases, the dynamic labor supply curve is increasingly inelastic so that the wage mark-down is increasingly high.

¹²Two functions f and g are said to be asymptotically equivalent if $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1$. We denote this relationship by $f \sim g$.

Employment concentration. The aggregate labor share is determined by (1) the labor share of a firm conditional on its productivity and (2) the equilibrium allocation of workers to firms of different productivity. Having characterized the former in the previous section, I now turn to the latter. An important feature of the model is that the stationary allocation of workers across productivity ranks is determined only by labor market frictions (i.e. $\lambda, \lambda_0, \delta, \chi_s, \chi_x, \chi_e$).

Proposition 3. *The equilibrium allocation of workers across firm productivity ranks is invariant to the underlying productivity distribution Γ .*

Proof. The mass of workers working at firm with productivity no greater than z (i.e. $P(z)$) depends on z only through its rank in the productivity distribution $\Gamma(z)$ $\square$

The intuition behind Proposition 3 is that (1) the decision of workers to move to a new firm depends only on the rank of the wage offered in the wage distribution and (2) the rank of a firm in the wage distribution is equal to its rank in the productivity distribution. Figure 2a presents the model economy's employment-productivity Lorenz curve in the estimated model. It shows that employment is concentrated in high productivity firms. For instance, the top 1% and the bottom 25% of firms (in terms of productivity) employ each roughly 10% of the workforce. The high employment share of highly productive firms reflects the fact that, on average, these firms are larger. Although the model does not impose a one-to-one relationship between productivity z and size N , high productivity firms pay higher wages and thus grow faster. And since productivity differentials are persistent, today's high productivity firms have on average enjoyed longer periods of high growth and tend to be larger. The average firm size of a firm with current productivity z can be obtained by comparing the employment-weighted productivity density $dP(z)/dz$ to the unweighted productivity density $d\Gamma(z)/dz$.

$$\mathbb{E}(N|z) = \frac{1 - u}{F} \frac{dP(z)}{d\Gamma(z)} \quad (28)$$

Figure 2b contains the average firm size which is roughly linear over the range presented. Firms at the 99.9th percentile are on average 37 larger than those at the 1st percentile. In the data, firm size is often a poor proxy for current productivity (see Haltiwanger et al. [2017] for empirical evidence) due to the fact that employment is slow-moving and current firm size is highly history-dependent. The fact that the conditional mean function for firm size is increasing in productivity hides the fact that the model exhibits extreme dispersion in firm size within productivity ranks.

Firm size distribution. One of the most celebrated empirical law in economics is *Zipf's law*, which states that the size distribution of units obeys a power law with tail coefficient just above one (see Gabaix [1999] and Toda [2016] for theories related to firm and city sizes). Across countries and time periods, the upper tail of the size distribution of firms is well approximated by Zipf's law, meaning that for large values of firm size N , we have that

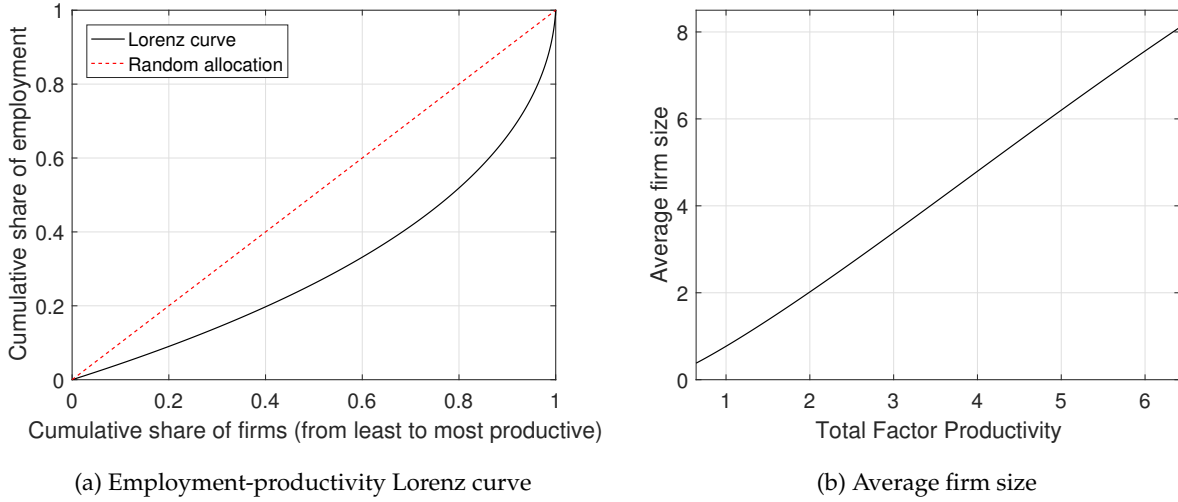


Figure 2: Firm size and productivity relationship (estimated model)

$$\mathbb{P}(N_i \geq N) = cN^{-\zeta} \text{ with } \zeta \approx 1$$

for some $c > 0$. The model developed in this paper contains two ingredients which are known to generate a power law: (1) *Gibrat's law*¹³ and (2) random death. Most of the theoretical literature studies processes where growth rates are i.i.d but in a recent contribution, [Beare and Toda \[2017\]](#) provide a theorem which can be used to compute the tail coefficient when the growth rates distribution depends on a state variable which follows a *Markov Chain*. The endogenous process for firm growth in this paper falls into the class of model they study and thus generates a Pareto upper tail for the firm size distribution. Does the model generate Zipf's law? The answer depends on the particular parameterization of the model (from Proposition 3, it cannot depend on Γ or α). When I estimate the model, I target a tail coefficient of 1.06 and I find that the model can easily accommodate this high level of employment concentration without compromising its ability to match other targeted and non-targeted moments. This is consistent with [Toda \[2016\]](#), who shows that in models with fixed labor supply, the average growth rate of a firm must be close to zero since a unit can only grow at the expense of another one shrinking. In his set-up, this in turn implies a tail coefficient close to one.

¹³A restricted form of Gibrat's law holds in the model because the growth rate is independent *conditional on productivity*.

4 Quantitative analysis

4.1 Estimation

The strategy will be to calibrate a subset of the parameters and the estimate the other ones using the *method of moments*. For interpretability, I first re-parametrize the model. Define $c = (\frac{z_l}{z_0})^{-\eta} > 0$ and notice that, conditional on (η, z_l) , we have that the transition parameters $(\chi_e, \chi_x, \chi_s)'$ defined in Equation 4 are equal to $(\chi_1 c, \chi_2(1 - c), \chi_2 c)'$. Which means that any $(\chi_e, \chi_x, \chi_s)' \in \mathbb{R}_+^3$ can be achieved by some values of $(z_0, \chi_1, \chi_2)' \in \mathbb{R}_+^3$. Since the parameters (z_0, χ_1, χ_2) have no obvious empirical counterpart, I will directly target $(\chi_e, \chi_x, \chi_s)'$ in the estimation process. I also normalize the outside option of the worker to $b = 1$. Finally, I normalize the mass of entrants χ_e to χ_x as it is not separately identified from the mass of initial meetings λ_0 . The reason is that the measure of active firms in steady-state $F = \frac{\chi_e}{\chi_x}$ plays no role in a model where there is a continuum of firms and no congestion in the labor market. By setting $\chi_e = \chi_x$, the mass of active firms is always equal to one.

Calibrated parameters. The net interest rate r is set to 2.5%, which is the difference between the effective business borrowing rate in Canada over the 2000-2015 of 4.4% as measured by the Bank of Canada¹⁴ and the average realized CPI inflation of 1.5%. To compute the user cost R of capital, as defined as in equation ??, I set the depreciation rate to 9% as in Barkai [2016] and expected inflation to 1.5% which was the average realized CPI inflation in Canada over that period. Then, the Cobb-Douglas exponent α is set to 0.142 as to imply a capital to output ratio of 1.06 as measured in the data (i.e. $\frac{K}{Y} = \frac{\alpha}{R}$ in the model). The firm exit rate χ_x is set to 2.0%, which is roughly the employment-weighted firm exit rate in Canada over the 2000-2008 period (measurement taken from Rollin [2012]). This value is much lower than the raw exit rate of 10% since there is a negative correlation between size and exit rate in the data while the model imposes a constant death rate. Finally, I target a ratio of 6.43 for the 99th to 50th percentiles of the labor productivity distribution. Table 7 summarizes these choices.

Estimated parameters. The remaining parameters $\theta = (\lambda, \lambda_0, \delta, \chi_s)'$ are jointly estimated using the *method of moment*. To ensure that Assumption 2 is satisfied, I restrict the parameters space to Θ

$$\Theta = \mathbb{R}_+^4 \cap \left\{ \lambda \leq \chi_s + \chi_x + \delta \right\} \quad (29)$$

¹⁴“The effective interest rate for businesses is a weighted-average borrowing rate for new lending to non-financial businesses, estimated as a function of bank and market interest rates. The weights are derived from business credit data. The business effective rate is a function of: short-term commercial paper and bankers’ acceptance rates, with terms of one and three months; the bank prime business lending rate, which is adjusted for movements in bank funding costs, so as to estimate the effective bank prime lending rate faced by new borrowers; and longer term borrowing rates, approximated using Merrill Lynch bond indices, which include both investment and non-investment grade companies (non-financial).”

Table 5: Calibrated parameters

Parameter	Value	Target	Source
r	0.025	Borrowing rate (4.4%) net of inflation (1.5%)	Bank of Canada
τ	0.09	Depreciation rate of 9%	Barkai [2016]
α	0.122	Capital output ratio of 1.06	NALMF
η	2.39	99 th / 50 th percentile ratio of labor prod. of 6.43	NALMF
χ_x	0.02	Employment-weighted firm exit rate of 1.6 %	Statistics Canada

To estimate the parameters, I target moments related to (1) worker and job flows (2) cross-sectional moments and (3) the aggregate labor share (see Table 6 for details). For each value of $\theta \in \Theta$, I solve the model (see Appendix C for a description of the solution method) and compute the model implied moments $\{m_i(\theta)\}_{i=1}^M$. The method of moment seeks to minimize the sum of squared differences between model-implied and data moments:

$$\Theta^* \in \arg \min_{\theta \in \Theta} \sum_{i=1}^M (m_i(\theta) - m_i)^2 \quad (30)$$

Although I cannot establish identification, I now provide intuition as to which parameters determine which moments. First, the mass of meetings per entrant λ_0 is the leading determinant of the job creation rate by entrant. Similarly, the meeting rate λ essentially scales the growth function g and thus increases job reallocation (job creation and job destruction). At the same time, high values for λ_0 and λ affect the unemployment rate by reducing the expected length of an unemployment spell. In addition, λ has a direct effect on the level of employment concentration. When the meeting rate is high, workers are reallocated at a faster rate towards the highest paying firms. Similarly, a higher value for the exogenous job destruction rate δ increases the unemployment rate. Finally, the arrival rate of productivity shocks is identified indirectly through its effect of on employment concentration. When χ_s is high, firm productivity changes frequently, which reduces the persistence of productivity differentials (and therefore wage differentials), which reduces the level of steady-state employment concentration.

Table 6: Moments

Moment	Data	Model
Unemployment rate	0.071	0.088
Job creation rate	0.110	0.127
– contribution of entrants	0.020	0.028
Firm size Pareto exponent	1.060	1.066

Table 7: Estimated parameters

λ	λ_0	δ	χ_s
0.153	3.47	0	0.143

4.2 Model validation

Non targeted moments. I now compare the model-implied joint distribution of firm-level variables (i.e. value-added Y , employment N , LS). I simulate the model (see Appendix C.2) and proceed to sort firm-level observations by value-added and bin them into 4 groups which contribute a quarter of total value-added (exactly as in Section 2). I then compute the conditional mean of employment and value-added within each group as well as the labor share within each group. For comparability, I normalize the average employment and output in group 1 in both the model and in the data.

Table 8: Contributions by groups sorted on value-added

Group	%	Y	N	LS
1	83	1.0	1.0	0.80
2	14	5.9	3.9	0.65
3	2.7	31	17.4	0.6
4	0.3	304	129.8	0.45

(a) Statistics by group (data)

Group	%	Y	N	LS
1	91	1.0	1.0	0.76
2	7	12.9	9.0	0.60
3	2.1	42	23	0.53
4	0.3	302	157	0.52

(b) Statistics by group (model)

Mechanism verification. The novel mechanism emphasized in this paper - *monopsony power* - implies that firms face a trade-off between wages and employment growth, where high-wage firms are able to attract workers at a higher pace and retain them more effectively. The model attributes all cross-sectional differences in wages to a firm-specific TFP component. In this section, I indirectly test the proposed mechanism. Although TFP is unobserved, I will use the fact that, through the lens of the model, log TFP and log labor productivity are collinear¹⁵ and use labor productivity to proxy for TFP.

First, I verify that the mechanism which generates wage dispersion is the present in the data. The hypothesis is that employment at high-paying firms should grow faster than at low-paying firms. I estimate the k -year ahead elasticity of employment growth to wages on a balanced panel of firms who were in operation

¹⁵Labor productivity is given by $LP \equiv zk^\alpha - \tau k$ and the first-order condition for capital implies that $k = (\frac{\alpha z}{R})^{\frac{1}{1-\alpha}}$ which means that labor productivity is equal to $LP = \left((\frac{\alpha}{R})^{\frac{\alpha}{1-\alpha}} - \tau (\frac{\alpha}{R})^{\frac{1}{1-\alpha}} \right) z^{\frac{1}{1-\alpha}}$

$$\log LP = c + \frac{1}{1-\alpha} \log z$$

throughout the 2000-2010 period. I estimate the following auxiliary model on annual data

$$\Delta^k \log N_{i,t} = \mu + \alpha_j + \lambda_t + \beta \log x_{i,t} + \varepsilon_{i,t} \quad \mathbb{E}_{i,t} \varepsilon_{i,t} = 0 \quad (31)$$

where α_j and λ_t are industry and year fixed effects. The independent variable $x_{i,t}$ is either log wages or log labor productivity while the dependent variable is the k^{th} forward difference of log employment (i.e. $\log N_{i,t+k} - \log N_{i,t}$). I present the result in Table where the standard errors are clustered at the firm level. I find that firms who pay 10% higher wages grow at an annual rate of 0.4 percentage point faster over the following year. Moreover, these effects are persistent. After five year, firm who were paying 10% higher wages grew at a 1.2 percentage point faster pace. Similarly, firms who had 10% high labor productivity grow at an annual rate 3.5% faster. These findings are consistent with worker reallocation towards high labor productivity, high wage firms. The results are highly significant which supports the prediction of the model yet differentials in wages and labor productivity explain only a small fraction of the total variance in growth rates (less than 5%). Firm growth distribution are known to exhibit very high kurtosis¹⁶ and the low R^2 implies that there are factors beyond wages and productivity the explain the long tails of the growth rate distribution.

In Section 2, I document size effects in the cross-section of firms: high value-added firms tend to have a lower labor share, pay higher wages and have more capital per worker (see Appendix A.2 for results in regression format). By construction, firm-level value added Y is the product of the number of employees N with the value-added per worker LP . In the model, the causal variable which determines firm-level outcome variables (LS, w, k) is total factor productivity. The hypothesis is that, conditional on TFP, employment size N should be irrelevant. Consider the following predictive model:

$$y_{i,t} = \mu + \alpha_j + \lambda_t + x'_{i,t} \beta + \varepsilon_{i,t} \quad \mathbb{E}_{i,t} \varepsilon_{i,t} = 0 \quad (32)$$

where α_j and λ_t are industry and year fixed effects. I estimate the model for each of the three independent variables $y_{i,t} \in \{\log LS_{i,t}, \log w_{i,t}, \log k_{i,t}\}$ with $x_{i,t}$ being either $\log N_{i,t}$ or $(\log N_{i,t}, \log LP_{i,t})'$ (see Table 13). For the labor share, the employment size elasticity is estimated to be -1.7% and the R^2 is 0.03. When I include labor productivity as a predictor, the effect of employment size vanishes (in fact it becomes positive) and the estimated elasticity of labor productivity is large at -36% . The in-sample predictive power of labor productivity is large: the R^2 increases to 0.40. Similarly, the predictive model for firm wages including only employment implies an employment elasticity of 5.9% (the well-known *firm size wage premium*) with a R^2 of 0.45. When we include labor productivity, the effect of employment becomes very small and the labor productivity elasticity is large at 64% and the R^2 increases to 81%. Finally, the result for capital intensity are similar yet less unequivocal. The employment elasticity reduces by half when we include labor productivity (8.7% versus 13.6%) and the R^2 increases from 0.13 to 0.19. Of course, labor productivity and the

¹⁶Multiple studies have documented a “tent-shaped” firm growth distributions. See [Dixon and Rollin](#) for Canadian data.

labor share are related by construction so this is not surprising but these findings support the correlation structure implied by the model. Labor productivity has a large effect on firm-level outcome variables and predicts a large fraction of the variation. In contrast, employment has some predictive power but does not provide much information beyond what is contained in labor productivity, consistently with the idea that employment is a noisy proxy for productivity.

5 Counterfactual analysis

Multiple studies document the existence of important differences in (revenue) productivity across firms, even within narrowly defined industries, and recent evidence points to be a trend towards increased productivity dispersion. For instance, [Berlingieri et al. \[2017b\]](#) document a widespread increase in productivity dispersion, measured as the 90/10 percentile ratio of the labor productivity distribution, across OECD countries over the 2001-2012 period. How do changes in the productivity distribution affect the labor share? In perfectly competitive markets with constant returns to scale technology, the distribution of total factor productivity across firms should be irrelevant for the labor share as, in equilibrium, wages reflect the marginal product of labor, which is equalized across firms. In this section, I use the estimated model to simulate the impact of an increase in the productivity dispersion across firms on the aggregate labor share.

5.1 The experiment

The experiment consists in decreasing the shape parameter η of the distribution of new TFP draws Γ_0 while changing the lower bound of the support z_0 as to maintain the equilibrium entry and exit rates χ_s, χ_x constant (see Equations 4). Recall that selection on entry and exit implies that the distribution of TFP for active firms is a truncated version of Γ_0 given by $\Gamma(z) = 1 - (\frac{z}{z_l})^{-\eta}$ over $z > z_l$. When η decreases, the distribution of TFP (and labor productivity) across active firms spreads out and firms at the top of the productivity distribution gain proportionally more than those at the bottom.

For the experiment, I consider a range of values for η around the estimated value of 2.39 in the Canadian data and compute productivity dispersion, wage dispersion, and the labor share. For interpretability, I will label firms below the median productivity levels as low productivity (L firms) while those above as high productivity (H firms). The labor productivity and wages at each productivity groups are defined the *employment-weighted* average¹⁷ in order to obtain exact aggregation. Denoting w_i, LP_i, μ_i the wage, labor productivity and output share of each firm type $i \in \{L, H\}$, we have that the aggregate labor share admits the following identity.

$$LS = \mu_L LS_L + \mu_H LS_H \quad (33)$$

¹⁷For instance, the employment-weighted labor productivity of firms in a range $z \in A$ is defined as $LP_A = \frac{\int_A zk(z)^k P(dz)}{\int_A P(dz)}$

I define productivity dispersion as the the differential (in log points) between the labor productivity at H and L firms. For the sake of the experiment, consider a counterfactual economy which is identical to the Canadian economy except that productivity dispersion is 20% higher (log difference of 0.2). Figure 3a presents the equilibrium relationship between the aggregate labor share and productivity dispersion. The negative relationship is strong, with a counterfactual labor share of 0.60 versus 0.63 in the Canadian economy. The intuition for the main mechanism behind the decline of the labor share can be understood from 3b. While productivity differentials between high and low productivity firms (H and L firms from now on) increase by 20%, wage differentials adjust by only 3% which leads to an increasingly weak relationship between productivity and wages. As a result, gains in productivity at H -firms only weakly translate into higher wages, leading to a lower labor share at H -firms firms.

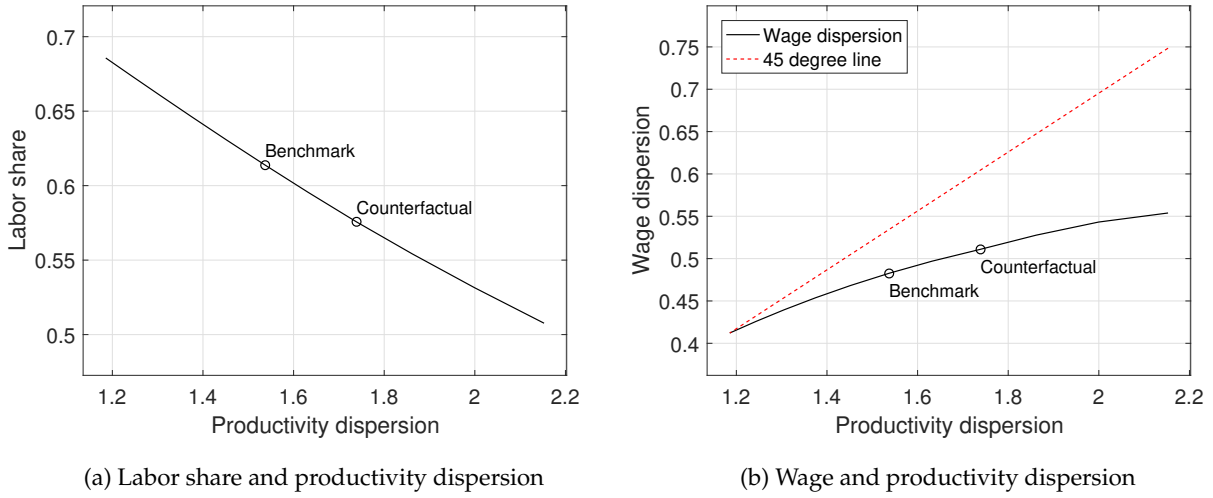


Figure 3: Productivity dispersion, wage dispersion and the labor share (counterfactual in estimated model)

The full mechanism is more complicated and involves two channels (1) a change in the labor share *within* firms and (2) a reallocation of value-added *between* firms. Denoting by Δx the counterfactual change of variable x , we obtain the typical *Olley-Pakes* decomposition.

$$\Delta LS = \underbrace{\mu_L \Delta LS_L + \mu_H \Delta LS_H}_{\text{within component}} + \underbrace{\Delta \mu_L LS'_L + \Delta \mu_H LS'_H}_{\text{between component}} \quad (34)$$

Within component. While the labor share at H -firms decreases from 54% to 51%, the labor share at L -firms increases from 1.56% to 1.73% (see Table 9). The reason why L -firms become increasingly unprofitable is that the productivity threshold z_l decreases in the counterfactual economy. Firms now tolerate lower levels of productivity without exiting in the hope of later collecting the increasingly high profit margins associated with being at the top of the productivity distribution. At the other end of the spectrum, H -firms compete for workers with firms which are significantly less productive than them which allows them to pay wages

largely below the marginal product of labor. The distribution of labor shares thus exhibit *polarization*, with more very low and more very high labor shares. Overall, the decline in the labor share at *H*-firms dominate due to their output share of 0.93 and the within component accounts for a 2.1 pp. decline in the aggregate labor share.

Between component. As established in Proposition 3, the employment shares across productivity ranks are invariant to changes in the productivity distribution. Table 9 confirms the theoretical result and shows that, in the counterfactual economy, *H* firms increase their share of output by 1.4 pp. through productivity increases only. The reallocation of value-added towards *H*-firms mechanically decreases the aggregate labor share by 1.4 pp.

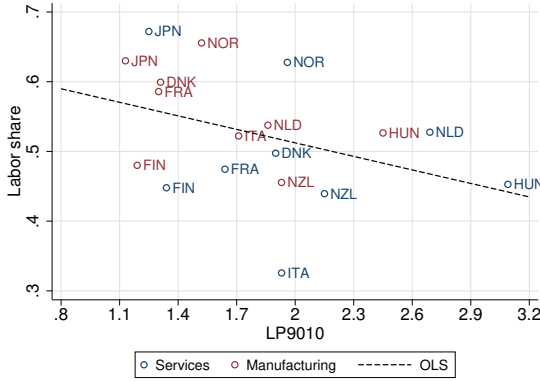
Table 9: Model experiment (percentage point unless stated otherwise)

Variable	Benchmark	Counter-factual	Diff.
Labor share	61.4	57.6	-3.8
within component			-2.1
between component			-1.7
Productivity dispersion (log ratio)	1.54	1.74	+0.2
Wage dispersion (log ratio)	0.48	0.51	+0.03
<i>H</i> -firm output share	92.9	94.3	+1.4
<i>H</i> -firm employment share	74.0	74.0	0
Unweighted labor share	130.0	141.6	+11.6
<i>L</i> -firm labor share	155.6	172.7	+17.1
<i>H</i> -firm labor share	54.2	50.6	-3.6

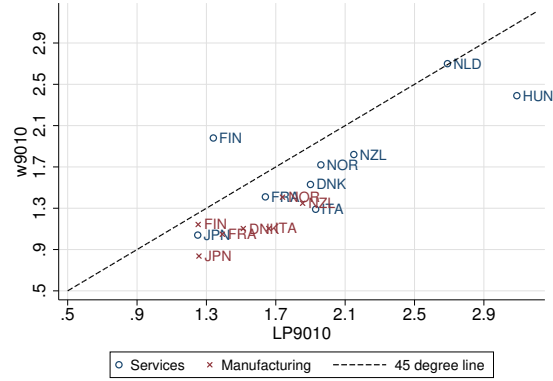
5.2 Cross-country evidence

The measurement of the productivity dispersion requires the use of firm-level data on value-added and employment. The *OECD* has developed a project which seeks to provide harmonized cross-country microdata on productivity and wage dispersion (see [Berlingieri et al. \[2017a\]](#) for a description of the *Multiprod* project). The data they collect are obtained by running a standardized routine on individual country's production surveys and business registers. The variables which I will use, labor productivity dispersion *LP9010* and wage dispersion *w9010*, are taken from [Berlingieri et al. \[2017b\]](#) and are constructed by computing the log difference between the 90th and 10th percentiles of the labor productivity and wage distribution across firms. The measures are computed at the two-digit industry level and then averaged with employment weights to provide, for each country-year, a measure for the manufacturing and non-financial services sec-

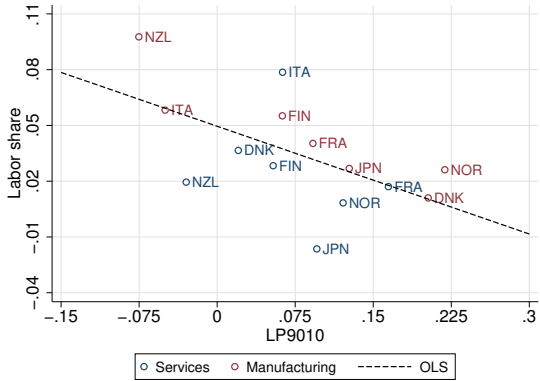
tors.¹⁸ I then merge the productivity and wage dispersion data with the *Penn World tables 9.0* (see [Feenstra et al. \[2015\]](#)) which contains a country-year measure of the aggregate labor share (i.e. the share of labor income of employees and self-employed workers in GDP).¹⁹ Due to the limited coverage of the *Multiprod* database, I obtain a balanced panel of 7 countries for the period 2001-2011.



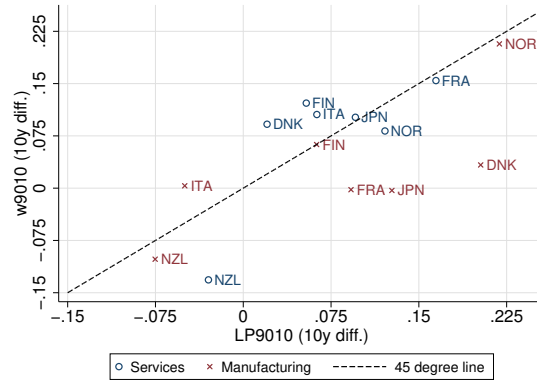
(a) Relationship in level (2001)



(b) Relationship in level (2001)



(c) Relationship in changes (2001-2011)



(d) Relationship in changes (2001-2011)

Figure 4: Cross-country relationship between (a,c) the labor share and productivity dispersion and (b,d) productivity and wage dispersion. All figures contain country-industry observations.

Figure 4a plots the cross-sectional relationship between productivity dispersion and the labor share

¹⁸I aggregate the manufacturing and services indices using 20% and 80% weights to account for the fact that the service sector employs roughly 80% of the workforce. The results are robust to the choice of weights as both measures have a correlation of above 0.9.

¹⁹As a robustness check, I use the cross-country dataset compiled by [Karabarbounis and Neiman \[2014\]](#) which includes a measure of the corporate labor share. The benefit is that, in the corporate sector, there is not ambiguity as to what constitute a payment to labor. In opposition, in the unincorporated sector, one has to make assumptions in order to divide *mixed income* into payments to labor versus payments to capital. All results are robust to the use of the corporate labor share, but at the cost of being available for a more limited set of countries.

across countries at the beginning of the sample in 2001. There is a clear negative relationship where countries with higher productivity dispersion tend to have a lower labor share. Moreover, as predicted by the model, wage dispersion is lower than productivity dispersion, a feature of the data which is evident from Figure 4b which exploits variation across countries and industries (services and manufacturing) in 2001. The same pattern holds when we consider changes over time. Figure 4c plots the relationship between productivity dispersion and the labor share over 10 year differences (2001-2011) and . Again, the relationship is negative, meaning that countries who experienced larger increases/decreases in productivity dispersion also saw their labor share decrease/increase the most. Figure 4d highlights the fact the data supports the model's mechanism: wage dispersion comoves positively with productivity dispersion, but less strongly (the data points tend to lie *below* the 45 degree line).

I now formalize the analysis and exploit both variations across countries and over time in the balanced panel in a regression framework. I estimate five specifications: (1) in levels, (2) in levels with two lags of the dependent and independent variable, (3) in levels with year and country fixed-effects, (4) in short differences (2001-2006 and 2006-2011) with year fixed-effects, (5) in short differences with year and country fixed-effects, and (6) in long differences (2001-2011). Standard errors are clustered at the country level and the results are summarized in Table 14. Despite the small sample, the estimated elasticity is statistically significant at the 5% level (except for specification 5) and quite stable across specifications. The range of point estimates is roughly from -0.1 to -0.3, meaning that when the differential in labor productivity between firms at the 90th and 10th percentiles increase by 10%, the labor share decreases is predicted to decrease by 1 to 3 percentage points. Those magnitudes are consistent with the model which, in the experiment, was predicting a 3 pp. response of the labor share to a 20 pp. increase in productivity dispersion²⁰, which implies a local elasticity of -0.15.

I also estimate the relationship between wage dispersion and productivity dispersion at the industry level using the same 6 specifications as for Table 15 with the difference that each of them includes industry fixed-effects (manufacturing and services). Consistently with the mechanism in the model, the elasticity between wage and productivity dispersion is positive and below one. Across specifications, the point estimate ranges roughly from 0.1 to 0.7 with large confidence intervals. The magnitudes estimated are somehow larger than those implied by the model's experiment with an implied local elasticity of 0.15.

²⁰Note that the measure of productivity dispersion in the model is slightly different than a 90/10 labor productivity ratio. But in the experiment, the implied change in the log 90/10 ratio is 0.21 which is nearly identical to the model's log change in productivity dispersion of 0.20.

6 Concluding remarks

In this paper I present a theory of the labor share in which workers extract surplus through their ability to quit to higher paying firms. In equilibrium, firm-level wages are positively correlated with the marginal product of labor but are much less dispersed than labor productivity. I find that, in economies with high productivity dispersion across firms, the labor share is lower as a result of high profit margins at the top of the firm productivity distribution. Counterfactual analysis in the estimated model highlights the particular transition path from low to high productivity dispersion which involves the *polarization* of labor shares across firms and an increase in output concentration. The model's predictions have important consequence for growth theory as it highlights the fact that labor market imperfections can weaken the pass-through of productivity into wages when productivity gains are achieved only at the top.

References

- Yves Achdou, Francisco Buera, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. Pde models in macroeconomics. *Proceedings of the Royal Society of London. Series A, Mathematical and physical sciences*, 2014.
- David Autor, David Dorn, Lawrence F Katz, Christina Patterson, John Van Reenen, et al. The fall of the labor share and the rise of superstar firms. Technical report, Centre for Economic Performance, LSE, 2017.
- Jesper Bagger and Rasmus Lentz. An empirical model of wage dispersion with sorting. Technical report, National Bureau of Economic Research, 2014.
- Jesper Bagger, Bent Jesper Christensen, and Dale T Mortensen. Productivity and wage dispersion: Heterogeneity or misallocation. Technical report, Working paper, 2014.
- Simcha Barkai. Declining labor and capital shares. 2016.
- Erling Barth, Alex Bryson, James C Davis, and Richard Freeman. It’s where you work: Increases in earnings dispersion across establishments and individuals in the us. Technical report, National Bureau of Economic Research, 2014.
- Brendan K Beare and Alexis Akira Toda. Geometrically stopped markovian random growth processes and pareto tails. *arXiv preprint arXiv:1712.01431*, 2017.
- Giuseppe Berlingieri, Patrick Blanchenay, Sara Calligaris, and Chiara Criscuolo. The multiprod project. 2017a.
- Giuseppe Berlingieri, Patrick Blanchenay, Chiara Criscuolo, et al. The great divergence (s). Technical report, OECD Publishing, 2017b.
- Giuseppe Berlingieri, Sara Calligaris, and Chiara Criscuolo. Who profits from patents? rent-sharing at innovative firms. Technical report, AER P&P, 2018.
- Christian Bontemps, Jean-Marc Robin, and Gerard J Van den Berg. Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review*, 41(2):305–358, 2000.
- Katarína Borovicková. Job flows, worker flows and labor market policies. Technical report, Technical report, 2016.
- Benjamin Bridgman. Is labor’s loss capital’s gain? gross versus net labor shares. *Macroeconomic Dynamics*, pages 1–18, 2017.

- Kenneth Burdett and Dale T Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273, 1998.
- Kenneth Burdett and Tara Vishwanath. Balanced matching and labor market equilibrium. *Journal of Political Economy*, 96(5):1048–1065, 1988.
- Pierre Cahuc, Fabien Postel-Vinay, and Jean-Marc Robin. Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, 74(2):323–364, 2006.
- David Card, Ana Rute Cardoso, Jörg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70, 2018.
- Melvyn G Coles and Dale T Mortensen. Equilibrium labor turnover, firm growth, and unemployment. *Econometrica*, 84(1):347–363, 2016.
- Mai Dao, Mitali Das, Zsoka Koczan, and Weicheng Lian. Why is labor receiving a smaller share of global income? theory and empirical evidence. Technical report, IMF Working Paper, 2017.
- Jan De Loecker and Jan Eeckhout. The rise of market power. 2017.
- Jay Dixon and Anne-Marie Rollin. The distribution of employment growth rates in canada: The role of high-growth and rapidly shrinking firms.
- Timothy Dunne, Lucia Foster, John Haltiwanger, and Kenneth R Troske. Wage and productivity dispersion in united states manufacturing: The role of computer investment. *Journal of Labor Economics*, 22(2):397–429, 2004.
- Chris Edmond, Virgiliu Midrigan, and Daniel Yi Xu. Competition, markups, and the gains from international trade. *The American Economic Review*, 105(10):3183–3221, 2015.
- Andrea L Eisfeldt and Adriano A Rampini. Leasing, ability to repossess, and debt capacity. *The Review of Financial Studies*, 22(4):1621–1657, 2008.
- Michael WL Elsby, Bart Hobijn, and Ayşegül Şahin. The decline of the us labor share. *Brookings Papers on Economic Activity*, 2013(2):1–63, 2013.
- Robert C Feenstra, Robert Inklaar, and Marcel P Timmer. The next generation of the penn world table. *The American Economic Review*, 105(10):3150–3182, 2015.
- Xavier Gabaix. Zipf’s law for cities: an explanation. *The Quarterly journal of economics*, 114(3):739–767, 1999.
- Douglas Gollin. Getting income shares right. *Journal of political Economy*, 110(2):458–474, 2002.
- John Haltiwanger, Ron S Jarmin, Robert B Kulick, and Javier Miranda. High growth young firms: Contribution to job, output and productivity growth. 2016.

- John Haltiwanger, Henry Hyatt, Lisa B Kahn, and Erika McEntarfer. Cyclical job ladders by firm size and firm wage. 2017.
- Barney Hartman-Glaser, Hanno N Lustig, and Mindy X Zhang. Capital share dynamics when firms insure managers. 2017.
- Hugo A Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1127–1150, 1992.
- Leo Kaas and Philipp Kircher. Efficient firm dynamics in a frictional labor market. *The American Economic Review*, 105(10):3030–3060, 2015.
- Loukas Karabarbounis and Brent Neiman. The global decline of the labor share. *The Quarterly Journal of Economics*, 129(1):61–103, 2014.
- Matthias Kehrig. The cyclical nature of the productivity distribution. *Working Paper*, 2015.
- Matthias Kehrig and Nicolas Vincent. Growing productivity without growing wages: The micro-level anatomy of the aggregate labor share decline. 2017.
- Dongya Koh, Raül Santaaulàlia-Llopis, and Yu Zheng. Labor share decline and intellectual property products capital. 2016.
- Jeremy Lise and Jean-Marc Robin. The macrodynamics of sorting between workers and firms. *The American Economic Review*, 107(4):1104–1135, 2017.
- Erzo GJ Luttmer. Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3):1103–1144, 2007.
- Erzo GJ Luttmer. On the mechanics of firm growth. *The Review of Economic Studies*, 78(3):1042–1068, 2011.
- Costas Meghir, Renata Narita, and Jean-Marc Robin. Wages and informality in developing countries. *American Economic Review*, 105(4):1509–46, 2015.
- Merton H Miller and Charles W Upton. Leasing, buying, and the cost of capital services. *The Journal of Finance*, 31(3):761–786, 1976.
- Giuseppe Moscarini and Fabien Postel-Vinay. Wage posting and business cycles: A quantitative exploration. *Review of Economic Dynamics*, 19:135–160, 2016.
- Christian Moser, Niklas Engbom, et al. Earnings inequality and the minimum wage: Evidence from brazil. In *2016 Meeting Papers*, number 72. Society for Economic Dynamics, 2016.
- Walter Y Oi and Todd L Idson. Firm size and wages. *Handbook of labor economics*, 3:2165–2214, 1999.

- Thomas Piketty and Gabriel Zucman. Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics*, 129(3):1255–1310, 2014.
- Fabien Postel-Vinay and Jean-Marc Robin. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6):2295–2350, 2002.
- Matthew Rognlie. Deciphering the fall and rise in the net capital share: accumulation or scarcity? *Brookings papers on economic activity*, 2015(1):1–69, 2016.
- Anne-Marie Rollin. Measures of employment turnover post 2000: Gross employment gains and losses versus net employment change. 2012.
- Jae Song, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. Firming up inequality. Technical report, National Bureau of Economic Research, 2015.
- Chad Syverson. What determines productivity? *Journal of Economic literature*, 49(2):326–365, 2011.
- Alexis Akira Toda. Zipf’s law: A microfoundation. 2016.
- Randall Wright, Philipp Kircher, Benoit Julien, and Veronica Guerrieri. Directed search: A guided tour. Technical report, National Bureau of Economic Research, 2017.

A Appendix - Data

A.1 Variable construction

I now describe the methodology used to construct the main firm-level variables: value-added, employment, wage and capital stock.

Value-added. Denote the yearly value-added of firm i at time t as $Y_{i,t}$. To compute value-added, I sum Net income (i.e. revenues minus expenses) and Payroll (i.e. the sum of wages and salaries paid) which are found on the income statement. Income includes revenues from the sale of goods and services as well as investment revenues and realized gains/losses on the disposal of asset while expenses include production costs, changes in inventory, wages and salaries, as well as interests paid and depreciation expenses. The annual payroll at a firm during a given tax year is obtained by summing the employment income from the T4 filed during that year.

$$Y_{i,t} \equiv \text{payroll}_{i,t} + \text{net income}_{i,t} \quad (35)$$

The payroll and net income variable corresponds, respectively, to the *wages and salaries* and *net operating surplus of corporations* components of the *National Accounts*.

Employment, wages and capital stock. Employment $N_{i,t}$ is obtained by averaging the number of employees declared on the monthly PD7 forms filed throughout the year. This procedure avoids attributing larger employment levels to firms with high turnover. If for instance, we had summed the number of yearly T4 forms filed, we would be imputing a large workforce to firms with a high labor turnover (i.e. a high share of employees who only during part of the year). The average wage at the firm i throughout year t is the ratio of payroll to employment.

$$N_{i,t} = \text{employment}_{i,t} \quad (36)$$

$$w_{i,t} = \frac{\text{payroll}_{i,t}}{\text{employment}_{i,t}} \quad (37)$$

The book value of the capital stock $K_{i,t}$ is obtained from the balance sheet by summing the value of tangible capital assets (i.e. land, buildings, machinery and equipment, etc.) net of accumulated amortization. Capital intensity $k_{i,t}$ is defined as the stock of capital per worker.

$$k_{i,t} = \frac{\text{tangible capital assets}_{i,t}}{\text{employment}_{i,t}} \quad (38)$$

Labor share. The labor share is defined as the payroll share of valued-added or alternatively as the ratio of average wage to labor productivity:

$$LS_{i,t} = \frac{w_{i,t}N_{i,t}}{Y_{i,t}} = \frac{w_{i,t}}{LP_{i,t}} \quad (39)$$

where labor productivity $LP_{i,t} = \frac{Y_{i,t}}{N_{i,t}}$ is defined as the value-added per worker.

A.2 Size effects

Table 10: Size relationships (NALMF, 2000-2015)

	$\log LS$	$\log LS$	$\log w$	$\log w$	$\log k$	$\log k$
$\log N$	-0.017*** (0.001)		0.059*** (0.002)		0.136*** (0.006)	
$\log Y$		-0.133*** (0.001)		0.254*** (0.002)		0.304*** (0.004)
N	871,246	871,246	871,246	871,246	871,246	871,246
R^2	0.03	0.18	0.45	0.62	0.13	0.17

Standard errors clustered at the firm level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

A.3 Firm size wage premium and sorting on observables

I now provide evidence that the wage premium is not merely the result of sorting on observable characteristics. Using the microdata underlying the *Labor Force Survey* over the 2000-2015, I apply the same restrictions which were applied on the NALMF dataset and conduct a series of cross-sectional regressions controlling for worker and firm characteristics. The survey is a repeated cross-section, which prevents the addition of a worker fixed effect to account for unobserved heterogeneity. Nevertheless, the richness of the dataset allows me to assess the importance of multiple factors on the firm-size wage premium. I compute the log hourly wage $\log w_{i,t}$ and estimate the following equation

$$\log w_{j,t} = \mu + x'_{i,t}\beta + \varepsilon_{i,t} \quad (40)$$

For each specification, $x_{i,t}$ contains an indicator for the size of the firm at which the employee works and year fixed-effects. There are three categories: $20 \leq N < 100$, $100 \leq N < 500$ and $N > 500$. I omit

the first group so that the results can be interpreted as a wage premium in difference (log points) from the $20 \leq N < 100$ group. The first set of controls is the industry of the firm (two digit NAICS), the second includes worker characteristics (age, location, unionized status as well as detailed education and occupation) while the last one includes the tenure of the worker at the firm. I include the tenure variable in the last specification but the interpretation is complicated by the fact that, in job-ladder models like the one developed in this paper, average tenure at the firm is positively correlated with firm wage premium. Table 11 contains the results for different combinations of the control variables.

Table 11: Firm-size wage premium (LFS, 2000-2015)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$100 \leq N < 500$	0.072*** (0.001)	0.052*** (0.001)	0.090*** (0.008)	0.028*** (0.001)	0.028*** (0.001)	0.041*** (0.007)	0.036*** (0.007)
$N \geq 500$	0.078*** (0.001)	0.122*** (0.001)	0.207*** (0.007)	0.068*** (0.001)	0.078*** (0.001)	0.104*** (0.005)	0.085*** (0.005)
Worker charact.				✓	✓	✓	✓
Industry		✓	✓		✓	✓	✓
Industry $\times$ size			✓			✓	✓
Tenure							✓
N	4,781,953	4,781,953	4,781,953	4,684,057	4,684,057	4,684,057	4,684,057
R^2	0.057	0.312	0.327	0.561	0.584	0.589	0.608

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

As can be seen from Table 11, the firm size wage premium is statistically significant and economically large in all specifications. In my preferred specification (column 6), where I control for worker fixed-effects and interact the size variable with firm industry dummies, the wage premium at the largest firms ($N > 500$) is on average 10% larger than at firms with less than 20 employees while the premium is 4% at the firm with $100 \leq N < 500$.

A.4 Mechanism verification regressions

Tables 12 and 13 present the mechanism verification regression results.

Table 12: Employment growth, wages and labor productivity (NALMF, 2000-2015)

	$k = 1$	$k = 2$	$k = 5$	$k = 1$	$k = 2$	$k = 5$
$\log w$	0.040*** (0.001)	0.070*** (0.002)	0.120*** (0.003)			
$\log LP$				0.035*** (0.001)	0.061*** (0.002)	0.106*** (0.003)
N	332,103	332,103	332,103	332,103	332,103	332,103
R^2	0.03	0.04	0.05	0.03	0.04	0.04

Standard errors clustered at the firm level in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 13: Size relationships controlling for labor productivity (NALMF, 2000-2015)

	$\log LS$	$\log LS$	$\log w$	$\log w$	$\log k$	$\log k$
$\log N$	-0.017*** (0.001)	0.010*** (0.001)	0.059*** (0.002)	0.010*** (0.001)	0.136*** (0.001)	0.087*** (0.006)
$\log LP$		-0.360*** (0.002)		0.640*** (0.003)		0.647*** (0.009)
N	871,246	871,246	871,246	871,246	871,246	871,246
R^2	0.03	0.40	0.45	0.81	0.13	0.19

Standard errors clustered at the firm level in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.5 Cross-country regressions

Tables 14 and 14 present the cross-country regression results.

B Appendix - Proofs and derivations

B.1 Derivation of stationary unemployment rate

The law of motion is given by

Table 14: Productivity dispersion and the labor share: cross-country panel regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	Labor share	Labor share	Labor share	Labor share (5y change)	Labor share (5y change)	Labor share (10y change)
LP9010	-0.084** (0.033)	-0.127** (0.052)	-0.177*** (0.046)			
LP9010 (5y change)				-0.271** (0.078)	-0.297 (0.184)	
LP9010 (10y change)						-0.235** (0.082)
Year FE			✓	✓	✓	
Country FE			✓		✓	
Lags		✓				
N	81	65	81	14	14	7
R ²	0.260	0.926	0.941	0.634	0.796	0.363

Standard errors clustered at the country level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 15: Productivity and wage dispersion: cross-country panel regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	w9010	w9010	w9010	w9010 (5y change)	w9010 (5y change)	w9010 (10y change)
LP9010	0.288 (0.359)	0.420*** (0.066)	0.095 (0.555)			
LP9010 (5y change)				0.501** (0.161)	0.310** (0.104)	
LP9010 (10y change)						0.717** (0.281)
Industry FE	✓	✓	✓	✓	✓	✓
Year FE			✓	✓	✓	
Country FE			✓		✓	
Lags		✓				
Observations	162	130	162	28	28	14
R ²	0.550	0.991	0.860	0.385	0.637	0.534

Standard errors clustered at the country level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

$$\dot{u} = (1 - u)(\delta + \chi_x) - u\lambda(1 - u) - u\lambda_0\chi_e$$

In steady-state

$$0 = (\delta + \chi_x) - u(\delta + \chi_x + \lambda_0\chi_e + \lambda) + \lambda u^2$$

Which is a quadratic equation with coefficients

$$a = \lambda, \quad b = -(\delta + \chi_x + \lambda_0\chi_e + \lambda), \quad c = \delta + \chi_x$$

and discriminant

$$\Delta = (\delta + \chi_x + \lambda_0\chi_e + \lambda)^2 - 4\lambda(\delta + \chi_x)$$

Notice that the following inequality

$$\Delta = (\delta + \chi_x + \lambda_0\chi_e + \lambda)^2 - 4\lambda(\delta + \chi_x) \geq (\delta + \chi_x + \lambda)^2 - 4\lambda(\delta + \chi_x) = (\delta + \chi_x - \lambda)^2 \quad (41)$$

implies that $\Delta \geq 0$ and so that we have two candidate solutions

$$u^{(1)} = \frac{\delta + \chi_x + \lambda_0\chi_e + \lambda - \sqrt{\Delta}}{2\lambda}$$

$$u^{(2)} = \frac{\delta + \chi_x + \lambda_0\chi_e + \lambda + \sqrt{\Delta}}{2\lambda}$$

We can discard the solution $u^{(2)}$ since it always implies an unemployment rate above one:

$$u^{(2)} \geq 1 \iff \delta + \chi_x + \lambda_0\chi_e - \lambda + \sqrt{\Delta} \geq 0$$

But using the inequality for Δ (Equation 41)

$$\delta + \chi_x + \lambda_0\chi_e - \lambda + \sqrt{\Delta} \geq \delta + \chi_x + \lambda_0\chi_e - \lambda + |\delta + \chi_x - \lambda|$$

$$= \begin{cases} \lambda_0\chi_e + 2(\delta + \chi_x - \lambda) & \text{if } \delta + \chi_x - \lambda > 0 \\ \lambda_0\chi_e & \text{if } \delta + \chi_x - \lambda \leq 0 \end{cases}$$

which is positive in both cases. If a solution exist must then be $u^{(1)}$. One can show that $u \geq 0$ since

$$u \geq 0 \iff -b - \sqrt{\Delta} \geq 0 \iff -b \geq \sqrt{\Delta}$$

But since $-b \geq 0$ and $\Delta \geq 0$, this is equivalent to

$$b^2 \geq b^2 - 4ac$$

which is true since $ac \geq 0$. Now, I show that $u \leq 1$

$$u \leq 1 \iff -b - \sqrt{\Delta} \leq 2a \iff -b \leq 2a + \sqrt{\Delta}$$

Both sides are non-negative so this is equivalent to

$$b^2 \leq 4a^2 + b^2 - 4ac + 4a\sqrt{\Delta} \iff c - a \leq \sqrt{\Delta}$$

But using the inequality derived earlier (Equation 41), we have that $\sqrt{\Delta} \geq |c - a|$ which concludes the proof.

B.2 Proof of Lemma 1

$$v(z_0, N_0) = \max_{\tau, \{w_s, K_s\}_{s=0}^\tau} \mathbb{E}_0 \int_0^\tau e^{-rs} \left(x_s z_s K_s^\alpha N_s^{1-\alpha} - w_s N_s - r K_s \right) ds \quad (42)$$

s.t.

$$\dot{N}(w_t) = \tilde{g}(w_t) N_t \quad (43)$$

The variational inequality of the obstacle type²¹ representation of the *Hamilton-Jacobi-Bellman* equation for the value function is given by

$$\min \left\{ rv(z, N) - \max_{w \geq b, K} \left\{ z K^\alpha N^{1-\alpha} - w N - RK + \frac{\partial}{\partial N} v(z, N) \tilde{g}(w) N \right\} \right. \\ \left. - \chi_2 \left(\int v(\xi, N) \Gamma(d\xi) - v(z, N) \right), v(z, N) \right\} = 0 \quad (44)$$

First, I show that the value function is homogeneous in N . Divide equations 42 and 43 by N_0 and define $N_s = \frac{\tilde{N}_s}{N_0}$ and $\tilde{K}_s = \frac{K_s}{N_0}$:

$$\frac{v(z_0, N_0)}{N_0} = \max_{\tau, \{w_s, K_s\}_{s=0}^\tau} \mathbb{E}_0 \int_0^\tau e^{-rs} \left(z_s \tilde{N}_s^\alpha \tilde{N}_s^{1-\alpha} - w_s \tilde{N}_s - r \tilde{K}_s \right) ds$$

$$\dot{\tilde{N}}(w_t) = \tilde{g}(w_t) \tilde{N}_t$$

But since choosing K_s or $\tilde{K}_s$ yields the same outcome, we have that $\frac{v(z_0, N_0)}{N_0} = v(z_0, 1)$ which concludes the proof. I now show that the exit decision is to exit whenever $z < z_l$.

²¹See Achdou et al. [2014] for an economic application of the variational inequality of the obstacle type representation of an optimal stopping time model.

B.3 Second order condition

We need to verify that $v(z)\tilde{g}''(w(z)) < 0$ for all $z > z_l$ but since $v(z) > v(z_l)$ in that range, we only need to verify that $\tilde{g}''(w(z)) < 0$.

$$g'(z) = \tilde{g}'(w(z))w'(z) \implies g''(z) = \tilde{g}''(w(z))(w'(z))^2 + \tilde{g}'(w(z))w''(z)$$

And from the first-order condition, we have that $w'(z) = g'(z)v(z) \implies w''(z) = g''(z)v(z) + g'(z)v'(z)$.

Putting together

$$\begin{aligned} g''(z) &= \tilde{g}''(w(z))(w'(z))^2 + \tilde{g}'(z) \left(g''(z)v(z) + g'(z)v'(z) \right) \\ \implies g''(z) \left(1 - \tilde{g}(w(z))v(z) \right) &= \tilde{g}''(w(z))(w'(z))^2 + \tilde{g}'(w(z))g'(z)v'(z) \end{aligned}$$

But from the first-order condition, we have that $1 - \tilde{g}(w(z))v(z) = 0$ so that

$$\tilde{g}''(w(z)) = -\frac{\tilde{g}'(w(z))g'(z)v'(z)}{(w'(z))^2} = -\frac{(g'(z))^2v'(z)}{w'(z)} < 0$$

B.4 Derivation of the employment-weighted productivity distribution

Denote the employment-weighted productivity density function by $p(z)$ (i.e. the measure of workers working at firms with productivity $z \in A \subset [z_l, +\infty)$ is given by $(1-u) \int_A p(\xi) d\xi$). The labor productivity of an employed worker who remains employed can change for two reasons: (1) the worker receives a job offer from a more productive firms and (2) the productivity at his employer changes. The productivity process of an employed worker who remains employed is thus given by

$$dz = \underbrace{\int dJ_1(\xi) \left(\max\{z, \xi\} - z \right) d\xi}_{\text{job offers}} + \underbrace{\int dJ_2(\xi) \left(\xi - z \right) d\xi}_{\text{productivity shocks}}$$

where $\{J_1(\xi), J_2(\xi)\}_{\xi \geq 0}$ are Poisson processes with intensities which depend on ξ :

$$\begin{aligned} J_1(\xi) &\sim \mathcal{P} \left((1-u)\lambda p(\xi) + \chi_e \lambda_0 \Gamma(\xi) \right) \\ J_2(\xi) &\sim \mathcal{P} \left(\chi_s \Gamma(\xi) \right) \end{aligned}$$

At rate $(1-u)\lambda p(\xi)$, an employed worker receives a job offer from an existing firm with productivity ξ and at rate $\chi_e \lambda_0 \Gamma(\xi)$ from an entering firm with productivity ξ . At rate $\chi_s \Gamma(\xi)$, the productivity of the employer is reset to ξ . The CDF of $\max\{z, \xi\}$ where $z \sim P$ is a random variable and ξ a scalar is given by $\mathbb{P}(\max\{z, \xi\} \leq x) \equiv P(x)1_{\xi \leq x}$. Using the formula for the *Kolmogorov Forward Equation* (KFE) for jump-processes, the employment-weighted productivity density $p(z)$ obeys the following KFE:

$$\begin{aligned} \dot{p}(z)(1-u) &= (1-u)^2\lambda \int p(\xi) \left(\partial_z P(z) 1_{\xi \leq z} - p(z) \right) d\xi + (1-u)\chi_e\lambda_0 \int \Gamma(\xi) \left(\partial_z P(z) 1_{\xi \leq z} - p(z) \right) d\xi \\ &\quad + (1-u)\chi_s(\Gamma(z) - p(z)) + \underbrace{(1-u)u\lambda p(z) + u\chi_e\lambda_0\Gamma(z) - (\delta + \chi_x)(1-u)p(z)}_{\text{employment inflow/outflows}} \end{aligned}$$

Integrating from z_l to z , we obtain the equivalent equation for the CDF $P(z)$

$$\begin{aligned} \dot{P}(z)(1-u) &= \underbrace{(1-u)^2\lambda P(z)(P(z)-1)}_{\text{Job offers to employed (by existing firm)}} + \underbrace{(1-u)\chi_e\lambda_0 P(z)(\Gamma(z)-1)}_{\text{Job offers to employed (by entrant)}} \\ &\quad + \underbrace{(1-u)\chi_s(\Gamma(z) - P(z))}_{\text{productivity shocks}} + \underbrace{u(1-u)\lambda P(z)}_{\text{Job creation (by existing firms)}} + \underbrace{u\chi_e\lambda_0\Gamma(z)}_{\text{Job creation (by entrants)}} - \underbrace{(1-u)(\delta + \chi_x)P(z)}_{\text{Job destruction (by existing firm)}} \end{aligned}$$

In steady-state, we have

$$\begin{aligned} 0 &= (1-u)^2\lambda P(z)(P(z)-1) + (1-u)\chi_e\lambda_0 P(z)(\Gamma(z)-1) \\ &\quad + (1-u)\chi_s(\Gamma(z) - P(z)) + u(1-u)\lambda P(z) + u\chi_e\lambda_0\Gamma(z) - (1-u)(\delta + \chi_x)P(z) \end{aligned}$$

dividing by $1-u$ on both sides

$$\begin{aligned} 0 &= (1-u)\lambda P(z)(P(z)-1) + \chi_e\lambda_0 P(z)(\Gamma(z)-1) \\ &\quad + \chi_s(\Gamma(z) - P(z)) + u\lambda P(z) + \frac{u}{1-u}\chi_e\lambda_0\Gamma(z) - (\delta + \chi_x)P(z) \end{aligned}$$

Which is a quadratic equation in $P(z)$

$$0 = (1-u)\lambda P(z)^2 + \left[-\lambda(1-u) - \chi_e\lambda_0(1-\Gamma(z)) + u\lambda - (\delta + \chi_x + \chi_s) \right] P(z) + \left[\frac{u}{1-u}\chi_e\lambda_0 + \chi_s \right] \Gamma(z)$$

with coefficients given by

$$\begin{aligned} a &= (1-u)\lambda \\ b(z) &= -\lambda(1-2u) - \chi_e\lambda_0(1-\Gamma(z)) - (\delta + \chi_x + \chi_s) \\ c(z) &= \left(\frac{u}{1-u}\chi_e\lambda_0 + \chi_s \right) \Gamma(z) \\ \Delta(z) &= b^2(z) - 4ac(z) \end{aligned}$$

so that a real solution of the form

$$P(z) = \frac{-b(z) \pm \sqrt{\Delta(z)}}{2a}$$

exists whenever $\Delta(z) \geq 0$. First, I show that $b(z) < 0$

$$b(z) \leq b(z_l) \leq \lambda - (\delta + \chi_x + \chi_s) < 0$$

where the first inequality comes from the fact that $b(z)$ is increasing, the second uses the fact that $u \leq 1$ and the last one uses Assumption 2. To show that $\Delta(z) \geq 0$, I will first show that $\Delta(z_l) \geq 0$ and then that $\Delta'(z) \geq 0$. Notice that

$$c(z_l) = 0 \implies \Delta(z_l) = b(z_l)^2 \geq 0$$

And since $b(z) \leq 0$, $b'(z) \geq 0$ and $c(z) \geq 0$, we have that

$$\Delta'(z) = 2b(z)b'(z) - 4ac'(z) \geq 0$$

Which concludes the proof that $\Delta(z) \geq 0$ for all $z \in [z_l, \bar{z}]$. To be a valid solution, the distribution $P(z)$ must satisfy (1) $\lim_{z \rightarrow z_l} P(z) = 0$, (2) $\lim_{z \rightarrow \bar{z}} P(z) = 1$, and (3) $P'(z) \geq 0$. Let us now verify condition (1) on both roots. Using the fact that $\Delta(z_l) = b(z_l)$, we have that

$$P(z_l) = \frac{-b(z_l) \pm |b(z_l)|}{2a}$$

but since $b(z_l) < 0$, we have that $-b(z_l) + |b(z_l)| > 0$ which means that the positive root is not a valid solution. I now verify conditions two on the negative root solution. First, notice that

$$\lim_{z \rightarrow \bar{z}} \Delta(z) = (\lambda(1-u) + \delta + \chi_x + \chi_s - u\lambda)^2 - 4\lambda(1-u)(\delta + \chi_x + \chi_s - u\lambda) = (\delta + \chi_x + \chi_s - \lambda)^2$$

$$\implies \lim_{z \rightarrow \bar{z}} P(z) = \frac{2\lambda(1-u) + \delta + \chi_x + \chi_s - \lambda - |\delta + \chi_x + \chi_s - \lambda|}{2\lambda(1-u)} = 1 + \frac{\delta + \chi_x + \chi_s - \lambda - |\delta + \chi_x + \chi_s - \lambda|}{2\lambda(1-u)}$$

So that Assumption 2 ensures that $\lim_{z \rightarrow \bar{z}} P(z) = 1$.

B.5 Joint distribution of size and productivity

In a stationary equilibrium, the joint density of productivity z and size N must solve the following *Kolmogorov Forward Equation*

$$\dot{\varphi}(z, N) = \underbrace{\chi_s \left(\int_{z_l}^{\infty} \varphi(\xi, N) d\xi \Gamma(z) - \varphi(z, N) \right)}_{\text{productivity shocks}} - \underbrace{g(z) \frac{\partial}{\partial N} N \varphi(z, N)}_{\text{firm growth}} + \underbrace{\chi_x \left(\Gamma(z) \psi_{N_0^{-1}(z)}(N) - \varphi(z, N) \right)}_{\text{firm turnover}} \quad (45)$$

where the measure of firms in steady-state is given by $F = \frac{\chi_e(z_l)}{\chi_x(z_l)}$.

C Appendix - Solution method

C.1 Value function

I now present a solution method which I use to obtain discrete approximation of the value function $v(z)$ and the wage schedule $w(z)$ over an arbitrary grid $G_z = \{z_1, \dots, z_{N_z}\}$ where $dz_i \equiv z_{i+1} - z_i$. From now on, denote $v(z_i)$, $w(z_i)$, $k(z_i)$, $\Gamma(z_i)$ and $g(z_i)$ by v_i , w_i , k_i and g_i . The equations from the model which I will use are:

$$(r + \chi_x)v(z) = \max_{w \geq b, k} \left\{ zk^\alpha - w - rk + v(z)\tilde{g}(w) \right\} + \chi_s \left(\int v(\xi)\Gamma(d\xi) - v(z) \right)$$

$$w(z) = b + \int_0^z v(\xi)g'(\xi)$$

I use a finite difference, explicit scheme in order where I start with an initial guesses $\{v_i^0\}_i, \{w_i^0\}_i$ and update according the the following rules. The value function and wage schedule are iteratively updated according to

$$\frac{v_i^{n+1} - v_i^n}{\Delta} + (r + \chi_x + \chi_s)v_i^{n+1} =$$

$$z_i k_i^\alpha - w_i^n - rk_i + g_i v_i^{n+1} + \chi_s \sum_{j=i}^{N_z} \frac{v_j^{n+1}\Gamma_j + v_{j+1}^{n+1}\Gamma_{j+1}}{2} dz_j$$

$$w_1^{n+1} = b$$

$$w_i^{n+1} = b + \sum_{j=1}^i \frac{v_j^{n+1}g'_j + v_{j+1}^{n+1}g'_{j+1}}{2} dz_j \quad \text{for } i \geq 2$$

where $g'_j \equiv \frac{g_{j+1} - g_{j-1}}{z_{j+1} - z_{j-1}} = \frac{g_{j+1} - g_{j-1}}{dz_{j+1} + dz_j}$ is a central difference and $\Delta > 0$ is the step size. In matrix form, the system can be solved as

$$v^{n+1} = B^{-1}b^n$$

$$w^{n+1} = Cv^{n+1}$$

where the matrices B and vector b^n are given by

$$B = \left(\frac{1}{\Delta} + \chi_x + \chi_s \right) I - M_g - \chi_s E$$

$$b^n = \frac{1}{\Delta} v^n + z_i k_i^\alpha - w_i^n - rk_i$$

and C is the *cumulative integral matrix* implied by the trapezoidal quadrature. I iterate until

$$\max_{i=1, \dots, N_z} |v_i^b - v_i^{b-1}| < \varepsilon$$

where $\varepsilon > 0$ is a pre-specified tolerance level. The initial value threshold z_1 (i.e. the endogenous productivity threshold z_l) is chosen so that $|v_1 - 0| < \varepsilon$.

C.2 Distribution

I approximate the distribution φ by simulating M firms over B periods of length Δ

$$z_{t+\Delta} = \begin{cases} z_t & \text{with prob. } 1 - \Delta\chi_s \\ z' \sim \Gamma & \text{with prob. } \Delta\chi_s \end{cases} \quad (46)$$

$$N_{t+\Delta} = \begin{cases} (1 + \Delta g(z_t))N_t & \text{with prob. } 1 - \Delta\chi_x \\ N_0(z_t) & \text{with prob. } \Delta\chi_x \end{cases} \quad (47)$$

M is chosen to be a large number such as 10^7 and Δ to be a small number such as 0.1. B is chosen to allow for convergence (i.e. when $\sup |\varphi_{i,j}^b - \varphi_{i,j}^{b-1}| < 10^{-8}$).

C.3 Pareto exponent

In this section, I show that the discrete approximation of the endogenous process for firm growth is covered by Theorem 3.1 in Beare and Toda [2017] which provide a formula to compute the tail coefficient of the stationary cross-sectional size distribution.

First, consider a discrete approximation of the log firm size process for surviving firms

$$\log N_{t+\Delta} = \log N_t + g_t$$

and consider a discrete approximation of the productivity distribution $\Gamma_i \equiv \{\Gamma_1, \dots, \Gamma_N\}$ over a finite set $\{z_1, \dots, z_N\}$ with $\sum_{i=1}^N \Gamma_i = 1$ and define $g_i \equiv g(z_i)$ to be the endogenous growth rate of firms with productivity z_i . Since firms die at rate χ_x , $\log N_T - \log N_0$ can be expressed as a *geometrically stopped random walk*:

$$\log N_T - \log N_0 \equiv \sum_{t=1}^T g_t$$

where T is a geometric random variable with success probability $1 - \Delta\chi_x$ and the growth rate process has the following law:

$$g_t = \begin{cases} g_{t-1} & \text{with prob. } 1 - \Delta\chi_s \\ g_1 & \text{with prob. } \Delta\chi_s q_1 \\ \vdots & \vdots \\ g_N & \text{with prob. } \Delta\chi_s q_N \end{cases}$$

Notice that $\{g_t\}$ can be represented as an *Irreducible Markov Chain* with transition matrix Π .²² We can write the $N \times N$ matrix Π as

$$\Pi = \Delta\chi_s I + (1 - \Delta\chi_s)Q$$

where I is the identity the rows of the matrix Q are $(\gamma_1, \dots, \gamma_I)$ so that the element $\Pi_{i,j}$ represents the probability of going from state $z = z_i$ to z_j in a period of length Δ . Then, consider a matrix-valued function $D(s)$ which has diagonal elements $(e^{s\Delta g_1}, \dots, e^{s\Delta g_N})$. Define the *Perron root* (i.e. its largest eigenvalue in absolute term) of a square matrix A as $\rho(A)$. Theorem 3.1 in Beare and Toda [2017] states that there can be at most one $\zeta > 0$ such that

$$\rho(\Pi D(\zeta)) = \frac{1}{1 - \Delta\chi_x}$$

If such a ζ exists, it satisfies

$$\lim_{x \rightarrow \infty} \frac{1}{x} \log \mathbb{P}(\log N_T - \log N_0 < x) = -\zeta$$

which implies that the distribution of $\log N_T - \log N_0$ has an exponential upper tail or, equivalently, that the distribution of N_T/N_0 has a Pareto upper tail with Pareto exponent ζ .

$$\mathbb{P}(N_T/N_0 > n) \sim cn^{-\zeta}$$

for some $c > 0$.

²²A Markov chain is said to be irreducible if it is possible to get to any state from any state. In this case, g_t can take any values in its domain $\{z_1, \dots, z_N\}$ with respective probabilities $\{\Delta\chi_s q_1, \dots, \Delta\chi_s q_N\}$ which are independent of g_{t-1} .