IFS

Institute for Fiscal Studies

Shubhdeep Deb
Jan Eeckhout
Aseem Patel
Lawrence Warren

22/40

Working paper

Market power and wage inequality

Economic and Social Research Council

Market Power and Wage Inequality*

Shubhdeep Deb† Jan Eeckhout‡ Aseem Patel§
Lawrence Warren¶

September 8, 2022

Abstract

We propose a theory of how market power affects wage inequality. We ask how goods and labor market power jointly affect the level of wages, the Skill Premium, and wage inequality. We then use detailed microdata from the US Census between 1997 and 2016 to estimate the parameters of labor supply, technology and the market structure. We find that a less competitive market structure lowers the wage level, contributes 7% to the rise in the Skill Premium and accounts for half of the increase in between-establishment wage variance.

Keywords. Market Power. Wage Inequality. Skill Premium. Technological Change. Market Structure. Endogenous Markups. Endogenous Markdowns.

JEL. C6. D3. D4. D5. L1.

---

*We thank colleagues and seminar audiences for many useful comments and insightful discussions. Eeckhout gratefully acknowledges support from the ERC, Advanced grant 882499 and Deb from "la Caixa" Foundation (ID 100010434) fellowship (code LCF/BQ/DR19/11740003). We have benefited from superb research assistance by Renjie Bao and Wei Hua. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Data Management System (DMS) number: P-7083300, Subproject number 7508369. Disclosure Review Board number: CBDRB-FY20-CED006-0001, CBDRB-FY20-CED006-0032, CBDRB-FY21-202,CBDRB-FY22-360.

†UPF Barcelona, `shubhdeep.deb@upf.edu`

‡UPF Barcelona, `jan.eeckhout@upf.edu` – ICREA-BSE-CREI

§University of Essex, `aseem.patel@essex.ac.uk`

¶US Census Bureau, `lawrence.fujio.warren@census.gov`

1 Introduction

Wage inequality in the United States has risen sharply since the 1980s. The skill premium, the ratio of the average wage of workers with college education or more over the average wage of workers with up to a high school education, has risen from 50% in 1980 to nearly 100% in recent years.[1] Furthermore, recent work has highlighted the significant role played by heterogenous firms in shaping the evolution of wage inequality. Most of the rise in wage inequality is due to the increase in between-firm inequality.[2] Over the same period, there has been a corresponding rise in market power.[3]

In this paper, we set out to answer the question: How does market power affect wage inequality? The answer to this question has far-reaching welfare implications and is not merely an intellectual curiosity. If we attribute a substantial role to market power, then wage inequality is inefficient – there is too much inequality – and there is a role for inequality reducing policy that is Pareto improving and that raises welfare for all. Instead, if there was no market power, the amount of wage inequality would be Pareto efficient and there would only be a role for policy based on equity grounds and redistribution, without any scope for efficiency enhancing intervention.

We augment the benchmark supply and demand framework of Katz and Murphy (1992) in two dimensions. First, we depart from a representative firm framework adopted by the literature and explicitly account for the role of firm heterogeneity in technology. This setup permits us to study the evolution of wage inequality within *and* between establishments. Second, our economy incorporates oligopolistic output markets as well as oligopsonistic input markets with heterogenous markups and markdowns. In doing so we develop a tractable, quantitative general equilibrium model where a finite number of firms compete in a market that each own a set of heterogeneous establishments. This allows us to measure the macroeconomic implication of market power on the *level* of wages as well as wage *inequality*. To the best of our

[1] See Acemoglu and Autor (2011).
[2] See Song et al. (2018)
[3] See De Loecker et al. (2020), Hall (2018) and Hershbein et al. (2022).

knowledge, this is the first paper to study the implications of firm heterogeneity, output market power *and* input market power on wage inequality.

Each of these two modifications is crucial for the results we get. First, we adjust the technology with the objective to build a model that can account for the heterogeneity of skill ratios across establishments that we see in the microdata. To that effect, we assume a non-Hicksian, Constant Returns to Scale (CES) production function where each establishment has skill-specific productivities. For example, some establishments are highly productive with low-skilled workers but not the high-skilled (cleaning and security companies for example); other establishments are disproportionately productive with high-skilled workers (such as biotech firms); and yet other establishments are productive with workers of both skill types.

Second, those firms owning heterogeneous establishments exert market power by competing in both goods and labor markets with few competitors. Our setup builds on Atkeson and Burstein (2009) to model the goods market and on Berger et al. (2022) for the labor market, where the market structure crucially depends on a finite number of firms competing in a market. Our theoretical and computational contribution is to solve the structural model with *both* goods and labor market power. This gives rise to endogenous, establishment-specific markups and markdowns; therefore, market power in our setup depends not only on the a) household substitutability/preference parameters but also on b) the market structure as well as on c) the dispersion of the technology among competitors. Employment of high and low-skilled workers, together with their wages, is determined in general equilibrium.

Market power in the input and the output market will have implications for both the wage levels and wage inequality. On the one hand, the presence of monopsony power will induce firms to hire workers at wages lower than their marginal revenue product. On the other hand, even output market power will have implications for wages. A firm with market power in the output market will set its price above its marginal cost. This higher price, in conjunction with a downward sloping product demand curve, will imply that the equilibrium quantities demanded will be lower, which in turn will reduce the demand for labor. Therefore, through a general equilibrium effect, wages decline when economy-wide output market power increases. We estimate each of these determinants of market power using rich establishment level data from the U.S Census Bureau. We combine data from the US Longitudinal Business Database (LBD) and the Longitudinal Employer-Household Dynamics (LEHD) to construct a database that contains establishment level employment by skill, wages, and revenue between 1997 and 2016.

One of the main novelties of our approach is that we estimate a stochastic model of the market structure jointly with the technology. It is common practice to use observables such as inputs and outputs of production in order to estimate the unobservable technology while imposing a model structure even though it is virtually impossible to measure directly how units of input are transformed into quantities of output. Similarly, at a macroeconomic level, it is impossible to measure how firms compete, how many competitors there are, and who competes against whom. Therefore, we take a similar approach to the estimation of the market structure as we do to the estimation of technology. Our model shows a systematic relationship between market structure, revenue, and the wage bill. Both revenue and the wage bill are directly observed in our data. We exploit this structural link by relying on our stochastic model of competition to estimate the market structure.

Our approach of randomly assigning establishments within an industry to compete is a clear shortcut to the standard IO approach that diligently measures and models the identity of the competitors, how they compete, what actions they take and which prices they set. Unfortunately, we cannot apply a similar approach to the macroeconomy with a vast variety of industries, markets and technologies. The market for dry-cleaning services or coffee shops is a block, whereas for a furniture retailer like IKEA it is the entire metropolitan area. Our stochastic approach to measuring the market structure is therefore more akin to measuring the economy-wide Solow residual via growth accounting than to the direct measurement of the number of cars produced per worker in an assembly plant.

The main results from our estimation are the following. First, our estimates of market structure highlight declining competition, as measured by the number of firms competing in a market. Second, we find strong evidence of Skill-Biased Technological Change (SBTC). Together, these changes result in an increase in market power. The implied markup distribution shows a sharp increase in the upper tail and a rise in the sales-weighted markup from 1.682 to 2.160 between 1997 and 2016. Meanwhile, the markdowns for high-skill and low-skill workers during the same period increased from 1.420 to 1.435 and 1.419 and 1.437 respectively.

In our counterfactual exercise we find that a change in the market structure accounts for 7% of the increase in the aggregate skill premium and 56% of the increase in between establishment inequality. The decline in competition also leads to a decline in equilibrium wages by about 11%.[4] Consistent with Katz and Murphy (1992), we also find strong evidence of SBTC's contribution to aggregate skill premium and wage inequality even when firms are heterogeneous.

Related Literature. A growing literature highlights the role of firms and establishments in the rise of wage inequality.[5] Song et al. (2018) show the increase in the dispersion of earnings *between* firms accounts for two thirds of the increase in wage inequality in the US. Similarly, Barth et al. (2016) find that much of earnings inequality is due to increased dispersion of earnings among establishments. In our setup, in addition to the role of increasing technological differences between establishments in affecting wage inequality, we have skill-specific wages that vary by establishment due to monopsony power. As a result, while changes in technology will have profound implications for wage inequality, our setup also allows us to study how the extent of competition or market structure in the economy affects within and between establishment inequality.[6]

We borrow heavily from the work on markups and markdowns in macroeconomic

[4]In related work, De Loecker et al. (2018) and Deb et al. (2022) find similar effects on the wage level from an increase in market power.

[5]See Card et al. (2013) for Germany, Barth et al. (2016) and Song et al. (2018) for the US, and Håkanson et al. (2021) for Sweden.

[6]Our method using firm-level technologies builds on Patel (2021), who relies on similar tools to analyze Job Polarization in France.

equilibrium, both theoretical and empirical. Our model accounts for both output and input market power.[7] The main feature of our model is that markups and markdowns are variable and endogenous, as in Atkeson and Burstein (2009), Melitz and Ottaviano (2008), Edmond et al. (2015), Edmond et al. (2018), Amiti et al. (2019), De Loecker et al. (2018) and Baqaee and Farhi (2017) for markups, and Berger et al. (2022) and Azkarate-Askasua and Zerecero (2020) for markdowns. Moreover, markups and markdowns in our model are heterogeneous and the distribution of productivities has aggregate implications as in the literature on the granular origins.[8] A key innovation of our model is to solve for heterogeneous markups and markdowns jointly with strategic interaction, in general equilibrium.

One of the challenges of the competitive markets explanation where technological change is the sole driver of wage inequality is that it cannot easily account for the decline or stagnation of real wages. In the last decades, wages for the lowest skilled workers have fallen. SBTC increases the demand for skills, and if SBTC means that there is technological *progress* – skilled workers do not only become more skilled *relative* to unskilled workers, all workers become more skilled in *absolute* terms – this must necessarily lead to an increase in real wages for all, though relatively more so for the skilled. It is unlikely that technology has *regressed* and workers have become less productive, especially in the current decades of fast technological innovation. In a model with rising market power, the general equilibrium effect on wages naturally results in a decline in real wages, even though there is an increase in the skill premium.

In contrast to our explanation based on the rise of market power, complementary work has focused on the role of technological change in a competitive setting to explain the fall in real wages and the rise of skill premium. Those explanations build not only on a change in Total Factor Productivity (TFP), but also posit changes in the output elasticities of labor, of low-skilled labor in particular, often due to changing capital

[7]See also Mertens (2021), Hershbein et al. (2022), Azar and Vives (2021). The latter proposes a theory of the impact of common ownership in the presence of input and output market power.

[8]See Gabaix (2011), Grassi et al. (2017), Baqaee and Farhi (2017), Acemoglu et al. (2012), Carvalho and Tahbaz-Salehi (2019), Carvalho and Grassi (2019), and Burstein et al. (2019).

prices or automation.[9]

2 Model Setup

Environment. Time is discrete. There are two types of agents: a representative household and heterogeneous establishments. The representative household supplies labor in an oligopsonistic labor market and consumes goods produced in an oligopolistic goods market. Establishments are organized in a continuum of markets indexed j; the measure of markets is J. Each market contains a finite number of establishments I_j indexed by $i \in \{1, \ldots, I_j\}$ that are owned by N firms indexed by $n \in \{1, \ldots, N_j\}$. The set of establishments i owned by each firm n in market j is denoted as: $\mathcal{I}_{nj} = \{i \mid i \text{ in firm } n, \text{ in market } j\}$.[10] Goods and jobs are differentiated between markets and within markets, both in output and input markets. An establishment hires two inputs: high-skilled, H_{inj}, and low-skilled, L_{inj}, workers to produce final goods, Y_{inj}, where subscripts i, n, and j identify the establishment, firm, and market, respectively.

Preferences. The representative household chooses consumption and its supply of labor to both high and low-skill labor markets. The utility of consumption as in Atkeson and Burstein (2009) and the disutility of labor supply as in Berger et al. (2022) have a double nested Constant Elasticity of Substitution (CES) aggregator from quantities within and across markets. Goods i within a market are close substitutes with elasticity η; goods between markets j are relatively less substitutable with elasticity θ. These elasticities are ranked $\eta > \theta$ indicating that the household is more willing to substitute goods within a market (Pepsi vs. Coke) than across markets (soda vs. laundry detergent). Similarly in the labor market, the household has CES preferences over employment in the high-skill and low-skill labor markets.[11] The elasticities of sub-

[9]See Krusell et al. (2000), Acemoglu and Restrepo (2018), Acemoglu and Restrepo (2019), and Acemoglu and Restrepo (2021).

[10]Rather than the strict interpretation, we think of this multi-establishment setup as a metaphor for different ways of modeling market power, including collusion, common ownership, firms with a changing product mix... The modeling choice to have multi-establishment firms is for practical reasons. This setup allows us, first, to change the market structure without changing preferences, and second, to randomly assign establishments under different market structures without changing the number of them.

[11]In what follows, we use employment and jobs interchangeably.

stitution within the market are given by $\{\hat{\eta}_L, \hat{\eta}_H\}$ and between the markets are given by $\{\hat{\theta}_L, \hat{\theta}_H\}$, with $\hat{\eta}_L > \hat{\theta}_L$ and $\hat{\eta}_H > \hat{\theta}_H$, indicating that jobs within a market (barista at two coffee stores) are more substitutable than jobs in different markets (barista vs mechanic). The household maximizes its static utility:

$$\max_{C_{inj}, L_{inj}, H_{inj}} U\left(C - \frac{1}{\bar{\phi}_L^{\frac{1}{\phi_L}}} \frac{L^{\frac{\phi_L+1}{\phi_L}}}{\frac{\phi_L+1}{\phi_L}} - \frac{1}{\bar{\phi}_H^{\frac{1}{\phi_H}}} \frac{H^{\frac{\phi_H+1}{\phi_H}}}{\frac{\phi_H+1}{\phi_H}}\right) \quad \text{s.t. } PC = LW_L + HW_H + \Pi$$

where C, H and L are the CES indices for aggregate consumption and employment of high and low-skilled workers respectively. $P,\ W_H$ and W_L are the CES aggregated indices for the prices of output and wages of skill groups H and L respectively.[12] Observe that the aggregate and the market specific quantities are normalized by the size of the market to neutralize the love-for-variety effects in the model.

$$C = \left(\int_j J^{-\frac{1}{\theta}} C_j^{\frac{\theta-1}{\theta}} dj\right)^{\frac{\theta}{\theta-1}} \quad , \quad C_j = \left(\sum_i I^{-\frac{1}{\eta}} C_{inj}^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}$$

$$S = \left(\int_j J^{\frac{1}{\hat{\theta}_S}} S_j^{\frac{\hat{\theta}_S+1}{\hat{\theta}_S}} dj\right)^{\frac{\hat{\theta}_S}{\hat{\theta}_S+1}} \quad , \quad S_j = \left(\sum_i I^{\frac{1}{\hat{\eta}_S}} S_{inj}^{\frac{\hat{\eta}_S+1}{\hat{\eta}_S}}\right)^{\frac{\hat{\eta}_S}{\hat{\eta}_S+1}} , \quad S \in \{H, L\}.$$

Technology. The starting point is Katz and Murphy (1992), but with a heterogeneous technology that is specific to the establishment and skill type:

$$Y_{inj} = \left[(A_{Linj} L_{inj})^{\frac{\sigma-1}{\sigma}} + (A_{Hinj} H_{inj})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{1}$$

where A_{Hinj}, A_{Linj} is the factor-augmenting technology jointly distributed according to $G(A_{Hinj}, A_{Linj})$ and σ is the elasticity of substitution. In our framework, the composition of workers across establishments varies for two reasons: 1) technology is factor-specific, and 2) there is monopsony power in both labor markets.

Market Structure. Each establishment with productivity (A_{Hinj}, A_{Linj}) belongs to a particular market j and there are I_j establishments in each market j. We define the

[12] We denote aggregate high and low-skilled labor computed by adding workers as : $\mathcal{S} = \int_j \sum_i S_{inj} dj,\ S \in \{H, L\}$.

market structure, N, as the total number of firms competing in a market. Since firms have market power in all three markets: the output market, low-skill and high-skill labor markets, we need to define what is the relevant set of firms competing in each market. A key assumption that makes our model tractable is that the set of firms competing in the goods market and the two labor markets are exactly the same.[13] Finally, we assume that each firm n in market j owns $\mathcal{I}_{nj}$ establishments that are assigned to a firm stochastically. The key idea is that despite this random assignment of ownership of establishments to firms, the model preserves some key properties as we vary N. Since N measures the extent of competition in a market, a decline in N would translate to an increase in market power in both the output and input markets.

3 Solution

Solution of the household's problem. Given product prices, P_{inj}, and wages, W_{Linj} and W_{Hinj}, the household chooses consumption bundles, C_{inj}, and the labor supply, L_{inj} and H_{inj}, to maximize utility subject to the budget constraint. The household's optimal solution for consumption and labor supply is:

$$C_{inj}(P_{inj}, P_{-inj}, P, C) = \frac{1}{J}\frac{1}{I}P_{inj}^{-\eta}P_j^{\eta-\theta}P^{\theta}C \tag{2}$$

$$S_{inj}(W_{Sinj}, W_{S,-inj}, W_S, S) = \frac{1}{J}\frac{1}{I}W_{Sinj}^{\hat{\eta}_S}W_{Sj}^{\hat{\theta}_S-\hat{\eta}_S}W_S^{-\hat{\theta}_S}S \tag{3}$$

where $S \in \{H, L\}$. Note that these equilibrium demand and supply functions not only depend on the price (wage) set by the establishment i, but also on its relative magnitude to the market price (wage) index. The aggregate and market price indices

[13] This implies that the N firms in market j compete with each other in the output and the two labor markets simultaneously. In reality one can imagine a firm n having a different set of competitors in the output market and each of the two labor markets. For instance, a soft drink producer may compete with another soft drink producer in the output market, but it may compete with retail stores for low-skilled workers and with a software company for high-skilled workers. However, this assumption would be reasonable for markets where skills in the labor market are closely tied to the output markets.

are defined as follows:

$$P = \left(\int_j \frac{1}{J} P_j^{1-\theta} dj \right)^{\frac{1}{1-\theta}} \;,\; P_j = \left(\sum_i \frac{1}{I} P_{inj}^{1-\eta} \right)^{\frac{1}{1-\eta}} \tag{4}$$

$$W_S = \left(\int_j \frac{1}{J} W_{Sj}^{1+\hat{\theta}_S} dj \right)^{\frac{1}{1+\hat{\theta}_S}} \;,\; W_{Sj} = \left(\sum_i \frac{1}{I} W_{Sinj}^{1+\hat{\eta}_S} \right)^{\frac{1}{1+\hat{\eta}_S}} \tag{5}$$

From the solutions in equations (2) and (3) we can write the inverse demand function and inverse labor supply functions as

$$P_{inj}(Y_{inj}, Y_{-inj}, P, Y) = \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} Y_{inj}^{-\frac{1}{\eta}} Y_j^{\frac{1}{\eta}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P \tag{6}$$

$$W_{Sinj}(S_{inj}, S_{-inj}, W_S, S) = \frac{1}{J}^{\frac{-1}{\hat{\theta}_S}} \frac{1}{I}^{\frac{-1}{\hat{\eta}_S}} S_{inj}^{\frac{1}{\hat{\eta}_S}} S_j^{\frac{1}{\hat{\theta}_S}-\frac{1}{\hat{\eta}_S}} S^{-\frac{1}{\hat{\theta}_S}} W_S \tag{7}$$

Solution of the firm's problem. Taking as given the inverse demand function in equation (6) and the inverse labor supply function for each type of worker in equation (7), firm n in market j chooses the optimal production plan for each of its establishments with the choice of the quantity of inputs H_{inj} and L_{inj} to maximize profits:

$$\Pi_{nj} = \max_{H_{inj}, L_{inj}} \sum_{i \in \mathcal{I}_{nj}} \left(P_{inj} Y_{inj} - W_{Hinj} H_{inj} - W_{Linj} L_{inj} \right). \tag{8}$$

There are three important features of the firm's maximization problem. First, as in models of monopolistic and monopsonistic competition, firms internalize the effect of their own quantity choices on their prices and wages. Second, given the multi-establishment setup, firms internalize the ownership structure and take into account interactions between quantity choices across the different establishments owned by it and its effect on prices. Finally, given Cournot competition, firms also internalize the quantity choices of the other$-n$ firms in the market and strategically choose their quantities, such that our equilibrium is characterized by an intersection of best response functions.[14]

[14]Because there is a continuum of other markets $-j$, j is infinitesimally small and there is no strategic interaction across markets

The first order conditions with respect to a given skill, S_{inj}, $S \in \{H, L\}$ is:

$$\underbrace{\left[P_{inj} + \frac{\partial P_{inj}}{\partial Y_{inj}} Y_{inj} + \sum_{i' \in \mathcal{I}_{nj} \setminus i} \left(\frac{\partial P_{i'nj}}{\partial Y_{inj}} Y_{i'nj} \right) \right] \frac{\partial Y_{inj}}{\partial S_{inj}}}_{\text{Marginal Revenue Product of Labor}} = \underbrace{\left[W_{Sinj} + \frac{\partial W_{sinj}}{\partial S_{inj}} S_{inj} + \sum_{i' \in \mathcal{I}_{nj} \setminus i} \left(\frac{\partial W_{Si'nj}}{\partial S_{inj}} S_{i'nj} \right) \right]}_{\text{Marginal Cost of Labor}} \tag{9}$$

Where $\mathcal{I}_{nj} \setminus i$ is the set of all other establishments owned by firm n, except establishment i. Factoring out P_{inj} and W_{Sinj}, we can express the above equation as

$$P_{inj} Y_{ij}^{\frac{1}{\sigma}} A_{Sinj}^{\frac{\sigma-1}{\sigma}} S_{inj}^{-\frac{1}{\sigma}} \left[1 + \varepsilon_{inj}^{P}\right] = W_{Sinj} \left[1 + \varepsilon_{inj}^{S}\right] \tag{10}$$

where ε_{inj}^{P} is the inverse demand elasticity and ε_{inj}^{S} denotes the inverse labor supply elasticity for skill S. In Appendix A.2, we derive each of these elasticities. We further show that the inverse demand elasticity is equal to:

$$\varepsilon_{inj}^{P} \equiv \frac{\partial P_{inj}}{\partial Y_{inj}} \frac{Y_{inj}}{P_{inj}} + \sum_{i' \in \mathcal{I}_{nj} \setminus i} \left(\frac{\partial P_{i'nj}}{\partial Y_{inj}} \frac{Y_{i'nj}}{P_{inj}} \right) = - \left[\frac{1}{\theta} s_{nj} + \frac{1}{\eta} (1 - s_{nj}) \right] \tag{11}$$

where $s_{nj} = \sum_{i \in \mathcal{I}_{nj}} s_{inj}$ is the sales share of the firm in market j and $s_{inj} = \frac{P_{inj} Y_{inj}}{\sum_i P_{inj} Y_{inj}}$ is the sales share of establishment i in market j.[15] Similarly, in the labor markets, the inverse labor supply elasticity for each skill satisfies

$$\varepsilon_{inj}^{S} \equiv \frac{\partial W_{Sinj}}{\partial S_{inj}} \frac{S_{inj}}{W_{Sinj}} + \sum_{i' \in \mathcal{I}_{nj} \setminus i} \left(\frac{\partial W_{Si'nj}}{\partial S_{inj}} \frac{S_{i'nj}}{W_{Sinj}} \right) = \frac{1}{\hat{\theta}_S} e_{Snj} + \frac{1}{\hat{\eta}_S} (1 - e_{Snj}) \tag{12}$$

where $e_{Snj} = \sum_{i \in \mathcal{I}_{nj}} e_{Sinj}$ is the wage bill share of firm n and $e_{Sinj} = \frac{W_{Sinj} S_{inj}}{\sum_i W_{Sinj} S_{inj}}$ is the wage bill share of the establishment in market j for each input $S \in \{H, L\}$.

The firm's inverse demand elasticity $\varepsilon_{inj}^{p} < 0$ directly determines the markup μ_{inj} which is the ratio of the price over the marginal cost. Similarly, we can define our markdowns δ_{Sinj} for each skill as the ratio of the wage to marginal revenue product of skill S which is pinned down by the inverse labor supply elasticity ε_{inj}^{S}.

$$\mu_{inj} = \frac{P_{inj}}{MC_{inj}} = \frac{1}{1 + \varepsilon_{inj}^{P}} \ , \ \delta_{Sinj} = \frac{MRPL_{Sinj}}{W_{Sinj}} = 1 + \varepsilon_{inj}^{S} \tag{13}$$

[15]Throughout, we use capital S to index high and low-skill and small s to refer to sales-share of a firm or an establishment.

Note that the markup (markdown) is the same for all the establishments owned by a given firm and is determined by the sum of sales shares (payroll share) of each establishment. The firm faces a non-zero residual inverse demand elasticity, ε^{p}_{inj}, and inverse labor supply elasticity, ε^{S}_{inj}, because it has market power. Under perfect competition, ε^{p}_{inj} and ε^{S}_{inj} are zero and the firm sets marginal product equal to the wage. Here, firms that have a large share s_{nj} of revenue in their market j face an inverse demand elasticity $\varepsilon^{p}_{inj} \approx -\frac{1}{\theta}$. The residual inverse demand is steep as the firm faces virtually no competition within the market and only from goods in other markets, which are not very substitutable. As a result, those firms have high market power. Instead, firms that have a small market share s_{nj} face a relatively flat residual inverse demand with inverse elasticity $\varepsilon^{p}_{inj} \approx -\frac{1}{\eta}$ (recall that $\theta < \eta$). Those firms face steep competition from firms that produce close substitutes. As a result, their market power is limited. Similar arguments apply in the labor market: firms with a large employment share e_{Snj} for skill S will have a steeper inverse labor supply function with $\varepsilon^{S}_{inj} = \frac{1}{\hat{\theta}_S}$. While for firms with low employment share the inverse labor supply function will be flatter with a elasticity $\varepsilon^{S}_{inj} = \frac{1}{\hat{\eta}_S}$ as $\hat{\eta}_S > \hat{\theta}_S$.

The skill premium in our model is defined as the ratio of the high-skill wage over the low-skill wage. In order to assess how market power affects the skill premium we take the log-ratio of the first order conditions and get the following equation:

$$\ln\left(\frac{W_{Hinj}}{W_{Linj}}\right) = \ln\left(\frac{\delta_{Linj}}{\delta_{Hinj}}\right) + \frac{\sigma-1}{\sigma}\ln\left(\frac{A_{Hinj}}{A_{Linj}}\right) - \frac{1}{\sigma}\ln\left(\frac{H_{inj}}{L_{inj}}\right) \tag{14}$$

Equation (14) expresses the establishment level skill premium, defined as the ratio of high-skill to low-skill wages paid at each establishment. Note that there is no direct role of ε^{P}_{inj}, and therefore of markups μ_{inj}, in affecting the establishment-specific skill premium. At face value, this equation looks very similar to the skill premium equation that Katz and Murphy (1992) estimate. In particular, with no labor market power, $\delta_{Linj} = \delta_{Hinj} = 1$, and no heterogeneity, it looks exactly identical:

$$\ln\left(\frac{W_H}{W_L}\right) = \frac{\sigma-1}{\sigma}\ln\left(\frac{A_H}{A_L}\right) - \frac{1}{\sigma}\ln\left(\frac{H}{L}\right)$$

However, there are fundamental conceptual differences. First, we explicitly account for heterogeneity in the productivity of skills at each establishment in our framework. Second, equation (14) holds at the establishment level. Third, we allow for input markets to be imperfectly competitive. This implies that in addition to the race between the technology ratio, A_H/A_L, and the skill ratio, H/L, in determining the evolution of the skill premium as postulated by Tinbergen (1974) and later formalized by Katz and Murphy (1992), our model features an additional force that may influence the evolution of the skill premium. The term δ_L/δ_H measures the markdown for low-skill workers relative to that of high-skill workers. The joint implication of these differences is that we have an entire distribution of establishment-specific skill premia in our model, with the additional force of differential monopsony power affecting the evolution of the skill premium.

Finally, in order to calculate the aggregate skill premium, we define the input share-weighted average wages for each skill as $\mathcal{W}_S = \int_j \sum_i S_{inj} W_{Sinj} dj / \mathcal{S}$ and $\mathcal{S} = \int_j \sum_i S_{inj} dj$ denotes the aggregate workers of a given skill. Hence, we define the aggregate skill premium as follows:

$$\kappa = \frac{\mathcal{W}_H}{\mathcal{W}_L} = \frac{\mathcal{L}}{\mathcal{H}} \times \frac{\int_j \sum_i H_{inj} W_{Hinj} dj}{\int_j \sum_i L_{inj} W_{Linj} dj} \tag{15}$$

The fundamental insight here is that wages $\mathcal{W}_H$ and $\mathcal{W}_L$ adjust in equilibrium to changes in the market structure as well as technology.

Computing the Equilibrium. This large economy with heterogeneous establishments, market power and non-Hicks-neutral technology does not have an analytical solution. We therefore solve the economy computationally. The equilibrium consists of a set of wages W_L, W_H and aggregate output Y such that the first order conditions as well as market clearing conditions are satisfied. Observe that as usual, there is indeterminacy in the price level, and we therefore use the aggregate price index $P = 1$ as the numeraire. Practically, we assume a large number of markets J to mimic the continuum that we assumed in the theory. We use the algorithm specified in the Appendix A.3 to calculate the model equilibrium. The algorithm fully specifies the equilibrium allocation of establishment level quantities, H_{inj}, L_{inj} and Y_{inj}, and establishment level prices W_{Hinj}, W_{Linj} and P_{inj}, and aggregates them to economy-wide prices and quantities. In addition, it allows us to compute establishment-level markups μ_{inj} and markdowns δ_{Linj} and δ_{Hinj}, as well as aggregate them to economy-wide measures of market power.

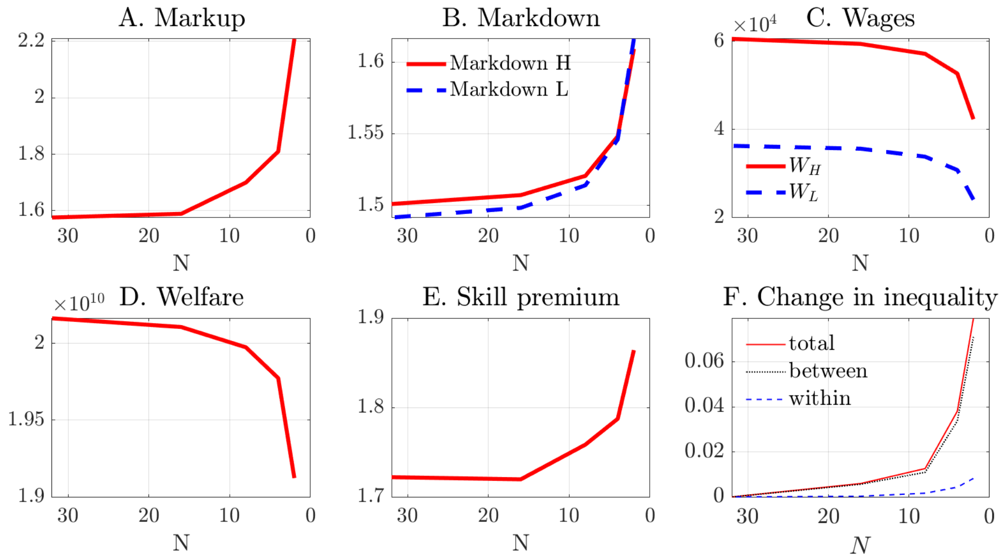
Comparative Statics. We compute the economy for a series of comparative statics exercises where we change market structure N and evaluate the impact this has on the key equilibrium features of the economy.[16]

In Figure 1, we report 6 panels: In panel A and B we show that as the number of competitors declines, the average (sales-weighted) markup and markdowns increase. As the number of competitors declines, the sales and the wage bill shares of the establishments in the market approach 1 and markups and markdowns approach their respective upper bounds. Panel C shows the average (worker-weighted) wages of high and low-skilled workers, $\mathcal{W}_H$ and $\mathcal{W}_L$, respectively. The decline in wages is a result of an increase in both markups and markdowns. For both skills, when markdowns increase, establishment-specific wages decline as establishments charge a larger markdown over wages relative to the marginal revenue product of labor. Meanwhile, an increase in markups leads to a decline in wages through the general equilibrium effect through a reduction in aggregate demand for labor as in De Loecker et al. (2018) and Deb et al. (2022). The combined effect of an increase in markups and markdowns in our model is that the average wages of both skills decline. Panel D shows the decline in welfare as an increase in market power reduces the utility from aggregate consumption more than the increase in utility from supplying lower labor in response to the decline in wages.

In panel E, we see that a reduction in the number of competitors N leads to a rise in the aggregate skill premium κ. Similar to the canonical model, an increase in the technology ratio, A_{Hinj}/A_{Linj}, increases the skill premium and an increase in the skill ratio, H_{inj}/L_{inj}, reduces it. However, in addition to these two competing forces our model

[16] In the comparative static exercise we assume $I_j = I = 32\ \forall j$ and $N_j = N\ \forall j$. In addition, we consider $N \in \{2, 4, 8, 16, 32\}$ such that each firm owns the same numbers of establishments given by I/N as we vary N.

Figure 1: Comparative Static



also allows for market power, such that an increase in the relative monopsony power, $\delta_{Linj}/\delta_{Hinj}$, also increases the skill premium. This increase in the relative monopsony power of firms may come from one of three sources: 1. changes in the technology $G(A_{Hinj}, A_{Linj})$; 2. changes in the substitutability parameters $(\hat{\eta}_S, \hat{\theta}_S)$; 3. changes in market structure N. Furthermore, how a change in N leads to a change in the skill premium will depend on its interaction with the underlying substitutability parameters and productivity distribution.

We first isolate the interaction between N and substitutability parameters in determining the skill premium. We consider a homogeneous setup where $A_{Hinj} = A_H$ and $A_{Linj} = A_L$ for all establishments while varying only N. Given this, in Proposition 1, we derive a closed form expression for the aggregate skill premium, which is a function of productivity ratio A_H/A_L and skill specific labor supply substitutability parameters and wage bill shares and constants $\sigma, \bar{\phi}_S, \phi$. Specifically, we use the fact that in the homogeneous case the wage bill shares for each skill can be expressed solely as a function of the number of competitors, given by $1/N$.

Proposition 1. *In homogeneous case, the skill premium is:*

$$\kappa = \left[\left(\frac{A_H}{A_L} \right)^{\frac{\sigma-1}{\sigma+\phi}} \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H} \right)^{\frac{1}{\sigma+\phi}} \right] \cdot \left[\frac{1 + \frac{1}{\hat{\theta}_L}\frac{1}{N} + \frac{1}{\hat{\eta}_L}(1 - \frac{1}{N})}{1 + \frac{1}{\hat{\theta}_H}\frac{1}{N} + \frac{1}{\hat{\eta}_H}(1 - \frac{1}{N})} \right]^{\frac{\sigma}{\sigma+\phi}} \tag{16}$$

When $N > 1$, a sufficient condition for $\frac{\partial \kappa}{\partial N} / \left(\frac{\kappa}{N}\right) < 0$ is $\hat{\eta}_H < \hat{\eta}_L$ and $\frac{1}{\hat{\theta}_H} - \frac{1}{\hat{\eta}_H} < \frac{1}{\hat{\theta}_L} - \frac{1}{\hat{\eta}_L}$.

Proof. In Appendix A.5. □

In Proposition 1, we show that under certain conditions for the substitutability parameters, an increase in market power through a decline in N leads to an increase in skill premium. The intuition is that as the number of competitors declines, firms increase the markdown for both skills as they constitute a larger share of the labor market for both skills. If firms exert relatively higher monopsony power over low-skilled workers compared to high-skilled workers, this leads to an increase in the skill premium.[17]

To gain intuition about how a change in N affects between-establishment inequality, consider a simplified version of our general model developed earlier. Suppose all sectors are identical and within each sector there are two establishments. Assume that in period t, these establishments compete against each other and in period $t+1$, both these establishments merge to become one firm. In other words, N declines from 2 to 1. Assume that establishment 1 is the dominant firm in the market and establishment 2 is the fringe controlling a very small (but positive) share of sales in the market. Finally, assume that $\hat{\eta}_H = \hat{\eta}_L = \hat{\eta}$ and $\hat{\theta}_H = \hat{\theta}_L = \hat{\theta}$. This last assumption effectively shuts down the within-establishment inequality channel and allows us to focus on the between-establishment inequality instead.

Given this environment, the markdown charged by the dominant establishment in the market will be close to the upper bound $\frac{\hat{\theta}+1}{\hat{\theta}}$. In contrast, the markdown charged by the fringe establishment is close to the lower bound of $\frac{\hat{\eta}+1}{\hat{\eta}}$. As N goes from 2 to 1, the markdown in establishment 2 increases (the wedge between wages and the

[17] With heterogeneous establishments, in addition to the substitutability parameters $\{\hat{\eta}_S, \hat{\theta}_S\}$ the underlying heterogeneity in A_H and A_L within each market also plays an important role in determining the direction of change in skill premium as N declines.

marginal revenue product of labor increases). This reduces the wages of workers in establishment 2 and increases wage inequality between the two establishments.

Finally, panel F in Figure 1, shows the change in the total wage inequality as N declines, and how much is within and between establishment inequality. We see that as competition declines, total wage inequality increases. In addition, both within and between establishment inequality also increase.

4 Quantitative Analysis

We proceed with the quantitative analysis, estimating the model parameters and analyzing the determinants of market power, namely the skill-specific substitutability parameters in the labor market, the technology distributions, and market structure. We then assess their role in the evolution of wage inequality.

Data. The data we use to estimate our model combines establishment-level data from the Longitudinal Business Database (LBD) with characteristics of the workers at these establishments from Longitudinal Employer-Household Dynamics (LEHD) data.[18] We use the LEHD to construct measures of the skill composition of each establishment in our data. LEHD provides information on the linkage of workers and firms in each state at quarterly frequency from unemployment insurance records. This data allows us to observe about 96% of workers and the identities of their employers (via tax identifiers) for a sample of 21 states, going back to 1997.[19]

For our exercise, we use the LEHD to derive measures of the composition of skill types and wages within each firm. The characteristics we use from LEHD are the total employment and wages of workers at a firm by level of educational attainment. We

[18] The frame of the LBD comes from the Census Business Register, which is populated from the quinquennial economic census and from tax data. LBD is an establishment level dataset containing information on payroll, employment, sales, geography, and industry. In LEHD, we observe the matching of employers and employees, including earnings and the characteristics of workers.

[19] Our sample includes CA, CO, CT, HI, ID, IL, KS, LA, ME, MD, MN, MO, MT, NJ, NM, NC, OR, RI, TX, WA, and WI. The LEHD infrastructure files include links to Decennial Census survey data and the American Community Survey as well as administrative records to provide demographic information on workers. On the firm side, links to the Census Business Register and Longitudinal Business Database provide information on industry, age, geography, organization, and other characterstics.

split workers into categories of high education (we will refer to as "skill") as those who attained some college education or above and low skill as those who attain a high school education or less.[20] We calculate average full-quarter earnings by skill type, giving us a measure of employment and payroll by skill for a SEIN (employer identifier) within LEHD. We take the firm-level ratio of high to low-skill employment and payroll per worker from LEHD and use these measures to split LBD employment and payroll into the same skill-specific ratios, but at the establishment level. This breaks up total payroll and employment in LBD into a measure of skill-specific average wages and employment in LBD.

Sample selection. Our sample is comprised of the subset of our LBD sample of establishments where the firm links to at least one SEIN in our 21 state LEHD sample. We drop establishments with missing or zero employment or payroll. When we estimate market structure, we drop establishments with missing sales and drop establishments above the 99th percentile of sales. We drop establishments with five or fewer employees, and for which we do not have at least one high or low-skill employee. This data set provides us with a measure of employment and earnings for each establishment by skill type, along with measures of total revenue, industry classification (NAICS), and geography (MSA) from 1997-2016.[21] We deflate all values to 2002 dollars.

Market definition. In order to estimate the model, we need to define a market. In the Industrial Organizations literature, this is the key ingredient. Given our interest in the macroeconomics of market power, it is impossible to observe the market structure for each individual firm in different industries and geography.[22] We therefore use a stochastic notion of the market definition. In the knowledge that we cannot use de-

[20]The educational attainment variable in LEHD is not directly observed for all individuals and we use the Census imputation when education is not observed. Our estimated elasticities are qualitatively similar when we restrict to only using observed educational attainment. We have also established robustness of our findings with different categorizations of skills.

[21]In what follows, we refer to NAICS 2 as a sector, NAICS 6 as an industry and the market as a collection of 32 establishments randomly assigned within each NAICS 6.

[22]There is too much variation in the market structure across industries and geography and there is mechanical variation over time. For a discussion of the problems with using NAICS codes and geographical areas to pin down the market definition, see Eeckhout (2020).

tailed information to define a market, instead we use the structure of the model and the random assignment of firms as competitors where firms within the same industry are equally likely to compete against each other. Yet, we determine the number of competitors N independently. Thus, even if an industry contains a large number of firms, if N is small, the extent of the competition is weak. While this approach to defining a market is much less detailed than the traditional approach, it does allow us to make progress in studying market power in the macroeconomy. The main idea is that we remain agnostic about which firms compete and that is something we cannot observe, just like Total Factor Productivity (TFP). But if we observe revenue and costs, we derive the number of competitors consistent with the model that gives rise to those revenues and costs, and hence profits and markups. Just like the Solow residual, we derive the number of competitors as an outcome.

Practically, we start by defining a broad set of potential competitors as a NAICS 6 industry.[23] Now depending on the industry, there can be a lot of establishments within each industry. In order to define a market within each NAICS 6 industry, we first randomly assign establishments to markets of size I. Once we select those I establishments that form a market, thereafter we randomly establish the identity of the firms that compete, and how many firms N are active within a market by randomly assigning these I establishments into N subsets of size I/N.

With this random assignment, if the number of firms N is smaller, the model predicted markups and markdowns will be higher, firm revenue will be higher, and wages will be lower. The objective is to use the observed revenue and wages from the data to estimate N.[24] As mentioned above, we also make the assumption that the market structure is the same for both the input and output markets.

Quantifying the model. We quantify our model in two steps and estimate it sepa-

[23] In Appendix C, we report results where we condition on geography and we define the broad set of competitors as those within NAICS 3 industry x MSA.

[24] We restrict our sample of establishments in these randomly assigned markets to those with nonmissing revenue. Revenue is a firm-level measure so for establishments in multi-establishment firms, we allocate revenues to establishments by their share of payroll within the firm. We truncate the revenue distribution by dropping establishments above the 99th percentile in revenue by year.

Table 1: **Externally chosen or calibrated parameters**

Variable	Value	Description	Source
θ	1.2	Between sector elasticity	De Loecker et al. (2018)
η	5.75	Within sector elasticity	De Loecker et al. (2018)
σ	2.9	Elasticity of substitution	Acemoglu and Autor (2011)
ϕ_H	0.25	High-skilled labor supply elasticity	Chetty et al. (2011)
ϕ_L	0.25	Low-skilled labor supply elasticity	Chetty et al. (2011)
I	32	Total number of estb.	Externally set

rately for 1997 and 2016. First, we estimate the parameters that determine the labor supply elasticity for high and low-skilled workers, namely, $\hat{\eta}_S$ and $\hat{\theta}_S, S \in \{H, L\}$, using the microdata and an instrumental variable strategy. Second, we *jointly* estimate the non-parametric distribution of technology $G(A_{Hinj}, A_{Linj})$ and our measure of competition in the model N. To estimate the $G(A_{Hinj}, A_{Linj})$, we rely on the structure of our model which provides a link between these unobservable productivities and employment (H_{inj}, L_{inj}), a quantity directly observed in the microdata, through the first-order conditions. We estimate N such that it matches the moments of the sales-weighted revenue over wage bill distribution between the data and the model using the method of moments. We calibrate some parameters externally (Table 1) and hold them fixed throughout our quantification exercise.

Step 1. Estimating labor market elasticities. In the first step we estimate $(\hat{\eta}_S, \hat{\theta}_S)$ separately for each of the two skills. The labor supply elasticity $\varepsilon^S_{inj} = \left(1/\hat{\theta}_S\right) e_{Snj} + \left(1/\hat{\eta}_L\right)\left(1 - e_{Snj}\right)$ is a function of a) the within $(\hat{\eta}_S)$ and the between-market $(\hat{\theta}_S)$ labor substitutability parameters and b) the skill-specific establishment-level employment (S_{inj}) in each market j as $e_{Sinj} = S_{inj}^{\frac{1+\eta}{\eta}} / \sum_i S_{inj}^{\frac{1+\eta}{\eta}}$. While establishment-level employment is directly observed in the data, we need to estimate the two labor substitutability parameters to calculate the elasticity. We estimate these parameters by relying on the inverse labor supply equation of our model in equation (7). Given that the theory provides us a deterministic relationship between wages and employment, it will not hold for all the establishments in the data. In order to take the model to the data, we augment it by adding to it an error term and a time subscript, t.[25] Re-writing the expression by taking logs on both sides, we get

$$\ln W^*_{Sinjt} = k_{jt} \; + \; \left(\frac{1}{\hat{\theta}_S} - \frac{1}{\hat{\eta}_S}\right) \ln S_{jt} \; + \; \frac{1}{\hat{\eta}_S} \ln S_{injt} \; + \; \varepsilon_{Sinjt} \tag{17}$$

where $\ln W^*_{Sinjt} = \ln W_{Sinjt} + \varepsilon_{Sinjt}$ and $k_{jt} = \ln J_t^{\frac{1}{\hat{\theta}_S}} I_{jt}^{\frac{1}{\hat{\eta}_S}} S_t^{-\frac{1}{\hat{\theta}_S}} W_t$.[26]

The error term potentially captures misspecification in our model that leads to measurement error in the observed value of establishment-specific wages. This misspecification could be either due to non-pecuniary match factors (such as distance to work, interactions with co-workers and supervisors as argued by Card et al. (2018)) or due to the impact of labor market institutions that are not in our model, such as the minimum wage. While we remain agnostic about the true source of this misspecification, we account for the fact that the error term is potentially correlated with employment. To address the bias stemming from this correlation, we devise an instrumental variable strategy to estimate our parameters of interest. We build on the recent work of Berger, Herkenhoff, and Mongey (2022) and Giroud and Rauh (2019) and exploit state-level corporate taxes as a source of exogenous variation shifting the demand curve in our model. We provide further details about the instrument below. Closest to our approach is the recent work of Felix (2021), who also relies on a similar strategy to estimate the labor substitutability parameters using the labor supply equation directly.[27] We make the following set of assumptions to identify our parameters of interest.

Assumption 1. *The error term is correlated with log of employment:*

$$\mathbb{E}(\varepsilon_{Sinjt} \times \ln S_{injt}) \neq 0 \tag{18}$$

Assumption 2. *The error term in equation (17) can be decomposed as follows:*

$$\varepsilon_{Sinjt} = \alpha_{Sinj} + \epsilon_{Sinjt}, \tag{19}$$

[25] We add the time subscripts since we will exploit time-series variation in wages and employment at the establishment-level and taxes at the state-level in our estimation. More details below.

[26] In our estimation exercise, we let the total number of establishments in a market to change, as observed in the data.

[27] In Appendix B, we show identication of the labor substitutability parameters in the simpler case without endogeneity. We also provide results from Monte Carlo experiments that demonstrates the ability of our estimator to parse out the true structural parameters in simulations.

where ϵ_{Sinjt} is assumed to have mean 0 and variance σ^2_ϵ. We treat α_{Sinj} as fixed unknown parameters.

Assumption 3. *Let $X(i)$ denote the geographical state of establishment i. Denote by $\tau_{X(i)t}$ the corporate tax faced by an establishment i in state X at time t. We assume that variation in taxes across state and over time is independent of the error term and correlated with the log of employment:*

$$\varepsilon_{Sinjt} \perp\!\!\!\perp \tau_{X(i)t}, \quad \mathbb{E}(S_{injt} \times \tau_{X(i)t}) \neq 0 \tag{20}$$

Under these assumptions, we can identify $\hat{\eta}_S$ and $\hat{\theta}_S$ using the following moments in the data:

$$\hat{\eta}_S = \frac{\mathbb{E}(\widetilde{S}_{injt} \times \tau_{X(i)t})}{\mathbb{E}(\widetilde{W^*}_{Sinjt} \times \tau_{X(i)t})} \tag{21}$$

$$\hat{\theta}_S = \left[\frac{\mathbb{E}(\{\overline{\Omega}_{Sjt} - c_{jt}\} \times \bar{\tau}_{jt})}{\mathbb{E}(\ln S_{jt} \times \bar{\tau}_{jt})} + \frac{\mathbb{E}(\widetilde{W^*}_{Sinj} \times \tau_{X(i)t})}{\mathbb{E}(\widetilde{S}_{inj} \times \tau_{X(i)t})} \right]^{-1} \tag{22}$$

where we denote[28]

$$\widetilde{S}_{injt} = \ln S_{injt} - \overline{\ln S_{jt}}, \qquad \widetilde{W^*}_{Sinjt} = \ln W^*_{Sinjt} - \overline{\ln W^*_{Sjt}}, \qquad \overline{\ln X_{jt}} = \frac{1}{I} \sum_i \ln X_{injt}$$

$$\overline{\Omega}_{Sjt} = \mathbb{E}_{jt}(\Omega_{Sinjt}), \qquad \Omega_{Sinjt} = \ln W^*_{Sinjt} - \frac{1}{\hat{\eta}_S} \ln S_{injt}, \qquad \bar{\tau}_{jt} = \frac{1}{I} \sum_{i \in j} \tau_{X(i)t}.$$

Estimation. We use Two-Stage Least Squares (2SLS) on the following equations to get the estimate of $\hat{\eta}_S$ and $\hat{\theta}_S$.

$$\ln W^*_{Sinjt} = k_{jt} + \gamma_S \ln S_{jt} + \beta_S \ln S_{injt} + \underbrace{\alpha_{Sinj} + \epsilon_{Sinjt}}_{\varepsilon_{Sinjt}} \tag{23}$$

[28]To derive equation (21) and equation (22), we start with the following moment conditions implied by Assumption 3.

$$\mathbb{E}[(\varepsilon_{Sinjt} - \bar{\varepsilon}_{Sjt}) \times \tau_{X(i)t}] = 0,$$
$$\mathbb{E}(\bar{\varepsilon}_{Sjt} \times \bar{\tau}_{jt}) = 0$$

where we replace $(\varepsilon_{Sinjt} - \bar{\varepsilon}_{Sjt})$ and $\bar{\varepsilon}_{Sjt}$ by their values provided in equation (23) and equation (24). In the application, we replace k_{jt} in equation (22) by a sector and year fixed effect. Further details are provided in the Estimation sub-section below.

where we define $\beta_S = \frac{1}{\hat{\eta}_S}$ and $\gamma_S = \left(\frac{1}{\hat{\theta}_S} - \beta_S\right)$. From equation (23), we notice that while we observe wages and employment in the data, we do not directly observe the establishment fixed effect α_{Sinj} and sector-year specific constants, k_{jt} and S_{jt}, which are both functions of our structural parameter $\hat{\eta}_S$ and $\hat{\theta}_S$. We need to control for these unobserved variables to avoid omitted variable bias stemming from them. We control for α_{Sinj} by including establishment fixed-effects in our estimation. To control for k_{jt} and S_{jt}, we include an interaction of sector and year fixed-effects. Together these two controls allow us to exploit within-establishment variation while controlling for time shocks that vary by sector. Finally, to control for endogeneity arising from correlation between the log of employment and the error term, we instrument $\ln S_{injt}$ with state corporate taxes, $\tau_{X(i)t}$. We think of the time-series variation in taxes as an exogenous shock to a firm's labor demand which help us identify the parameters of firm's labor supply equation.

Once we get an estimate of β_S (and implicitly $\hat{\eta}_S$) from equation (23), we proceed to estimate γ by relying on the following equation:[29]

$$\overline{\Omega}_{Sjt} = k_{jt} + \gamma_S \ln S_{jt} + \bar{\varepsilon}_{Sjt} \tag{24}$$

where $\bar{\varepsilon}_{Sjt} = \mathbb{E}_{jt}(\varepsilon_{Sinjt})$. As before, k_{jt}, which is itself a function of $\hat{\theta}$, is unobserved to the econometrician. To address the issue of omitted variable bias stemming from the unobservability of k_{jt}, we control for it by including a sector and a year fixed effect (separately) in equation (24).[30] Finally, to address the issue of endogeneity due to potential correlation between $\ln S_{jt}$ and $\bar{\varepsilon}_{Sjt}$, we instrument $\ln S_{jt}$ by $\bar{\tau}_{jt}$, the average tax-rate in a given market j.[31] Intuitively, we exploit within-sector time-series variation

[29]To get to equation (24), we start by taking $\beta_S \ln S_{injt}$ to the LHS in equation (23) to get $\ln W^*_{injt} - \beta \ln L_{injt}$. We take sectoral average on both sides to get to equation (24).

[30]Notice that if we were to control for k_{jt} by including an interaction of sector-year fixed-effects, we would no longer be able to identify γ_S as there will not be any variation in S_{jt}. Given that k_{jt} contains I_{jt}, the number of establishments within a market, we implicitly assume that the tax variation is uncorrelated with the size of the market. Giroud and Rauh (2019) have argued that there can be a non-zero correlation between market size (I_j) and taxes, which can be a threat to the identification of γ_S in our framework. However, Giroud and Rauh (2019) in their analysis define a market as a state which is different from the interpretation that we have adopted in our model.

[31]In practice, when we estimate equation (24), we weigh each sector by its size to limit the effect of outliers on the estimate of $\hat{\theta}_S$.

in average tax rates to estimate γ_S. Next, to estimate the labor disutility parameter, we rely on the aggregate labor supply equation of the household for each skill as follows written in logs:[32]

$$\ln W_{st} = \frac{1}{\phi_S} \ln \frac{1}{\bar{\phi}_{St}} + \frac{1}{\phi_S} \ln S_t \tag{25}$$

We calibrate the value of the Frisch elasticity, ϕ_S, to be equal to 0.25 (see Chetty et al. (2011)) for both High and low-skilled workers. This allows us to estimate the value of $\bar{\phi}_{St}$, one for each year, by inverting equation (25).

Finally, once all the key parameters of interest are estimated, and given the skill-specific employment observed in the microdata, we calculate wages by using equation (7). The difference between the model implied wages and the ones observed in the data is precisely the measurement error denoted in equation (17).

Estimation Sample. To estimate the within and between-market substitution parameters, we rely on the panel dimension of our merged LBD-LEHD data. We estimate these parameters for the tradeable sector between 1997 and 2011.[33] We extend our stochastic assignment procedure to account for the panel dimension of our data. To do so, we first randomly assign establishments to markets, conditional on NAICS 6, in 1997 such that there are at most 32 establishments in each market. Once assigned to a market, the establishment always remains in it as long as we observe it in the data. For every subsequent year starting from 1997, we again randomly assign the establishment unobserved previously (i.e., the new entrants) to one of the existing markets created in 1997. As a result, the size, and the composition of the markets evolve randomly over time given the entry and exit of establishments from markets. Our baseline estimates are based on this sample. Finally, we perform two robustness exercises without random assignment of establishments to markets: 1) by re-estimating them on the same panel with national labor markets (i.e., by defining a market as the NAICS 6 industry); 2) by defining a local labor market (i.e., by defining a market as the NAICS 3 industry x MSA). These results are in Appendix C.

[32] We assume there is no measurement error in aggregate wages, i.e. $\ln W^*_{st} = \ln W_{st}$.

[33] We do not have state tax data beyond 2011.

Step 2. Estimating the distribution of technologies and N. Equipped with the estimates of within and across market labor substitutability parameters, we proceed to estimate N, the total number of firms competing in a market and distribution of technologies, $G(A_{Hinj}, A_{Linj})$. To do so, we rely on the first-order conditions (FOCs) for each skill.

$$P_{inj} Y_{inj}^{\frac{1}{\sigma}} A_{Sinj}^{\frac{\sigma-1}{\sigma}} S_{inj}^{-\frac{1}{\sigma}} \left[1 + \varepsilon_{inj}^{P}\right] = W_{Sinj} \left[1 + \varepsilon_{inj}^{S}\right] \tag{26}$$

Our approach to estimating the technology distribution non-parametrically, using equation (26) for each skill is motivated by the observation that we can re-write the FOCs solely in terms of employment and structural parameters, objects that we either directly observe in the microdata or we have estimated, along with the technology distribution; our unobserved parameters of interest. To do so, we first replace output market elasticity (ϵ_{inj}^{P}) and input market elasticities (ϵ_{inj}^{S}) in the FOCs by equation (12) and equation (12), respectively. These elasticities are functions of the revenue share and the wage bill share which can be expressed as a function of output and employment, as follows:

$$s_{nj} = \frac{\sum_{i \in \mathcal{I}_{nj}} P_{inj} Y_{inj}}{\sum_i P_{inj} Y_{inj}} = \frac{\sum_{i \in \mathcal{I}_{nj}} Y_{inj}^{\frac{\eta-1}{\eta}}}{\sum_i Y_{inj}^{\frac{\eta-1}{\eta}}}, \quad e_{Snj} = \frac{\sum_{i \in \mathcal{I}_{nj}} W_{Sinj} S_{inj}}{\sum_i W_{Sinj} S_{inj}} = \frac{\sum_{i \in \mathcal{I}_{nj}} S_{inj}^{\frac{\hat{\eta}_S+1}{\hat{\eta}_S}}}{\sum_i S_{inj}^{\frac{\hat{\eta}_S+1}{\hat{\eta}_S}}}$$

Finally, we substitute out prices (P_{inj}), wages (W_{Sinj}) and output (Y_{inj}) in the two FOCs as well as in the expressions for the revenue share and the wage-bill share by equations (6), (7) and (1). This allows us to get the a system of implicit functions for each establishment i for each skill $S \in \{H, L\}$.

$$F_{iS}(\mathbf{A}_{Hj}, \mathbf{A}_{Lj}, \mathbf{H}_j, \mathbf{L}_j, Y;\ \boldsymbol{\zeta}) = 0, \quad \forall i \tag{27}$$

where we define (the superscript T denotes the transpose)

$$\mathbf{A}_{Sj}^T = [A_{S1nj}, A_{S2nj}, \ldots, A_{SInj}]$$
$$\mathbf{S}_j^T = [S_{1nj}, S_{2nj}, \ldots, S_{Inj}]$$
$$\boldsymbol{\zeta}^T = [\eta,\ \theta,\ \sigma,\ \hat{\eta}_H,\ \hat{\eta}_L,\ \phi_H,\ \phi_L,\ \bar{\phi}_H,\ \bar{\phi}_L,\ ,I,\ J,\ N]$$

Note that the only economy-wide aggregate that appears in equation (27) is aggregate output, Y. This is because we can calculate aggregate wages, W_S, and aggregate employment, S, conditional on the estimates of labor substitutability parameters and observed employment in the microdata at the end of Step 1. Given that we normalize aggregate prices in our theoretical setup, we set the value of $P = 1$ in our estimation. In order to estimate the technology parameters, $\mathbf{A}_j^T = [\mathbf{A}_{Hj}^T, \mathbf{A}_{Lj}^T]$, for all $j \in [1, \ldots, J]$, we treat employment from our microdata as the endogenous observed input in the FOCs to back out a distribution of technology for each establishment in the economy that is consistent with the model. More specifically, for each market j, we need to pin down a $2 \times I$ vector of the unobserved technology $\mathbf{A}_j^T$. For each establishment i, we have two first-order conditions, one for each skill, and since we have I establishments in each market, we have a system of equations with $2 \times I$ equations and $2 \times I$ unknowns. As a result, given structural parameters $\boldsymbol{\zeta}$, we solve this system of equation for each market j to pin down an estimate of $G(A_{Hinj}, A_{Linj})$. The algorithm that we use in practice that helps us achieve this objective is outlined in the Appendix A.4.

To initialize the algorithm and solve the system of equations for each market j, we need to know $\boldsymbol{\zeta}$, which contains N, the total number of firms competing in each market. To do so, within each NAICS 6 industry, we first randomly assign establishments into markets of size I. We then guess a value of $N \in \{2, 4, 8, 16, 32\}$ and randomly assign establishments within each market to firms such that there are N firms competing in each market. In order to optimally pick a N, we lean on the observation that for every guess of N, the model produces a distribution of revenues and wage-bill. Conditional on employment and wages, our model produces a monotonically declining relationship between the ratio of the revenue over wage-bill and N. To see this, note that the revenue over wage bill in the model can be written as:

$$\frac{R_{inj}}{W_{Hinj}H_{inj} + W_{Linj}L_{inj}} \equiv \psi_{inj} = \left[\omega_{Hinj} \times \mu_{inj} \times \delta^{H}_{inj}\right] + \left[\omega_{Linj} \times \mu_{inj} \times \delta^{L}_{inj}\right] \tag{28}$$

where $\omega_{Sinj} = \frac{W_{Sinj}S_{inj}}{W_{Hinj}H_{inj}+W_{Linj}L_{inj}}$ denotes the share of wages of skill S in the total wage bill. Intuitively speaking, equation (28) says that holding employment and wages fixed at their equilibrium level at each establishment, a decline in N leads to an increase in the revenue share s_{nj} and the skill-specific wage bill share e^{S}_{nj} of each firm since it owns a greater number of establishments in each market. For any given values of within and across-market substitutability in the product and the labor market, this increase will lead to an increase in the market power of firms in both the input and the output markets and increases the wedge between revenue and wage bill. We exploit this link in our estimation of N by relying on Simulated Method of Moments to minimize the distance of the sales-weighted mean of the revenue over wage bill between our model and the data:

$$N^{*} = \min_{N \in \{2, 4, 8, 16, 32\}} \left[\mathbb{E}\left(\widehat{sw}^{D}_{inj}\psi^{D}_{inj}\right) - \mathbb{E}\{\widehat{sw}^{M}_{inj}(N)\psi^{M}_{inj}(N)\}\right]^{2} \tag{29}$$

where $\widehat{sw}^{D}_{inj} = \frac{R_{inj}}{\int_j \sum_i R_{inj} dj}$ denotes the sales-share of establishment i in the data while $\widehat{sw}^{M}_{inj}$ denotes the same quantity in the model.

Finally, as our production function abstracts from capital and intermediate inputs, we adjust the revenue in the data to make it comparable to our model. To do so, we multiply the revenue in the data by a constant α_N such that $R^{Adj,Data}_{inj} = \alpha_N \times R^{Data}_{inj}$, where $R^{Adj,Data}_{inj}$ denotes the adjusted revenue in the data and R^{Data}_{inj} is the unadjusted revenue in the data.[34] We pin down the value of α_N such that $N = 16$ in 1997.[35] In

[34]We assume that the gross revenues are produced using a cobb-douglas production function of the form $Y_{ij} = N^{\alpha_N}_{ij} K^{\alpha_K}_{ij} M^{\alpha_M}_{ij}$ where $N = \left[\left(A_{Hij}H_{ij}\right)^{\frac{\sigma-1}{\sigma}} + \left(A_{Lij}L_{ij}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$ and $\alpha_N + \alpha_K + \alpha_M = 1$. Then the gross revenue can be written as $R_{ij} = \mu_{ij} W_{Lij} L_{ij}(1 + \epsilon^{w_L}_{ij}) + \mu_{ij} W_{Hij} H_{ij}(1 + \epsilon^{w_H}_{ij}) + \mu_{ij} P^{M} M_{ij} + \mu_{ij} P^{K} K_{ij} = \alpha_N R_{ij} + \alpha_M R_{ij} + \alpha_K R_{ij}$. While in our model with only labor N as input the revenue $R^{N}_{ij} = \mu_{ij} W_{Lij} L_{ij}(1 + \epsilon^{w_L}_{ij}) + \mu_{ij} W_{Hij} H_{ij}(1 + \epsilon^{w_H}_{ij})$ such that $R^{N}_{ij} = \alpha_N R_{ij}$

[35]Given the monotonic relation between revenue in the model and N, there exists an α_N such that

the following years we hold the value of α_N constant and estimate N by matching the sales-weighted distribution of revenue over wage bill in the data and the model. Our estimate of α_N is 0.314 which is in line with the estimates found in the literature on production function estimation.[36]

5 Results

In this section, we report the results of our quantification exercise. First, we report the estimates of the labor market substitutability parameters for each skill from Step 1. Second, we report the results for the distribution of establishment-specific technology that we estimated in Step 2. Finally, we report the estimates of our market structure N as well as the average markups and markdowns.

Estimates of Labor Substitutability Parameters. In Table 2, Panel A, we present the OLS and the IV estimates of our reduced form parameters $\beta_S = \frac{1}{\eta_S}$ and $\gamma_S = \frac{1}{\theta_S} - \frac{1}{\eta_S}$. For both the skills and for both $\hat{\eta}_S$ and $\hat{\theta}_S$, we find that OLS estimate of parameters is biased downward compared to the IV. More importantly, the OLS estimate for β_S is not consistent with the theory as it shows a negative relationship between wages and employment in the establishment's labor supply curve. The IV corrects for the bias and shows that the corresponding structural parameters in Panel B of Table 2 are in line with the theory: $\hat{\eta}_L > \hat{\theta}_L$ and $\hat{\eta}_H > \hat{\theta}_H$, i.e., within-market substitutability is greater than the between-market substitutability.

We find that the estimate of the within-market substitutability parameter for high-skilled workers, $\hat{\eta}_H$, is 2.53 while that of low-skilled workers, $\hat{\eta}_L$, is 2.42. These estimates imply that jobs within a market have similar substitutability for high and low-skilled workers. Furthermore, we find that the estimate of between-market substitutability for the high-skilled worker, $\hat{\theta}_H$, is 2.02 while that of the low-skilled worker,

the sales-weighted revenue of wage bill in the data (after adjustment using α_N) exactly equals the sales-weighted revenue over wage bill in the model.

[36] Closest to our specification is the work of De Loecker (2011) and Doraszelski and Jaumandreu (2013), both of whom rely on a Cobb-Douglas production function with capital, intermediate inputs, and labor and both find that the output elasticities to be in the range of 0.17 and 0.334.

$\hat{\theta}_L$, is 1.85. This implies that jobs across markets are less substitutable for low-skill workers, which can be interpreted as indicating that the mobility cost for low-skilled workers to move across markets is relatively high compared to that of high-skilled workers.

To the best of our knowledge, this is the first paper that separately provides structural estimates of labor substitutability parameters for high and low-skilled workers. Berger et al. (2022) also estimate a model of oligopsony in the labor market without the distinction between high and low skill types. Their estimate for the within-market substitutability is equal to 10.85 while for between-market substitutability is 0.42. To get to these estimates, they rely on Indirect Inference.[37]

In contrast, we take a different approach. While we use the same instrument, we exploit the log-linearity of the labor supply function to estimate the substitutability parameters. This is like the approach adopted by Felix (2021) who relies on the import tariff reductions as an exogenous variation to estimate the within-market substitutability parameter and cross-market variation in import competition to estimate the between-market substitutability parameter in a model of oligopsonistic labor markets.

Finally, Table 2, Panel C provides the first-stage estimates of our IV. In both cases we find that the first-stage is negative and statistically significant. In the case of the estimation of β_S, when we use taxes as an instrument for changes in labor demand we find that taxes are negatively correlated with employment at the establishment level. This reduced-form relationship between employment and taxes is consistent with the evidence presented in Giroud and Rauh (2019) and Berger et al. (2022). We also find a similar relationship when we estimate γ_S where in the first stage we find a negative

[37] Apart from the methodological difference in the estimation of labor substitutability parameters, three additional differences lead to different estimates of the labor substitutability parameters between our work and the results in Berger et al. (2022). First, because we have no information on the market, we randomly assign our firms to markets drawn from industry classifications instead of assuming the market is a particular industry classification. Second, our estimates of the labor supply function are at the establishment level while Berger et al. (2022) estimates it at the firm level. Lastly, in our baseline, our labor markets are considered to be national while Berger et al. (2022) consider local labor markets defined by NAIXS 3 x MSA. In Appendix C we show that when we consider the same market definition of NAICS 3 x MSA we find a higher estimate of $\hat{\eta}_S$ indicating that once we condition on geography jobs are more substitutable within a market.

correlation between average market employment and average market-level taxes.

Table 2: **Estimates of reduced-form parameters: Tradeables with Random Sampling**

A. OLS and Second-Stage IV Estimates					
	OLS	IV		OLS	IV
	(1)	(2)		(3)	(4)
β_H	-0.180***	0.396***	γ_H	-0.079***	0.100***
SE	0.0007	0.062	SE	0.0003	0.005
Market-Year SE	(0.001)	(0.095)	Market SE	(0.003)	(0.041)
β_L	-0.110***	0.414***	γ_L	-0.095***	0.127***
SE	0.0007	0.057	SE	0.0003	0.005
Market-Year SE	(0.002)	(0.089)	Market SE	(0.003)	(0.042)
Market x Year FE	Yes	Yes	Market FE	Yes	Yes
Establishment FE	Yes	Yes	Year FE	Yes	Yes
B. Structural Parameters					
η_H	-5.56	2.53	θ_H	-3.87	2.02
η_L	-9.09	2.42	θ_L	-4.87	1.85
C. First-stage Regressions for the IV					
$\tau^H_{X(i)t}$	-	-0.012***	$\bar{\tau}^H_{jt}$	-	-0.061***
SE		0.001	SE		0.001
Market-Year SE		(0.002)	Market SE		(0.009)
$\tau^L_{X(i)t}$	-	-0.014***	$\bar{\tau}^L_{jt}$	-	-0.066***
SE		0.001	SE		0.001
Market-Year SE		(0.002)	Market SE		(0.009)
Market x Year FE	-	Yes	Market FE	-	Yes
Establishment FE	-	Yes	Year FE	-	Yes
No. of obs (High-Skilled)	1,147,000	1,147,000		70,000	70,000
No. of obs (Low-Skilled)	1,147,000	1,147,000		70,000	70,000

Notes: Standard errors clustered at the market-year level for the first stage and at the sector level at the second stage are reported in the parenthesis. Non-clustered standard errors are reported without parenthesis. *** p < 0.01, ** p<0.05, *p<0.1. The significance stars correspond to clustered standard errors. Estimates of γ_S in columns 3 and 4 are conditional on the estimates of columns 1 and 2, respectively. Number of observations are common for both the first and the second-stage. The number of observations reflects rounding for disclosure avoidance. $\tau^S_{X(i)t}$ denotes the co-efficient infront of taxes in the first-stage regression for the estimate of β_S. The same instrument is used separately, first to estimate β_H and then to estimate β_L.

Clustering. In Table 2, we provide two sets of estimates for the standard error. The first estimate does not cluster the standard error at any level. The second estimate clusters the standard error at the market-year level for the estimate of $\hat{\eta}_S$ and the market level for the estimate of $\hat{\theta}_S$. The clustering for $\hat{\eta}_S$ accounts for the fact that unobserved shocks may be correlated within a market-year and plausibly uncorrelated across markets. Given that we include an establishment fixed effect in our baseline specification, we also cluster our standard errors at the establishment level. None of the results about the significance of our estimates is affected when we cluster at the establishment level. Finally, we cluster the estimates of the standard error of $\hat{\theta}_S$ at the sector level to account for the potential correlation of market-specific shock over time.

Robustness of elasticity estimates. In the baseline version of the model, we estimated the labor substitutability parameters by randomly assigning establishments to markets within a given NAICS 6. Moreover, we did not include interaction between geography (say, MSA) with NAICS prior to making the random assignment. These choices could lead to two concerns: First, the estimates of labor substitutability parameters were influenced by the random assignment and second, that labor markets were not correctly specified. For instance, for the tradables sector one can argue that while NAICS 6 is a good proxy for competition between firms in the product market it is not a good indicator of competition in the labor market. This could lead to potential misspecification of competition in the labor market and could affect the estimates of labor supply. To address these concerns, we perform two exercises to demonstrate that the estimates of our labor substitutability parameters are robust to these extensions. In the first robustness exercise, we re-estimate the substitutability parameters by defining a market as NAICS 6. In other words, we do not randomly assign establishments to markets conditional on NAICS 6. In the second exercise, we condition on geography define a market as NAICS 3 x MSA instead of simply NAICS 6 and make no random assignment after defining the market. This is the same definition used by Berger et al. (2022) to estimate the labor supply elasticities. The result of this exercise is provided in the Appendix C. We find that the estimates of within and across market substitutability are very similar compared to our baseline results.

TFP Distribution. In Figures 2a and 2b, we plot the density of $\ln A_{Hinj}$ and $\ln A_{Linj}$.

Table 3: **Moments of Technology Distribution**

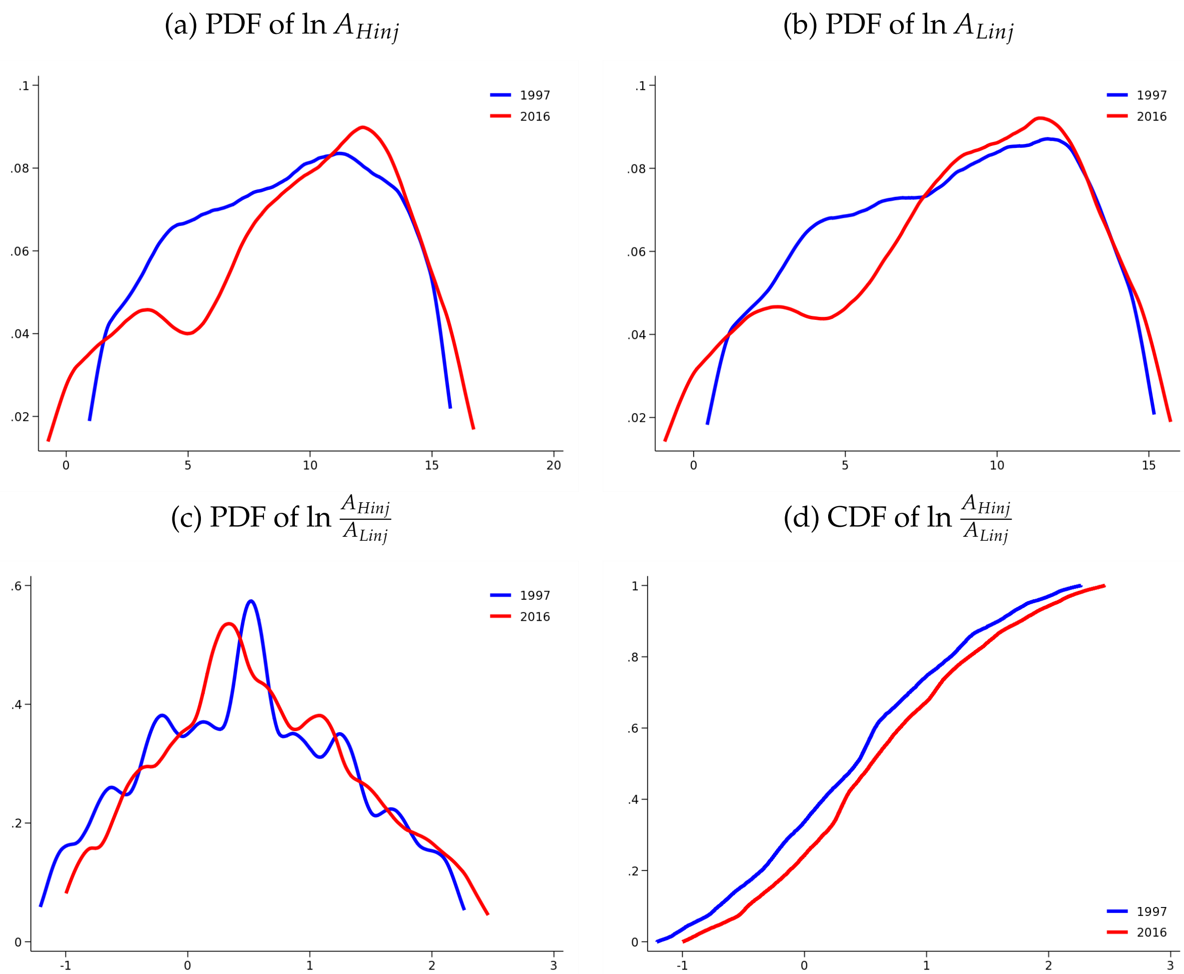
	$\ln A_{Hinj}$		$\ln A_{Linj}$		$\ln \frac{A_{Hinj}}{A_{Linj}}$	
	Mean	Variance	Mean	Variance	Mean	Variance
1997	8.69	21.71	8.20	21.32	0.49	1.11
2016	8.89	28.54	8.26	26.11	0.63	1.12

By looking at the levels of the technology, separately for each skill, two things become evident. First, there is a substantial amount of heterogeneity across establishments in their technology for high and low-skilled workers. Second, there is an increase in the variance of these technologies over time, for both high and low-skill workers. The variance of the distribution of productivities for high-skilled workers is *higher* compared to low-skilled workers in both years. This heterogeneity and its increase over time have an important implication for heterogeneity in establishment-level markup and markdowns as well as wage inequality over time. We explore the quantitative implications of these results in our counterfactual experiments in Section 6.

Consistent with the literature, we find strong evidence in support of skill-biased technological change. In Figure 2c we show that the mean of the distribution of relative productivities in 2016 compared to 1997. The mean of $\ln A_{Hinj}/A_{Linj}$ has increased from 0.49 to 0.63 and the variance has remained effectively unchanged as shown in table 3. Meanwhile, in Figure 2d, we show that the 2016 cdf of relative productivities first-order stochastically dominates the distribution in 1997.

Estimated market structure, markups and markdowns. Table 4 reports our estimated value of N has declined substantially between the two endpoints of our data: N was 16 in 1997 while it has declined to 4 in 2016, implying that any given firm competes with fewer other firms, on average, in a market. We remain agnostic about the source of this decline. For example, this decline in N can be due to a rise in common ownership – large investors owning shares in competing firms. In their recent work, Ederer and Pellegrino (2022) show that in the US the "network of common ownership has a hub-and-spoke structure with a large proportion of firms sharing significant overlap and the remainder of largely unconnected firms at the periphery." This evidence is in line

Figure 2: **Estimated Distribution of Skill-Specific Technology**



(a) PDF of $\ln A_{Hinj}$ (b) PDF of $\ln A_{Linj}$

(c) PDF of $\ln \frac{A_{Hinj}}{A_{Linj}}$ (d) CDF of $\ln \frac{A_{Hinj}}{A_{Linj}}$

Notes: Panels (a) and (b) show the probability density function of productivities of $\ln A_{Hinj}$ and A_{Linj}, respectively, for 1997 and 2016. Panels (c) and (d) show the probability density function and the cumulative density function of the ratio of $\ln \frac{A_{Hinj}}{A_{Linj}}$, respectively.

with the declining estimate of N that we document in the paper.

We find that the estimated N in 2016 is low relative to 1997. We rationalize this finding as follows. Our model has two forces that can drive the wedge between revenue over the wage-bill, which has increased over time, as shown in Table (5). These forces are technological change and N. Note that if markets were perfectly competitive, the ratio of revenue over wage bill would equal one for all firms.

While there has been an increase in the variance of the distribution of technology

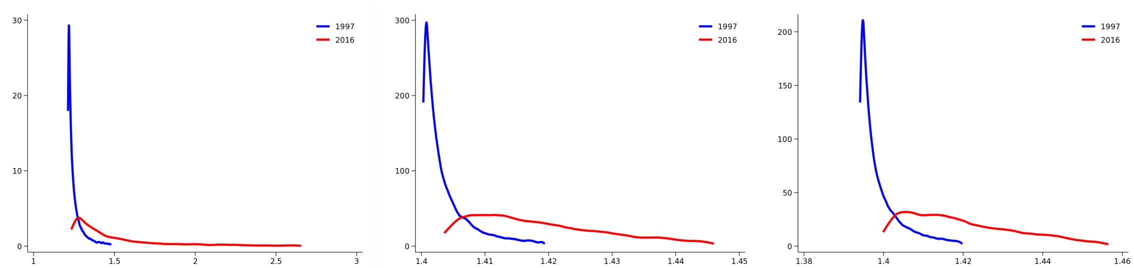
Table 4: **Estimates of the Market Power and Labor Supply Parameters**

	N	$\bar{\phi}_H$	$\bar{\phi}_L$	**Average Markup**	**Average Markdown**	
					High-Skilled	**Low-Skilled**
1997	16	166900	180800	1.682	1.420	1.419
2016	4	96430	64760	2.160	1.435	1.437

Notes: The average markup is the sales-weighted average markup estimated from our model. The average markdown is the sales-weighted markdowns for high and low-skilled.

Figure 3: **Estimated Markup and Markdown distribution**

(a) Distribution of μ_{inj} (b) Distribution of δ_{Hinj} (c) Distribution of δ_{Linj}



over time, the underlying heterogeneity cannot fully explain the increase in the wedge between revenue and the wage bill.[38] The residual increase in this wedge is explained by a decline in N which leads to higher market power for firms.[39] In a recent paper, De Loecker et al. (2018) also estimates a model of imperfect competition with strategic interactions in the output market and show that competition in the aggregate economy has declined.

This decline in competition leads to an increase in sales-weighted average markup has increased from 1.68 in 1997 to 2.16 in 2016, while markdowns for both skills increase only marginally.[40] Figure 3a, 3b and 3c show that the distribution of the un-

[38] In our estimation strategy, the distribution of technology is a function of both the underlying employment distribution in the data and the market structure N.

[39] The effect of N of the wedge is highly non-linear in a model with Cournot competition. In other words, the increase in the wedge, when N moves from 16 to 8, is lower than its increase when N moves from 8 to 4. Consequently, N needs to be as low as 4 for our model to match the observed wedge in the data.

[40] Qualitatively speaking, this increase in markup is consistent with the rise of markup documented by De Loecker et al. (2020), who use Compustat data and rely on the production function estimation to get their results. With regards to markdowns, we observe a marginal increase in the sales-weighted markdowns, while Hershbein et al. (2022), using the Census of Manufacturers (CMF), find that a more pronounced increase in average markdowns since 1997.

Table 5: **Model Fit**

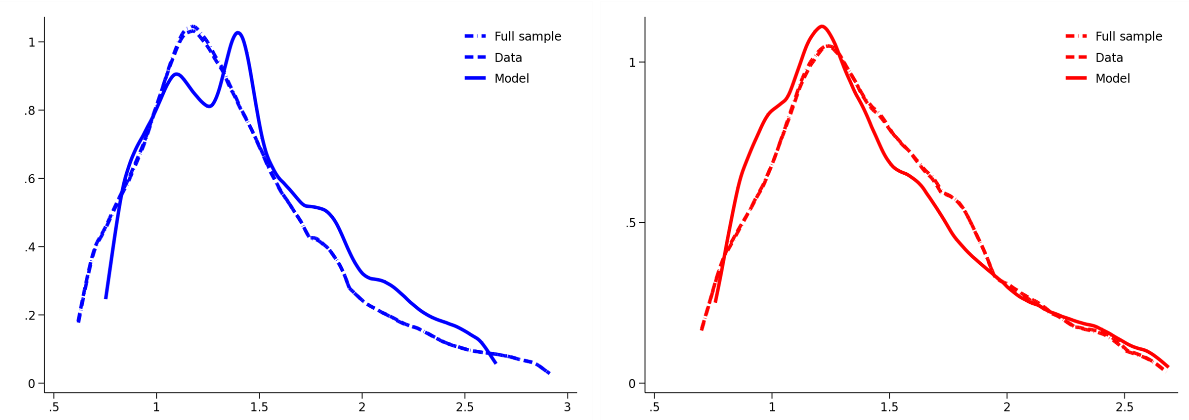
	1997		**2016**		**Δ**	
	Data	**Model**	**Data**	**Model**	**Data**	**Model**
Skill Premium	1.510	1.468	1.710	1.643	0.200	0.210
$\widehat{sw}_{inj}\frac{R_{inj}}{\sum_S W_{Sinj}S_{inj}}$	2.524	2.524	3.290	3.444	0.766	0.920

Notes: $\widehat{sw} = \frac{R_{inj}}{\int_j \sum_i R_{inj}}$ and $\frac{R_{inj}}{\sum_S W_{Sinj}S_{inj}}$ denotes the sales weight and revenue over the wage bill share for a given establishment.

weighted markups and markdowns have shifted to the right in 2016 compared to 1997, with a much more substantial shift for markups compared to the markdowns. The main insight is that the variance of markups has increased substantially.

Model Fit. Our model does reasonably well matching the level and the change of skill premium between 1997 and 2016 (Table 5). The model underpredicts the levels slightly, but tracks the data closely when it comes to the change over time. Furthermore, in Figure 4 we show that the model skill premium distribution has a close fit to the data in both 1997 and 2016.

Figure 4: Skill Premium Distribution



(a) 1997 (b) 2016

For the sales-weighted average of the revenue over the wage bill, the relevant comparison is for the year 2016 since we match this quantity between the data and the model in 1997 by construction to estimate α_N, the output elasticity of labor that we apply to revenue in the data to account for the absence of capital and intermediate inputs in our model. This is the key moment that we target to estimate N. As shown earlier, the wedge between revenue and the wage bill informs us about the market power of firms in their market. We find that the model provides a reasonable fit for this moment in the data.

6 Counterfactuals

Given the estimated parameters of the model, we perform a set of counterfactual experiments to quantify the contribution of market structure and technological change in driving the aggregate skill premium and within and between establishment inequality in Tables 6 and 7, respectively.

Quantifying the effect of N. To quantify the effect of N, we perform the following experiment: we hold fixed all parameters of the model fixed to their estimated values in 1997 and change N from its value of 16 in 1997 to its estimated value of 4 in 2016. We find that the skill premium goes up from 1.468 to 1.480 in this counterfactual, implying that the change in the market structure accounts for approximately 7% of the rise in the skill premium. Intuitively, as a result of the decline in N, even though the average markdowns for high and low-skilled workers have increased only marginally, the average markdown for low-skilled workers has increased relatively more compared to that of high-skilled workers, leading to a 7% increase in the skill premium.

We also find that the decline in competition in the output and the input market has substantial effects on the levels of the wages of the high and low-skilled workers. The wages of high-skilled workers drop by approximately 11.1% for high-skilled workers and 11.8 % for low-skilled workers. The level of wages drops despite the small changes in the average markdown of high and low-skilled workers because of the general equilibrium effect of the rise in the market power of firms in the product market. Since firms are exerting monopoly power in the goods market, the resulting increase in markups leads to a fall in the demand for goods and therefore labor. In Deb et al.

Table 6: **Results of the Counterfactual Exercises**

	Skill Premium			
	Level	Contribution	W_H	W_L
1997	1.468	–	100.00	100.00
N	1.480	6.86%	88.90	88.19
A_{Hinj}, A_{Linj}	1.939	269.14 %	236.01	178.81
$\bar{\phi}_H, \bar{\phi}_L$	1.245	-127.43 %	93.78	110.58
A_{Hinj}, A_{Linj} and N	1.942	270.86%	164.08	124.07
A_{Hinj}, A_{Linj} and $\bar{\phi}_H, \bar{\phi}_L$	1.625	89.71%	222.91	200.24
N and $\bar{\phi}_H, \bar{\phi}_L$	1.255	-121.71%	83.39	97.52
2016	1.643	100.00 %	155.23	138.68

Notes: The third column (titled Contribution) is constructed as follows: $(\frac{\kappa_{\text{CF}} - \kappa_t}{\kappa_{t+1} - \kappa_t}) \times 100$, where κ denotes the level of the skill premium, $t = 1997$, $t + 1 = 2016$ and CF denotes the counterfactual under consideration. Columns 4 and 5 are constructed as follows: each row is normalized to 100 in 1997 (i.e. each row is divided by the value in 1993 and multiplied by 100).

(2022), we take this insight forward and show that the rise in the output market power of firms accounts for 80% of the wage stagnation, and can account for the decoupling of productivity and wage growth in the US.

Quantifying the effect of A_{Hinj} and A_{Linj}. As before, we fix all parameters of the model to their values in 1997 and feed the technology distribution estimated in 2016 in the model. We find that this shift in the technology distribution accounts for approximately 269% of the total change in the skill premium. Changes in the productivity distributions are an important source of wage growth for both high and low-skilled workers, and relatively more for high-skilled. This evidence is in line with the previous literature highlighting the role of skill-biased technological change as being an important driver of the rise in the skill premium.

With regards to the level of wages, our counterfactual exercise shows that the average wage for high-skilled workers would have increased by 136% and that of low-skilled workers would have increased by 78%. The wages for both high and low-skilled workers increase because the productivity of both skills have improved as in Table 3, improving their marginal products.

Table 7: **Within and Between-Establishment Decomposition**

	Levels			Percentage Terms		
	Total	**Within**	**Between**	**Total**	**Within**	**Between**
1997	0.308	0.047	0.261	0.0	0.0	0.0
N	0.323	0.048	0.275	53.6	33.3	56.0
A_{Hinj}, A_{Linj}	0.400	0.087	0.313	328.6	1333.3	208.0
$\bar{\phi}_H, \bar{\phi}_L$	0.287	0.027	0.260	-75.0	-666.7	-4.0
A_{Hinj}, A_{Linj} and N	0.380	0.087	0.293	257.1	1333.3	128.0
A_{Hinj}, A_{Linj} and $\bar{\phi}_H, \bar{\phi}_L$	0.356	0.050	0.306	171.4	100.0	180.0
N and $\bar{\phi}_H, \bar{\phi}_L$	0.302	0.028	0.274	-21.4	-633.3	52.0
2016	0.336	0.050	0.286	100.0	100.0	100.0

Notes: The columns under percentage terms are calculated as follows: $(\frac{d_{\text{CF}}-d_t}{d_{t+1}-dt}) \times 100$, where $d \in \{\text{Total, Within, Between}\}$, $t = 1997$, $t+1 = 2016$ and CF denotes the counterfactual under consideration.

However, the actual increase in wages for high and low-skilled workers in 2016 relative to 1997 is 55% and 38%, respectively. This difference in wages between 2016 and the counterfactual economy with only technological change stems from two sources: the shift in the N and the change in the aggregate supply of high and low-skilled workers in the economy. To understand which of the two forces are important, we perform two counterfactuals. In the first counterfactual, we shift both technology and N jointly (row 5) and in the second counterfactual we jointly shift technology and $\bar{\phi}_S$ (row 6). The first counterfactual isolates the role of market power while the second counterfactual isolates the role of the labor supply disutility parameter. The results show that the shift in the N explains a substantial part of this decline. The shift in the labor supply lowers the wages of high-skilled workers but increases that of low-skilled workers. Hence, in the absence of market power, the effect of technological change on wages would have been higher than the one observed in 2016. In other words, market power impedes gains from technological change accruing to workers.

Within and Between Establishment Inequality. We perform the same decomposition as Bloom et al. (2016) at the establishment level to quantify how much of within and between establishment inequality can be attributed to the rising market power of a firm. In the notation of our model, the decomposition is written as follows

$$\mathbb{V}ar_k(w_t^{ki}) = \underbrace{\sum_i \omega_i \times \left\{ \frac{H_i(W_{Hi} - \overline{W}_t^i)^2 + L_i(W_{Li} - \overline{W}_t^i)^2}{H_i + L_i} \right\}}_{\text{Within establishment}} + \underbrace{\sum_i \omega_i [\overline{W}_t^i - \overline{W}_t^A]^2}_{\text{Between establishment}} \quad (30)$$

where k denotes an individual, ω_i is the employment share of establishment i in the entire economy, $\overline{W}_t^i$ is the average establishment wage and $\overline{W}_t^A$ is the average wage in the economy.[41]

In our model, we find that variance of (log) earnings increases over time. Roughly 11% of the total increase in the variance of earnings is due to an increase in within establishment inequality while the remaining 89% of the earnings are due to an increase in between establishment inequality.

To isolate the role of N and technology distribution in explaining the rise in within and between establishment inequality, we perform a series of counterfactuals by individually changing the values of these parameters from their estimated level in 1997 to 2016 while holding all other parameters constant. We find that the decline in N can explain 33% of the total change in within establishment inequality and approximately 56% of the total change in between-establishment inequality. We also find that skill biased technological change has substantially increased within establishment inequality by 1333% and between establishments inequality by 208% contributing to the bulk of the observed increase.[42]

The mechanism through which a decline in N affects within establishment inequality is reminiscent of the effect of N on the skill premium in the homogeneous firm case we showed above. As N declines, a differential increase in monopsony power of firms over high and low-skilled workers leads to an increase in within-establishment wage inequality.

[41] In equation (30) we sum over all establishments in the economy.

[42] In recent work, Cortes and Tschopp (2020) argue that an increase in the price sensitivity of consumer demand can lead to an increase in between firm wage inequality. They show that an increase in price sensitivity leads to a reallocation of consumer demand towards more productive firms. This leads to the exit of less productive firms from the economy and a rise in variance of productivity of the surviving firms.

7 Conclusion

In this paper, we address the question of how the rise in market power affects wage inequality. We provide a theoretical framework that augments the canonical supply-demand framework of Katz and Murphy (1992) to incorporate rich heterogeneity between firms, as well as market power through strategic interaction in the product market along with monopsony power in the labor markets. In addition to the race between technology ratio and relative skill supply as postulated by Tinbergen (1974), our model highlights an additional channel that affects the skill premium: the relative monopsony power for different skills. This enables us to show how an increase in market power, through declining competition, affects the skill premium and wage inequality.

To quantify the effect of market power, we take our model to microdata from the US Census Bureau. Here, we make a methodological contribution in showing how practitioners can estimate the parameters pertaining to within and between market substitutability of workers directly from the upward-sloping labor supply equation faced by establishments.

Furthermore, we remain agnostic regarding the true definition of a market. The key restriction we face is that it is impossible to observe which firms are competing with whom in the macroeconomy. To address these issues, we estimate a stochastic model of competition by randomly assigning establishments to markets and firms within industry classification. In addition, applying our framework to the micro data allows us to estimate an economy-wide productivity distribution consistent with the observed employment distribution. Our estimates provide evidence of increased dispersion in technology, Skill-Biased Technological Change, and a less competitive market structure between 1997 and 2016.

Our counterfactual exercise shows that a less competitive market structure alone explains approximately 7% of the rise in the skill premium as well as approximately an 11% decline in the level of wages. Finally, we also find that it explains approximately half of the total change in between-establishment inequality.

References

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): "The network origins of aggregate fluctuations," *Econometrica*, 80, 1977–2016.

ACEMOGLU, D. AND P. RESTREPO (2018): "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 108, 1488–1542.

——— (2019): "Robots and jobs: Evidence from US labor markets," *Journal of Political Economy*.

——— (2021): "Tasks, automation, and the rise in US wage inequality," Tech. rep., National Bureau of Economic Research.

AMITI, M., O. ITSKHOKI, AND J. KONINGS (2019): "International shocks, variable markups, and domestic prices," *The Review of Economic Studies*, 86, 2356–2402.

ATKESON, A. AND A. BURSTEIN (2009): "The race between education and technology," *American Economic Review*.

AZAR, J. AND X. VIVES (2021): "General equilibrium oligopoly and ownership structure," *Econometrica*, 89, 999–1048.

AZKARATE-ASKASUA, M. AND M. ZERECERO (2020): "The aggregate effects of labor market concentration," *Unpublished Working Paper*.

BAQAEE, D. R. AND E. FARHI (2017): "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," Tech. rep., National Bureau of Economic Research.

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2016): "It's where you work: Increases in the dispersion of earnings across establishments and individuals in the United States," *Journal of Labor Economics*, 34, S67–S97.

Berger, D., K. Herkenhoff, and S. Mongey (2022): "Labor market power," *American Economic Review*, 112, 1147–93.

Bloom, N., F. Guvenen, D. Price, J. Song, and T. Wachter (2016): "Firming Up Inequality," *NBER Working Paper*, 21199.

Burstein, A., V. Carvalho, and B. Grassi (2019): "Bottom-up markup fluctuations," in *2019 Meeting Papers*, Society for Economic Dynamics, 505.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36, S13–S70.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

Carvalho, V. M. and B. Grassi (2019): "Large firm dynamics and the business cycle," *American Economic Review*, 109, 1375–1425.

Carvalho, V. M. and A. Tahbaz-Salehi (2019): "Production networks: A primer," *Annual Review of Economics*, 11, 635–663.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins," *American Economic Review*, 101, 471–75.

Cortes, G. M. and J. Tschopp (2020): "Rising concentration and wage inequality," *Working Paper*.

De Loecker, J. (2011): "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 79, 1407–1451.

De Loecker, J., J. Eeckhout, and S. Mongey (2018): "Quantifying Market Power," Mimeo.

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

Deb, S., J. Eeckhout, A. Patel, and L. Warren (2022): "What drives wage stagnation: Monopsony or Monopoly?" Tech. rep., UPF mimeo.

Doraszelski, U. and J. Jaumandreu (2013): "R&D and productivity: Estimating endogenous productivity," *Review of economic studies*, 80, 1338–1383.

Ederer, F. and B. Pellegrino (2022): "A tale of two networks: Common ownership and product market rivalry," Tech. rep., National Bureau of Economic Research.

Edmond, C., V. Midrigan, and D. Y. Xu (2015): "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 105, 3183–3221.

——— (2018): "How Costly are Markups," NYU mimeo.

Eeckhout, J. (2020): "Comment on: Diverging Trends in National and Local Concentration," in *NBER Macroeconomics Annual 2020, Volume 35*, NBER.

Felix, M. (2021): "Trade, Labor Market Concentration, and Wages," *Job Market Paper*.

Gabaix, X. (2011): "The granular origins of aggregate fluctuations," *Econometrica*, 79, 733–772.

Giroud, X. and J. Rauh (2019): "State taxation and the reallocation of business activity: Evidence from establishment-level data," *Journal of Political Economy*, 127, 1262–1316.

Grassi, B. et al. (2017): "IO in IO: Competition and volatility in input-output networks," *Unpublished Manuscript, Bocconi University*.

HÅKANSON, C., E. LINDQVIST, AND J. VLACHOS (2021): "Firms and Skills The Evolution of Worker Sorting," *Journal of Human Resources*, 56, 512–538.

HALL, R. E. (2018): "New evidence on the markup of prices over marginal costs and the role of mega-firms in the us economy," Tech. rep., National Bureau of Economic Research.

HERSHBEIN, B., C. MACALUSO, AND C. YEH (2022): "Monopsony in the US Labor Market," *American Economic Review*, 112, 2099–2138.

KATZ, L. F. AND K. M. MURPHY (1992): "Changes in relative wages, 1963–1987: supply and demand factors," *The Quarterly Journal of Economics*, 107, 35–78.

KRUSELL, P., L. E. OHANIAN, J.-V. RÍOS-RULL, AND G. L. VIOLANTE (2000): "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 68, 1029–1053.

MELITZ, M. J. AND G. I. OTTAVIANO (2008): "Market size, trade, and productivity," *The review of economic studies*, 75, 295–316.

MERTENS, M. (2021): "Labor Market Power and Wage (In)Equality," *Working Paper*.

PATEL, A. (2021): "The Role of Firms in Shaping Job Polarization," Essex mimeo.

SONG, J., D. J. PRICE, F. GUVENEN, N. BLOOM, AND T. VON WACHTER (2018): "Firming Up Inequality*," *The Quarterly Journal of Economics*, 134, 1–50.

TINBERGEN, J. (1974): "Substitution of graduate by other labour," *Kyklos: international review for social sciences*.

Online Appendix

A Derivations

A.1 Household's optimization

OPTIMUM CONSUMPTION FUNCTIONS:

$$\max_{C_{inj},L_{inj},H_{inj}} U\left(C - \frac{1}{\bar{\phi}_L^{\frac{1}{\phi_L}}} \frac{L^{\frac{\phi_L+1}{\phi_L}}}{\frac{\phi_L+1}{\phi_L}} - \frac{1}{\bar{\phi}_H^{\frac{1}{\phi_H}}} \frac{H^{\frac{\phi_H+1}{\phi_H}}}{\frac{\phi_H+1}{\phi_H}}\right) \quad \text{s.t. } PC = LW_L + HW_H + \Pi$$

The solution to household's market-level demand function is a solution to

$$\max_{Y_j} \quad \left(\int_j J^{\frac{-1}{\theta}} Y_j^{\frac{\theta-1}{\theta}} dj\right)^{\frac{\theta}{\theta-1}} \text{ s.t} \int_J P_j Y_j dj \leq Z \tag{A31}$$

Then the optimal allocation is given by ;

$$\frac{\theta}{\theta-1}\left(\int_j J^{\frac{-1}{\theta}} Y_j^{\frac{\theta-1}{\theta}} dj\right)^{\frac{\theta}{\theta-1}-1} J^{\frac{-1}{\theta}} \frac{\theta-1}{\theta} Y_j^{\frac{\theta-1}{\theta}-1} = \lambda P_j \tag{A32}$$

This can be simplified as $J^{\frac{-1}{\theta}} Y^{\frac{-1}{\theta}} Y_j^{\frac{-1}{\theta}} = \lambda P_j$. Next multiply each side by Y_j and integrate across J to get $Y = \lambda \int_j P_j Y_j dj$.We define the market price index P such that $PY = \int_j P_j Y_j dj$ which would imply that $\lambda = P^{-1}$. Then plugging this into the first order condition delivers the market specific demand function.

$$Y_j = \left(\frac{1}{J}\right)\left(\frac{P_j}{P}\right)^{-\theta} Y \tag{A33}$$

The aggregate wage index can be recovered by multiplying both sides by P_j and integrating across markets.

$$P = \left(\left(\frac{1}{J} \right) \int_J P_j^{1-\theta} dj \right)^{\frac{1}{1-\theta}} \tag{A34}$$

We can apply a similar formulation to derive the establishment specific demand function; $Y_{inj} = \frac{1}{I} \left(\frac{P_{inj}}{P_j} \right)^{-\eta} Y_j$ and the market price index is;$P_j = \left(\frac{1}{I} \sum_i P_{inj}^{1-\eta} \right)^{\frac{1}{1-\eta}}$. Then the establishment specific demand function is given by;

$$Y_{inj} = \left(\frac{1}{J} \right) \left(\frac{1}{I} \right) \left(\frac{P_{inj}}{P_j} \right)^{-\eta} \left(\frac{P_j}{P} \right)^{-\theta} Y \tag{A35}$$

To derive the market specific inverse demand function we can write; $P_j = J^{\frac{-1}{\theta}} \left(\frac{Y_j}{Y} \right)^{\frac{-1}{\theta}} P$ and similarly at the establishment level as; $P_{inj} = I^{\frac{-1}{\eta}} \left(\frac{Y_{inj}}{Y_j} \right)^{\frac{-1}{\eta}} P_j$. Combining the last two equations we can get the establishment specific inverse demand curve as;

$$P_{inj} = \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} Y_{inj}^{\frac{-1}{\eta}} Y_j^{\frac{1}{\eta} - \frac{1}{\theta}} Y^{\frac{1}{\theta}} P \tag{A36}$$

Optimum Labor Supply functions : To derive equation (3), we follow Berger et al. (2022) and adjust for the love for variety by scaling the utility function. The household's aggregate labor supply function for each skill $S \in \{H, L\}$ can be derived from

$$\max_S \quad U \left(C - \frac{1}{\bar{\phi}_L^{\frac{1}{\phi_L}}} \frac{L^{\frac{\phi_L+1}{\phi_L}}}{\frac{\phi_L+1}{\phi_L}} - \frac{1}{\bar{\phi}_H^{\frac{1}{\phi_H}}} \frac{H^{\frac{\phi_H+1}{\phi_H}}}{\frac{\phi_H+1}{\phi_H}} \right) \quad \text{s.t. } PC = LW_L + HW_H + \Pi$$

Then the first order condition for $S \in \{H, L\}$ is

$$\frac{W_S}{P} = \overline{\phi}_S^{-\frac{1}{\phi_S}} S^{\frac{1}{\phi_S}} \iff S = \overline{\phi}_S \left(\frac{W_S}{P} \right)^{\phi_S}$$

which gives the aggregate labor supply function. The households optimum choice of allocation of labor across markets can be written as the solution to;

$$\min_{S_j} \quad \left(\int_j \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}_S}} S_j^{\frac{\hat{\theta}_S+1}{\hat{\theta}_S}} dj \right)^{\frac{\hat{\theta}_S}{\hat{\theta}_S+1}} \text{ s.t} \int_J W_{Sj} S_j dj \geq Z \tag{A37}$$

Then the optimal allocation is given by ;

$$\frac{\hat{\theta}_S}{\hat{\theta}_S+1} \left(\int_j \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}_S}} S_j^{\frac{\hat{\theta}_S+1}{\hat{\theta}_S}} dj \right)^{\frac{\hat{\theta}_S}{\hat{\theta}_S+1}-1} \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}_S}} \frac{\hat{\theta}_S+1}{\hat{\theta}_S} S_j^{\frac{\hat{\theta}_S+1}{\hat{\theta}_S}-1} = \lambda W_{Sj} \tag{A38}$$

This can be simplified as $\frac{1}{J}^{\frac{-1}{\hat{\theta}_S}} S^{\frac{-1}{\hat{\theta}_S}} S_j^{\frac{1}{\hat{\theta}_S}} = \lambda W_{Sj}$. Next multiply each side by S_j and integrate across J to get $S = \lambda \int_j W_{Sj} S_j dj$.We define the aggregate wage index W such that $WS = \int_j W_j S_j dj$ which would imply that $\lambda = W^{-1}$. Then plugging this into the first order condition delivers the market specific labor supply equation as a function of wage levels and aggregate labor supply.

$$S_j = \left(\frac{1}{J}\right) \left(\frac{W_{Sj}}{W_S}\right)^{\hat{\theta}_S} S \tag{A39}$$

The aggregate wage index can be recovered by multiplying both sides by W_j and integrating across markets.

$$W_S = \left(\left(\frac{1}{J}\right) \int_J W_{Sj}^{1+\hat{\theta}_S} dj \right)^{\frac{1}{1+\hat{\theta}_S}} \tag{A40}$$

We can apply a similar formulation to derive the establishment level labor supply; $S_{inj} = \left(\frac{1}{I}\right) \left(\frac{W_{Sinj}}{W_{Sj}}\right)^{\hat{\eta}_S} S_j$ and the market specific wage index is;$W_{Sj} = \left(\left(\frac{1}{I}\right) \sum_i W_{Sinj}^{1+\hat{\eta}_S} \right)^{\frac{1}{1+\hat{\eta}_S}}$. Then the establishment level labor supply curve is given by

$$S_{inj} = \left(\frac{1}{J}\right) \left(\frac{1}{I}\right) \left(\frac{W_{Sinj}}{W_{Sj}}\right)^{\hat{\eta}_S} \left(\frac{W_{Sj}}{W_S}\right)^{\hat{\theta}_S} S \tag{A41}$$

To derive the market specific inverse labor supply function we can write; $W_{Sj} =$

$\left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}_S}} \left(\frac{S_j}{S}\right)^{\frac{1}{\hat{\theta}_S}} W_S$ and similarly at the establishment level as; $W_{inj} = \left(\frac{1}{I}\right)^{\frac{-1}{\hat{\eta}_S}} \left(\frac{S_{inj}}{S_j}\right)^{\frac{1}{\hat{\eta}_S}} W_{Sj}$. Combining the last two equations we can get the establishment level inverse labor supply curve as;

$$W_{Sinj} = \frac{1}{J}^{\frac{-1}{\hat{\theta}_S}} \frac{1}{I}^{\frac{-1}{\hat{\eta}_S}} S_{inj}^{\frac{1}{\hat{\eta}_S}} S_j^{\frac{1}{\hat{\theta}_S} - \frac{1}{\hat{\eta}_S}} S^{-\frac{1}{\hat{\theta}_S}} W_S \tag{A42}$$

A.2 Solving the equilibrium

Optimal Firm Solution: There are N firms indexed by n in each sector. A firm owns I/N establishments. An establishment's sales share and wage bill share are denoted by s_{inj} and e_{Linj}, e_{Hinj}, respectively. As a result, the firm's sales share and wage bill share can be expressed as $s_{nj} = \sum_{i \in \mathcal{I}_{nj}} s_{inj}$ and $e_{Lnj} = \sum_{i \in \mathcal{I}_{nj}} e_{Linj}$ for the low-skilled and $e_{Hnj} = \sum_{i \in \mathcal{I}_{nj}} e_{Hinj}$ for the high-skill, respectively. Firm's problem here is to choose an employment level L_{inj}, H_{inj} for each establishment i simultaneously to maximize its profit. The FOC for input L_{inj} is derived here,

$$\left[P_{inj} + \frac{\partial P_{inj}}{\partial Y_{inj}} Y_{inj} + \sum_{i' \in \mathcal{I}_{nj} \setminus i} \left(\frac{\partial P_{i'nj}}{\partial Y_{inj}} Y_{i'nj}\right)\right] \frac{\partial Y_{inj}}{\partial L_{inj}} = \left[W_{Linj} + \frac{\partial W_{Linj}}{\partial L_{inj}} L_{inj} + \sum_{i' \in \mathcal{I}_{nj} \setminus i} \left(\frac{\partial W_{Li'nj}}{\partial L_{inj}} L_{i'nj}\right)\right] \tag{A43}$$

Note that $\frac{\partial P_{inj}}{\partial Y_{inj}} Y_{inj} = \left[-1/\eta + (1/\eta - 1/\theta) s_{inj}\right] P_{inj}$ and

$$\begin{aligned} \frac{\partial P_{i'nj}}{\partial Y_{inj}} Y_{i'nj} &= \frac{\partial P_{i'nj}/P_{i'nj}}{\partial Y_{inj}/Y_{inj}} \frac{P_{i'nj} Y_{i'nj}}{P_{inj} Y_{inj}} P_{inj} \\ &= \frac{\partial \log P_{i'nj}}{\partial \log Y_{inj}} \frac{s_{i'nj}}{s_{inj}} P_{inj} \\ &= \left[\left(\frac{1}{\eta} - \frac{1}{\theta}\right) s_{inj}\right] \frac{s_{i'nj}}{s_{inj}} P_{inj} \\ &= \left(\frac{1}{\eta} - \frac{1}{\theta}\right) s_{i'nj} P_{inj} \end{aligned} \tag{A44}$$

and similarly, $\frac{\partial W_{L,inj}}{\partial L_{inj}} L_{inj} = \left[1/\hat{\eta}_L + (1/\hat{\theta}_L - 1/\hat{\eta}_L) e_{L,inj}\right] W_{L,inj}$ and

$$\frac{\partial W_{Li'nj}}{\partial L_{inj}} L_{i'nj} = \left(\frac{1}{\hat{\theta}_L} - \frac{1}{\hat{\eta}_L}\right) e_{Li'nj} W_{Linj}. \tag{A45}$$

Combining these the FOC can be rewritten into

$$\left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right] P_{inj} \frac{\partial Y_{inj}}{\partial L_{inj}} = \left[1 + \frac{1}{\hat{\theta}_L} e_{nj} + \frac{1}{\hat{\eta}_L}(1 - e_{Lnj})\right] W_{Linj}, \tag{A46}$$

where markup and markdown are defined as

$$\begin{aligned} \mu_{inj} &\equiv \frac{P_{inj}}{MC_{inj}} = \left(1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}\left(1 - s_{nj}\right)\right)^{-1} \\ \delta_{Linj} &\equiv \frac{MRPL_{inj}}{W_{Linj}} = \left(1 + \frac{1}{\hat{\theta}_L} e_{Lnj} + \frac{1}{\hat{\eta}_L}\left(1 - e_{Lnj}\right)\right). \end{aligned} \tag{A47}$$

We can similarly derive the FOC for H_{inj} to get

$$\delta_{Hinj} \equiv \frac{MRPH_{inj}}{W_{Hinj}} = \left(1 + \frac{1}{\hat{\theta}_H} e_{Hnj} + \frac{1}{\hat{\eta}_H}\left(1 - e_{Hnj}\right)\right) \tag{A48}$$

Solving the Model. Start from the first order condition for high-skilled worker:

$$Y_{inj}^{\frac{1}{\sigma}} A_{L,inj}^{\frac{\sigma-1}{\sigma}} L_{inj}^{-\frac{1}{\sigma}} \cdot \left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right] P_{inj} = \left[1 + \frac{1}{\hat{\theta}_L} e_{L,nj} + \frac{1}{\hat{\eta}_L}(1 - e_{L,nj})\right] W_{L,inj} \tag{A49}$$

Similarly, we have a similar equation for a low-skilled worker

$$Y_{inj}^{\frac{1}{\sigma}} A_{H,inj}^{\frac{\sigma-1}{\sigma}} H_{inj}^{-\frac{1}{\sigma}} \cdot \left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right] P_{inj} = \left[1 + \frac{1}{\hat{\theta}_H} e_{H,nj} + \frac{1}{\hat{\eta}_H}(1 - e_{H,nj})\right] W_{H,inj}. \tag{A50}$$

By plugging into the inverse labor supply and inverse demand functions, we can rewrite each of these two conditions into:

$$\frac{1}{J}^{\frac{1}{\theta}}\frac{1}{I}^{\frac{1}{\eta}}(Y_{inj})^{-\frac{1}{\eta}}\left[\left(\frac{1}{I}^{\frac{1}{\eta}}\sum_{i}(Y_{inj})^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}\frac{(\theta-\eta)}{\eta\theta}}\right]\left[1-\frac{1}{\theta}\frac{\sum_{i'\in\mathcal{I}_{nj}}(Y_{inj})^{\frac{\eta-1}{\eta}}}{\sum_{i}(Y_{inj})^{\frac{\eta-1}{\hat{\eta}}}}-\frac{1}{\eta}\left(1-\frac{\sum_{i'\in\mathcal{I}_{nj}}(Y_{inj})^{\frac{\eta-1}{\eta}}}{\sum_{i}(Y_{inj})^{\frac{\eta-1}{\eta}}}\right)\right]\frac{\partial Y_{inj}}{\partial S_{inj}}Z_S \tag{A51}$$

$$=\frac{1}{J}^{\frac{-1}{\hat{\theta_S}}}\frac{1}{I}^{\frac{-1}{\hat{\eta_S}}}(S_{inj})^{\frac{1}{\hat{\eta_S}}}\left[\left(\frac{1}{I}^{\frac{-1}{\hat{\eta_S}}}\sum_{i}(S_{inj})^{\frac{\hat{\eta_L}+1}{\hat{\eta_S}}}\right)^{\frac{\hat{\eta_S}}{\hat{\eta_S}+1}\frac{(\hat{\eta_S}-\hat{\theta_S})}{\hat{\eta_S}\hat{\theta_S}}}\right]\left[1+\frac{1}{\hat{\theta_S}}\frac{\sum_{i'\in\mathcal{I}_{nj}}(S_{inj})^{\frac{\hat{\eta_S}+1}{\hat{\eta_S}}}}{\sum_{i}(S_{inj})^{\frac{\hat{\eta_S}+1}{\hat{\eta_S}}}}+\frac{1}{\hat{\eta_S}}\left(1-\frac{\sum_{i'\in\mathcal{I}_{nj}}(S_{inj})^{\frac{\hat{\eta_S}+1}{\hat{\eta_L}}}}{\sum_{i}(S_{inj})^{\frac{\hat{\eta_S}+1}{\hat{\eta_S}}}}\right)\right]$$

where $S\in\{H,L\}$, $Z_S=W_S^{-1}L_S^{1/\hat{\theta}_S}Y^{1/\theta}$ is the skill specific aggregate and the aggregate price P is normalized to 1. We will use these two equations to solve the model computationally as follows.

A.3 Algorithm to solve the model

We want to solve the economy with TFP $A_{H,inj}$ and $A_{L,inj}$.

- Guess three aggregates: W_H^k, W_L^k and Y^k, where k is the index of iteration.
- Given those three initial values, solve the first order conditions and learn H_{inj}, L_{inj} and Y_{inj} for each establishment.
- Compute $W_{H,inj}$, $W_{L,inj}$ and P_{inj} for each establishment using the inverse labor supply function for each skill and inverse demand function. Then aggregate the establishment wages $W_{H,inj}$, $W_{L,inj}$ into W_H^{k+1}, W_L^{k+1} and establishment output Y_{inj} to Y^{k+1} using the respective CES aggregators.
- Update the initial guess and iterate until all three aggregates converge $W_H^{k+1}=W_H^k$, $W_L^{k+1}=W_L^k$ and $Y^{k+1}=Y^k$ to get the equilibrium aggregates W_H^*, W_L^* and Y^*.
- Finally, compute everything else through the first order condition.

A.4 Algorithm to back out technology shocks

Algorithm

- Given that we can express the two first order conditions for each establishment only as a function of $A_{Sinj}, S_{inj} \forall i \in j$ in section (A.2), we begin by solving for $Z_S = W_S^{-1} S^{1/\hat{\theta}} Y^{1/\theta}$. We first use the aggregate labor supply function to substitute out W_S as a function of S using $W_S = \frac{S^{1/\varphi}}{\overline{\varphi}_S}$.

- Given our estimation of the labor supply function from step 1 of our estimation, we know $\hat{\eta}_S, \hat{\theta}_S, \overline{\varphi}_S$. Now $Z_S = S^{1/\hat{\theta}-1/\varphi} Y^{1/\theta} \overline{\varphi}_S$ where we only need to solve for Y. To do so, we use a two step procedure.

- **Step 1:** We guess $Y = \widetilde{Y}$ and solve for the $A_{Sinj} \forall i$. At this stage we identify the $\mu^*_{inj}, \delta^*_{Sinj}, W^*_{Sinj}, S^*_{inj}$. S^*_{inj} are establishment level skill specific employment which we use from the data, W^*_{Sinj} are model wages from the labor supply function and $\mu^*_{inj}, \delta^*_{Sinj}$ are independent of aggregate Y as they only depend on the relative A_{Sinj} within a sector.

- **Step 2:** In step 1 we identify $Y^* = \int_j \sum_i P_{inj} Y_{inj} dj$ as the firm level revenue's are independent of the guess $\widetilde{Y}$. Therefore, we can solve the model a second time using Y^* to retrieve the estimated A^*_{Sinj} distribution.[43]

A.5 Proofs

Proof of proposition 1

[43] An alternate way to solve for the aggregate Y^* would be to loop over guess $\widetilde{Y}$ until the goods market is in equilibrium.

Proof. From first order conditions, we know:

$$
\begin{aligned}
\kappa \equiv \kappa_{ij} &= \frac{A_{H,ij}^{\frac{\sigma-1}{\sigma}} H_{ij}^{-\frac{1}{\sigma}} \delta_{L,ij}}{A_{L,ij}^{\frac{\sigma-1}{\sigma}} L_{ij}^{-\frac{1}{\sigma}} \delta_{H,ij}} \\
&= \left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma}} \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L} e_{L,nj} + \frac{1}{\hat{\eta}_L}(1-e_{L,nj})}{1+\frac{1}{\hat{\theta}_H} e_{H,nj} + \frac{1}{\hat{\eta}_H}(1-e_{H,nj})}\right] \cdot \left(\frac{H}{L}\right)^{-\frac{1}{\sigma}} \\
&= \left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma}} \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L} e_{L,nj} + \frac{1}{\hat{\eta}_L}(1-e_{L,nj})}{1+\frac{1}{\hat{\theta}_H} e_{H,nj} + \frac{1}{\hat{\eta}_H}(1-e_{H,nj})}\right] \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H}\right)^{\frac{1}{\sigma}} \cdot \left(\frac{W_L}{W_H}\right)^{\frac{\phi}{\sigma}}
\end{aligned}
$$

By rearranging, we get the aforementioned expression.[44]

Now we have following properties:

1. From equation (16), when $N > 1$ it is clear that: $\partial\kappa/\partial\hat{\theta}_L < 0$, $\partial\kappa/\partial\hat{\eta}_L < 0$, $\partial\kappa/\partial\hat{\theta}_H > 0$ and $\partial\kappa/\partial\hat{\eta}_H > 0$.

2. Below, we show that $\partial\kappa/\partial A_H > 0$ and $\partial\kappa/\partial A_L < 0$.

3. About the increment in skill premium when changing N, we have:

$$
\frac{\partial\kappa}{\partial N} / \left(\frac{\kappa}{N}\right) = \frac{\sigma}{\sigma+\phi} \frac{N\left[\left(1+\frac{1}{\hat{\eta}_L}\right)\left(\frac{1}{\hat{\theta}_H}-\frac{1}{\hat{\eta}_H}\right) - \left(1+\frac{1}{\hat{\eta}_H}\right)\left(\frac{1}{\hat{\theta}_L}-\frac{1}{\hat{\eta}_L}\right)\right]}{\left[N\left(1+\frac{1}{\hat{\eta}_H}\right)+\frac{1}{\hat{\theta}_H}-\frac{1}{\hat{\eta}_H}\right]\left[N\left(1+\frac{1}{\hat{\eta}_L}\right)+\frac{1}{\hat{\theta}_L}-\frac{1}{\hat{\eta}_L}\right]}.
$$

   A sufficient condition for this term to be negative is: $\hat{\eta}_H < \hat{\eta}_L$ and $\frac{1}{\hat{\theta}_H} - \frac{1}{\hat{\eta}_H} < \frac{1}{\hat{\theta}_L} - \frac{1}{\hat{\eta}_L}$.

To get exact form for partial for the other results:

$$
\frac{\partial\kappa}{\partial\hat{\theta}_L} = \frac{\sigma}{\sigma+\phi}\left[\left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma+\phi}} \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H}\right)^{\frac{1}{\sigma+\phi}}\right] \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})}{1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})}\right]^{\frac{-\phi}{\sigma+\phi}} \frac{-\hat{\theta}_L^2}{N+\frac{1}{\hat{\theta}_H}+\frac{1}{\hat{\eta}_H}(N-1)}
$$

[44] We denote $\kappa = \frac{W_H}{W_L}$ and assume that $\phi_L = \phi_H = \phi$.

$$\frac{\partial \kappa}{\partial \hat{\eta}_L} = \frac{\sigma}{\sigma+\phi}\left[\left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma+\phi}} \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H}\right)^{\frac{1}{\sigma+\phi}}\right] \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})}{1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})}\right]^{\frac{-\phi}{\sigma+\phi}} \frac{-(1-\frac{1}{N})\hat{\eta}_L^2}{[1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})]}$$

$$\frac{\partial \kappa}{\partial \hat{\theta}_H} = \frac{\sigma}{\sigma+\phi}\left[\left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma+\phi}} \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H}\right)^{\frac{1}{\sigma+\phi}}\right] \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})}{1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})}\right]^{\frac{-\phi}{\sigma+\phi}} \frac{[1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})]N^{-1}\hat{\theta}_H^{-2}}{[1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})]^2}$$

$$\frac{\partial \kappa}{\partial \hat{\eta}_H} = \frac{\sigma}{\sigma+\phi}\left[\left(\frac{A_H}{A_L}\right)^{\frac{\sigma-1}{\sigma+\phi}} \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H}\right)^{\frac{1}{\sigma+\phi}}\right] \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})}{1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})}\right]^{\frac{-\phi}{\sigma+\phi}} \frac{[1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})](1-\frac{1}{N})\hat{\eta}_H^{-2}}{[1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})]^2}$$

$$\frac{\partial \kappa}{\partial A_H} = \frac{\sigma-1}{\sigma+\phi}A_H^{\frac{-\phi-1}{\sigma+\phi}}\left[\left(\frac{1}{A_L}\right)^{\frac{\sigma-1}{\sigma+\phi}} \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H}\right)^{\frac{1}{\sigma+\phi}}\right] \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})}{1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})}\right]^{\frac{\sigma}{\sigma+\phi}}$$

$$\frac{\partial \kappa}{\partial A_L} = -\frac{\sigma-1}{\sigma+\phi}A_L^{\frac{1-\phi-2\sigma}{\sigma+\phi}}\left[(A_H)^{\frac{\sigma-1}{\sigma+\phi}} \cdot \left(\frac{\bar{\phi}_L}{\bar{\phi}_H}\right)^{\frac{1}{\sigma+\phi}}\right] \cdot \left[\frac{1+\frac{1}{\hat{\theta}_L}\frac{1}{N}+\frac{1}{\hat{\eta}_L}(1-\frac{1}{N})}{1+\frac{1}{\hat{\theta}_H}\frac{1}{N}+\frac{1}{\hat{\eta}_H}(1-\frac{1}{N})}\right]^{\frac{\sigma}{\sigma+\phi}}$$

□

B Identification without endogeneity

In this Appendix we show that, under the assumption that the error term is uncorrelated with employment, $\hat{\eta}_S$, $\hat{\theta}_S$ can be identified from the following moments.

$$\hat{\eta}_S = \left(\frac{\mathbb{C}ov(\widetilde{S}_{inj},\ \widetilde{W^*}_{Sinj})}{\mathbb{V}ar(\widetilde{S}_{inj})}\right)^{-1} \tag{A52}$$

$$\hat{\theta}_S = \left[\left(\frac{\mathbb{C}ov(\ln S_j,\ \overline{\Omega}_{Sj})}{\mathbb{V}ar(\ln S_j)}\right) + \left(\frac{\mathbb{C}ov(\widetilde{S}_{inj},\ \widetilde{W^*}_{Sinj})}{\mathbb{V}ar(\widetilde{S}_{inj})}\right)\right]^{-1} \tag{A53}$$

where we denote

$$\widetilde{S}_{inj} = \ln S_{inj} - \overline{\ln S_j}, \qquad \widetilde{W^*}_{Sinj} = \ln W^*_{Sinj} - \overline{\ln W^*_{Sj}}, \qquad \overline{\ln X_j} = \frac{1}{I}\sum_i \ln X_{inj}$$

$$\overline{\Omega}_{Sj} = \mathbb{E}(\Omega_{Sinj}|J=j), \qquad \Omega_{Sinj} = \ln W^*_{Sinj} - \frac{1}{\hat{\eta}_S}\ln S_{inj}.$$

The moment condition in equation (A52) is equivalent to regressing the difference of log-employment from the mean of sectoral log-employment on difference of log-wages from the mean of sectoral log-employment. The moment condition in equation (A53) is equivalent to regressing the sectoral employment CES index on average sectoral wages (after removing the effect of average sectoral employment). Given that the sectoral CES index is a function of $\hat{\eta}_S$, we need to construct moments in equation (A52) and equation (A53) sequentially, starting with first retrieving the estimate of $\hat{\eta}_S$.

Deriving the moment conditions. To derive the moment conditions in equation (A52) and equation (A53), start by differencing out the sector-specific mean wages and mean employment from equation (17) to get the following expression:

$$\ln W^*_{Sinj} - \overline{\ln W^*_{Sj}} = \frac{1}{\hat{\eta}_S}(\ln S_{inj} - \overline{\ln S_j}) + (\varepsilon_{Sinj} - \overline{\varepsilon_{Sj}}) \tag{A54}$$

An OLS regression of equation (A54) helps us retrieve $\hat{\eta}_S$ and equation (A52) specifies the moments that helps us pin it down. Equipped with the estimate of $\hat{\eta}_S$, we can construct S_j, the CES index of sector-level employment. In the second step, we can then estimate the between-market substitution parameter $\hat{\theta}_S$ by relying on equation (17) and subtracting $\frac{1}{\hat{\eta}_S}\ln S_{inj}$ from $\ln W^*_{Sinj}$.

$$\ln W^*_{Sinj} - \frac{1}{\hat{\eta}_S}\ln S_{inj} \equiv \Omega_{Sinj} = \mathrm{c} + \left(\frac{1}{\hat{\theta}_S} - \frac{1}{\hat{\eta}_S}\right)\ln S_j + \varepsilon_{Sinj} \tag{A55}$$

To construct the moment in equation (A53), take sector-specific averages of both sides on equation (A55) and regress $\ln S_j$ on $\overline{\Omega}_{Sj}$ to retrieve the estimate of θ.

$$\overline{\Omega}_{Sj} = \mathrm{c} + \left(\frac{1}{\hat{\theta}_S} - \frac{1}{\hat{\eta}_S}\right)\ln S_j + \bar{\varepsilon}_{Sj} \tag{A56}$$

Simulation. To see the performance of the estimation strategy, we perform the following experiment. We simulate an economy with one skill and add a measurement error to the resulting inverse labor supply equation in equation (**??**). Following the strategy outlined above, we use equation (A54) and equation (A56) to estimate the

Table A1: Simulation results

	$\hat{\eta}_S$	$\hat{\theta}_S$	$\bar{\phi}_S$	$\hat{\eta}_S$	$\hat{\theta}_S$	$\bar{\phi}_S$
	$\varepsilon_{Sinj} \sim N(0, 0.2)$			$\varepsilon_{Sinj} \sim N(0, 2)$		
True Value	2.00	1.50	10.00	2.00	1.50	10.00
OLS	1.99	1.49	9.99	1.99	1.48	9.98
NLS	2.00	1.49	9.96	2.00	1.48	9.88
GMM	1.95	1.50	10.05	2.00	1.48	9.77

within and across market substitutability parameters. The results of the estimation are presented in Table (A1).

In summary, we show that under the assumption of zero correlation between the error term and log of employment, one can retrieve the structural parameters $\hat{\eta}_S$ and $\hat{\theta}_S$ using cross-sectional data on employment and wages in a two-step procedure by running OLS on equation (A54) and equation (A56).

Monte Carlo Simulation. In order to understand the bias that would stem if we ran OLS to estimate $\hat{\eta}_S, \hat{\theta}_S$ and $\bar{\phi}_S$ if the error term was correlated with employment, we run the following Monte Carlo simulation. We simulate the labor supply equation of the model, i.e. equation 17, as follows:[45]

$$\begin{aligned} \ln S^*_{inj} &= \ln S_{inj} + \rho \times \epsilon_{Sinj}, \\ \ln S_{inj} &\sim \mathbb{N}(\mu_S, \sigma_S^2), \\ \epsilon_{Sinj} &\sim \mathbb{N}(\mu, \sigma^2). \end{aligned} \tag{A57}$$

This implies that we can re-write equation 17 as follows:

$$\ln W_{Sinj} = c_j^* + \left(\frac{1}{\hat{\theta}_S} - \frac{1}{\hat{\eta}_S} \right) \ln S_j^* + \frac{1}{\hat{\eta}_S} \ln S^*_{inj} + \tilde{\epsilon}_{Sinj}. \tag{A58}$$

In the case that $\rho \neq 0$, equation A57 will lead to non-zero correlation between log

[45]For simplicity, we simulate only a cross-section. Consequently, we suppress the time notation.

Table A2: Monte Carlo Simulation: Bias from using the OLS

	$\hat{\eta}_S$	$\hat{\theta}_S$	$\bar{\phi}_S$
True Value	3.00	1.50	10.00
$\rho = -1.5$	9.75	6.80	963.69
	(0.115)	(0.383)	(198.87)
$\rho = -0.5$	3.75	2.01	16.84
	(0.015)	(0.042)	(0.64)
$\rho = 0.5$	3.75	2.01	16.80
	(0.014)	(0.047)	(0.67)
$\rho = 1.5$	9.75	6.82	969.96
	(0.121)	(0.408)	(480.34)

of employment and the error term in equation A58. To study the bias, we pick four different values of $\rho \in \{-1.5, -0.5, 0.5, 1.5\}$. For each of these values, we simulate an economy for which the labor supply equation holds.[46] Next, we try to recover the estimates of $\hat{\eta}_S, \hat{\theta}_S$ and $\bar{\phi}_S$ using OLS as specified in Appendix **??**. The results of this exercise is presented in Table A2. To get these results we assumed that $J = 500, I = 32, \mu_S = \mu = 0, \sigma_S^2 = \sigma^2 = 1, \hat{\eta}_S = 3.00, \hat{\theta}_S = 1.50, \bar{\phi}_S = 10.00$. As can be seen from Table A2, absolute deviation of ρ from 0 leads to an upward bias in the estimates of $\hat{\eta}_S, \hat{\theta}_S$ and $\bar{\phi}_S$. This bias increases substantially as ρ increases for all the three estimates.

C Robustness of the estimates of Labor Substitutability parameters

In this section, we provide the results of our robustness exercise. In the main text, we showed the results for the labor substitutability parameters where we randomly

[46] Since we are interested in understanding the sign and the magnitude of the bias, we do not solve our full economic model.

assigned establishments to market to constitute a market. In Table A3, we relax this assumption. Instead of assigning establishments randomly, we define a market as NAICS 6. Given this definition, we re-estimate our econometric model and the results are provided in Table A3.

Finally, we provide additional results to look at the role of geography on our estimated parameters. To do so, we define a market as NAICS 3 x MSA. The results of this exercise are provided in Table A4. With regards to our estimation that includes geography, we only focus on the markets that include establishments that belong to different states. The reason to do so is that if all establishments belong to the same state then there is no variation in taxes within a market and the market does not contribute anything to the overall identification of $\hat{\eta}_S$.

When we do not rely on random sampling within NAICS 6, the estimates of the η_H and η_L are 2.70 and 2.51 respectively, as compared to our baseline values of 2.53 and 2.42. On the other hand, the estimates of θ_H and θ_L are 1.93 and 1.87, respectively, as compared to 2.02 and 1.85. With regards to changing the definition of the labor market from NAICS 6 to NAICS 3 x Geo, we find the value of the substitutability parameters for both high and low-skilled workers increases (relative to the benchmark) and the difference between $\eta_S - \theta_S$ widens. For instance, the estimates of η_H and η_L are 5.39 and 6.41, respectively and that of θ_H and θ_L are 2.92 and 3.43. The only difference with regards to the baseline specification is that we do not include establishment fixed-effects in our regression.[47] These results imply that the firm's market power in the input market is lower compared to the benchmark since the elasticity of substitution is higher both within and across markets. Given this difference in estimates, we use these elasticities to measure the effect of N on the skill premium and between-establishment wage inequality and we find similar results.

[47] When we include the establishment fixed effect, we find that the estimates of labor substitutability parameters are theory inconsistent.

Table A3: **Estimates of Labor Substitutability Parameters: Tradeables without Random Sampling**

A. OLS and Second-Stage IV Estimates					
	OLS	IV		OLS	IV
	(1)	(2)		(3)	(4)
β_H	-0.177***	0.371***	γ_H	-0.0025***	0.148***
SE	0.0007	0.057	SE	0.0002	0.001
Market-Year SE	(0.002)	(0.075)	Market SE	(0.023)	(0.043)
β_L	-0.1075***	0.399***	γ_L	-0.011***	0.136***
SE	0.0007	0.051	SE	0.0003	0.001
Market-Year SE	(0.002)	(0.065)	Market SE	(0.025)	(0.041)
Market x Year FE	Yes	Yes	Market FE	Yes	Yes
Establishment FE	Yes	Yes	Year FE	Yes	Yes
B. Structural Parameters					
η_H	-5.65	2.70	θ_H	-5.73	1.93
η_L	-9.30	2.51	θ_L	-8.41	1.87
C. First-stage Regressions for the IV					
$\tau^H_{X(i)t}$	-	-0.013***	$\bar{\tau}^H_{jt}$	-	-0.015***
SE		0.001	SE		0.001
Sector-Year SE		(0.075)	Market SE		(0.001)
$\tau^L_{X(i)t}$	-	-0.261***	$\bar{\tau}^L_{jt}$	-	-0.276***
SE		0.001	SE		0.001
Market-Year SE		(0.059)	Market SE		(0.059)
Market x Year FE	-	Yes	Market FE	-	Yes
Establishment FE	-	Yes	Year FE	-	Yes
No. of obs (High-Skilled)	1,166,000	1,166,000		5900	5900
No. of obs (Low-Skilled)	1,166,000	1,166,000		5900	5900

Notes: Standard errors clustered at the market-year level for the first stage and at the sector level at the second stage are reported in the parenthesis. Non-clustered standard errors are reported without parenthesis. *** p < 0.01, ** p<0.05, *p<0.1. The significance stars correspond to clustered standard errors. Estimates of γ_S in columns 3 and 4 are conditional on the estimates of columns 1 and 2, respectively. Number of observations are common for both the first and the second-stage. The number of observations reflects rounding for disclosure avoidance. $\tau^S_{X(i)t}$ denotes the co-efficient infront of taxes in the first-stage regression for the estimate of β_S. The same instrument is used separately, first to estimate β_H and then to estimate β_L.

Table A4: **Estimates of Labor Substitutability Parameters: NAICS 3 x MSA**

A. OLS and Second-Stage IV Estimates					
	OLS	IV		OLS	IV
	(1)	(2)		(3)	(4)
β_H	0.079***	0.185***	γ_H	0.063***	0.157***
SE	0.0006	0.063	SE	0.0004	0.002
Sector-Year SE	(0.002)	(0.074)	Sector SE	(0.013)	(0.044)
β_L	-0.029***	0.156***	γ_L	-0.063***	0.136***
SE	0.0007	0.086	SE	0.0004	0.001
Sector-Year SE	(0.001)	(0.092)	Sector SE	(0.013)	(0.044)
Sector x Year FE	Yes	Yes	Sector FE	Yes	Yes
Establishment FE	Yes	Yes	Year FE	Yes	Yes
B. Structural Parameters					
η_H	12.62	5.39	θ_H	7.05	2.92
η_L	34.98	6.41	θ_L	9.23	3.43
C. First-stage Regressions for the IV					
$\tau^H_{X(i)t}$	-	0.031***	$\bar{\tau}^H_{jt}$	-	-0.110***
SE		0.004	SE		0.001
Sector-Year SE		(0.005)	Sector SE		(0.022)
$\tau^L_{X(i)t}$	-	-0.024***	$\bar{\tau}^L_{jt}$	-	-0.127***
SE		0.004	SE		0.001
Sector-Year SE		(0.004)	Sector SE		(0.023)
Sector x Year FE	-	Yes	Sector FE	-	Yes
Establishment FE	-	Yes	Year FE	-	Yes
No. of obs (High-Skilled)	497,000	497,000		5800	5800
No. of obs (Low-Skilled)	497,000	497,000		5800	5800

Notes: Standard errors clustered at the market-year level for the first stage and at the sector level at the second stage are reported in the parenthesis. Non-clustered standard errors are reported without parenthesis. *** p < 0.01, ** p<0.05, *p<0.1. The significance stars correspond to clustered standard errors. Estimates of γ_S in columns 3 and 4 are conditional on the estimates of columns 1 and 2, respectively. Number of observations are common for both the first and the second-stage. The number of observations reflects rounding for disclosure avoidance. $\tau^S_{X(i)t}$ denotes the co-efficient infront of taxes in the first-stage regression for the estimate of β_S. The same instrument is used separately, first to estimate β_H and then to estimate β_L.