Barcelona School of Economics

WORKING PAPERS

What Drives Wage Stagnation: Monopsony or Monopoly?

BSE Working Paper 1361| September 2022

Shubhdeep Deb, Jan Eeckhout, Aseem Patel, Lawrence Warren

What drives Wage Stagnation: Monopsony or Monopoly?*

Shubhdeep Deb — UPF Barcelona†
Jan Eeckhout — UPF Barcelona‡
Aseem Patel — University of Essex§
Lawrence Warren — US Census Bureau¶

September 22, 2022

Abstract

Wages for the vast majority of workers have stagnated since the 1980s while productivity has grown. We investigate two coexisting explanations based on rising market power: 1. Monopsony, where dominant firms exploit the limited mobility of their own workers to pay lower wages; and 2. Monopoly, where dominant firms charge too high prices for what they sell, which lowers production and the demand for labor, and hence equilibrium wages economy-wide. Using establishment data from the US Census Bureau between 1997 and 2016, we find evidence of both monopoly and monopsony, where the former is rising over this period and the latter is stable. Both contribute to the decoupling of productivity and wage growth, with monopoly being the primary determinant: in 2016 monopoly accounts for 75% of wage stagnation, monopsony for 25%.

Keywords. Market Power. Monopsony. Monopoly. Markdowns. Markups. Wage Stagnation. Concentration. HHI.

---

*We benefited from feedback from several seminar audiences, from a detailed discussion by Monica Morlacco as well as from comments by Peter Neary, Christian Moser, and Roman Zarate. Renjie Bao and Wei Hua provided excellent research assistance. Eeckhout gratefully acknowledges support from the ERC, Advanced grant 882499, and from ECO2015-67655-P and Deb from "la Caixa" Foundation (ID 100010434) fellowship (code LCF/BQ/DR19/11740003). 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-FY22-CED006-0027.

†shubhdeep.deb@upf.edu

‡jan.eeckhout@upf.edu – ICREA-BSE-CREI

§aseem.patel@essex.ac.uk

¶lawrence.fujio.warren@census.gov

1 Introduction

With the rise of market power by dominant firms, researchers have recognized the effect on the economy as a whole (such as the decline in the startup rate and business dynamism), and on the labor market in particular, with a declining labor share and wage stagnation. Dominant firms affect wages in two ways: through monopsony power in the labor market and through monopoly power in the goods market.

In the absence of sufficient competition by other employers where workers can get jobs, dominant firms exert *monopsony* power and can hire their own workers at wages below their productivity. This is the reverse of monopoly power in the goods market (see Robinson, 1933). Due to mobility frictions across geography and sectors, captive workers cannot exert their outside options easily. As a result, a dominant firm faces an upward sloping labor supply function, which would be flat in a competitive labor market. Exploiting their market power, firms hire workers at wages below the marginal revenue product of workers, where the gap between marginal revenue product of labor and wages is the markdown. More monopsony power thus leads to lower wages.

There is also a negative effect on wages resulting from goods market power, even if the labor market is perfectly competitive. If firms exert *monopoly* power in the goods market, and there are enough of those dominant firms, then there is also a general equilibrium effect on wages. A firm that has market power in its own market sets higher prices relative to cost, denoted by the markup. As a result of higher prices, demand falls and therefore so does production. This does not directly affect wages, because even though a firm has market power in its narrowly defined market, that market is small relative to the economy. However, when there is an overall increase in market power in many goods markets, we see an aggregate effect on wages. The decline in wages follows from the economy-wide decline in the demand for labor, which results in falling wages for workers in the aggregate, not just those employed by the firms that charge higher prices.

The objective of this paper is double. First, we lay out a model of the economy where labor market power (monopsony) and goods market power (monopoly) coexist. This permits us to determine the total effect of market dominance on wages. The economic mechanism establishes how wages become decoupled from productivity as a result of the rise in market power: wages stagnate even as productivity continues to grow. Most importantly, with this mechanism we can decompose the total effect of market power on wages into the sources that are due to goods market power and those that are due to labor market power. The theoretical model builds on the framework of Deb et al. (2021) that analyzes how market power affects wage inequality and the skill premium. In this tractable general equilibrium model of the macroeconomy, a small number of heterogeneous firms compete in markets with goods and labor market power jointly, and both markups and markdowns are simultaneously determined.

The second objective is to quantify and measure the effect of market power on wages, decomposed into monopsony and monopoly power. We use establishment-level data from the US Census Bureau – the Longitudinal Business Database (LBD) – to estimate both markups and markdowns simultaneously. This is challenging because both are a function of marginal revenue and marginal cost, which we typically do not directly observe in the data. In addition, while the concept of market power is very clear, the practical problem is that we do not easily observe it.[1] We therefore use the structure of our macroeconomic model as well as data on wages, employment and revenue to estimate the labor supply elasticities, the firm-level productivities and the market structure.

Our quantitative exercise yields the following results. First, we find a clear increase in the estimated parameter for market power economy-wide between 1997 and 2016. The number of firms competing in the market drops, thus leading to more concentration. Second, the estimated average markup increases from 1.69 to 2.2, while average markdowns have increased only marginally from 1.37 to 1.4. The markup trend is consistent with the findings in De Loecker et al. (2020), with the increase mainly driven by the upper percentiles of the markup distribution. Third, the increase in market power leads to wage stagnation and can explain the rising disconnect between productivity and wages. Fourth, in a series of counterfactual exercises to decompose the contribution to wage stagnation, we find that goods market power contributes to the majority of the wage stagnation. In 2016, the relative contribution of monopoly power to the reduction in wages was 75%, with 25% due to monopsony. When we consider wage growth, the share of monopoly is even higher. This leads us to conclude that monopoly is the main determinant of wage stagnation. There is monopsony power – workers are paid below their marginal revenue product – but it is virtually constant over time, and as a

[1] In the absence of direct observation, researchers have relied on indirect measures such as concentration ratios, most commonly the Herfindahl-Hirschman Index (HHI). The problem is that concentration ratios are often inadequate measures of market power, especially in a macroeconomic setting, and can result in misleading conclusions (see for example Berry, Gaynor, and Scott Morton (2019a), Syverson (2019), and Eeckhout (2020)).

result, it contributes little to the widening gap between wages and productivity.

Methodologically, we borrow heavily from the approach in Deb et al. (2021). In the absence of detailed data on the demand system of each individual market and in our quest to measure market power economy-wide, we model the market structure in a stochastic manner. Our notion of the market structure is stochastic in the sense that we randomly assign establishments from the same industry. The key parameter that captures the extent of market power is the number of competitors in a market, expressed as the number of competing firms operating the establishments within each market. Fewer competitors give rise to a systematic change in the distribution of markups/markdowns, revenue, wages and output. We then obtain an estimate of the number of competitors as well as technology parameters by matching the revenue and wage bill distribution observed in the data to our model. While this approach is certainly far less detailed than the demand approach for a specific, narrowly defined market (see Berry et al., 1995), our approach does allow us to get an estimate of the extent to which there is market power at the aggregate, macroeconomic level.

RELATED LITERATURE. Our approach to use a macroeconomic model with endogenous market power in the output market and the general equilibrium effect on wages builds on earlier work by Atkeson and Burstein (2008) and De Loecker et al. (2021). We combine these models with insights from Berger, Herkenhoff, and Mongey (2022) who model market power in the labor market. Our model thus combines output and input market power in one framework, building on our earlier paper Deb et al. (2021) where we study the contribution of different sources of market power in explaining the rise in skill premium and wage inequality. Market power in our model has three main components: 1. The underlying heterogeneity in the establishment productivity distribution; 2. The extent of competition as measured by the number of firms competing within markets; and 3. The extent of frictions faced by the household in the goods and labor market.[2]

The way we estimate markups using an economy-wide demand system and a random market structure is complementary to the production approach for measuring markups, as in Hall (1988), De Loecker and Warzynski (2012) and De Loecker et al. (2020).[3] With sufficiently de-

[2]Our model is also related to Azar and Vives (2021) who have a finite number of firms competing in both input and output markets and where an increase in common ownership leads to an increase in concentration.

[3]This approach typically estimates a production function in order to back out the output elasticities, see Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015) and De Loecker et al. (2020).

tailed data, that approach can also be used to jointly estimate markups and markdowns, as in De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), Hershbein, Macaluso, and Yeh (2022) and Morlacco (2017). Our approach to use the structure of our model has the added advantage that it allows us to calculate welfare, do counterfactuals, and most importantly, it allows us to decompose the joint effect of goods and labor market power on wage stagnation, the primary objective of this paper.

We use micro data at the establishment-level, and the structure of our model allows us to back out the individual productivity for each establishment. This approach builds on Patel (2021) who uses micro data to measure establishment productivity and analyze the role of firms in driving job polarization. The estimated productivities and the model's tractability in general equilibrium allow for the derivation of prices, revenue and wages at the micro level. In our case, the distribution of revenues and wage bill implied by our model is used to estimate the market structure in the economy by matching these equilibrium outcomes with the micro data. This allows us to estimate the market structure for the goods and labor markets in the US and to track its evolution over time.

Our paper is related to a large literature on monopsony and the measurement of markdowns. The objective of this literature is to estimate to what extent a firm can set the wage below the worker's marginal revenue product. The literature has measured labor market power in four distinct ways. The first approach measures labor market power by estimating the elasticity of the labor supply curve faced by an individual firm, which when significantly less than infinity indicates monopsony power. Early quasi-experimental studies by Staiger, Spetz, and Phibbs (2010), Falch (2010) and Matsudaira (2014) find mixed evidence on the extent of monopsony power.[4] However, recent studies by Dube, Jacobs, Naidu, and Suri (2018), Azar, Marinescu, and Steinbaum (2019b) and Azar, Berry, and Marinescu (2019a) find estimates that indicate the presence of pervasive monopsony power. Two recent papers focus attention on the identification techniques from the Industrial Organization literature: Berry, Gaynor, and Scott Morton (2019b) and Goolsbee and Syverson (2019). Goolsbee and Syverson (2019) uses data on the academic labor market and interpret the frictions as caused by the inability to substitute between occupations. They find variation in monopsony power across ranks, between tenured faculty whose high paying outside options are limited, and lecturers.

[4]While this could be a result of the intrinsic nature of specific markets analyzed in each study, Manning (2011) suggests that the large variance in estimates could also stem from the use of the simple models of monopsony.

A second approach is to establish a negative relationship between the level of employer concentration in the labor market and wages in that market as in Azar, Marinescu, and Steinbaum (2017) and Rinz (2018). Using this method, several papers find diverging trends between local concentration and national concentration (mostly HHI), both in the output market and the labor market (see amongst others Rossi-Hansberg, Sarte, and Trachter (2018), Rinz (2018) and Hershbein, Macaluso, and Yeh (2022)).[5] For articles that point out the limitations of using HHI, see amongst others, Syverson (2019), Eeckhout (2020), Berry, Gaynor, and Scott Morton (2019b), and Miller et al. (2021). Eeckhout (2020) illustrates that the decline in local concentration measures is mechanical: as population grows, more firms locate in a given area, which automatically decreases the denominator of the HHI formula, irrespective of whether competition increases or decreases. Furthermore, Berry, Gaynor, and Scott Morton (2019b) highlight that this strand of literature suffers from "severe measurement problems, and worse conceptual problems" and suggest that the studies that do not use measures of concentration (HHI), but instead use alternative approaches such as the production function approach, can mitigate some of these limitations. In this paper, we go in that direction by using a structural model to estimate a production function in an environment with variable market structure. This is an alternate way of measuring market power that circumvents the thorny issue of static market definitions.

A third approach uses the production function estimation approach to measure markdowns from detailed firm-level balance sheet data as in Hershbein et al. (2022), Mertens (2021), Azkarate-Askasua and Zerecero (2020),Morlacco (2017) and Rubens (2021), in addition to the papers mentioned above. Specifically, Hershbein et al. (2022) use data from U.S. manufacturers and find an average markdown of 1.53 and sharply rising monopsony power since the early 2000's.

Finally, several papers use structural models to measure monopsony power. Like ours, this approach assumes a labor supply mechanism with frictions. When workers cannot costlessly move to another job, the employer can exert monopsony power. In one strand of the literature, the source of the rents are search frictions. Manning (2003, 2011) formulates a "generalized model of monopsony", which builds on the on-the-job search model of Burdett and Mortensen (1998). The match surplus inherent in the search frictions permits firms to extract some of the rents and pay workers below their marginal product.[6] Instead of search frictions, here we

[5] Though Ganapati (2019) finds increasing concentration at all levels, both national and local.

[6] A variation of a model with a different search technology is by Jarosch, Nimczik, and Sorkin (2019).

model the frictions due to the imperfect ability to substitute among differentiated jobs. We build directly on Berger et al. (2022) which allows us to model and measure both goods and labor market power simultaneously.

This paper also contributes to the literature studying the decoupling of wages from productivity. Machin (2016), Stansbury and Summers (2017), Eeckhout (2021), and Greenspon, Stansbury, and Summers (2021) document the divergence between productivity and pay in the United States. Our model offers a novel mechanism and new insights regarding why wages stagnate in the absence of technological regress. In a world of perfect competition, productivity growth mirrors the growth of wages. After all, workers are paid their marginal revenue product and any growth in technology must show up in wage growth. In the presence of market power, however, this no longer holds. Market power drives a wedge between the real wage paid and the productivity of the worker. As a result, as market power increases, this wedge increases, leading to the de-coupling of productivity and wages over time.

OUTLINE. The remainder of this article is organized as follows. In Section 2, we lay out the theoretical framework followed by the quantitative analysis in Section 3. In Section 4, we present our estimation results. In Section 5, we perform counterfactual experiments to quantify the contribution of monopoly and monopsony in explaining wage stagnation in the US. We conclude in Section 6.

2 The Model

Our model builds on Deb et al. (2021) where firms have market power both in the product market and labor market and where they hire both high and low skilled workers. Instead, in the current model all workers are homogeneous. Market power results from three forces: 1. Differentiated products and jobs in the goods and labor markets, respectively; 2. Heterogeneity in the productivity of establishments; and 3. A finite number of firms competing in a market. For tractability, we assume that the market definition of goods coincides with the market definition of the labor inputs, implying that the same set of firms compete in both the product and input market simultaneously.

ENVIRONMENT. We consider a static economy that consists of two types of decision makers, representative households containing a continuum of workers/consumers, and a continuum of heterogeneous establishments. There is a continuum of markets indexed by subscript j with total measure J and a finite number of establishments equal to I in each market j. Establishments are indexed by i and are heterogeneous in their productivity. Each market also has a finite number of firms equal to N that are indexed by n. We assume that the number of establishments I in each market is constant, and each firm owns I/N establishments. We denote the set of all the establishments i that are owned by firm n in market j by $\mathcal{I}_{nj} = \{i \mid i \text{ in firm } n, \text{ in sector } j\}$. The main advantage of this multi-establishment setup is that as the number of competing firms N changes, the preference structure remains constant as the number of varieties I within each market is constant.[7] Firms within each market j have market power due to imperfect competition in both the goods and labor market between firm n and the remaining $-n$ firms in the market. A representative household consumes the bundle of goods C_{inj} and supplies labor L_{inj} to establishments in each market.

HOUSEHOLDS. The representative household chooses the demand for the establishment's output as well as its labor supply to each establishment to maximize utility. The household preferences for consumption of the differentiated final goods is modeled as in Atkeson and Burstein (2008) and De Loecker et al. (2021) while the households preferences over differentiated jobs is modeled as in Berger et al. (2022).[8] The household solves the following problem

$$V = \max_{C_{inj}, L_{inj}} \left(C - \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{L^{\frac{\phi+1}{\phi}}}{\frac{\phi+1}{\phi}} \right) \quad \text{s.t.} \quad PC = LW + \Pi \tag{1}$$

[7]We do not think of this multi-establishment setup as a strict representation in the data, but rather as a modeling tool to measure market power that may stem from collusion, common ownership, firms with a changing product mix... This choice to model multi-establishment firms has two practical advantages: we can change the market structure without changing preferences, and we can randomly assign establishments under different market structures without changing the number of them. For an alternative approach with single-establishment firms where the preferences do change as N changes, see amongst many others De Loecker et al. (2021).

[8]In order to keep preferences constant as market structure N changes we eliminate the love for variety by using J and I as scalars.

where the aggregate and market specific consumption and labor indices are:

$$C = \left(\int_j J^{-\frac{1}{\theta}} C_j^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}} \ , \ C_j = \left(\sum_i I^{-\frac{1}{\eta}} C_{inj}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \tag{2}$$

$$L = \left(\int_j J^{\frac{1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj \right)^{\frac{\hat{\theta}}{\hat{\theta}+1}} \ , \ L_j = \left(\sum_i I^{\frac{1}{\hat{\eta}}} L_{inj}^{\frac{\hat{\eta}+1}{\hat{\eta}}} \right)^{\frac{\hat{\eta}}{\hat{\eta}+1}} \tag{3}$$

and Π are the aggregate profits redistributed lump sum to the household. For the preferences over goods, the within-market elasticity of substitution is η, and the between-market elasticity is θ. We assume that $\eta > \theta$, so goods within a market are more substitutable than goods between markets. For the labor market, $\hat{\eta}$ and $\hat{\theta}$ denote the within and between-market elasticities of substitution for jobs. We assume $\hat{\eta} > \hat{\theta}$, which implies that jobs are more substitutable within a market than between markets.

FIRMS AND MARKET STRUCTURE. Firms make production decisions according to Cournot quantity competition.[9] There are N firms that compete within each market and own I/N heterogeneous establishments. Establishments operate under a linear, single input production technology $Y_{inj} = A_{inj} L_{inj}$. Each firm n in market j chooses the quantity of production Y_{inj} for each establishment it owns in set $\mathcal{I}_{nj}$. In their optimal decision, they take into account the quantity decisions of all the other the firms $-n$ in its market. In addition, given our multi-establishment setup, firms also internalize the interaction between the establishments that it owns. Since there is a continuum of markets, there is no strategic interaction between firms from different markets, only within markets. In our framework, the aggregate price P and wage W, also affect the individual firms' optimal decisions of quantity supplied and labor demanded.

Moreover, given imperfect substitutability of goods and labor inputs, firms have market power in both the goods and the labor market and therefore optimize subject to a downward sloping demand function and an upward sloping labor supply function faced by each of its establishments.

We solve for the static Cournot-Nash equilibrium in this economy. For firm n in market j, the objective is to maximize profits by choosing output for all its establishments, taking as

[9] All our results immediately extend to Bertrand price competition with differentiated goods. Everything is identical except for the residual demand elasticity and labor supply elasticity that establishments face.

given the behavior of all competing firms $-n$ in the market:

$$\Pi_{nj} = \max_{Y_{inj}} \sum_{i \in \mathcal{I}_{nj}} \left[P_{inj}(Y_{inj}, Y_{-inj}) Y_{inj} - W_{inj}(L_{inj}, L_{-inj}) L_{inj} \right] \tag{4}$$

$$\text{s.t. } Y_{inj} = A_{inj} L_{inj}.$$

The strategic interaction between firms acts through the demand for goods $P_{inj}(Y_{inj}, Y_{-inj})$ as well as through the supply for labor $W_{inj}(L_{inj}, L_{-inj})$. We now first solve for the optimal household consumption and labor supply decision.

HOUSEHOLD OPTIMAL SOLUTION. Taking product prices P_{inj} and wages W_{inj} as given, the household chooses optimal consumption bundles C_{inj} and labor supply L_{inj} to maximize utility subject to the household budget.

The first order conditions for consumption C_{inj} of each good and of labor supply L_{inj} for each job satisfy:

$$C_{inj}(P_{inj}, P_{-inj}, P, C) = \frac{1}{J}\frac{1}{I} P_{inj}^{-\eta} P_j^{\eta-\theta} P^{\theta} C \tag{5}$$

$$L_{inj}(W_{inj}, W_{-inj}, W, L) = \frac{1}{J}\frac{1}{I} W_{inj}^{\hat{\eta}} W_j^{\hat{\theta}-\hat{\eta}} W^{-\hat{\theta}} L \tag{6}$$

where the market-specific price and wage indices P_j, W_j and aggregate indices P and W are given by:

$$P_j = \left(\sum_i \frac{1}{I} P_{inj}^{1-\eta} \right)^{\frac{1}{1-\eta}} \ , \ P = \left(\int_j \frac{1}{J} P_j^{1-\theta} dj \right)^{\frac{1}{1-\theta}} \tag{7}$$

$$W_j = \left(\sum_i \frac{1}{I} W_{inj}^{1+\hat{\eta}} \right)^{\frac{1}{1+\hat{\eta}}} \ , \ W = \left(\int_j \frac{1}{J} W_j^{1+\hat{\theta}} dj \right)^{\frac{1}{1+\hat{\theta}}}. \tag{8}$$

Market clearing in the goods and labor markets imply that the aggregate price P and wage index W satisfy:

$$PC = \int_J \sum_i P_{inj} C_{inj} dj \ , \ WL = \int_J \sum_i W_{inj} L_{inj} dj. \tag{9}$$

FIRM OPTIMAL SOLUTION. An establishment's sales share and wage bill share are denoted by s_{inj} and e_{inj}, 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 $e_{nj} = \sum_{i\in\mathcal{I}_{nj}} e_{inj}$, respectively. The firm's solution to the optimization problem (4) with respect to the output Y_{inj} of each of its i establishments satisfies:

$$P_{inj} + \frac{\partial P_{inj}}{\partial Y_{inj}} Y_{inj} + \sum_{i'\in\mathcal{I}_{nj}\setminus i} \left(\frac{\partial P_{i'nj}}{\partial Y_{inj}} Y_{i'nj} \right) = \frac{1}{A_{inj}} \left[W_{inj} + \frac{\partial W_{inj}}{\partial L_{inj}} L_{inj} + \sum_{i'\in\mathcal{I}_{nj}\setminus i} \left(\frac{\partial W_{i'nj}}{\partial L_{inj}} L_{i'nj} \right) \right] \quad (10)$$

where $\mathcal{I}_{nj} \setminus i$ is the set of all other establishments owned by firm n (except for establishment i), and where prices P_{inj} and wages W_{inj} are a function of the actions of the competitors Y_{i-nj}. Notice that the firm solves this condition for each establishment i, while at the same time taking into account the effect that the choice in establishment i has on establishments i' within the same firm n. In other words, the firm jointly maximizes over all its establishments. At the extreme, where $N = 1$, the firm solves for the outcome with perfect collusion between all establishments in the market.

Cournot competition in the input and output market gives us closed form solutions for the inverse demand elasticity ϵ^P_{inj} and inverse labor supply elasticity ϵ^W_{inj}, which can be expressed in terms of market shares in the goods market and labor market, respectively (see the Appendix (A.2) for the derivation). Because the firm optimizes over all of its establishments simultaneously, the relevant market share is the firm's *total* market share s_{nj} in the goods market and e_{nj} in the labor market. The first-order condition can then be written as;

$$P_{inj} \Big[1 \underbrace{- \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj})}_{\epsilon^P_{inj}} \Big] A_{inj} = W_{inj} \Big[1 + \underbrace{\frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj})}_{\epsilon^W_{inj}} \Big] \quad (11)$$

We define our markup $\mu_{inj} = \frac{P_{inj}}{MC_{inj}}$ and markdown $\delta_{inj} = \frac{MRPL_{inj}}{W_{inj}}$ for each establishment as

$$\mu_{inj} = \frac{1}{1 + \epsilon^P_{inj}} = \left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj}) \right]^{-1} \quad \text{and} \quad \delta_{inj} = 1 + \epsilon^W_{inj} = \left[1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj}) \right]. \quad (12)$$

where MC_{inj} is the marginal cost and $MRPL_{inj}$ is the marginal product of labor at each establishment.

LIMIT CASES. The limit cases of our model conveniently nest a spectrum of competition frameworks and provide intuition for how firm heterogeneity and market structure affect market power in the model.

Homogeneous Establishments. If there is no heterogeneity in productivity, then the sales shares and wage bill shares are identical for all firms and equal to $\frac{1}{N}$. Without heterogeneity, the terms for the markup and markdown are identical for all firms and given by:

$$\mu_{inj} = \left[1 - \frac{1}{\theta}\frac{1}{N} - \frac{1}{\eta}\left(1 - \frac{1}{N}\right)\right]^{-1} \quad \text{and} \quad \delta_{inj} = \left[1 + \frac{1}{\hat{\theta}}\frac{1}{N} + \frac{1}{\hat{\eta}}\left(1 - \frac{1}{N}\right)\right] \tag{13}$$

Monopolistic and Monopsonistic competition. We can increase competition in the economy by increasing the number of firms competing in each market. As $N \to \infty$, the sales share and wage bill share converges to 0 for all firms. The notion of differentiated markets also disappears, leaving one elasticity of substitution for each term. The resulting markups and markdowns are:

$$\mu_{inj} = \frac{\eta}{\eta - 1} \quad \text{and} \quad \delta_{inj} = \frac{\hat{\eta} + 1}{\hat{\eta}}. \tag{14}$$

This is similar to Melitz (2003), where there is a continuum of heterogeneous firms, yet despite this heterogeneity each firm has a constant homogenous markup. Note that even with $N \to \infty$ markups and markdowns are strictly larger than 1 because goods and labor are not perfect substitutes (they are, only when $\eta, \hat{\eta} \to \infty$).

Alternatively, we can also consider a case where $N = 1$ in all markets. In this case, there is only substitutability across markets, and we reach the upper bound for markups and markdowns in the model:

$$\mu_{inj} = \frac{\theta}{\theta - 1} \quad \text{and} \quad \delta_{inj} = \frac{\hat{\theta} + 1}{\hat{\theta}}. \tag{15}$$

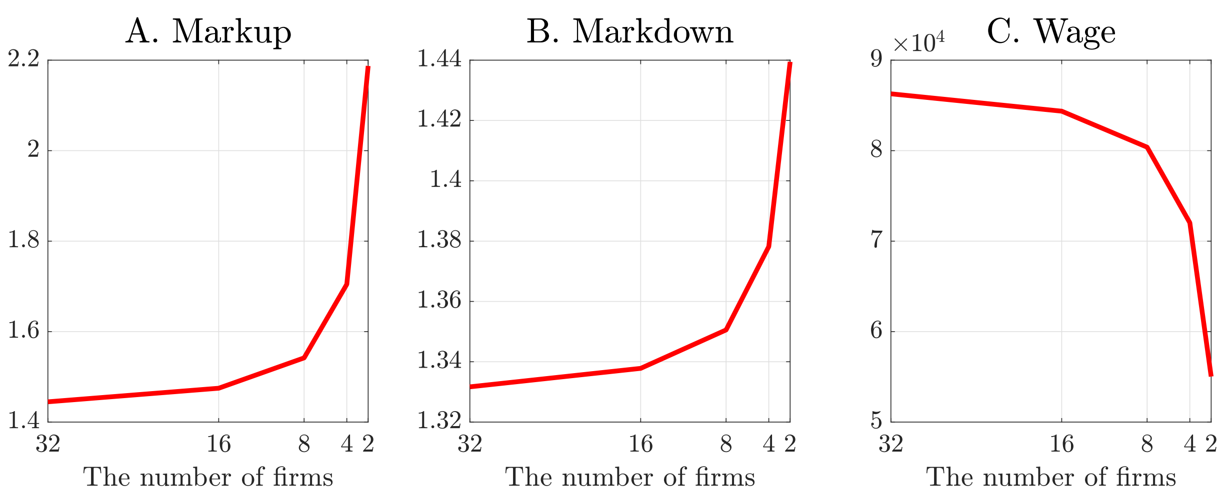
Perfect competition. Finally, competition also increases when the elasticity of substitution of goods and jobs increases within and between markets. Moving to the perfect substitutability case, we have (1) $\eta \to \infty$, (2) $\theta \to \infty$, (3) $\hat{\eta} \to \infty$, (4) $\hat{\theta} \to \infty$ and firms become price takers.

Therefore, the markup and markdown in this case converge to 1.

$$\mu_{inj} = 1 \quad \text{and} \quad \delta_{inj} = 1 \tag{16}$$

COMPARATIVE STATICS. Figure 1 shows how markups, markdowns, and the average wage in the economy change as we change market structure.[10] As the number of competitors in a local market N declines, markets become more concentrated and as a result, markups and markdowns increase as seen in panels A and B. In panel C we see that the average wage in the economy declines as the number of competitors declines. As markets become more concentrated, firms charge higher markups and higher markdowns. Monopsonistic firms pay lower wages and in the aggregate, a decline in labor demand further reduces the economy-wide wage.

Figure 1: **Comparative Statics**



Notes: We use the structural parameters that we estimate in Section 4 to construct the comparative statics plot above.

3 Quantitative Analysis

DATA. For our analysis, we use establishment-level micro data from the Census Bureau's Longitudinal Business Database (LBD). The LBD combines Economic Census, survey, and administrative data sources on employer businesses and covers the universe of employer estab-

[10] In the comparative statics we exogenously set $I = 32$, and consider an ownership structure such that each firm owns an equal number of establishments as we vary N. As a result we show the results for $N \in \{2, 4, 8, 16, 32\}$, such that each firm own I/N establishments.

lishments in the United States. The LBD provides information on ownership and organization, employment, payroll, revenue, industry (NAICS), and geography. We use annual data from 1997 to 2016, during which revenue information is available at the firm level from the Revenue Enhanced LBD. For multi-establishment firms, we impute revenues to each establishment by the establishment's payroll share within the firm. From this frame, our sample consists of firms in tradeable sectors as outlined in Delgado et al. (2014).[11] We further restrict the sample to C Corporations in the continental US (excluding AK, HI, and US territories). We drop all establishments with missing establishment, firm, or geographic (county and MSA) identifiers as well as missing employment or payroll. We winsorize establishment-level employment and average wages at the 1st and 99th percentiles, and additionally drop establishments with 5 or fewer employees. Wages and Revenue are deflated to 2002 dollars.

MARKET DEFINITION. A key object in our model is the market structure, N, which governs the extent of competition in the economy. A high N implies a large number of competing firms in a market, and therefore low market power while a low N implies high market power with high markups, high markdowns and lower wages in equilibrium. In order to estimate this notion of competition we need to define a market, a key ingredient in the Industrial Organization literature. In a macroeconomic setting with firms across different industries and geographies, it is virtually impossible to identify a market in order to define the set of competitors for each firm. As a result, we adopt a stochastic notion of market structure in which firms are equally likely to compete with each other in a narrowly defined NAICS 6 industry. To do so, we start by randomly assigning establishments within narrowly defined NAICS 6 industries into markets of size I. These I establishments within each market are then assigned ownership stochastically to N firms that each own I/N establishments.

Despite this random assignment, the model preserves some key predictions as we vary N which allows us to use data on revenue and wage bill at establishments to estimate the extent of competition in the economy. Just like the measurement of Total Factor Productivity (TFP) as a Solow residual, we interpret our estimation of market structure N as a residual that explains the evolving relationship between revenue and wage bill in the data. As argued in Deb et al. (2021), although our approach is much less detailed than the traditional approach of identifying com-

[11] We use the following 2-digit NAICS sectors as our sample of tradeable goods sectors: 11, 21, 31, 32, 33, and 55.

petitors in a narrowly defined industry, our stochastic notion of market structure does make progress in identifying the extent of competition in the economy using rich establishment-level micro data.

QUANTIFYING THE MODEL. Our quantification exercise closely follows Deb et al. (2021). In this section we provide a summary of the key arguments. We refer the interested reader to Deb et al. (2021) for technical details related to identification.

We estimate the model in two steps. In the first step, we estimate the labor substitutability parameters $\hat{\eta}$ and $\hat{\theta}$ that are key to estimating markdowns and the labor supply elasticity. In the second step, we jointly estimate N, the number of firms in a market and the distribution of establishment-level productivity. To do so, we first guess a value of N and estimate the distribution of productivity in the economy that is consistent with the employment distribution observed in the Census micro data. We then estimate the market structure N to match the sales-weighted average of the ratio of revenue over wage bill in the data. We calibrate the preference parameters (η, θ, ϕ, I) exogenously and keep them constant over time (see Table 1).

Table 1: Exogenous Parameters

Variable	**Value**	**Definition**	**Paper**
θ	1.2	Product market: Between-market elasticity	De Loecker et al. (2021)
η	5.75	Product market: Within market elasticity	De Loecker et al. (2021)
ϕ	0.25	Aggregate Labor Supply Elasticity	Chetty et al. (2011)
I	32	Establishments in each market	Externally set

STEP 1: ESTIMATING LABOR SUBSTITUTABILITY PARAMETERS. To estimate the labor substitutability parameters, we rely on the labor supply equation (17) that contains information on both $\hat{\eta}$ and $\hat{\theta}$.

$$W_{inj} = \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} W. \tag{17}$$

To ease the exposition of our estimation strategy, we begin by re-writing equation (17) in logs

$$\ln W^*_{injt} = k_{jt} \ + \ \gamma \ln L_{jt} \ + \ \beta \ln L_{injt} \ + \ \underbrace{\alpha_{inj} \ + \ \epsilon_{injt}}_{\varepsilon_{injt}} \tag{18}$$

where we define $\ln W^*_{injt} = \ln W_{injt} + \varepsilon_{injt}$, $k_{jt} = \ln J_t^{-\frac{1}{\theta}} I_{jt}^{-\frac{1}{\eta}} L_t^{-\frac{1}{\theta}} W_t$, $\beta = \frac{1}{\eta}$ and $\gamma = \left(\frac{1}{\theta} - \beta\right)$.[12]

The error term, ε_{injt}, captures misspecification in wages between the data and the model. We further assume that the error term has a permanent establishment-specific component that we denote by α_{inj}. This assumption will allow us to exploit within-establishment variation to estimate the parameters of interest. Note that the labor supply equation does not depend on N, the total number of firms competing in a market. This is critical since it allows us to estimate the labor substitutability parameters without knowing the value of N in Step 1.[13]

We use Two-Stage Least Squares (2SLS) to estimate the parameters β and γ, sequentially. Equipped with the estimates of these parameters, we retrieve our structural parameters of interest. We proceed by outlining our strategy to estimate β, followed by γ.[14]

To control for endogeneity arising from the correlation between the log of employment and the error term, we instrument $\ln L_{injt}$ with state corporate taxes, $\tau_{X(i)t}$. We think of variation in taxes as an exogenous shock to an establishment's labor demand function which allows us to identify the labor substitutability parameters that characterize the labor supply function. This is similar to the approach adopted by Felix (2021) who instead relies on import tariff reductions in Brazil in the 1990s as the exogenous variation to estimate the labor substitutability parameters in a model of oligopsonistic competition.

In practice, we exploit the longitudinal structure of the LBD and merge state-level corporate income tax rates from Giroud and Rauh (2019), giving us an unbalanced panel from 1997-2011.

[12] Note that we have introduced the time subscript t as we rely on the panel dimension of our data to estimate the key parameters of interest. Furthermore, we have added a jt subscript to I in the expression for k_{jt}. This is because with the random assignment, we allow the size of the market j to evolve over t as establishments enter and exit our sample.

[13] Our estimation strategy is such that once we have estimated the labor substitutability parameters from Step 1, backed out the establishment-level productivities and estimated N from Step 2, these primitives will endogenously generate L_{inj} in the model that matches exactly the L_{inj} of each establishment observed in the micro data. These employment levels would then generate wages through the upward-sloping labor supply function in equation (17). The difference between wages in the data and the model will be equal to ε_{injt}.

[14] From the equation (18), observe that while we observe wages and employment in the data, we do not directly observe the establishment fixed effect α_{inj}, market-year specific constant, L_{jt} and k_{jt}. We control for α_{inj} by including establishment fixed effects in our estimation. To control for k_{jt} and L_{jt}, we include an interaction of market and year-fixed effects. These two controls allow us to exploit within-establishment variation while controlling for time shocks that vary by market.

We estimate our time-invariant labor elasticity parameters using this panel.

Once we get an estimate of β (and implicitly $\hat{\eta}$) from Eq. (18), we proceed to estimate γ by relying on the following equation derived from Eq. (18),

$$\overline{\Omega}_{jt} = k_{jt} + \gamma \ln L_{jt} + \bar{\varepsilon}_{jt} \tag{19}$$

where we define,[15]

$$\overline{\Omega}_{jt} = \mathbb{E}_{jt}\left[\ln W^{*}_{injt} - \frac{1}{\hat{\eta}} \ln L_{injt}\right] \text{ and } \bar{\varepsilon}_{jt} = \mathbb{E}_{jt}\left[\varepsilon_{injt}\right].$$

Like in the estimation of β, we control for potential endogeneity between $\ln L_{jt}$ and $\bar{\varepsilon}_{jt}$ by relying on an instrument. Specifically, we instrument for $\ln L_{jt}$ by $\bar{\tau}_{jt}$, the average tax-rate within a given market j. Intuitively, we exploit the time variation in market-level employment and wages to estimate γ while controlling for year-specific shocks that are common across markets.

Last, we use the aggregate labor supply equation to estimate the labor disutility parameter, $\bar{\phi}_t$, where ϕ denotes the Frisch elasticity.

$$\ln W_t = \frac{1}{\phi} \ln \frac{1}{\bar{\phi}_t} + \frac{1}{\phi} \ln L_t. \tag{20}$$

We calibrate the value of the Frisch elasticity, ϕ, to be equal to 0.25, from Chetty et al. (2011). This allows us to estimate the value of $\bar{\phi}_t$, one for each year, by inverting Eq. (20).[16]

ESTIMATION SAMPLE FOR STEP 1. We closely follow the steps outlined in Deb et al. (2021) to construct the sample that we use to estimate the labor substitutability parameters. To extend the stochastic assignment of establishments to markets wile retaining the panel dimension of our data, we proceed as follows. First, we randomly assign establishments to markets, conditional on NAICS 6 in the year 1997. We ensure that there are at most 32 establishments in each market. An establishment assigned to a given market will remain in the same market for all years that

[15]We cannot control for k_{jt} by including a sector and a year fixed effect, separately, in the equation (19). The sector-year-specific constant, k_{jt}, includes the total number of establishments within a market. An implicit assumption that we make is that $\mathbb{E}(I_{jt}\bar{\tau}_{jt}) = 0$, which implies that there is no correlation between state-level taxes and the size of a market. Recent work by Giroud and Rauh (2019) has argued that market size may potentially be correlated with taxes. However, this is unlikely to be true in our case as our definition of a market is NAICS-6 which straddles establishments across multiple states.

[16]Like in Deb et al. (2021), we implicitly assume that there is no measurement error in aggregate wages. Hence, we assume that $\ln W^{*}_{t} = \ln W_t$.

we observe it in the data. For years after 1997, we randomly assign the new establishments entering the data to one of the pre-existing markets created in 1997. As a result, the size and the composition of the market will evolve randomly over time depending on the entry and exit of establishments in our sample. The baseline results of the estimation of labor substitutability parameters are based on the sample of random assignment using the panel data. We perform a robustness exercise where we estimate the same parameters without random assignment. These results are presented in Appendix A.6.

STEP 2: BACKING OUT THE ESTABLISHMENT'S PRODUCTIVITIES AND ESTIMATING N. For any given N, in order to back out the technology distribution, we use the establishment's first-order condition.[17] To solve for the productivity parameters, A_{inj}, we reformulate the inverse demand function, the inverse labor supply function and the production function along with the sales share and wage bill share only as a function of the technology and employment and other exogenous parameters of the model. This gives us a system of I equations and I unknowns within each market.

In order to back out the TFP, we first need to define a market. To do so, we rely on the methodology we described under Market Definition in Section 3. Within each industry of 6-digit NAICS, we randomly assign establishments to markets of size I.[18] Given the distribution of L_{inj}^{data} within each market and a value of $N \in \{2, 4, 8, 16, 32\}$, the system of non-linear equations allow us to back out A_{inj} for each establishment in each market, and gives us a distribution of productivities $G(A_{inj}; N)$. The solution gives us $Y_{inj} = A_{inj} L_{inj}$ for all establishments, which is aggregated to Y. Once we solve for the aggregate Y, we pin down the level of the economy which gives us establishment-specific revenue R_{inj}.

This model-generated distribution of productivities $G(A_{inj}; N)$ is conditional on market structure N, and as a result, so is the revenue distribution. We can show that the revenue in the model is monotonically declining in N. Revenue can be written as $R_{inj} = \mu_{inj} \delta_{inj} W_{inj} L_{inj}$. To see this, note that the distribution $G(A_{inj}; N)$ maps to the same employment distribution in the data for each N, and given the estimates of $\hat{\eta}$ and $\hat{\theta}$ from Step 1, the employment distribu-

[17] In appendix (A.4) we describe in detail the derivation of the first-order condition (A27) only as a function of the observed employment L_{inj}, productivities A_{inj} and aggregates and explain our algorithm to solve for the aggregates.

[18] We restrict our sample of establishments in these randomly assigned markets to those with non-missing revenue. We truncate the revenue distribution by dropping establishments above the 99th percentile in revenue by year.

tion maps to the same wage distribution making both W_{inj} and L_{inj} independent of N at this stage. At the same time, both μ_{inj} and δ_{inj} are strictly decreasing in N, implying that the revenue R_{inj} predicted by the model is strictly decreasing in N. Equivalently, as markets become more concentrated (as N declines), the ratio of revenue to the wage bill increases in the model. As in Deb et al. (2021), we use this monotonicity of $R_{inj}/(W_{inj}L_{inj})$ with respect to N to estimate market structure 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:

$$N^* = \min_{N \in \{2, 4, 8, 16, 32\}} \left[\mathbb{E}\left(\widehat{sw}^D_{inj}\psi^D_{inj}\right) - \mathbb{E}\{\widehat{sw}^M_{inj}(N)\psi^M_{inj}(N)\} \right]^2 \tag{21}$$

where $\widehat{sw}^D_{inj} = \frac{R_{inj}}{\int_j \sum_i R_{inj} dj}$ denotes the sales-share and ψ^D_{inj} is the revenue over wage bill ratio of establishment i in the data while $\widehat{sw}^M_{inj}$ and ψ^M_{inj} denote the same quantities in the model.

Finally, we adjust the revenue in the data using $R^{\text{Adjusted}}_{inj} = \alpha_L R^{\text{data}}_{inj}$ to make it comparable to our model with labor as the only input.[19] We pin down α_L in 1997 such that market structure N is 16 in 1997.[20] In the following years we hold the value of α_L constant and estimate N by matching the sales weighted average of revenue over wage bill in the data and the model.

4 Results

We now present the results of our estimation: the labor substitutability parameters, the estimated market structure, and the evolution of aggregate markups and markdowns as well as the kernel densities.

LABOR SUPPLY ELASTICITIES. We present the OLS and IV estimates of our reduced form parameters $\beta = \frac{1}{\hat{\eta}}$ and $\gamma = \frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}$ in Table 2. Our results show that our OLS estimates display a substantial downward bias relative to our IV estimates. In fact, the negative value of our OLS coefficient for β implies that wages and employment are inversely related, yielding a downward-sloping labor supply curve which is inconsistent with theory. Our instrumental

[19]We model output only as a function of labor, while in the data output could be a function of labor, capital and materials $Y_{inj} = A_{inj}L^{\alpha_L}_{inj}K^{\alpha_K}_{inj}M^{\alpha_M}_{inj}$ such that the revenue in the data is a function of all inputs and not just labor. To make our revenue in the model comparable to that in the data we adjust the revenue in the data $R^{\text{Adjusted}}_{inj} = \alpha_L R^{\text{data}}_{inj}$.

[20]Given the monotonic relation between revenue in the model and N, there exists an α_L such that the sales weighted average of revenue of wage bill in the data (after adjustment using α_L) exactly equals the sales weighted average of revenue over wage bill in the model.

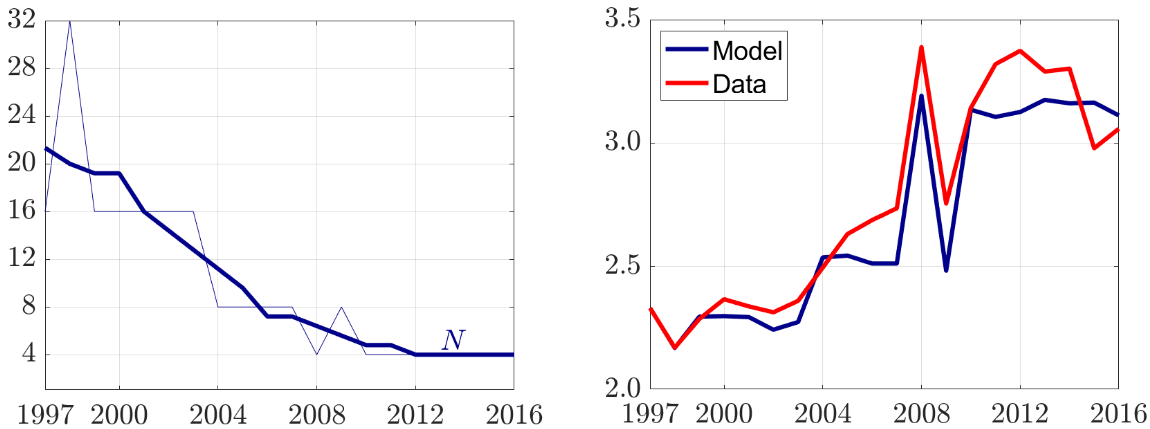
variables estimates should be free of bias, and it is reassuring that they imply that labor supply is upward sloping. Further, looking at the structural parameters $\hat{\eta}$ and $\hat{\theta}$, we see that the degree of substitutability of labor within markets, $\hat{\eta}$, is higher than the substitutability of labor across markets, $\hat{\theta}$, and both are positive. Panel B of Table 2 displays the value of our structural parameters from our estimation. The estimates of $\hat{\eta}$ and $\hat{\theta}$ are tightly linked to the distribution of markdowns in the model, as they define the upper and lower bounds of the support of the distribution. The lower bound of the markdown distribution is $(\hat{\eta}+1)/\hat{\eta} = 1.32$ and the upper bound is $(\hat{\theta}+1)/\hat{\theta} = 1.53$, implying that workers' wages can be anywhere between $1/1.53 = 0.65$ and $1/1.32 = 0.76$ percent of their marginal revenue product of labor.

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	-0.197***	0.323***	γ	0.110***	0.207***
SE	0.0005	0.051	SE	0.0002	0.006
Market-Year SE	(0.001)	(0.053)	Market SE	(0.002)	(0.048)
Market x Year FE	Yes	Yes	Market FE	Yes	Yes
Establishment FE	Yes	Yes	Year FE	Yes	Yes
B. Structural Parameters					
η	-5.08	3.10	θ	-11.47	1.89
C. First-stage Regressions for the IV					
$\tau_{X(i)t}$	-	-0.003***	$\bar{\tau}_{jt}$	-	-0.023***
SE		0.0002	SE		0.0003
Market-Year SE		(0.0002)	Market SE		(0.004)
Market x Year FE	-	Yes	Market FE	-	Yes
Establishment FE	-	Yes	Year FE	-	Yes
No. of obs	2,559,000			2,674,000‡	

Notes: Standard errors clustered at the market-year level for the first stage and at the market 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 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. $\tau_{X(i)t}$ denotes the co-efficient infront of taxes in the first-stage regression for the estimate of β.
‡ Denotes the number of weighted observations.

Figure 2: **Estimated Market Structure and Model Fit**



(a) Estimated value of N

(b) $\widehat{sw}_{inj}\psi_{inj}$: Data vs. Model

Table 2, Panel C also provides the first-stage estimates of our IV estimation. In both the establishment-level regression for β and the market-level regression for γ, we find a negative relationship between taxes and employment. For β, this relationship is at the establishment level while for γ, the relationship is between the average market-level tax and corresponding market-level employment and both coefficients are statistically significant. This negative relationship between employment and taxes is consistent with the findings of Giroud and Rauh (2019) and is also used in the instrumental variables approach employed by Berger et al. (2022).[21]

Market structure, markups and markdowns. Figure 2a plots the evolution of the estimated market structure N for each year.[22] We calibrate $\alpha_L = 0.3$ in order to fix $N = 16$ in 1997. Holding this value of α_L fixed in the subsequent years, we find that the number of competing firms within a market decreases gradually from 32 in 1998 to 4 in 2016. This is consistent with the evidence on increasing concentration at the national level as well as the recent work

[21]Despite using the same underlying data and obtaining the same reduced-form estimate for β, our estimate of $\hat{\theta}$ is higher and the estimate of $\hat{\eta}$ is lower than in Berger et al. (2022). However, there are three important differences between our methodology and theirs that can explain this difference. First, they estimate these parameters on local labor markets, which they define as 3-digit NAICS industry groups within a Commuting Zone. Second, they rely on Indirect Inference to estimate these parameters while we take the theory-derived labor supply equation directly to the data. Finally, the labor supply function is at the level of the establishment in our framework while it is at the level of the firm in theirs.

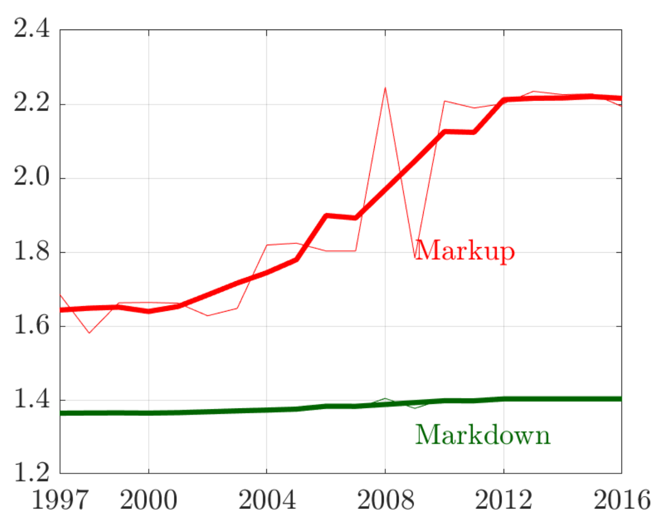
[22]Throughout the paper when we plot two lines for the same variable, the thick lines correspond to 5 year centered moving averages and thinner lines correspond to estimated or model values.

of De Loecker et al. (2021), who also estimate a model of imperfect competition with strategic interactions in the output market and show that competition in the aggregate economy has declined.

While we remain agnostic about the source of this decline in N, possible explanations to rationalize it are common ownership (Ederer and Pellegrino (2022)), laxer antitrust enforcement, technological change that leads to higher returns to scale.

In Figure 2b, we plot the sales weighted average of revenue over wage bill in the data and in the model. This increasing moment in the data can be explained by two competing forces in our model, an increase in the dispersion of the productivity distribution within markets and a decline in N. Note that if markets were perfectly competitive, the ratio of revenue over wage bill would equal one for all establishments. While an increase in the dispersion of productivities across establishments explains some of this increasing wedge, the residual wedge is explained by a declining estimate of N which further leads to higher market power for establishments. Our estimated value of N in 2016 is low compared to its value in 1997. This is because the effect of N on the wedge is highly non-linear in a model of Cournot competition. When $N = 32$, the model approaches a competitive economy. However, as N moves from 16 to 8 the increase in the wedge is lower as compared to its increase when N moves from 8 to 4. In other words, the model requires N to be as low as 4 for it to be able to match the data.

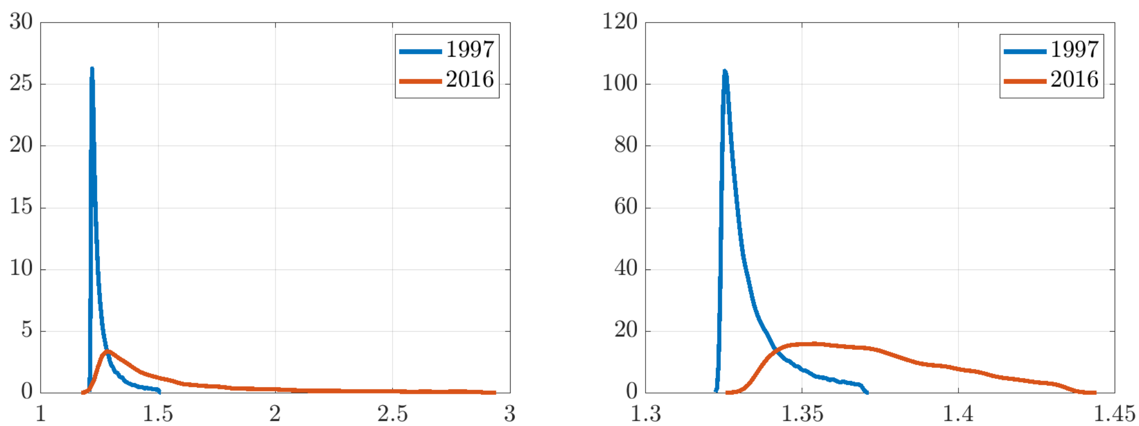
Figure 3: **Average (sales-weighted) Markups and Markdowns**



With the estimated elasticities $\hat{\eta}$ and $\hat{\theta}$, the underlying productivity distribution and the number of competitors N, we can now calculate the markup and markdown for each establishment as predicted by the model. In Figure 3 we plot the evolution of aggregate, sales-weighted markups, which have increased from 1.69 to 2.20 between 1997 and 2016.

Over the same period, markdowns have remained stable, increasing only marginally from 1.37 to 1.40. Establishments do exert monopsony power over workers, but the magnitude of the markdown does not change over time. Despite the fact that the market structure changes substantially (N decreases), markdowns do not reflect this change over time. The main reason is that the estimated labor supply elasticity $\hat{\theta}$ implies a smaller upper bound for markdowns (1.53) which is significantly lower than the upper bound for markups (6.0) given by θ. Therefore, this difference in elasticity estimates leads to only marginal increases in markdowns over the entire sample.

Figure 4: **Kernel density (unweighted): Markup and Markdown**



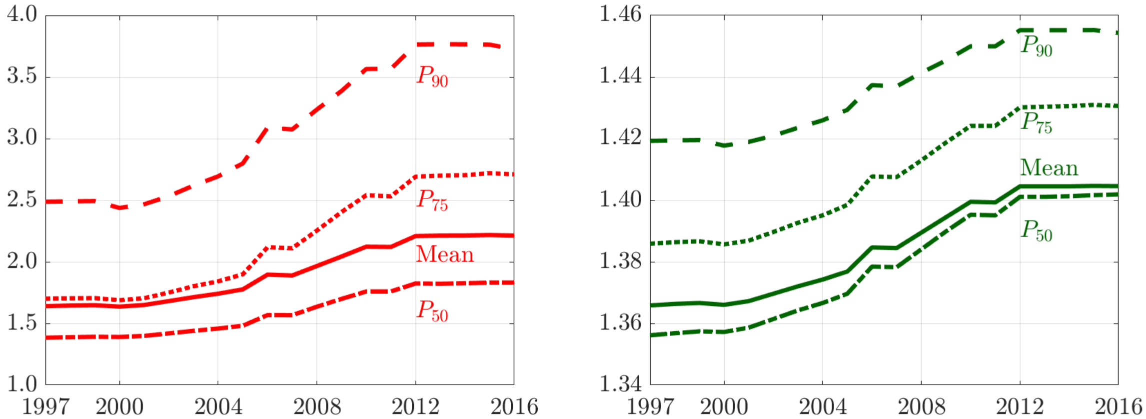
(a) Markup

(b) Markdown

The change in aggregate measures of market power is also driven by changes in the distribution of markups and markdowns across establishments. In Figure 4 we plot the distributional shifts in the unweighted markups and markdowns in 1997 and 2016. We find that the variance of markups has increased substantially and that the right tail has much higher mass in 2016 than in 1997. By contrast, the distribution of markdowns has much less variance across establishments, and as a result the outward shift in the tail translates only into marginal increases in aggregate markdowns.

We also analyze the changes in the percentiles of the sales-weighted markup and markdown

Figure 5: **Percentiles of (sales weighted) Markup and Markdown distribution**



(a) Percentiles of Markup distribution

(b) Percentiles of Markdown distribution

distributions in Figure 5.[23] We find that P_{90}, the ninetieth percentile of the markup distribution increases from 2.64 in 1997 to 3.62 in 2016. The increasing trend in markups, the substantial shift in the right tail of the markup distribution, and the sharp increase in the 90th percentile of the sales-weighted markup distribution are consistent with the findings in De Loecker et al. (2020).

5 Wage Stagnation

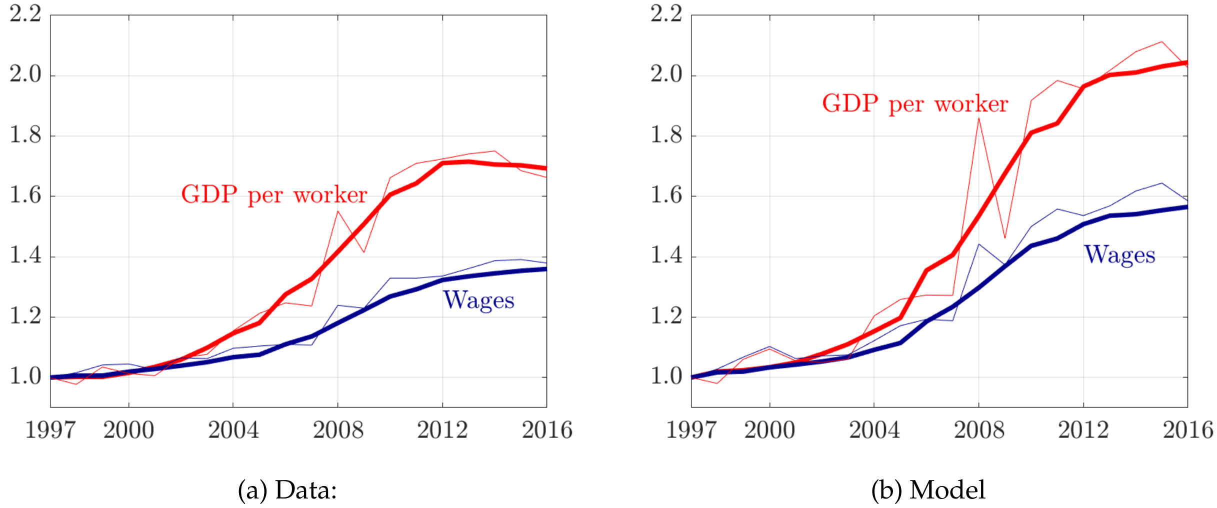
With the estimated key parameters of the model, we proceed to study the implications of goods and labor market power in explaining the decoupling between labor productivity and wages in the US, a phenomenon we refer to as wage stagnation.

PRODUCTIVITY-WAGE DECOUPLING. Figure 6 plots the GDP per worker and the average wage in the data and our model for the estimated market structure between 1997 and 2016, where we normalize their levels to 1 in 1997 in both figures. Figure 6a plots GDP per worker – real GDP divided by employment – and the average wage in the Census data. In the Census data, GDP per worker grew by 66% while wages only increased by 38%.[24] In Figure 6b, we plot

[23] We plot the 5 year centered moving averages for percentiles.

[24] We weigh the average wage at the establishment level by its employment to compute the average wage of workers.

Figure 6: **Productivity-Wage Decoupling: Data vs Model**



the model-implied GDP per worker as well as the model-implied wage for our sample. We see that wages increase by roughly 58 % while GDP per worker increases by 102% in the model.

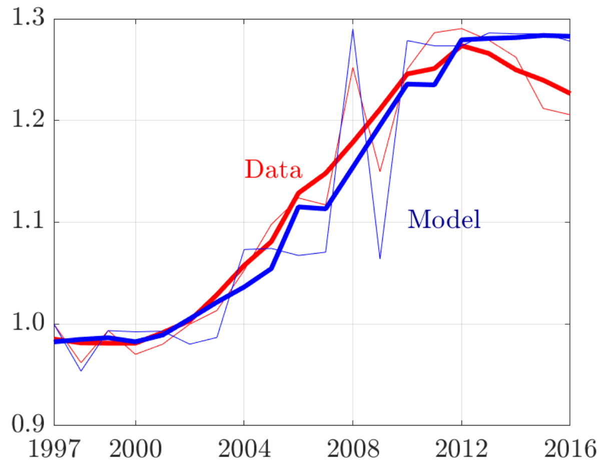
Our model generates the decoupling between GDP per worker and wages that we see in the data, but the increase in both GDP and wages is larger. However, the model does match the increase in the ratio of GDP per worker to the wage in our data when this ratio is normalized to 1 in 1997. As plotted in Figure 7, the ratio increases by 21% in the data and by 28% in model.

In a world with perfect competition in both goods and labor markets, the real wage is equal to the productivity of the worker, $A = W/P$. This implies that $\Delta_t \ln A = \Delta_t \ln(W/P)$, and as a result, over time, any growth in productivity leads to an equivalent growth in real wages. This is contrary to what we observe in the data, where the wedge between productivity and wages has increased. In our model, we attribute this rise in the wedge to the rise in the market power of firms in the goods and the labor markets. To see this at the micro and at the aggregate level, consider the first-order condition from the equation (11), which we rearrange as follows:

$$\underbrace{\frac{P_{inj} A_{inj}}{W_{inj}}}_{\text{Ratio of TFPR to Wage}} = \underbrace{\mu_{inj}}_{\text{Markup}} \times \underbrace{\delta_{inj}}_{\text{Markdown}} . \tag{22}$$

Equation (22) implies that the markups and markdowns form a wedge between the dollar value

Figure 7: **Normalized ratio of GDP per worker and Wage : Data and Model**



Notes: We construct the ratio of the average wage to GDP per worker in the data and in the model for each year between 1997 and 2016. We then normalize the series to 1 in 1997 for both series, which we plot in the figure above.

of an establishment's productivity and the wage paid to its workers.[25] The higher the market power in the economy, the greater is the wedge between establishment-level revenue and the wage bill. Aggregating this first-order condition gives us the following relationship

$$\underbrace{\left(\frac{\int_j \sum_i R_{inj} dj}{\int_j \sum_i L_{inj} dj}\right)}_{\text{GDP per worker}} \Big/ \underbrace{\left(\frac{\int_j \sum_i L_{inj} W_{inj} dj}{\int_j \sum_i L_{inj} dj}\right)}_{\text{Average wage}} = \underbrace{\left[\int_j \sum_i \left(\frac{R_{inj}}{R} \frac{1}{\mu_{inj}\delta_{inj}} dj\right)\right]^{-1}}_{\text{Aggregate Wedge}} \tag{24}$$

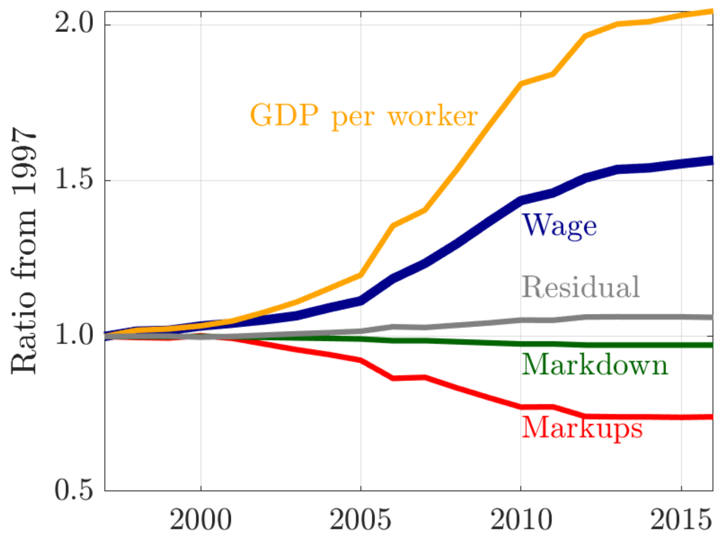
where we denote aggregate revenue as $R = PY = \int_j \sum_i R_{inj} dj$. Equation (24) shows that the ratio of GDP per worker to the average wage of workers can be expressed as the inverse of the sales-weighted average of the establishment-level wedge $\frac{1}{\mu_{inj}\delta_{inj}}$, which we refer to as the aggregate wedge in equation (24). Over time, if market power increases, especially among establishments that have a high sales-share in the economy, then the implication is a rise in the decoupling between GDP per worker and average wages in the economy.

[25] Moreover, the identity implies that the growth rate of each component must follow:

$$\underbrace{\Delta_t \ln(P_{injt} A_{injt})}_{\text{TFPR growth}} = \underbrace{\Delta_t \ln(W_{injt})}_{\text{Wage growth}} + \underbrace{\Delta_t \ln(\mu_{injt})}_{\text{Markup growth}} + \underbrace{\Delta_t \ln(\delta_{injt})}_{\text{Markdown growth}} . \tag{23}$$

This equation suggests that the growth in the total factor productivity revenue (TFPR) for each establishment can be decomposed into the sum of the growth in wages, markups and markdowns. As TFPR is a measure of revenue

Figure 8: **Decomposition of wages**



Decomposition of Wage growth As an alternative way to evaluate the role of aggregate markups and markdowns in the decoupling we consider a representative firm setup:

$$\mathcal{W} = \frac{\text{GDP per worker}}{\mu\delta}\Omega, \quad \text{where} \quad \Omega = \int_j \sum_i \left(\frac{R_{inj}}{R} \frac{\mu}{\mu_{inj}} \frac{\delta}{\delta_{inj}} \right) dj \tag{25}$$

and where we define $\mathcal{W} = \frac{\int_J \sum_i L_{inj} W_{inj} dj}{\int_j \sum_i L_{inj} dj}$ as the employment-weighted average wage in the economy which is equivalent to the average wage of workers. GDP per worker is given by $\frac{\int_j \sum_i R_{inj} dj}{\int_j \sum_i L_{inj} dj}$ and the aggregate markup, μ, and markdown, δ, are both sales-weighted. Ω in the expression is the residual which captures the heterogeneity that is not absorbed by the aggregates in the representative framework.[26]

Figure 8 plots the contribution of the rise of aggregate markups, markdowns and GDP per worker on the growth of wages. It shows that while growth in GDP per worker increases wages, the rise of aggregate markup leads to significant downward pressure on wages while the rise of aggregate markdown contributes only marginally to the stagnation of wages. The

per worker, we can re-write the first order condition as $R_{inj}/(L_{inj} W_{inj}) = \mu_{inj}\delta_{inj}$.

[26]Note that equation (25) follows immediately from equation (24). We can re-write equation (24) as follows:

$$\frac{R}{\mathcal{L}} \times \frac{1}{\mathcal{W}} = \frac{\mu\delta}{\Omega}, \quad \text{where} \quad \mathcal{L} = \int_j \sum_i L_{inj} dj$$

Re-arranging the above equation gets us to equation (25).

residual Ω has a marginally positive effect on the level of wages. The large negative contribution of markups to wages is because the increase in markups is substantially larger in comparison to the increase in markdowns.

COUNTERFACTUAL ECONOMIES. We now analyze several counterfactual economies to decompose the effect on wages that is due to goods and labor market power. In our model, wages change as a result of goods market power (monopoly) – a general equilibrium effect – and as a result of labor market power (monopsony) – a direct effect on wages. We analyze several solutions to the model where we shut down the different sources of market power. First, we analyze the planner's solution where all channels of market power are closed. Then we shut down either market power in the labor market only or market power in the goods market only. In each counterfactual economy, we analyze the effect on the wage level.

The social planner takes consumer preferences as given and maximizes consumer utility subject to the aggregate resource constraint. While the establishments still face a downward sloping demand function and an upward sloping labor supply function, they behave as atomistic price takers on both markets under the planner's allocation. Consequently, the social planner solves

$$V = \max_{C_{inj}, L_{inj}} \left(C - \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{L^{\frac{\phi+1}{\phi}}}{\frac{\phi+1}{\phi}} \right) \tag{26}$$

$$\text{s.t. } C_{inj} = Y_{inj} = A_{inj} L_{inj}$$

and also subject to the aggregation equations (2) and (3). This helps us reduce the planner's problem to the optimal allocation of labor and consumption. The first-order condition that gives the optimal allocation L_{inj}^{11} (where $\mu = 1$ and $\delta = 1$ for all establishments) is:[27]

$$[L_{inj}^{11}] \; : \; \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-\frac{1}{\eta}} C_j^{\frac{1}{\eta} - \frac{1}{\theta}} C^{\frac{1}{\theta}} A_{inj} = \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} L^{\frac{1}{\phi}}. \tag{27}$$

The planner directly chooses an allocation and there is no price system. However, if there

[27] This is equivalent to the allocation where the planner equates the marginal rate of substitution to marginal rate of transformation $\frac{U'(L_{inj})}{U'(C_{inj})} = f'(L_{inj})$ which is equivalent to $\frac{\frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} L_{inj}^{1/\hat{\eta}} L_j^{1/\hat{\theta} - 1/\hat{\eta}} L^{-1/\hat{\theta}} L^{1/\phi}}{\frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-1/\eta} C_j^{1/\eta - 1/\theta} C^{1/\theta}} = A_{inj}$. We derive this in Appendix (A.5)

were prices and we substitute the equilibrium goods demand and the labor supply functions in the planner's problem, equation (27) would satisfy $P_{inj}A_{inj} = W_{inj}$. The planner's allocation set marginal product equal to the marginal cost, without markup and markdown distortions. Equation (27) gives the planner's allocation of labor L^{11}_{inj}, and given that there are no distortions in the output or the input market, this also characterizes the first-best outcome in our model.

In what follows, we define $L^{\mu\delta}_{inj}$, where $\mu = 1$ denotes the planner's optimal solution in the goods market and $\delta = 1$ the planner's optimal solution in the labor market. Otherwise, the notation μ and δ denote the equilibrium outcome with market power. For instance, $L^{\mu 1}_{inj}$ is the labor allocation when there is goods market power but no labor market power, such that there is strategic interaction among firms in the goods market but firms behave as atomistic price takers in the labor market. Then, the decentralized allocation with market power in both output and input markets is given by

$$[L^{\mu\delta}_{inj}] \;:\; \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-\frac{1}{\eta}} C_j^{\frac{1}{\eta}-\frac{1}{\theta}} C^{\frac{1}{\theta}} \frac{A_{inj}}{\mu_{inj}} = \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} L^{\frac{1}{\phi}} \delta_{inj} \tag{28}$$

Note that this is the same equation as in our baseline model with establishment-level markups and markdowns. Similarly, we can define $L^{1\delta}_{inj}$ and $L^{\mu 1}_{inj}$ as follows

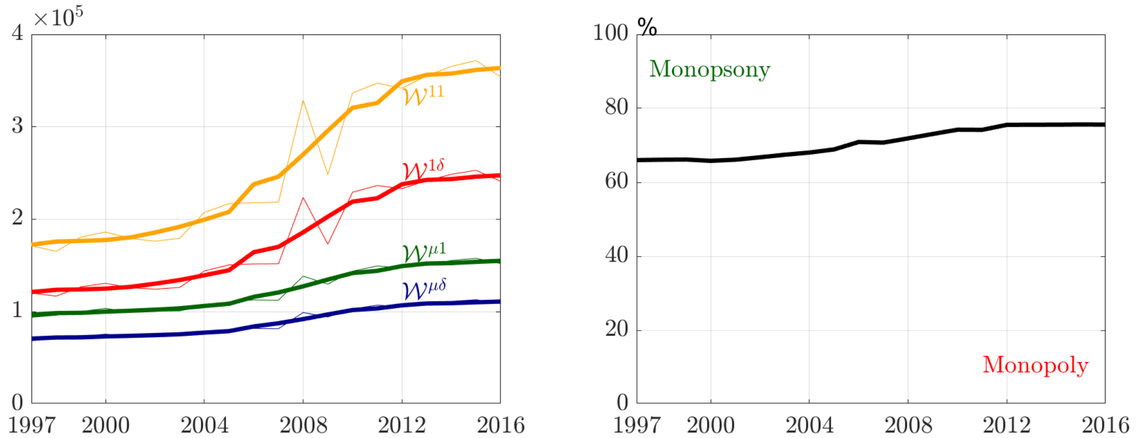
$$[L^{1\delta}_{inj}] \;:\; \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-\frac{1}{\eta}} C_j^{\frac{1}{\eta}-\frac{1}{\theta}} C^{\frac{1}{\theta}} A_{inj} = \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} L^{\frac{1}{\phi}} \delta_{inj} \tag{29}$$

$$[L^{\mu 1}_{inj}] \;:\; \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-\frac{1}{\eta}} C_j^{\frac{1}{\eta}-\frac{1}{\theta}} C^{\frac{1}{\theta}} \frac{A_{inj}}{\mu_{inj}} = \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} L^{\frac{1}{\phi}} \tag{30}$$

COUNTERFACTUAL WAGE LEVEL DECOMPOSITION. We can now solve the general equilibrium model under the four regimes and compare the evolution of wages over time. In Figure 9a, the blue line represents the evolution of wages in the baseline model with goods market power and labor market power. The green line represents the wage series, $\mathcal{W}^{\mu 1}$, when there is only market power in the goods market and no market power in the labor market. The red line represents the wage series, $\mathcal{W}^{1\delta}$, when there is only market power in the labor market and no market power in the goods market.

Both goods market power and labor market power decrease the wage. An increase in labor market power leads to a reduction in an establishment's wage as the labor supply curve slopes upward. At the same time, output market power reduces the level of the output of the estab-

Figure 9: **Contribution of Goods and Labor Market Power to Wage Level**



(a) Counterfactual wage levels

(b) % Contribution of Monopoly and Monopsony

lishment, which implies a reduction in employment. Due to the reduced demand for labor in the aggregate, the rise of output market power through the general equilibrium effect leads to a fall in the aggregate wage level $\mathcal{W}$.

The effect of eliminating labor market power (green) leads to a smaller increase in wages than the effect of eliminating goods market power (red). We find that $\mathcal{W}_t^{1\delta} - \mathcal{W}_t^{\mu\delta} > \mathcal{W}_t^{\mu 1} - \mathcal{W}_t^{\mu\delta}$ in every year $t \in [1997, 2016]$, indicating that eliminating goods market power has a bigger effect on wages than eliminating labor market power. When we calculate the percentage contribution in Figure 9b, we find that the contribution of output market power to the level of wages has increased from 67% in 1997 to 75% in 2016.[28]

Next, we quantify the contribution of goods and labor market power in explaining the *change* in wages between 1997 and 2016. The contribution of goods market power is defined as the fraction of the total change in wages explained by goods market power net of the change in wages observed in the decentralized economy.[29] We net out the change in wages observed in the decentralized economy $\mathcal{W}^{\mu\delta}$ to isolate the effect of market power from wage increases due to technological change between 1997 and 2016.

[28]We calculate these percentages as the contribution of goods market power by taking the level increase in wages by eliminating goods market power, $\mathcal{W}_t^{1\delta} - \mathcal{W}_t^{\mu\delta}$ as a share of the total increase in wages by sequentially eliminating both sources of market power, $(\mathcal{W}_t^{1\delta} - \mathcal{W}_t^{\mu\delta}) + (\mathcal{W}_t^{\mu 1} - \mathcal{W}_t^{\mu\delta})$. The relative percentage contribution of labor market power is then 100 minus the number we compute for goods market power. Note that the effect relative to the first best is non-linear in goods and labor market power and that both effects do not sum up to 100% of the increase in wages in the planner's allocation.

[29]The total change in wages includes the change due to both goods and labor market power.

In equations (31) and (32), we first define the counterfactual change in wages from goods and labor market power, respectively, net of wage changes in the decentralized economy.

$$\Delta\mathcal{W}^{1\delta} - \Delta\mathcal{W}^{\mu\delta} = \underbrace{\mathcal{W}^{1\delta}_{2016} - \mathcal{W}^{1\delta}_{1997}}_{\text{Wage change without GMP}} - \underbrace{\left(\mathcal{W}^{\mu\delta}_{2016} - \mathcal{W}^{\mu\delta}_{1997}\right)}_{\text{Wage change in the decentralized economy}} \tag{31}$$

$$\Delta\mathcal{W}^{\mu 1} - \Delta\mathcal{W}^{\mu\delta} = \underbrace{\mathcal{W}^{\mu 1}_{2016} - \mathcal{W}^{\mu 1}_{1997}}_{\text{Wage change without LMP}} - \underbrace{\left(\mathcal{W}^{\mu\delta}_{2016} - \mathcal{W}^{\mu\delta}_{1997}\right)}_{\text{Wage change in the decentralized economy}} \tag{32}$$

Next, using equations (31) and (32), we define the contribution of goods market power to changes in wages as

$$\text{Contribution of GMP} = \frac{\Delta\mathcal{W}^{1\delta} - \Delta\mathcal{W}^{\mu\delta}}{(\Delta\mathcal{W}^{1\delta} - \Delta\mathcal{W}^{\mu\delta}) + (\Delta\mathcal{W}^{\mu 1} - \Delta\mathcal{W}^{\mu\delta})} \tag{33}$$

In our counterfactual economies, we can calculate the changes in each aggregate wage series between 1997 and 2016. We find that monopoly power accounts for 81.8% of the change in wages between 1997 and 2016, and the remainder, 18.2%, is due to monopsony power.

6 Conclusion

In this paper we propose a general equilibrium model of the macroeconomy with the simultaneous determination of markups and markdowns. We take this model to the micro data and infer both, markups from goods market power (monopoly) and markdowns due to labor market power (monopsony). In the process, we estimate establishment-level productivity as well as the economy-wide market structure. We do so by using a novel way of determining market power by estimating market structure that best fits the micro data from the revenue and wage distributions, using a stochastic interpretation of market structure.

We find that the market structure has led to more market power over time, where the number of competitors in each market has declined over time. This has led to an increase in market power from 1997 until 2016, where markups have increased from 1.69 to 2.2. Instead, markdowns have increased only marginally from 1.37 to 1.4 over the same period. We find that the decline in competition in the economy can rationalize the decoupling between productivity and wages between 1997 and 2016, where the rise of markups puts a significant downward pressure on wages.

The presence of market power, both monopoly and monopsony, can account for lower wages relative to an efficient economy. We perform counterfactual experiments to decompose the effect of monopoly and monopsony on the wage level relative to the planner's solution. We find that both markups and markdowns reduce the level of wages relative to a planner's economy, but that the general equilibrium effect of monopoly power on real wages dominates the effect of monopsony power. Monopoly accounts for about 67% of the wage decline in 1997 and 75% in 2016.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification properties of recent production function estimators," *Econometrica*, 83, 2411–2451.

Atkeson, A. and A. Burstein (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98, 1998–2031.

Azar, J., S. Berry, and I. E. Marinescu (2019a): "Estimating Labor Market Power," Yale mimeo.

Azar, J., I. Marinescu, and M. Steinbaum (2019b): "Measuring Labor Market Power Two Ways," in *AEA Papers and Proceedings*, vol. 109, 317–21.

Azar, J., I. Marinescu, and M. I. Steinbaum (2017): "Labor market concentration," Tech. rep., National Bureau of Economic Research.

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*.

Berger, D., K. Herkenhoff, and S. Mongey (2022): "Labor Market Power," *American Economic Review*, forthcoming.

Berry, S., M. Gaynor, and F. Scott Morton (2019a): "Do Increasing Markups Matter? Lessons from empirical industrial organization," *Journal of Economic Perspectives*, 33, 44–68.

——— (2019b): "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization," *Journal of Economic Perspectives*, 33, 44–68.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890.

Burdett, K. and D. T. Mortensen (1998): "Wage differentials, employer size, and unemployment," *International Economic Review*, 257–273.

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.

De Loecker, J., J. Eeckhout, and S. Mongey (2021): "Quantifying Market Power and Business Dynamism in the Macroeconomy," Tech. rep., NBER Working Paper.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 135, 561–644.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, markups, and trade reform," *Econometrica*, 84, 445–510.

De Loecker, J. and F. M. P. Warzynski (2012): "Markups and Firm-level Export Status," *American Economic Review*, 102, 2437–2471.

Deb, S., J. Eeckhout, A. Patel, and L. Warren (2021): "The Contribution of Market Power to Wage Inequality," Tech. rep., UPF mimeo.

Delgado, M., R. Bryden, and S. Zyontz (2014): "Categorization of traded and local industries in the US economy," MIT mimeo.

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2018): "Monopsony in Online Labor Markets," Tech. rep., National Bureau of Economic Research.

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

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

——— (2021): *The Profit Paradox. How Thriving Firms Threaten the Future of Work*, Princeton, NJ: Princeton University Press.

Falch, T. (2010): "The elasticity of labor supply at the establishment level," *Journal of Labor Economics*, 28, 237–266.

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

GANAPATI, S. (2019): "Oligopolies, prices, output, and productivity," *Available at SSRN 3030966*.

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.

GOOLSBEE, A. AND C. SYVERSON (2019): "Monopsony Power in Higher Education: A Tale of Two Tracks," Tech. rep., National Bureau of Economic Research.

GREENSPON, J., A. M. STANSBURY, AND L. H. SUMMERS (2021): "Productivity and Pay in the US and Canada," Working Paper 29548, National Bureau of Economic Research.

HALL, R. (1988): "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96, 921–947.

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

JAROSCH, G., J. S. NIMCZIK, AND I. SORKIN (2019): "Granular search, market structure, and wages," Tech. rep., National Bureau of Economic Research.

LEVINSOHN, J. AND A. PETRIN (2003): "Estimating production functions using inputs to control for unobservables," *The review of economic studies*, 70, 317–341.

MACHIN, S. (2016): "Rising Wage Inequality, Real Wage Stagnation and Unions," in *Inequality: Causes and consequences*, Emerald Group Publishing Limited.

MANNING, A. (2003): *Monopsony in motion: Imperfect competition in labor markets*, Princeton University Press.

——— (2011): "Imperfect competition in the labor market," in *Handbook of labor economics*, Elsevier, vol. 4, 973–1041.

MATSUDAIRA, J. D. (2014): "Monopsony in the low-wage labor market? Evidence from minimum nurse staffing regulations," *Review of Economics and Statistics*, 96, 92–102.

MELITZ, M. (2003): "The impact of trade on aggregate industry productivity and intra-industry reallocations," *Econometrica*, 71, 1695–1725.

MERTENS, M. (2021): "Labour market power and between-firm wage (in) equality," Tech. rep., IWH-CompNet Discussion Papers.

MILLER, N., S. BERRY, F. S. MORTON, J. BAKER, T. BRESNAHAN, M. GAYNOR, R. GILBERT, G. HAY, G. JIN, B. KOBAYASHI, F. LAFONTAINE, J. LEVINSOHN, L. MARX, J. MAYO, A. NEVO, A. PAKES, N. ROSE, D. RUBINFELD, S. SALOP, M. SCHWARTZ, K. SEIM, C. SHAPIRO, H. SHELANSKI, D. SIBLEY, AND A. SWEETING (2021): "On The Misuse of Regressions of Price on the HHI in Merger Review," Tech. rep., Georgetown mimeo.

MORLACCO, M. (2017): "Market Power in Input Markets: Theory and Evidence from French Manufacturing," Tech. rep., Yale.

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 263–97.

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

RINZ, K. (2018): "Labor market concentration, earnings inequality, and earnings mobility," Mimeo.

ROBINSON, J. (1933): *The economics of imperfect competition*, MacMillan and Co.

ROSSI-HANSBERG, E., P.-D. SARTE, AND N. TRACHTER (2018): "Diverging trends in national and local concentration," Tech. rep., National Bureau of Economic Research.

RUBENS, M. (2021): "Market Structure, Oligopsony Power, and Productivity," *Oligopsony Power, and Productivity (March 8, 2021)*.

STAIGER, D. O., J. SPETZ, AND C. S. PHIBBS (2010): "Is There Monopsony in the Labor Market? Evidence from a Natural Experiment," *Journal of Labor Economics*, 28, 211–236.

STANSBURY, A. M. AND L. H. SUMMERS (2017): "Productivity and Pay: Is the link broken?" Working Paper 24165, National Bureau of Economic Research.

SYVERSON, C. (2019): "Macroeconomics and Market Power: Context, Implications, and Open Questions," *Journal of Economic Perspectives*, 33, 23–43.

APPENDIX

A Derivations

A.1 Household's optimization

OPTIMUM LABOR SUPPLY FUNCTIONS We follow Berger et al. (2022) and add adjustments for the love for variety by scaling the utility function the number of market J and establishment I in each market. The households optimum choice of allocation of labor across markets can be written as the solution to;

$$\text{Min}_{L_j}\left(\int_j \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj\right)^{\frac{\hat{\theta}}{\hat{\theta}+1}} dj \;\; \text{s.t} \int_J W_j L_j \geq Z \tag{A1}$$

Then the optimal allocation is given by ;

$$\frac{\hat{\theta}}{\hat{\theta}+1}\left(\int_j \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj\right)^{\frac{\hat{\theta}}{\hat{\theta}+1}-1} \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}}} \frac{\hat{\theta}+1}{\hat{\theta}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}-1} = \lambda W_j$$

$$\underbrace{\left(\int_j \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj\right)^{\frac{\hat{\theta}}{\hat{\theta}+1}-1}}_{L^{\frac{-1}{\hat{\theta}}}} \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}-1} = \lambda W_j$$

$$\frac{1}{J}^{\frac{-1}{\hat{\theta}}} L^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1}{\hat{\theta}}} = \lambda W_j$$

Next multiply each side by L_j and integrate across J

$$\frac{1}{J}^{\frac{-1}{\hat{\theta}}} L^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1+\hat{\theta}}{\hat{\theta}}} = \lambda W_j L_j$$

$$L^{\frac{-1}{\hat{\theta}}} \underbrace{\int_j \frac{1}{J}^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1+\hat{\theta}}{\hat{\theta}}} dj}_{L^{\frac{\hat{\theta}+1}{\hat{\theta}}}} = \lambda \int_j W_j L_j dj$$

$$L^{\frac{-1}{\hat{\theta}}} L^{\frac{\hat{\theta}+1}{\hat{\theta}}} = \lambda \int_j W_j L_j dj$$

which is equivalent to;

$$L = \lambda \int_j W_j L_j dj$$

We define the aggregate wage index W such that $WL = \int_j W_j L_j dj$ which would imply that $\lambda = W^{-1}$. Then plugging this into the first order condition delivers the market labor supply equation as a function of wage levels and aggregate labor supply.

$$\frac{1}{J}^{\frac{-1}{\hat{\theta}}} L^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1}{\hat{\theta}}} = \frac{W_j}{W}$$

$$L_j^{\frac{1}{\hat{\theta}}} = \frac{1}{J}^{\frac{1}{\hat{\theta}}} \frac{W_j}{W} L^{\frac{1}{\hat{\theta}}}$$

$$L_j = \left(\frac{1}{J}\right) \left(\frac{W_j}{W}\right)^{\hat{\theta}} L$$

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

$$\int_J W_j L_j dj = \left(\frac{1}{J}\right) \left(\frac{1}{W}\right)^{\hat{\theta}} L \int_J W_j^{1+\hat{\theta}} dj$$

$$WL = \left(\frac{1}{J}\right) \left(\frac{1}{W}\right)^{\hat{\theta}} L \int_J W_j^{1+\hat{\theta}} dj$$

$$W^{1+\hat{\theta}} = \left(\frac{1}{J}\right) \int_J W_j^{1+\hat{\theta}} dj$$

$$W = \left(\left(\frac{1}{J}\right) \int_J W_j^{1+\hat{\theta}} dj\right)^{\frac{1}{1+\hat{\theta}}}$$

We can apply a similar formulation to derive the establishment level labor supply;

$$L_{inj} = \left(\frac{1}{I}\right) \left(\frac{W_{inj}}{W_j}\right)^{\hat{\eta}} L_j$$

The market wage index is;.

$$W_j = \left(\left(\frac{1}{I}\right) \sum_i W_{inj}^{1+\hat{\eta}}\right)^{\frac{1}{1+\hat{\eta}}}$$

Then the establishment level labor supply curve is given by;

$$L_{inj} = \left(\frac{1}{J}\right) \left(\frac{1}{I}\right) \left(\frac{W_{inj}}{W_j}\right)^{\hat{\eta}} \left(\frac{W_j}{W}\right)^{\hat{\theta}} L \tag{A2}$$

To derive the inverse labor supply function we can write;

$$\left(\frac{1}{J}\right)^{-1}\frac{L_j}{L}=\left(\frac{W_j}{W}\right)^{\hat{\theta}}$$

$$W_j=\left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}}}\left(\frac{L_j}{L}\right)^{\frac{1}{\hat{\theta}}}W$$

similarly at the establishment level;

$$W_{inj}=\left(\frac{1}{I}\right)^{\frac{-1}{\hat{\eta}}}\left(\frac{L_{inj}}{L_j}\right)^{\frac{1}{\hat{\eta}}}W_j$$

Combining the last two equations we can get the establishment level inverse labor supply curve as a function of labor supplied by the household and aggregates.

$$W_{inj}=\frac{1}{J}^{\frac{-1}{\hat{\theta}}}\frac{1}{I}^{\frac{-1}{\hat{\eta}}}L_{inj}^{\frac{1}{\hat{\eta}}}L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}}L^{-\frac{1}{\hat{\theta}}}W \tag{A3}$$

OPTIMUM CONSUMPTION FUNCTIONS The household solves a static maximization problem

$$\max_{C_{inj},L_{inj}}\left(C-\frac{1}{\bar{\phi}^{\frac{1}{\phi}}}\frac{L^{\frac{\phi+1}{\phi}}}{\frac{\phi+1}{\phi}}\right)\text{s.t.}\quad PC=LW+\Pi \tag{A4}$$

The households optimum choice of allocation of consumption across markets can be written as the solution to;

$$\text{Max}_{C_j}\left(\int_j\left(\frac{1}{J}\right)^{\frac{1}{\theta}}C_j^{\frac{\theta-1}{\theta}}dj\right)^{\frac{\theta}{\theta-1}}\quad\text{s.t}\int_j P_jC_jdj\leq Z \tag{A5}$$

The optimal allocation is given by;

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

Which can be written as;

$$\left(\frac{1}{J}\right)^{\frac{1}{\theta}}C^{\frac{1}{\theta}}C_j^{\frac{-1}{\theta}}=\lambda P_j$$

Then multiplying each side by C_j and integrating across J we get

$$C = \lambda \int_j P_j C_j dj$$

We define the aggregate price index P such that $PC = \int_j P_j C_j dj$ implying that $\lambda = P^{-1}$. Then plugging this into the first order condition gives us the demand function at the market level.

$$C_j = \left(\frac{1}{J}\right)\left(\frac{P_j}{P}\right)^{-\theta} C$$

To derive the aggregate price index, we multiply both side by P_j and integrate across markets.

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

Finally, using similar steps we can derive the establishment level demand function as

$$C_{inj} = \frac{1}{I}\left(\frac{P_{inj}}{P_j}\right)^{-\eta} C_j$$

and the market price index as

$$P_j = \left(\frac{1}{I}\sum_i P_{inj}^{1-\eta}\right)^{\frac{1}{1-\eta}}$$

Plugging in the market demand function gives us the establishment's demand function as

$$C_{inj} = \left(\frac{1}{J}\right)\left(\frac{1}{I}\right)\left(\frac{P_{inj}}{P_j}\right)^{-\eta}\left(\frac{P_j}{P}\right)^{-\theta} C \tag{A6}$$

To derive the market inverse demand function we can write

$$P_j = J^{\frac{-1}{\theta}}\left(\frac{C_j}{C}\right)^{\frac{-1}{\theta}} P$$

and similarly we can write

$$P_{inj} = I^{\frac{-1}{\eta}}\left(\frac{C_{inj}}{C_j}\right)^{\frac{-1}{\eta}} P_j$$

Combining the above two equations gives us the establishment level inverse demand function.

$$P_{inj} = \frac{1}{J}^{\frac{1}{\theta}}\frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{\frac{-1}{\eta}} C_j^{\frac{1}{\eta}-\frac{1}{\theta}} C^{\frac{1}{\theta}} P \tag{A7}$$

A.2 Firm's Problem

SOLVING THE FIRM'S FIRST ORDER CONDITION. There are N firms indexed by n in each market. A firm owns I/N establishments. An establishment's sales share and wage bill share are denoted by s_{inj} and e_{inj}, 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 $e_{nj} = \sum_{i\in\mathcal{I}_{nj}} e_{inj}$, respectively. Firm's problem here is to choose an output level Y_{inj} for each establishment i simultaneously to maximize its profit:

$$\Pi_{nj} = \max_{Y_{inj}} \sum_{i\in\mathcal{I}_{nj}} \left(P_{inj} Y_{inj} - \frac{W_{inj}}{A_{inj}} Y_{inj} \right)$$

The FOC gives:

$$P_{inj} + \frac{\partial P_{inj}}{\partial Y_{inj}} Y_{inj} + \sum_{i'\in\mathcal{I}_{nj}\setminus i} \left(\frac{\partial P_{i'nj}}{\partial Y_{inj}} Y_{i'nj} \right) = \frac{1}{A_{inj}} \left[W_{inj} + \frac{\partial W_{inj}}{\partial L_{inj}} L_{inj} + \sum_{i'\in\mathcal{I}_{nj}\setminus i} \left(\frac{\partial W_{i'nj}}{\partial L_{inj}} L_{i'nj} \right) \right]$$

Note that $\frac{\partial P_{inj}}{\partial Y_{inj}} Y_{inj} = \left[-1/\eta + (1/\eta - 1/\theta)s_{inj}\right] P_{inj}$ and for the other establishments $i' \in \mathcal{I}_{nj} \setminus i$ owned by firm n we have

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

and similarly, $\frac{\partial W_{inj}}{\partial L_{inj}} L_{inj} = \left[1/\hat{\eta}_L + (1/\hat{\theta} - 1/\hat{\eta})e_{inj}\right] W_{inj}$ and for the other establishments $i' \in \mathcal{I}_{nj} \setminus i$ owned by firm n we have

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

The FOC can be rewritten into:

$$\left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj}) \right] P_{inj} = \left[1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj}) \right] \frac{W_{inj}}{A_{inj}}, \tag{A8}$$

where markup and markdown are relatively defined as:

$$\begin{aligned}\mu_{inj} &\equiv \frac{P_{inj}}{MC_{inj}} = \left(1 - \frac{1}{\theta}s_{nj} - \frac{1}{\eta}\left(1 - s_{nj}\right)\right)^{-1} \\ \delta_{inj} &\equiv \frac{MRPL_{inj}}{W_{inj}} = \left(1 + \frac{1}{\hat{\theta}}e_{nj} + \frac{1}{\hat{\eta}}\left(1 - e_{nj}\right)\right).\end{aligned} \tag{A9}$$

A.3 Solving the equilibrium

The firm's FOC (A8) has 4 unknowns; two levels P_{inj}, W_{inj} which are a function of the aggregates P, Y, W, L and two shares s_{inj}, e_{inj}. The objective is to reduce the FOC to a single unknown s_{inj} independent of aggregates and therefore the price levels. Once, given the productivity distribution we solve for the sales share s_{inj} distribution we recover the wage bill share distribution and then finally pin down the aggregates and therefore the level of prices and quantities in the economy. We proceed in 4 steps.

STEP 1: SOLVING THE FIRM'S PROBLEM IN SHARES. Rearranging equation (A8), we derive:

$$P_{inj} = \frac{\left[1 + \frac{1}{\hat{\theta}}e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj})\right]}{\left[1 - \frac{1}{\theta}s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right]}\frac{W_{inj}}{A_{inj}}. \tag{A10}$$

Plug in the inverse labor supply function (A3):

$$P_{inj} = \frac{\left[1 + \frac{1}{\hat{\theta}}e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj})\right]}{\left[1 - \frac{1}{\theta}s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right]}\frac{J^{\frac{1}{\hat{\theta}}} I^{\frac{1}{\hat{\eta}}} \left(\frac{L_{inj}}{L_j}\right)^{\frac{1}{\hat{\eta}}} \left(\frac{L_j}{L}\right)^{\frac{1}{\hat{\theta}}} W}{A_{inj}}. \tag{A11}$$

Finally, using the CES property $e_{inj} = I^{\frac{1}{\hat{\eta}}} \left(\frac{L_{inj}}{L_j}\right)^{1+\frac{1}{\hat{\eta}}}$ as

$$e_{inj} = \frac{W_{inj}L_{inj}}{W_j L_j} = \frac{L_{inj}^{1+\frac{1}{\hat{\eta}}}}{I^{-\frac{1}{1+\hat{\eta}}}\left(\sum_{i'} L_{i'nj}^{1+\frac{1}{\hat{\eta}}}\right)^{\frac{1}{1+\hat{\eta}}} L_j} = I^{\frac{1}{\hat{\eta}}}\left(\frac{L_{inj}}{L_j}\right)^{1+\frac{1}{\hat{\eta}}}$$

We use i to refer to the particular establishment we are optimizing for while $\sum_{i'}$ to refer to the summation over all establishments in market j. Then we can write P_{inj} in equation (A11) terms of s_{inj}, e_{inj}, A_{inj}, and other market or economy-level variables:

$$P_{inj} = \frac{\left[1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj})\right]}{\left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right]} \frac{J^{\frac{1}{\hat{\theta}}} I^{\frac{1}{1+\hat{\eta}}} e_{inj}^{\frac{1}{1+\hat{\eta}}} \left(\frac{L_j}{L}\right)^{\frac{1}{\hat{\theta}}} W}{A_{inj}}. \tag{A12}$$

STEP 2: MAPPING BETWEEN SALES AND WAGE BILL. We begin by using the definition of wage bill share $e_{inj} = I^{\frac{1}{\hat{\eta}}} \left(\frac{L_{inj}}{L_j}\right)^{1+\frac{1}{\hat{\eta}}}$ and plug in the production function and the inverse demand function to get

$$e_{inj} = \frac{\left(Y_{inj}/A_{inj}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}}{\sum_{i'} \left(Y_{i'nj}/A_{i'nj}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}} \tag{A13}$$

$$= \frac{\left(P_{inj}^{-\eta}/A_{inj}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}}{\sum_{i'} \left(P_{i'nj}^{-\eta}/A_{i'nj}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}} \tag{A14}$$

On the other hand, we have:

$$s_{inj} = \frac{1}{I}\left(\frac{P_{inj}}{P_j}\right)^{1-\eta} \quad \Leftrightarrow \quad P_{inj} = I^{\frac{1}{1-\eta}} s_{inj}^{\frac{1}{1-\eta}} P_j \tag{A15}$$

We then substitute the establishment level price P_{inj} above in the expression for wage bill share of an establishment e_{inj} to derive the mapping between sales and wage bill share for each establishment.

$$e_{inj} = \frac{\left(s_{inj}^{\frac{\eta}{\eta-1}}/A_{inj}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}}{\sum_{i'} \left(s_{i'nj}^{\frac{\eta}{\eta-1}}/A_{i'nj}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}} = \left[\sum_{i'} \left(\left(\frac{s_{i'nj}}{s_{inj}}\right)^{\frac{\eta}{\eta-1}} \frac{A_{inj}}{A_{i'nj}}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}\right]^{-1} \tag{A16}$$

STEP 3 : EQUATION IN SHARES. Given the mapping between sales and wage bill shares, we use the sales share expression to solve a system of I equations and I unknowns for each market.

$$s_{inj} = \frac{P_{inj}^{1-\eta}}{\sum_{i'} P_{inj}^{1-\eta}} = \frac{\left[\frac{1+\frac{1}{\hat{\theta}}e_{nj}+\frac{1}{\hat{\eta}}(1-e_{nj})}{1-\frac{1}{\theta}s_{nj}-\frac{1}{\eta}(1-s_{nj})} \frac{e_{inj}^{\frac{1}{1+\hat{\eta}}}}{A_{inj}}\right]^{1-\eta}}{\sum_{i'} \left[\frac{1+\frac{1}{\hat{\theta}}e_{n'j}+\frac{1}{\hat{\eta}}(1-e_{n'j})}{1-\frac{1}{\theta}s_{n'j}-\frac{1}{\eta}(1-s_{n'j})} \frac{e_{i'n'j}^{\frac{1}{1+\hat{\eta}}}}{A_{i'n'j}}\right]^{1-\eta}}, \tag{A17}$$

where the second equality uses equation (A12). In the above summation we refer to n' as the firm that establishment i' belongs to. Therefore, we can solve s_{inj} from equation (A17) using the mapping between sales and wage bill shares as in equation (A16). Note that at this stage, the solution to the wage bill shares and sales share are independent of the aggregates and only depend on the relative productivity levels among establishments in each market.

STEP 4 :SOLVING FOR THE LEVELS IN THE ECONOMY. The equilibrium system of equations is given as follows :

$$
\begin{aligned}
\text{FOC:} \quad & A_{inj}P_{inj} = \mu_{inj}\delta_{inj}W_{inj} \\
\text{Establishment-level LS:} \quad & W_{inj} = J^{\frac{1}{\widehat{\theta}}} I^{\frac{1}{\widehat{\eta}}} \left(\frac{L_{inj}}{L_j}\right)^{\frac{1}{\widehat{\eta}}} \left(\frac{L_j}{L}\right)^{\frac{1}{\widehat{\theta}}} W \\
\text{Aggregate LS:} \quad & L = \overline{\varphi} W^{\varphi} \\
\text{Establishment-level demand:} \quad & Y_{inj} = \frac{1}{J}\frac{1}{I}\left(\frac{P_{inj}}{P_j}\right)^{-\eta} \left(\frac{P_j}{P}\right)^{-\theta} Y \\
\text{Establishment-level inverse demand:} \quad & P_{inj} = J^{-\frac{1}{\theta}} I^{-\frac{1}{\eta}} \left(\frac{Y_{inj}}{Y_j}\right)^{-\frac{1}{\eta}} \left(\frac{Y_j}{Y}\right)^{-\frac{1}{\theta}} P
\end{aligned}
$$

- Besides, we have the relationship in share:

$$
\begin{aligned}
\frac{Y_{inj}}{Y_j} &= I^{\frac{1}{\eta-1}} s_{inj}^{\frac{\eta}{\eta-1}} \\
\frac{L_{inj}}{L_j} &= \left(\frac{1}{I}\right)^{\frac{1}{\widehat{\eta}+1}} e_{inj}^{\frac{\widehat{\eta}}{\widehat{\eta}+1}}
\end{aligned}
$$

- Hence, we can write FOC as:

$$
\begin{aligned}
Y_j &= \frac{1}{J} \underbrace{\left[\frac{I^{-\frac{(\widehat{\eta}+\eta)(\widehat{\theta}+1)}{\widehat{\theta}(\eta-1)(\widehat{\eta}+1)}} A_{inj}}{\mu_{inj}\delta_{inj} e_{inj}^{\frac{1}{\widehat{\eta}+1}} s_{inj}^{\frac{1}{\eta-1}} \left(\sum_i \left(\frac{s_{inj}^{\frac{\eta}{\eta-1}}}{A_{inj}} \right)^{\frac{\widehat{\eta}+1}{\widehat{\eta}}} \right)^{\frac{\widehat{\eta}}{\widehat{\eta}+1}\frac{1}{\widehat{\theta}}}} \right]^{\frac{\theta\widehat{\theta}}{\theta+\widehat{\theta}}}}_{\alpha_j} \left(\frac{Y^{\frac{1}{\theta}} L^{\frac{1}{\widehat{\theta}}} P}{W} \right)^{\frac{\theta\widehat{\theta}}{\theta+\widehat{\theta}}} \\
&= \frac{1}{J}\alpha_j \left(\frac{Y^{\frac{1}{\theta}} L^{\frac{1}{\widehat{\theta}}} P}{W} \right)^{\frac{\theta\widehat{\theta}}{\theta+\widehat{\theta}}}
\end{aligned}
$$

- Aggregate it into Y, we get:

$$Y = \left[\int_j \frac{1}{J} \left(\alpha_j\right)^{\frac{\theta-1}{\theta}} dj\right]^{\frac{\theta}{\theta-1}} \left(\frac{L^{\frac{1}{\widehat{\theta}}} P}{W}\right)^{\frac{\theta\widehat{\theta}}{\theta+\widehat{\theta}}} Y^{\frac{\widehat{\theta}}{\widehat{\theta}+\theta}}$$

and hence:

$$Y = \left[\int_j \frac{1}{J} \left(\alpha_j\right)^{\frac{\theta-1}{\theta}} dj\right]^{\frac{\widehat{\theta}+\theta}{\theta-1}} \left(\frac{P}{W}\right)^{\widehat{\theta}} L \tag{A18}$$

- Using this relationship, we can get:

$$Y_j = \frac{1}{J}\alpha_j \left[\int_j \frac{1}{J} \left(\alpha_j\right)^{\frac{\theta-1}{\theta}} dj\right]^{\frac{\widehat{\theta}}{\theta-1}} \left(\frac{P}{W}\right)^{\widehat{\theta}} L$$

$$Y_{inj} = I^{\frac{1}{\eta-1}} s_{inj}^{\frac{\eta}{\eta-1}} \frac{1}{J}\alpha_j \left[\int_j \frac{1}{J} \left(\alpha_j\right)^{\frac{\theta-1}{\theta}} dj\right]^{\frac{\widehat{\theta}}{\theta-1}} \left(\frac{P}{W}\right)^{\widehat{\theta}} L$$

thus

$$\begin{aligned} L_{inj} &= \frac{I^{\frac{1}{\eta-1}} s_{inj}^{\frac{\eta}{\eta-1}} \frac{1}{J}\alpha_j \left[\int_j \frac{1}{J} \left(\alpha_j\right)^{\frac{\theta-1}{\theta}} dj\right]^{\frac{\widehat{\theta}}{\theta-1}} \left(\frac{P}{W}\right)^{\widehat{\theta}} L}{A_{inj}} \\ &= \underbrace{\frac{s_{inj}^{\frac{\eta}{\eta-1}} \alpha_j \left[\int_j \frac{1}{J} \left(\alpha_j\right)^{\frac{\theta-1}{\theta}} dj\right]^{\frac{\widehat{\theta}}{\theta-1}}}{A_{inj}}}_{X_{inj}} \left[I^{\frac{1}{\eta-1}} \left(\frac{1}{J}\right) \left(\frac{P}{W}\right)^{\widehat{\theta}} L\right] \end{aligned}$$

- Finally, by aggregating L_{inj} into L, we got a function with only W unknown.

$$L_j = \underbrace{\left(\sum_i I^{\frac{1}{\widehat{\eta}}} X_{inj}^{\frac{\widehat{\eta}+1}{\widehat{\eta}}}\right)^{\frac{\widehat{\eta}}{\widehat{\eta}+1}}}_{X_j} \left[I^{\frac{1}{\eta-1}} \left(\frac{1}{J}\right) \left(\frac{P}{W}\right)^{\widehat{\theta}} L\right]$$

$$L = \underbrace{\left(\int_j J^{\frac{1}{\widehat{\theta}}} X_j^{\frac{\widehat{\theta}+1}{\widehat{\theta}}} dj\right)^{\frac{\widehat{\theta}}{\widehat{\theta}+1}}}_{X} \left[I^{\frac{1}{\eta-1}} \left(\frac{1}{J}\right) \left(\frac{P}{W}\right)^{\widehat{\theta}} L\right]$$

$$\left(\frac{W}{P}\right)^{\widehat{\theta}} = I^{\frac{1}{\eta-1}} \left(\frac{1}{J}\right) X \tag{A19}$$

Finally, we normalize $P = 1$ and use the 3 aggregation equations for the goods market clearing (A18), labor market clearing (A19) and the aggregate labor supply equation to pin down Y, W, L.

A.4 Backing out productivity distribution in levels

We use the following identities from the CES structure of preferences to rewrite the producer's first order condition:

$$e_{inj} = \frac{1}{I}\left(\frac{W_{inj}}{W_j}\right)^{(1+\hat{\eta})} = \left[\frac{W_{inj}}{\left(\sum_i W_{inj}^{1+\hat{\eta}}\right)^{\frac{1}{1+\hat{\eta}}}}\right]^{(1+\hat{\eta})} \tag{A20}$$

$$s_{inj} = \frac{1}{I}\left(\frac{P_{inj}}{P_j}\right)^{1-\eta} = \left[\frac{P_{inj}}{\left(\sum_i P_{inj}^{1-\eta}\right)^{\frac{1}{1-\eta}}}\right]^{(1-\eta)} \tag{A21}$$

$$e_{nj} = \frac{1}{I}\sum_i \left(\frac{W_{inj}}{W_j}\right)^{(1+\hat{\eta})} = \frac{\sum_{i\in\mathcal{I}_{nj}} W_{inj}^{1+\hat{\eta}}}{\sum_i W_{inj}^{1+\hat{\eta}}} \tag{A22}$$

$$s_{nj} = \frac{1}{I}\sum_i \left(\frac{P_{inj}}{P_j}\right)^{1-\eta} = \frac{\sum_{i\in\mathcal{I}_{nj}} P_{inj}^{1-\eta}}{\sum_i P_{inj}^{1-\eta}} \tag{A23}$$

Substituting these expressions into (11), we can now express the first order condition as:

$$P_{inj}\left[1 - \frac{1}{\theta}\left[\frac{\sum_{i\in\mathcal{I}_{nj}} P_{inj}^{1-\eta}}{\sum_i P_{inj}^{1-\eta}}\right] - \frac{1}{\eta}\left(1 - \left[\frac{\sum_{i\in\mathcal{I}_{nj}} P_{inj}^{1-\eta}}{\sum_i P_{inj}^{1-\eta}}\right]\right)\right] = \frac{W_{inj}}{A_{inj}}\left(1 + \frac{1}{\hat{\theta}}\left[\frac{\sum_{i\in\mathcal{I}_{nj}} W_{inj}^{1+\hat{\eta}}}{\sum_i W_{inj}^{1+\hat{\eta}}}\right] + \frac{1}{\hat{\eta}}\left(1 - \left[\frac{\sum_{i\in\mathcal{I}_{nj}} W_{inj}^{1+\hat{\eta}}}{\sum_i W_{inj}^{1+\hat{\eta}}}\right]\right)\right) \tag{A24}$$

To reduce the first order condition to a single unknown variable, we express the first order condition only in terms of the firm's employment and productivity. We know $P_{inj} = G(Y_{inj}) = F(L_{inj})$ where the first equality holds due to the inverse demand faced by a firm and the second through the production function. The firm-specific wage can be mapped to firm employment using the labor supply equation $L_{inj} = W(W_{inj})$.

Specifically, we use the following inverse demand curve and the labor supply curve

$$
\begin{aligned}
P_{inj} &= \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} Y_{inj}^{-\frac{1}{\eta}} Y_j^{\frac{1}{\eta}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P \\
&= \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} Y_{inj}^{-\frac{1}{\eta}} \left[\left(\sum_i Y_{inj}^{\frac{\eta}{\eta-1}} \right)^{\frac{\eta-1}{\eta}} \right]^{\frac{1}{\eta}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P \\
&= \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} (A_{inj}L_{inj})^{-\frac{1}{\eta}} \left[\left(\sum_i (A_{inj}L_{inj})^{\frac{\eta}{\eta-1}} \right)^{\frac{\eta-1}{\eta}} \right]^{\frac{1}{\eta}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P \qquad \text{(A25)} \\
W_{inj} &= \frac{1}{J}^{\frac{1}{-\hat{\theta}}} \frac{1}{I}^{\frac{1}{-\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} W \qquad \text{(A26)}
\end{aligned}
$$

Plugging equation (A25) and equation (A26) in (A24), gives us each firm's first order condition only in terms of A_{inj} and L_{inj}.

$$
\begin{aligned}
&\frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} (A_{inj}L_{inj})^{-\frac{1}{\eta}} \left[\left(\frac{1}{I}^{\frac{1}{\eta}} \sum_i (A_{inj}L_{inj})^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}\frac{(\theta-\eta)}{\eta\theta}} \right] \left[1 - \frac{1}{\theta} \frac{\sum_{i\in\mathcal{I}_{nj}} (A_{inj}L_{inj})^{\frac{\eta-1}{\eta}}}{\sum_i (A_{inj}L_{inj})^{\frac{\eta-1}{\hat{\eta}}}} - \frac{1}{\eta} \left(1 - \frac{\sum_{i\in\mathcal{I}_{nj}} (A_{inj}L_{inj})^{\frac{\eta-1}{\eta}}}{\sum_i (A_{inj}L_{inj})^{\frac{\eta-1}{\eta}}} \right) \right] Z \\
&= \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} \frac{(L_{inj})^{\frac{1}{\hat{\eta}}}}{A_{inj}} \left[\left(\frac{1}{I}^{\frac{-1}{\hat{\eta}}} \sum_i (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}} \right)^{\frac{\hat{\eta}}{\hat{\eta}+1}\frac{(\hat{\eta}-\hat{\theta})}{\hat{\eta}\hat{\theta}}} \right] \left[1 + \frac{1}{\hat{\theta}} \frac{\sum_{i\in\mathcal{I}_{nj}} (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}}{\sum_i (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}} + \frac{1}{\hat{\eta}} \left(1 - \frac{\sum_{i\in\mathcal{I}_{nj}} (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}}{\sum_i (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}} \right) \right] \qquad \text{(A27)}
\end{aligned}
$$

where $Z = W^{-1}L^{1/\hat{\theta}}Y^{1/\theta}$ and the aggregate price P is normalized to 1. Given these aggregate indices and I observed employment levels (L_{inj}), the system within each market with I establishments reduces to I equations in I unknown technology levels (A_{inj}). To back out the technology shocks using the above expression, we use a two step procedure described in Deb et al. (2021). During the estimation of the technology distribution we already know our previously estimated parameters $\hat{\eta}, \hat{\theta}, \overline{\varphi}$. In addition, as we use L_{inj} we know L_j, L and W from the aggregate labor supply function. To solve for aggregate out we use an initial guess $\widetilde{Y}$ in step one to solve the system of equations in each market. After solving the economy with this guess we identify the equilibrium aggregate output Y^* and in step 2 we solve the system of equations again with Y^* to identify the underlying productivity distribution.

A.5 Social Planners Solution

The planner's problem:

$$\max_{L_{inj}} U = C - \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{L^{\frac{\phi+1}{\phi}}}{\frac{\phi+1}{\phi}}$$

where

$$C = \left[\int_j J^{-\frac{1}{\theta}} C_j^{\frac{\theta-1}{\theta}} dj \right]^{\frac{\theta}{\theta-1}}$$

$$C_j = \left[\sum_i I^{-\frac{1}{\eta}} C_{inj}^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}$$

$$L = \left[\int_j J^{\frac{1}{\widehat{\theta}}} L_j^{\frac{\widehat{\theta}+1}{\widehat{\theta}}} dj \right]^{\frac{\widehat{\theta}}{\widehat{\theta}+1}}$$

$$L_j = \left[\sum_i I^{\frac{1}{\widehat{\eta}}} L_{inj}^{\frac{\widehat{\eta}+1}{\widehat{\eta}}} \right]^{\frac{\widehat{\eta}}{\widehat{\eta}+1}}$$

subject to

$$C_{inj} = A_{inj} L_{inj}$$

Lagrange function:

$$\mathcal{L}\left(C_{inj}, L_{inj}; \lambda_{inj}\right) = \left[C - \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{L^{\frac{\phi+1}{\phi}}}{\frac{\phi+1}{\phi}} \right] + \int_j \sum_i \left[\lambda_{inj} \left(C_{inj} - A_{inj} L_{inj} \right) \right] dj$$

FOC:

$$\frac{\partial \mathcal{L}}{\partial C_{inj}} = 0 = \frac{\partial C}{\partial C_{inj}} + \lambda_{inj}$$

$$\frac{\partial \mathcal{L}}{\partial L_{inj}} = 0 = -\frac{1}{\bar{\phi}^{\frac{1}{\phi}}} L^{\frac{1}{\phi}} \frac{\partial L}{\partial L_{inj}} - \lambda_{inj} A_{inj}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_{inj}} = 0 = C_{inj} - A_{inj} L_{inj}$$

From the first two FOCs, we get:

$$\lambda_{inj} = -\frac{\partial C}{\partial C_{inj}}$$
$$\lambda_{inj} A_{inj} = -\frac{1}{\bar{\phi}^{\frac{1}{\phi}}} L^{\frac{1}{\phi}} \frac{\partial L}{\partial L_{inj}}$$

$$\frac{1}{A_{inj}} = \frac{\frac{\partial C}{\partial C_{inj}}}{\frac{1}{\bar{\phi}^{\frac{1}{\phi}}} L^{\frac{1}{\phi}} \frac{\partial L}{\partial L_{inj}}}$$
$$\frac{1}{A_{inj}} = \frac{\frac{\partial C}{\partial C_j} \frac{\partial C_j}{\partial C_{inj}}}{\frac{1}{\bar{\phi}^{\frac{1}{\phi}}} L^{\frac{1}{\phi}} \left(\frac{\partial L}{\partial L_j} \frac{\partial L_j}{\partial L_{inj}} \right)}$$

$$\underbrace{I^{-\frac{1}{\eta}} J^{-\frac{1}{\theta}} \left(\frac{C_j}{C} \right)^{-\frac{1}{\theta}} \left(\frac{C_{inj}}{C_j} \right)^{-\frac{1}{\eta}}}_{P_{inj}} = \frac{1}{A_{inj}} \underbrace{I^{\frac{1}{\eta}} J^{\frac{1}{\theta}} L_{inj}^{\frac{1}{\eta}} L_j^{\frac{1}{\theta} - \frac{1}{\eta}} L^{-\frac{1}{\theta}} \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} L^{\frac{1}{\phi}}}_{W_{inj}}$$

A.6 Labor Market Elasticity Estimation

Results for Tradeables without Random Sampling. In Table (A1), we provide the results of our robustness exercise where we re-estimate the labor substitutability parameters from Section 4 without randomly assigning establishments to markets. In line with the tradeables sector, we find that the OLS estimate for both the reduced form parameter is downward biased compared to the IV. We find that the first-stage is negative and statistically significant for both the parameters. The structural estimates are also consistent with the prediction of the theory $\hat{\eta} > \hat{\theta} > 0$. Finally, we find that the estimates of $\hat{\eta}$ and $\hat{\theta}$ are lower compared to the sample when we rely on random sampling.

Table A1: **Estimates of reduced-form parameters: Tradeables without Random Sampling**

A. OLS and Second-Stage IV Estimates					
	OLS	IV		OLS	IV
	(1)	(2)		(3)	(4)
β	-0.192***	0.369***	γ	0.144***	0.281***
SE	0.0005	0.0457	SE	0.0002	0.0014
Market-Year SE	(0.003)	(0.055)	Market SE	(0.020)	(0.083)
Market x Year FE	Yes	Yes	Market FE	Yes	Yes
Establishment FE	Yes	Yes	Year FE	Yes	Yes
B. Structural Parameters					
η	-5.20	2.71	θ	-20.59	1.54
C. First-stage Regressions for the IV					
$\tau_{X(i)t}$	-	-0.003***	$\bar{\tau}_{jt}$	-	-0.109**
SE		0.0001	SE		0.0004
Market-Year SE		(0.0002)	Market SE		(0.048)
Market x Year FE	-	Yes	Market FE	-	Yes
Establishment FE	-	Yes	Year FE	-	Yes
No. of obs	2,603,000			2,676,000‡	

Notes: Standard errors clustered at the market-year level for the first stage and at the market 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 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. $\tau_{X(i)t}$ denotes the co-efficient infront of taxes in the first-stage regression for the estimate of β.

‡ Denotes the number of weighted observations.