What drives Wage Stagnation: Monopsony or Monopoly?*

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

May 2022

Abstract

Wages for the vast majority of workers have stagnated since the 1980s. We offer 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. We offer a novel way to jointly model and measure monopsony and monopoly and their contribution to wages. Our model provides a mechanism that explains the decoupling of productivity and wage growth. Using Census data for the period 1997-2016, we find that relative to an efficient economy, monopoly power contributes to 67% of the wage decline in 1997 and 77% in 2016 while monopsony power contributes to 53% of the wage decline in 1997 and 49% in 2016. While both monopoly and monopsony power lead to a substantial decline in wages relative to an efficient economy, the resulting decoupling between productivity and wages over time is primarily due to a rise in monopoly power.

Keywords. Market Power. Monopsony. Monopoly. Markdowns. Markups. Wage Stagnation. Wage Inequality. 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 provided superb 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. Disclosure Review Board number: CBDRB-FY22-CED006-0008.

†shubhdeep.deb@upf.edu
‡jan.eeckhout@upf.edu – ICREA-GSE-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 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 option 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 product 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 local 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 in itself does not 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 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 economy-wide, not just those of the firms that charge higher prices.

The objective of this paper is double. First, we propose 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 wage inequality and the skill premium. In this tractable general equilibrium model of the macroeconomy where heterogeneous firms compete in markets with goods and labor market power jointly, both markups and markdowns are simultaneously determined.

The second objective is to quantify and measure the effect on wages of market power, decomposed into monopsony and monopoly power. We use microlevel Census data – 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 observe jointly. In addition, while the concept of market power is very clear, the practical problem is that we do not easily observe market power.[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 productivity and the market structure.

The results of our estimation are as follows. 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 1.45 to 1.93, while average markdowns have increased only marginally from 1.33 to 1.38. The markup trend is consistent with the findings in De Loecker et al. (2020), with a the increase mainly driven by the upper percentiles of the markup distribution. Third, the effect of 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 majority of the wage stagnation. It contributes to 67% of wage decline in 1997 and 77% in 2016 while the contribution labor market power declines from 53% to 49% during the same period.

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. The key parameter that captures the extent of market power is the number of competitors in a local market. Fewer competitors give rise to a systematic change in the distribution of markups/markdowns, of revenue, and of output. Because we have no definition of a market, our notion of the market structure is stochastic in the sense that we randomly assign establishments from the same industry. We then obtain an estimate of the number of competitors as well as technology parameters by matching the firm level revenue over payroll and employment distribution observed in the data to our model. While this approach is certainly far less de-

[1] In the absence of direct observation, researchers have relied on indirect measures such as concentration ratios, most commonly the Herfindahl Hirschman Index. 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 Berry, Gaynor, and Scott Morton (2019a), Syverson (2019), and Eeckhout (2020)).

tailed than the demand approach for a specific, narrowly defined market (Berry, Levinsohn, and Pakes (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 the earlier paper by Deb et al. (2021) who study the contribution of different sources of market power in explaining the rise in skill premium. Market power in our model has three main components; 1) The underlying heterogeneity in the firm productivity distribution 2) The extent of competition as measured by the number of firms competing in local markets and 3) The extent of frictions faced by the household in the goods and labor market. Azar and Vives (2018) have a model with a finite number of firms competing in both input and output markets where an increase in common ownership leads to an increase in concentration.

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).[2] With sufficiently detailed 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 (2018) 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, 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 determining 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 distributions of revenues and wages implied by our model are 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 without assuming a time invariant and

[2] 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).

strict market definition as is necessary using HHI.

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.[3] 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 interesting variation across ranks, between tenured faculty whose high paying outside options are limited, and lecturers.

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 (2018)).[4] 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 shares, irrespective of whether competition increases or decreases. 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 estimating a production function in an environment with variable market structure. This is an alternate way of measuring local market power that circumvents the thorny issue of static

[3] 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.

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

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. (2018), Mertens (2021), Azkarate-Askasua and Zerecero (2020) and Morlacco (2017). De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) and Rubens (2021) use the production function approach to measure both input and output market markups jointly.

Finally, several papers use structural models to measure monopsony power. Like ours, this approach assumes a labor supply mechanism with frictions. When workers cannot costlesly move to another job, the employer can exert monopsony power. As a result, the firm extracts rents and pays a wage below the worker's marginal revenue product. In one strand of the literature, the source of the rents are search frictions. Manning (2003) and Manning (2011) formulate 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.[5] Instead of search frictions, here we model the frictions due to the imperfect ability to substitute among heterogeneous workers' labor input. This literature builds directly on Berger et al. (2022). We have market power in both the input and the output market, which allows us to decompose the effect on wages of each source. Our model uses the model developed in Deb et al. (2021).

The key objective of our paper is quantify the effect of the rise of market power on wage stagnation. 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 mechanism and new insights why wages stagnate in the absence of technological regress. In a model of perfect competition productivity growth should mirror the growth of wages, after all workers are paid their marginal revenue product and any growth in technology must show up in growth in wages. In the presence of market power, however, this may no longer hold. Market power drives a wedge between the real wage paid to workers and the productivity of the worker. As a result, as market power increases, this wedge increases leading to the de-linking of productivity and wages over time.

2 The Model

In this section we describe a model where firms have market power both in the product and labor market. In our model, market power results from three forces: 1) differentiated products/jobs in the goods/labor

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

markets; 2) Firm heterogeneity; and 3) A finite number of firms competing in a local market. We assume that the market definition of goods coincides with the market definition of the labor inputs; implying that the same set of firms in a local market compete in both the product and input market.

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 firms. There is a continuum of markets indexed by subscript j with total measure J and a finite number I establishments in each market, and N firms. Establishments, denoted by i are heterogeneous in their productivity. Firms are indexed by n. We assume that the number of establishment I in each market is constant, and the number of establishments per firm is the same for all firms in all markets, and equal to I/N. Denote the set of all the establishment i that are owned by firm n in market j as: $\mathcal{I}_{nj} = \{i \mid i \text{ in firm } n, \text{ in sector } j\}$. 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 is constant.[6] Firms within each market j have market power due to imperfect competition between firm n and the remaining $-n$ firms in the market. A representative household consumes the bundle of all goods C_{inj} produced, and it supplies labor L_{inj} in all establishments in each market.

HOUSEHOLDS. The 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 De Loecker et al. (2021) while the households preferences over differentiated jobs in modeled as in Berger et al. (2022)[7]. The household solves the following problem

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

$$\text{s.t.} \;\; PC = LW + \Pi \tag{2}$$

[6]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).

[7]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:

$$C = \left(\int_j J^{-\frac{1}{\theta}} C_j^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}} \ , \ C_j = \left(\sum_{i,n} I^{-\frac{1}{\eta}} C_{inj}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \tag{3}$$

$$L = \left(\int_j J^{\frac{1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj \right)^{\frac{\hat{\theta}}{\hat{\theta}+1}} \ , \ L_j = \left(\sum_{i,n} I^{\frac{1}{\hat{\eta}}} L_{inj}^{\frac{\hat{\eta}+1}{\hat{\eta}}} \right)^{\frac{\hat{\eta}}{\hat{\eta}+1}} \tag{4}$$

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. Let $\hat{\eta} > \hat{\theta}$ so that the jobs are more substitutable within a market than it is between markets.

FIRMS AND MARKET STRUCTURE. Firms make production decisions according to Cournot quantity competition.[8] There are N firms that compete within each local market. We model multi-establishment firms in our economy and each firm owns I establishments, where I is common across firms and across markets. 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 $\{Y_{inj}\}_{i=1}^{I}$ for each of its I establishments. In their optimal decision, they take into account the quantity decisions of all the other the firms $-n$ in its market. Since there is a continuum of markets, there is no strategic interaction between firms from different markets, only within markets. Of course, the aggregate prices P and W affect the individual firms' decisions.

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 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 I establishments, taking as given the behavior of all competing firms $-n$ in the market:

$$\Pi_{nj} = \max_{\{Y_{inj}\}_{i \in \mathcal{I}_{nj}}} \sum_{i \in \mathcal{I}_{nj}} \left[P_{inj}(Y_{inj}, Y_{-inj}) Y_{inj} - W_{inj}(L_{inj}, L_{-inj}) L_{inj} \right] \tag{5}$$

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

[8] All our results immediately extend to Bertrand price competition with differentiated goods. Everything is identical except for the residual demand elasticity that firms face. As a result, the results are qualitatively the same.

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 each establishment to maximize utility subject to the household budget constraint and feasibility constraints ($C_{inj} \leq Y_{inj}$).

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

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

$$L_{inj}(W_{inj}, W_{-inj}, W, L) = \frac{1}{J}\frac{1}{I}W_{inj}^{\hat{\eta}}W_j^{\hat{\theta}-\hat{\eta}}W^{-\hat{\theta}}L \tag{7}$$

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

$$P_j = \left(\sum_{i,n}\frac{1}{I}P_{inj}^{1-\eta}\right)^{\frac{1}{1-\eta}} \ , \ P = \left(\int_j \frac{1}{J}P_j^{1-\theta}dj\right)^{\frac{1}{1-\theta}} \tag{8}$$

$$W_j = \left(\sum_{i,n}\frac{1}{I}W_{inj}^{1+\hat{\eta}}\right)^{\frac{1}{1+\hat{\eta}}} \ , \ W = \left(\int_j \frac{1}{J}W_j^{1+\hat{\theta}}dj\right)^{\frac{1}{1+\hat{\theta}}}. \tag{9}$$

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

$$PC = \int_J \sum_{i,n} P_{inj}C_{inj}dj \ , \ WL = \int_J \sum_{i,n} W_{inj}L_{inj}dj. \tag{10}$$

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 sale share and wage bill share can be expressed as $s_{nj} = \sum_i s_{inj}$ and $e_{nj} = \sum_i e_{inj}$ respectively, where index i corresponds to all establishments owned by firm n. The firm's solution to the optimization problem (5) with respect to the output Y_{inj} of each of its I establishments satisfies:

$$P_{inj} + \frac{\partial P_{inj}}{\partial Y_{inj}}Y_{inj} + \sum_{-i\in\mathcal{I}_{nj}}\left(\frac{\partial P_{-inj}}{\partial Y_{inj}}Y_{-inj}\right) = \frac{1}{A_{inj}}\left[W_{inj} + \frac{\partial W_{inj}}{\partial L_{inj}^d}L_{inj} + \sum_{-i\in\mathcal{I}_{nj}}\left(\frac{\partial W_{-inj}}{\partial L_{inj}}L_{-inj}\right)\right] \tag{11}$$

where price and wages P_{inj} and W_{inj} are a function of the actions of the competitors Y_{i-nj}. Notice that the firm solves this condition I times 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 firms in the market.

Cournot competition in the input and output market gives us closed form solutions for these elasticities, which can be expressed in terms of market shares (see the Appendix for the derivation). Because the firm optimizes over all its establishment simultaneously, the relevant market share is the firm's *total* market share s_{nj} The first-order condition can then be written as;

$$P_{inj}\Big(1 \underbrace{-\frac{1}{\theta}s_{nj} - \frac{1}{\eta}(1-s_{nj})}_{\epsilon^P_{inj}}\Big)A_{inj} = W_{inj}\Big(1 + \underbrace{\frac{1}{\hat{\theta}}e_{nj} + \frac{1}{\hat{\eta}}(1-e_{nj})}_{\epsilon^W_{inj}}\Big) \tag{12}$$

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

$$\mu_{inj} = \frac{1}{1+\epsilon^P_{inj}} = \left(1 - \frac{1}{\theta}s_{nj} - \frac{1}{\eta}(1-s_{nj})\right)^{-1} \quad \text{and} \quad \delta_{inj} = 1 + \epsilon^W_{inj} = \left(1 + \frac{1}{\hat{\theta}}e_{nj} + \frac{1}{\hat{\eta}}(1-e_{nj})\right). \tag{13}$$

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.

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:

$$\mu_{inj} = \left(1 - \frac{1}{\theta}\frac{1}{N} - \frac{1}{\eta}\left(1 - \frac{1}{N}\right)\right)^{-1} \quad \text{and} \quad \delta_{inj} = \left(1 + \frac{1}{\hat{\theta}}\frac{1}{N} + \frac{1}{\hat{\eta}}\left(1 - \frac{1}{N}\right)\right) \tag{14}$$

We can also increase competition in the economy by increasing the number of firms competing in each market. As $N_j \to \infty$, 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 markup and markdown are:

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

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.

Alternatively, competition also increases when the elasticity of substitution of goods and labor increases within and between markets. Moving to the perfect substitutability case, $\eta, \theta, \hat{\eta}, \hat{\theta} \to \infty$ and firms become price takers. Thus, the markup and markdown in this case converge to 1.

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

Lastly, we can consider monopolistic competition, where $N_j = 1$ In this case, there is only substitability across markets, and we reach the upper bound for markups and markdowns in the model:

$$\mu_{inj} = \frac{\theta}{\theta - 1} \quad \text{and} \quad \delta_{inj} = \frac{\hat{\theta} + 1}{\hat{\theta}}. \tag{17}$$

Comparative Statics. Figure 1 shows how markups, markdowns and the median wage in the economy change as we change market structure. 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 median wage in the economy declines as the number of competitors

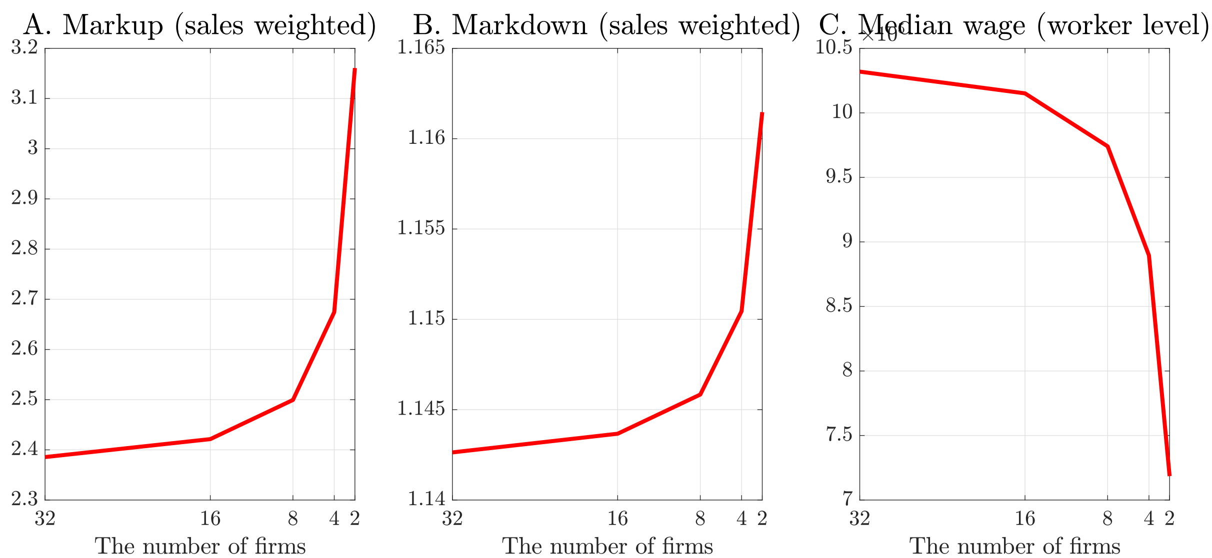


Figure 1: Comparative Statics (parameters from the estimated model below)

declines. As markets become more concentrated firms charge a higher markup and markdowns. Monopsonistic firms pay lower wages and in the aggregate a decline in the labor demand further reduces the economy wide wage.

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 establishments in the United States. The LBD provides information on ownership and organization, employment, payroll, revenue, industry (NAICS), and geography. We use data from 1997-2016, during which revenue information is available at the firm level from the Revenue Enhanced LBD. From this frame, our sample consists of firms in tradeable sectors as outlined in Delgado, Bryden, and Zyontz (2014). 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 identifiers as well as missing employment or payroll. We winsorize establishment-level employment and average wages at the 1 and 99th percentiles. Wages and Revenue are deflated to 2002 dollars.

MARKET DEFINITION. In order to solve the model, we need to define a market. In the Industrial Organizations literature, this is the key ingredient. Given our interest in the macroeconomics of market power, it is impossible to observe the market structure for each individual firm in different industries and geography. The NAICS code and the geographical definition is too broad and restrictive to define a market. There is too much variation in the market structure across industries and geography and there is mechanical variation over time. For a discussion of the problems with using NAICS codes and geographical areas to pin down the market definition, see Eeckhout (2020).

We therefore use a stochastic notion of the market definition. In the knowledge that we cannot use detailed information to define a market, instead we use the structure of the model and the random assignment of firms as competitors where firms within the same industry are equally likely to compete against each other. Yet, we determine the number of competitors N independently. Thus, even if an industry contains a large number of firms, if N is small, the extent of the competition is weak. While this approach to defining a market is much less detailed than the approach in IO, it does allow us to make progress in studying market power in the macroeconomy.

Practically, we start by defining a sector as a NAICS 6 industry. Now depending on the industry, there can be a lot of firms within each industry. We set $N \times I = 32$ and within each sector we pick 32 establishments such that we preserve the quantitative features of the entire sector.[9] Once we select those

[9] We draw 5000 random samples of 32 establishments in each sector and select the draw that best matches the joint distribution of Employment and Wages given by their mean and variance-covariance matrix within the sector.

32 establishments that form a sector, we randomly establish the identity of the firms that compete, and how many firms N are active within a market.

With this random assignment, if the number of firms N is smaller, the model predicted markups and markdowns will be higher, firm revenue will be higher, and wages will be lower. The objective is to use the observed revenue and wages from the data to estimate N. [10]

As mentioned above, we also make the assumption that the market structure is the same for both the input and output markets. In other words, if consumers view that an output market consists of products from 3 firms that are close substitutes, then in the input market the household will also view these 3 firms as a part of the same market for their labor supply decisions. The idea is that a good produced within in a market of close substitutes (say coffee shops in a geographical area) simultaneously determines a close substitutability of labor within this market. At the same time, the substitutability of goods across markets (car dealerships and coffee shops) also implies that it is much harder to switch jobs from car salesperson to barista. While we maintain a common market definition, we allow the elasticities for output goods and labor to differ: $\eta \neq \hat{\eta}$ and $\theta \neq \hat{\theta}$.

QUANTIFYING THE MODEL. Our quantification exercise closely follows Deb et al. (2021). We estimate the model in three steps. We first estimate the labor supply elasticities $\hat{\eta}$ and $\hat{\theta}$. Given our multi-establishment model note that at this stage we do not need information on the firm productivity distribution or market structure N. In the second step, for any given N, we identify the distribution of productivity in the economy that is consistent with the employment distribution observed in the data. The backed out distribution of productivities $G(A_{inj}; N)$ uses the firm-level employment L_{inj} from the data to back out the productivity. In the final step, we estimate the market structure N to match the revenue over wage bill distribution in the data. We calibrate the preference parameters $(\eta, \theta, \phi, I \times N)$ exogenously and keep them constant over time (see Table 1).

Step 1: Estimating Labor Supply Elasticities. Note first that in the model, the labor supply is independent of N. The establishment level wages W_{inj} only depend on L_{inj} and aggregates W, L, and not on the market structure N. This can be seen from the labor supply equation (18). Note that we observe both prices (wages) and quantities (employment) in the labor market separately, which allows us to use the

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

Variable	Value	Definition	Paper
θ	1.3	Output market: Between-market elasticity	DLEM(2021)
η	5.75	Output market: Within market elasticity	DLEM(2021)
ϕ	0.25	Elast. Aggregate LS	Chetty et. al(2011)
$I \times N$	32	Establishments in each market	Externally set

Table 1: Exogenous Parameters

labor supply function to estimate $\hat{\eta}$ and $\hat{\theta}$.

$$W_{inj}^{\text{model}} = \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I_j}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} W. \tag{18}$$

In logs our labor supply equation can be written as;

$$\ln W_{inj}^{\text{model}} = \underbrace{-\frac{1}{\hat{\theta}} \ln\left(\frac{1}{J}\right) \quad -\frac{1}{\hat{\theta}} \ln L + \ln W}_{k} \; \underbrace{-\frac{1}{\hat{\eta}} \ln\left(\frac{1}{I_j}\right) + \left(\frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}\right) \ln L_j}_{k_j} + \frac{1}{\hat{\eta}} \ln L_{inj} \tag{19}$$

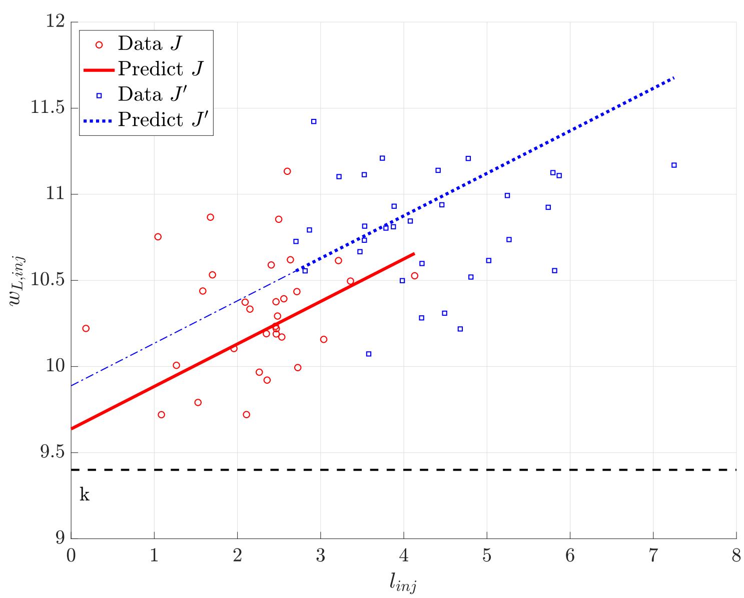


Figure 2: Estimating Labor Market Elasticities

The above equation gives us a linear relationship between employment in the data/model and the wages predicted by the model. The terms containing $I, J, \ln L$ and $\ln W$ are constants for all firms and give an economy wide constant intercept k on the wage predictions. The term $\left(\frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}\right) \ln L_j$ gives the market specific intercept k_j while the term $\frac{1}{\hat{\eta}} \ln L_{inj}$ gives the wage predictions with a constant slope of $\frac{1}{\hat{\eta}}$.

Note that at this stage W_{inj}^{model} is independent of A_{inj} and N and we only need to know the employment levels L_{inj} in the data[11]. Moreover, our estimates of productivity in step 2) and N in step 3) would lead to an endogenous distribution of L_{inj} in the model that exactly matches the employment level of each establishment in the data. As a result of this sequential estimation, the elasticities estimated in step 1 are independent of the steps ahead. We estimate the two labor market elasticities $\hat{\theta}, \hat{\eta}$ using state taxes as instruments as described below.

Estimation. To ease the exposition of our estimation strategy, we begin by re-writing Eq (19) as follows

$$\ln W_{injt}^{*} = \mathrm{k}_{jt} \ + \ \gamma \ln L_{jt} \ + \ \beta \ln L_{injt} \ + \ \underbrace{\alpha_{inj} \ + \ \epsilon_{Linjt}}_{\varepsilon_{injt}} \tag{20}$$

where we define $\beta = \frac{1}{\hat{\eta}}$ and $\gamma = \left(\frac{1}{\hat{\theta}} - \beta\right)$. We use Two-Stage Least Squares (2SLS) to estimate the reduced-form 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.

To estimate β we rely on Eq. (20). From the equation, we notice that while we observe wages and employment in the data, we do not directly observe the establishment fixed effect α_{inj} and sector-year specific constants, k_{jt} and L_{jt}, which are both functions of our structural parameter η and θ. We need to control for these unobserved variables to avoid omitted variable bias stemming from them. 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 sector and year fixed-effects. Together these two controls allow us to exploit within-establishment variation while controlling for time shocks that vary by sector. Finally, to control for endogeneity arising from correlation between the log of employment and the error term, we instrument $\ln L_{injt}$ with state corporate taxes, $\tau_{X(i)t}$. We think of the time-series variation in taxes as an exogenous shock to a firm's labor demand which help us identify the parameters of firm's labor supply equation. In order to implement our estimation strategy, we must 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. We estimate our time-invariant labor elasticity parameters using this panel.

Once we get an estimate of β (and implicitly η) from Eq. (20), we proceed to estimate γ by relying on

[11] Note that at the end of step 3, once we have estimated the elasticities (step 1), backed out the productivities (step 2) and estimated N (step 3), these primitives will generate endogenous L_{inj} in the model that match exactly the L_{inj} of each establishment in the micro data. These employment levels would then generate wages through the upward sloping labor supply function in equation (18).

the following equation derived from Eq. (20),[12]

$$\overline{\Omega}_{jt} = \mathrm{k}_{jt} + \gamma \ln L_{jt} + \bar{\varepsilon}_{jt} \tag{21}$$

As before, k_{jt}, which is itself a function of $\hat{\theta}$, is unobserved to the econometrician. To address the issue of omitted variable bias stemming from the unobservability of k_{jt}, we control for it by including a sector and a year fixed effect (separately) in Eq. (21).[13] Finally, to address the issue of endogeneity due to potential correlation between $\ln L_{jt}$ and $\bar{\varepsilon}_{jt}$, we instrument $\ln L_{jt}$ by $\bar{\tau}_{jt}$.[14] Intuitively, we exploit within sector variation and control for common time shocks to estimate γ. The results of our estimation of the reduced-form parameters are given in Table (2). Given these estimates, we calculate the value of our structural parameters $\hat{\eta}$ and $\hat{\theta}$ in Table (3).

Finally, to estimate the labor disutility parameter, we rely on the aggregate labor supply equation of the household for each skill as follows written in logs [15]:

$$\ln W_t = \ln \frac{1}{\bar{\phi}_t} + \frac{1}{\phi} \ln L_t \tag{22}$$

We calibrate the value of the Frisch elasticity, ϕ, to be equal to 0.25 by relying on the work of Chetty et al. (2011). This allows us to estimate the value of $\bar{\phi}_{St}$, one for each year, by inverting Eq. (22).

Step 2: Backing out the establishment's productivities. For any given N in order to back out the technology distribution we use the firm's first-order condition as in equation (A51). To do so we reformulate the inverse demand function, 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 $I \times N$ equations and $I \times N$ unknowns within each market. Given the L_{inj}^{data} distribution within each market 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

[12] To get to Eq. (21), start by taking $\beta \ln L_{injt}$ to the LHS in Eq. (20) to get $\ln W^*_{injt} - \beta \ln L_{injt}$. Take sectoral average on both sides to get to Eq. (21).

[13] Notice that if we were to control for k_{jt} by including an interaction of sector-year fixed-effects, we would no longer be able to identify γ as there will not be any variation in L_{jt}. Given that k_{jt} contains I_{jt}, the number of establishments within a market, we implicitly assume that the tax variation is uncorrelated with the size of the market. Giroud and Rauh (2019) have argued that there can be a non-zero correlation between market size (I_j) and taxes, which can be a threat to the identification of γ in our framework. However, Giroud and Rauh (2019) in their analysis define a market as a state which is different from the interpretation that we have adopted in our model. We define a market as NAICS-6 which straddles establishment across multiple states. Hence, we believe that in our framework we can effectively rule out any correlation between state-level taxes and the number of establishments within a given NAICS 6.

[14] In practice, when we estimate Eq. (21), we weigh each sector by its size to limit the effect of outliers on the estimate of $\hat{\theta}$.

[15] We assume that there is no measurement error in aggregate wages. Hence, we assume that $\ln W^*_t = \ln W_t$.

$G(A_{inj}; N)$.

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}.

Step 3: Estimating the Market Structure. 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. This is because $R_{inj} = \mu_{inj} \delta_{inj} W_{inj} L_{inj}$. 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 distribution 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. We use this monotonicity of $R_{inj}/(W_{inj} L_{inj})$ with respect to N to estimate market structure. We adjust the revenue in the data using $R_{inj}^{\text{Adjusted}} = \alpha_L R_{inj}^{\text{data}}$ to make it comparable to our model with labor as the only input[16]. We pin down α_N in 1997 such that market structure N is 16 in 1997[17]. In the following years we hold the value of α_N constant and estimate N by matching the sales weighted distribution of revenue over wage bill in the data and the model.

4 Results

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

LABOR SUPPLY ELASTICITIES. We estimate our model on the tradeables sector. In the Appendix, we provide results for the sample that includes both tradeables and the non-tradeables sectors. In both these cases, we define a market as NAICS-6.

Discussion of estimates. In Table (2), we present the OLS and the IV estimates of our reduced form parameters $\beta = \frac{1}{\hat{\eta}}$ and $\gamma = \frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}$. In both cases, we find that OLS is downward biased compared to

[16] We model output only as a function of labor, while in the data output could be a function of labor, capital and materials $Y_{inj} = A_{inj} L_{inj}^{\alpha_L} K_{inj}^{\alpha_K} M_{inj}^{\alpha_M}$ 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 $R_{inj}^{\text{Adjusted}} = \alpha_L R_{inj}^{\text{data}}$.

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

Table 2: **Estimates of reduced-form parameters: Tradeables**

A. OLS and Second-Stage IV Estimates					
	OLS	IV		OLS	IV
	(1)	(2)		(3)	(4)
$\frac{1}{\hat{\eta}}$	-0.187	0.287	$\frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}$	0.180	0.298
	(3.8e-4)	(0.048)		(1.3e-4)	(0.001)
Sector x Year FE	Yes	Yes	Sector FE	Yes	Yes
Establishment FE	Yes	Yes	Year FE	Yes	Yes
B. First-Stage Regressions for the IV					
$\tau_{X(i)t}$	-	-0.003	$\bar{\tau}_{jt}$	-	-0.138
		(1.9e-4)			(3.8e-4)
Sector x Year FE	-	Yes	Sector FE	-	Yes
Establishment FE	-	Yes	Year FE	-	Yes
No. of obs.	3,921,000	3,921,000	No. of obs.	3,921,000	3,921,000

Notes: Standard errors are reported in parenthesis. Estimates of γ in columns 3 and 4 are conditional on the estimates of columns 1 and 2, respectively. We define a market as all the establishments in the tradeables sector within NAICS-6. Further results pertaining to a sample that includes tradeables and non-tradeables are provided in the Appendix. Number of observations are common for both the first and the second-stage. The number of observations reflects rounding for disclosure avoidance.

Table 3: **Estimates of structural parameters: Tradeables**

Parameter	Description	Estimate IV
$\hat{\eta}$	Within-market elasticity	3.49
$\hat{\theta}$	Between-market elasticity	1.71

the IV. More importantly, the OLS estimate for β is not consistent with the theory as it shows a negative relationship between wages and employment in the firm's labor supply curve. Using the IV corrects for the bias and shows that the corresponding structural parameters η and θ provided in Table (3) are in line with the theory: $\eta > \theta > 0$, i.e., jobs within a market are more substitutable than jobs between markets. The structural estimates of η and θ also pin down that bounds on the distribution of markdowns in the model. We find that the lower bound of the markdown distribution is $\frac{\hat{\eta}+1}{\hat{\eta}} = 1.28$ and the upper bound is $\frac{\hat{\theta}+1}{\hat{\theta}} = 1.58$, implying that the wedge between the marginal revenue product of labor and the wage can

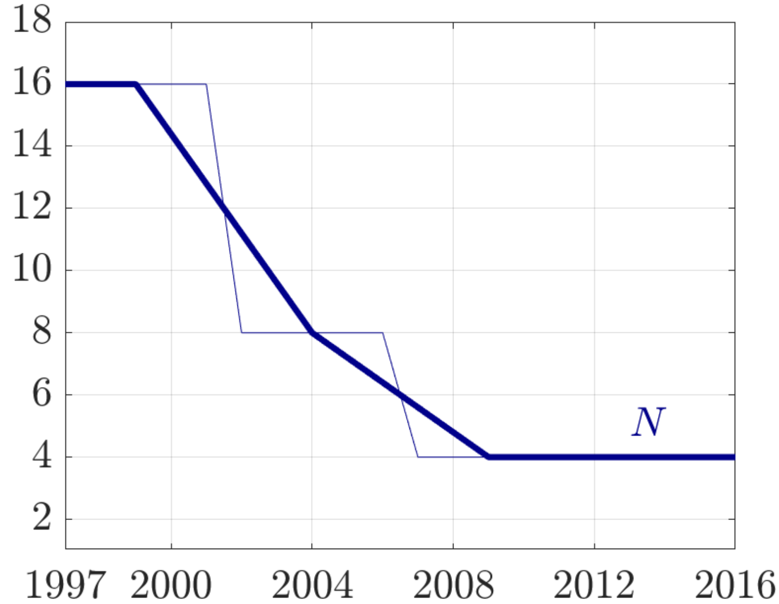


Figure 3: Estimated Market Structure

be anywhere between 28% and 58%.

Finally, in Table (2), we also provide the first-stage estimates of our IV. In both cases we find that the first-stage is negative and statistically significant. In the case of the estimation of β, we find that taxes are negatively correlated with employment at the establishment-level. For the estimation of γ, we also find a similar negative relationship between average sector-level employment and average market-level taxes in the data. This reduced-form relationship between employment and taxes is consistent with the evidence presented in Giroud and Rauh (2019) and Berger et al. (2022).

Despite the same underlying data our estimate of $\hat{\theta}$ is higher and the estimate of $\hat{\eta}$ is lower than in theirs. This would imply that the difference between the upper and lower bounds for markdowns we estimate $\{(\hat{\eta}+1)/\hat{\eta}, (\hat{\theta}+1)/\hat{\theta}\}$ are smaller. The difference could arise from three main sources. First, and perhaps the most important difference could arise from the difference in the coarseness of NAICS code. We use NAICS 6 while they use NAICS 3 x geography as a market. As a result, due to the absence of geography the jobs within our markets are more differentiated which leads to less substitutability within a market and a lower estimate of the within market elasticity $\hat{\eta}$. Second,Berger et al. (2022) use taxes as an instrument to compute the elasticity of the labor supply function and then use indirect inference in the model to estimate the $\hat{\eta}$ and $\hat{\theta}$ consistent with the empirical estimates. On the other hand we rely on the structure of the labor supply function and use the instrument to estimate the elasticities $\hat{\eta}$ and $\hat{\theta}$ directly given the observed employment and wages in the data. Third, our estimates of the labor supply function are at the establishment level while they estimate a firm level level labor supply function.

NUMBER OF COMPETITORS. Figure 3 plots the evolution of the estimated N for each year[18]. We find that number of competing firms within a market decreases gradually from 16 in 1997 to 4 in 2016. This is consistent with the evidence on increasing concentration at the national level. Through the lens of our model, given the employment levels in the data[19] and the wages predicted by the model, the increase in the wedge between revenue over wage bill in the data can only be reconciled by a decline in competition (N) within the model. As N declines, the model is able to match the increase in the revenue over wage bill ratio in the data which leads to our result on the estimation of market structure.

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 firm as predicted by the model. In Figure 4 we plot the evolution of aggregate, revenue-weighted markups, which have increased from 1.45 to 1.93 from 1997 to 2016. The increase in the trend of the markups is in line with the markups in De Loecker et al. (2020).

Over the same period, markdowns have remained stable and increased only marginally from 1.33 to 1.38. Firms do exert some 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), this is not reflected in the markdown over time. The main reason is that the estimated labor supply elasticity $\hat{\theta}$ implies a smaller upper bound for markdowns (1.58) which is significantly lower than the upper bound for markups (4.33) given by θ. Therefore, this difference in elasticity estimate leads to only marginal increases in markdowns over the entire sample.

The change in market power is driven by changes in the distribution across firms. In Figure 5 we plot the distributional shifts in the unweighted markups and markdowns in 1997 and 2016. We find that the the variance of markups has increased substantially and that the right tail has much higher mass in 2016 than in 1997, a fact that is consistent with the findings in De Loecker et al. (2020). Instead, the distribution of markdowns is very compact and with much less variance in markdowns across firms.

We also analyze the changes in the percentiles of the revenue-weighted markup and markdown distributions in Figure 6. We find that $P90$, the ninetieth percentile of the markup distribution increases

[18] Throughout the paper the thick lines correspond to 5 year moving averages and thinner lines correspond to estimated or model values

[19] The establishment level employment in the data and the model are the same given our backed out productivities in step 2 of our estimation procedure.

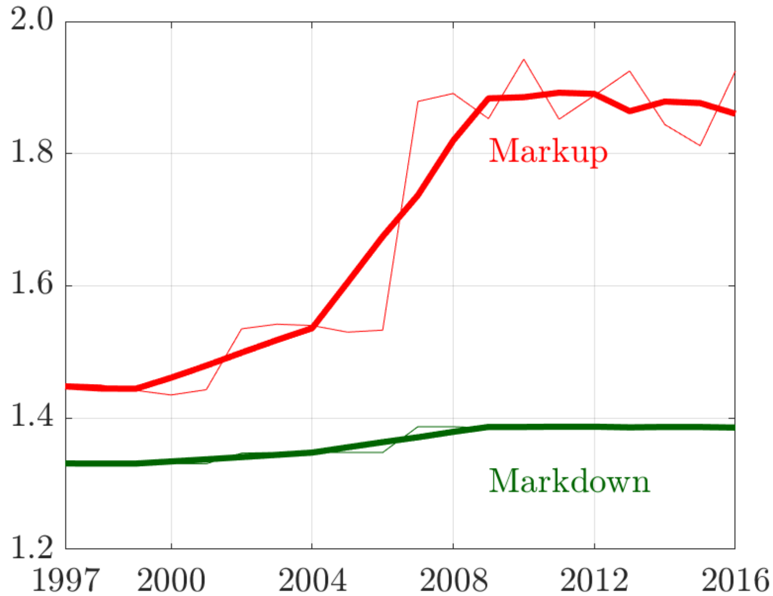
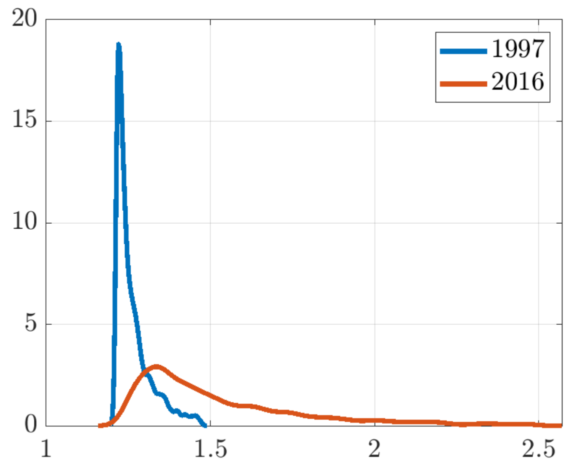
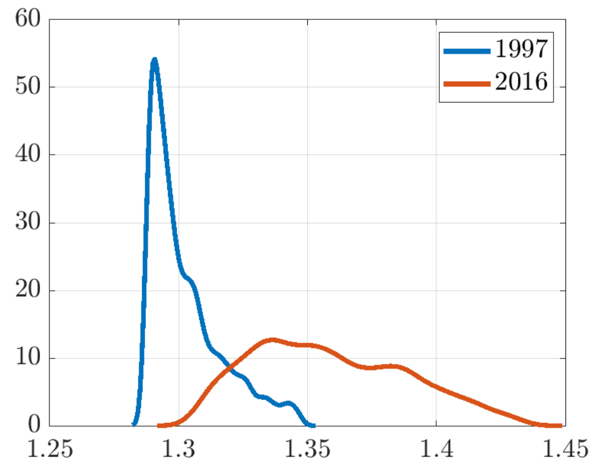


Figure 4: Average Markups and Markdowns. Average is revenue weighted.



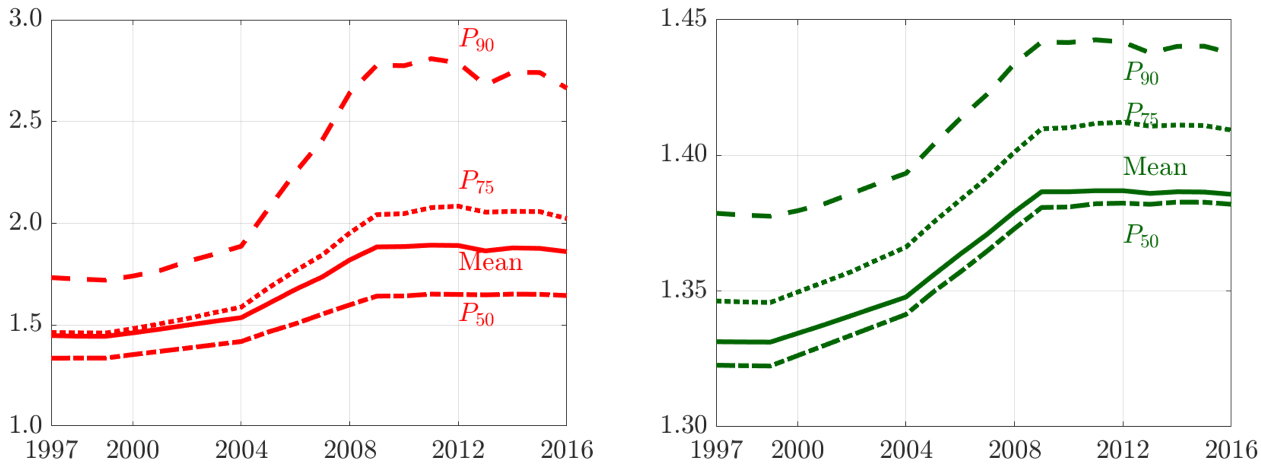
(a) Markup



(b) Markdown

Figure 5: Kernel density (unweighted): Markup and Markdown

from 1.74 in 1997 to 2.89 in 2016. This increase of 66 percent suggests that firms with high markups have increased their sales share in the economy, leading to an increase in the aggregate markup, consistent with evidence in De Loecker et al. (2020).



(a) Percentiles of Markup distribution

(b) Percentiles of Markdown distribution

Figure 6: Percentiles of (sales weighted) Markup and Markdown distribution

5 Wage Stagnation

PRODUCTIVITY-WAGE DECOUPLING. The US economy has witnessed a disconnect between productivity and the wage of a typical worker since the early 1980's. Figure (7) plots the wages of non-supervisory production workers and the GDP per worker from 1948 until 2020[20]. While GDP per worker and wages increase at the same rate until 1980, their growth rates start decoupling after 1980.

Figure (8) 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 both figures. Figure (8a) plots GDP per worker by dividing real GDP by employment and the average wages in the data. The solid lines correspond to the values in the Census (tradeables sector) data while the dashed lines correspond to the aggregate US economy as in Figure (7). In the Census data the GDP per worker grows by 43% points while wages only increase by 20% points[21]. During this period the GDP per worker in the US increased by roughly 30% points while wages for non-supervisory workers only increase by 10% points. In both these data sets the difference between the growth rate of GDP per worker and wages are similar, around 20% points.

In figure(8b) we plot the GDP per worker and wages implied by our model where the underlying data is the tradeable sectors within Census data. We see that wages increase by roughly 27 % in the

[20]Wages in the US in 2012 USD (source : Bureau of Labor Statistics-Current Employment Statistics). The GDP per worker in the US is calculated by dividing real GDP in 2012 USD (source: Bureau of Economic Analysis) by employed labor force (source : Bureau of Labor Statistics).

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

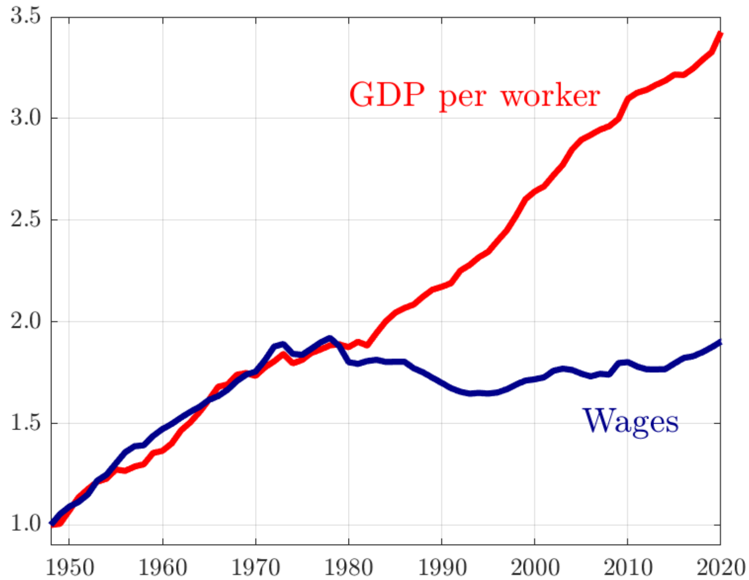
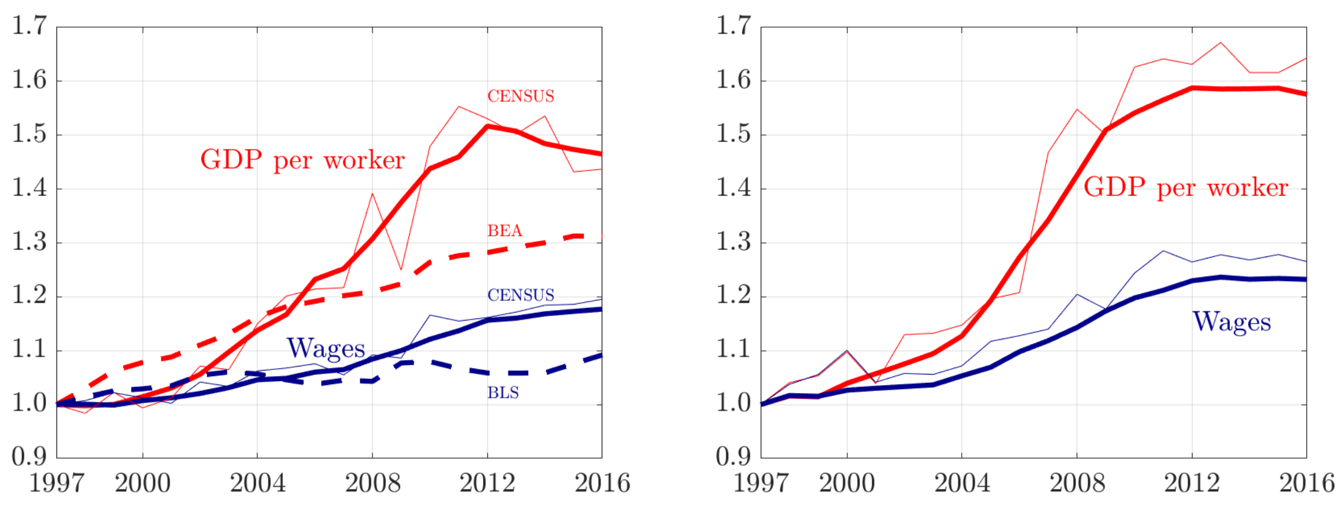


Figure 7: Productivity Wage decoupling since 1948

model which is comparable to the 20% increase in the data. The GDP per worker increases by 64% in the model and by 43% in the data[22]. Despite generating the decoupling between GDP per worker and wages, our model however overpredicts the growth of GDP per worker relative to the underlying data.



(a) Data:

(b) Model

Figure 8: Productivity-Wage Decoupling: Data vs Model

In a world with perfect competition in both goods and labor markets, real wage is equal to the pro-

[22]The thin lines represent model numbers and the thick lines represent the 5 year moving averages.

ductivity of the worker; $A = W/P$. This implies $\Delta_t \ln A = \Delta_t \ln(W/P)$ as a result over time any growth in productivity leads to an equivalent growth in wages. However with market power, the first order condition (12) for firms can be rearranged to be written as:

$$\underbrace{\frac{P_{inj}A_{inj}}{W_{inj}}}_{\text{Ratio of TFPR to Wage}} = \underbrace{\mu_{inj}}_{\text{Markup}} \times \underbrace{\delta_{inj}}_{\text{Markdown}} \tag{23}$$

which implies that the markups and markdowns form a wedge between the dollar value of a firm's productivity and the wage paid by that firm. The higher the market power in the economy, the greater is the wedge. Moreover, the identity implies that the growth rate of each component must follow:

$$\underbrace{\Delta_t \ln(P_{inj,t}A_{inj,t})}_{\text{TFPR growth}} = \underbrace{\Delta_t \ln(W_{inj,t})}_{\text{Wage growth}} + \underbrace{\Delta_t \ln(\mu_{inj,t})}_{\text{Markup growth}} + \underbrace{\Delta_t \ln(\delta_{inj,t})}_{\text{Markdown growth}} \tag{24}$$

The equation suggests that the growth in 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 per worker we can re-write the first order condition as;

$$\underbrace{\frac{R_{inj}}{L_{inj}}}_{\text{Revenue per worker}} = \underbrace{W_{inj}}_{\text{Wage}} \times \underbrace{\mu_{inj}}_{\text{Markup}} \times \underbrace{\delta_{inj}}_{\text{Markdown}} \tag{25}$$

While equation (25) shows how the market power drives a wedge between revenue over wage bill for a firm, we can aggregate the firms first order condition to derive an equivalent expression in the aggregate.

$$\underbrace{\Delta_t \ln\left(\frac{\int_j \sum_{i,n} R_{inj,t}}{\int_j \sum_{i,n} L_{inj,t}}\right)}_{\text{GDP per worker growth}} - \underbrace{\Delta_t \ln\left(\frac{\int_j \sum_{i,n} L_{inj,t}W_{inj,t}}{\int_j \sum_{i,n} L_{inj,t}}\right)}_{\text{Average worker wage growth}} = \underbrace{\Delta_t \ln\left(\int_j \sum_{i,n} \frac{R_{inj,t}}{R_t} \frac{1}{\mu_{inj,t}\delta_{inj,t}} dj\right)}_{\text{Aggregate Wegde growth}} \tag{26}$$

Equation (26) shows that the difference between GDP per worker and average wage of workers can be expressed as the sales weighted average of the firm level wedge $\frac{1}{\mu_{inj}\delta_{inj}}$. Over time, if firms increase their market power, especially, firms that have a high sales share in the economy, the implication of this would be a rise in the decoupling between GDP per worker and average wages in the economy.

Another way to evaluate the role of aggregate markups and markdowns in the decoupling would be

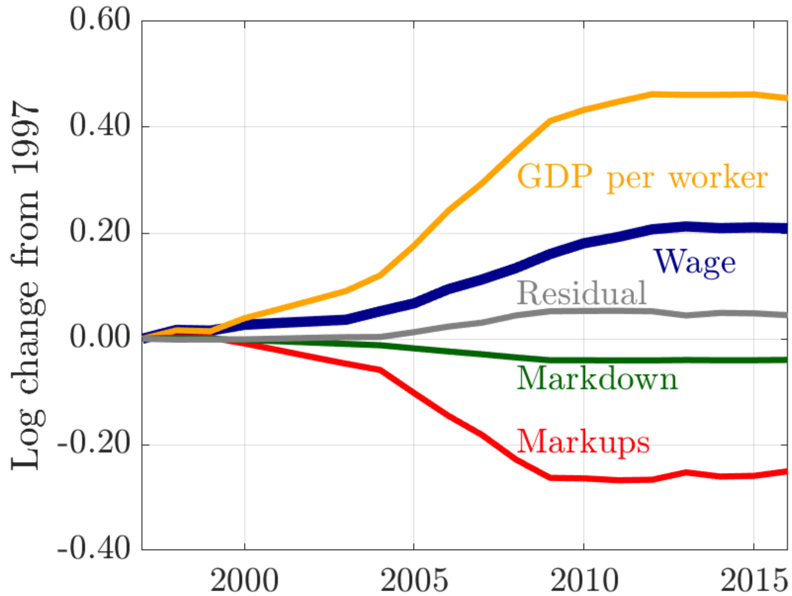


Figure 9: Decomposition of wages.

to consider;

$$W = \frac{\text{GDP per worker}}{\mu\delta}\Omega \tag{27}$$

where $W = \int_J \sum_{i,n} \frac{L_{inj}W_{inj}}{\int_j \sum_{i,n} L_{inj}}$ is the employment weighted establishment wage in the economy which is equivalent to the average wage of workers. GDP per worker is given by $\frac{\int_j \sum_{i,n} R_{inj}}{\int_j \sum_{i,n} L_{inj}}$ and the aggregate markup and markdown are sales weighted. Ω in the expression is the residual. Figure (9) plots 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 the wages, the rise of aggregate markup leads to a significant downward pressure on wage while the rise of aggregate markdown contributes only marginally to the stagnation of wages. This is because the increase in markups is substantial in comparison to markdowns since 1997.

COUNTERFACTUAL ECONOMIES. Wages change both due to the output market power of firms – a general equilibrium effect – or due monopsony – a direct effect on wages. We now analyze several counterfactual economies to decompose the effect on wages that is due output and input market power. To that effect, 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. Each time we analyze the effect on the wage level.

The social planner takes the consumer preferences as given and maximizes consumer utility subject to the aggregate resource constraint. The social planner solves

$$V = \max_{C_{inj}, L_{inj}} \sum_{0}^{\infty} \left(C - \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{L^{\frac{\phi+1}{\phi}}}{\frac{\phi+1}{\phi}} \right) \tag{28}$$

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

and also subject to the aggregation equations (3) and (4). This helps us reduce the planner's problem to the optimum allocation of labor and consumption and the first order condition is given by:[23]

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

Equation (29) gives a planner's allocation of labor L_{inj}^{oo}. If there exists a decentralized economy with the price index P with perfect competition in both input and output markets, then this decentralized economy would be first best efficient if $P = 1$.

In what follows, we define $L_{inj}^{s_1 s_2}$, $s_k \in \{\star, o\}$, where $\star$ denotes that the solution under market power and o denotes the planner's optimal solution. For, instance $L_{inj}^{\star o}$ is the labor allocation when there is goods market power but no labor/input market power. Then the decentralized allocation with market power in both output and input markets is given by:

$$[L_{inj}^{\star\star}] \ : \ \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-\frac{1}{\eta}} C_j^{\frac{1}{\eta}-\frac{1}{\theta}} C^{\frac{1}{\theta}} \Big(1 - \frac{s_{inj}}{\theta} - \frac{(1-s_{inj})}{\eta}\Big) A_{inj} = \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} L^{\frac{1}{\phi}} \Big(1 + \frac{e_{inj}}{\hat{\theta}} + \frac{(1-e_{inj})}{\hat{\eta}}\Big) \tag{30}$$

Note that, this is the same equation as in our baseline model with markups and markdowns. Similarly,

[23]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 $\frac{\frac{1}{\bar{\phi}^{\frac{1}{\phi}}} L_{inj}^{1/\hat{\eta}} L_j^{1/\hat{\theta}-1/\hat{\eta}} L^{-1/\hat{\theta}} L^{1/\phi}}{C_{inj}^{-1/\eta} c_j^{1/\eta-1/\theta} C^{1/\theta}} = A_{inj}$

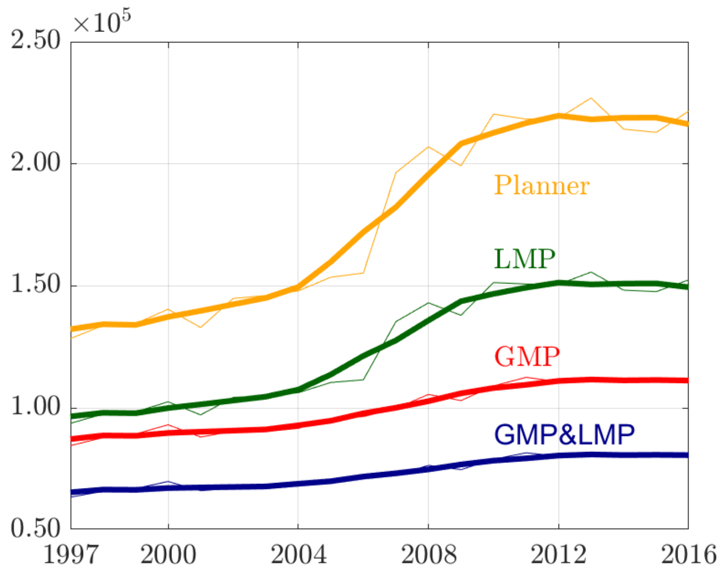


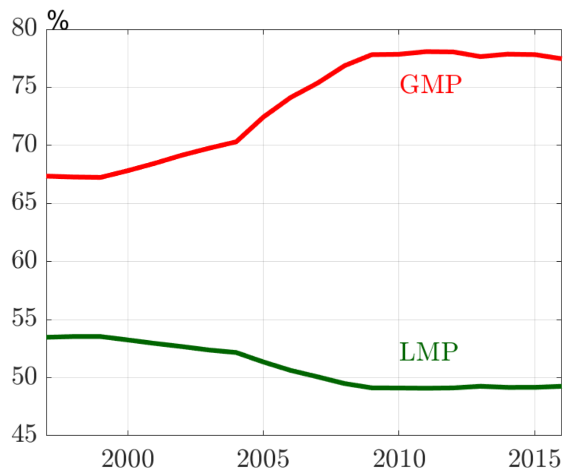
Figure 10: Market power and counterfactual wages

$L^{\star o}_{inj}$ and $L^{o\star}_{inj}$ are defined as

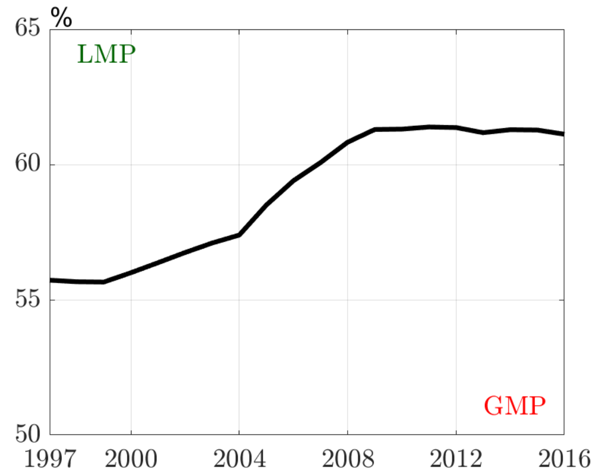
$$[L^{o\star}_{inj}] \;:\; \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-\frac{1}{\eta}} C_j^{\frac{1}{\eta}-\frac{1}{\theta}} C^{\frac{1}{\theta}} A_{inj} = \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} L^{\frac{1}{\phi}} \left(1 + \frac{1}{\hat{\theta}} e_{inj} + \frac{1}{\hat{\eta}}(1 - e_{inj})\right) \tag{31}$$

$$[L^{\star o}_{inj}] \;:\; \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} C_{inj}^{-\frac{1}{\eta}} C_j^{\frac{1}{\eta}-\frac{1}{\theta}} C^{\frac{1}{\theta}} \left(1 - \frac{1}{\theta} s_{inj} - \frac{1}{\eta}(1 - s_{inj})\right) A_{inj} = \frac{1}{\bar{\phi}^{\frac{1}{\phi}}} \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} L^{\frac{1}{\phi}} \tag{32}$$

Counterfactual Wage Level Decomposition. We can now solve the general equilibrium model under the four regimes and compare the distribution of wages over time. In Figure (10), the blue line represents the evolution of wages in the baseline model with goods market power and labor market power. The yellow line denotes the wage for the planner's solution with no market power in goods or labor inputs. The intermediate green and red series denote the wage when there is only market power in the labor and goods market, respectively. Both goods market power and labor market power decrease the wage relative to the planner's solution. Firm level wages are closely linked to a firm's employment in the model due to the labor supply curve. Therefore an increase in labor market power leads to a reduction in a firm's wage. At the same time, output market power reduces the level of output of the firm, 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 leads to a fall in the aggregate wage level W. The planner, by setting the output to the efficient level, substantially increases the level of employment and the resulting wage.



(a) Evolution of wages over time



(b) Contribution of market power to wage decline

Figure 11: Contribution of market power to wage decline

Figure (11a) shows the evolution of the percentage contribution of output market power and labor market power in explaining the wage decline relative to social planner. We find that output market power contributes to 67% and 77% of the decline in 1997 and 2016 respectively. Labor market power also contributes a wage decline of around 53% in 1997 but it declines to 49% in 2016. This implies that the dominant firms exercise their market power in the goods market in order to restrict labor demand more than their market power in the labor market. Note that the shares do not add up to exactly one hundred percent because the system is non-linear and in the counterfactual experiment we therefore have to take into account the role of joint interactions of goods and labor market power. Moreover, we can see that the relative importance of output market power in the decline of wages is increasing over time. To see this, we plot the relative contribution of output market power in wage decline as a fraction of the contribution of both market powers GMP/(GMP+LMP) in figure (11b). We find that the relative contribution of output market power has been increasing from 56% to 61%.

6 Conclusion

In this paper we propose a general equilibrium model of the macroeconomy with the simultaneous determination of markups and markdowns. We use this model to infer both markups from output goods market power and markdowns due to monopsony power in the labor markets. We take this model to micro data. In the process we estimate firm-level productivity as well as the economy-wide market structure, using a novel way of estimating 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 1.45 to 1.93. Instead, markdowns have increased only marginally from 1.33 to 1.38 over the entire period.

The presence of market power, both monopoly and monopsony, can account for lower wages relative to an efficient economy. We perform counterfactual experiments to explain 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. We find that the general equilibrium effect of monopoly power on real wages dominates the effect of monopsony power on wages directly. In 1997 monopoly accounts for 67% of the wage stagnation whereas monopsony accounts for 53%. Over time, the contribution of monopoly power to wage decline has increased to 77% while monopsony power has declined to 49% in 2016. As a result, while both monopoly and monopsony power substantially lower the equilibrium wage, the contribution of monopoly power in wage decline has increased over time, leading to the decoupling between productivity and wages between 1997 and 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 (2018): "Oligopoly, Macroeconomic Performance, and Competition Policy," Mimeo IESE.

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.

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2018): "Monopsony in Online Labor Markets," 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.

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 (2018): "Concentration in US local labor markets: evidence from vacancy and employment data," Cleveland Fed mimeo.

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.

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 CONSUMPTION FUNCTIONS In the economy with CES aggregation technology, total consumption within a household C can be written as ($\alpha(j) = \frac{1}{J}$ and $\beta(i) = \frac{1}{I}$ in our model):

$$C = \left(\int_j \left(\frac{1}{J}\right)^{\frac{1}{\theta}} \left(\sum_i \left(\frac{1}{I}\right)^{\frac{1}{\eta}} C_{inj}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}} \tag{A1}$$

To derive the demand function, we first solve a maximization problem such that C is maximized with chosen c_{inj} subject to the budget constraint

$$\int_j \sum_{i,n} P_{inj} C_{inj} dj \leq Z (= WL + \Pi) \tag{A2}$$

where Z is total amount of money spent. This optimization problem is equivalent to the lagarian (Maximizing the monotonic transformation of C is easier and gives the same results since C is strictly increasing in c_{inj}):

$$\mathcal{L} = \left(\int_j \left(\frac{1}{J}\right)^{\frac{1}{\theta}} \left(\sum_i \left(\frac{1}{I}\right)^{\frac{1}{\eta}} C_{inj}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}\frac{\theta-1}{\theta}} dj \right) - \Lambda \left(\int_j \sum_{i,n} P_{inj} C_{inj} dj - Z \right),$$

The first order condition is

$$C_{inj} = (\frac{P_{inj}}{P_{i'j}})^{-\eta} C_{i'j}, \forall j \tag{A3}$$

Multiply both sides of (A3) by P_{inj} take the sum over i we can get

$$\begin{aligned} \sum_{i,n} P_{inj} C_{inj} &= \sum_{i,n} P_{inj}^{1-\eta} P_{i'j}^{\eta} C_{i'j}, \forall j \\ \Rightarrow Z_j &= P_{i'j}^{\eta} C_{i'j} \sum_{i,n} P_{inj}^{1-\eta}, \forall j \\ \Rightarrow C_{inj} &= \frac{Z_j P_{inj}^{-\eta}}{\sum_{i,n} P_{inj}^{1-\eta}}, \forall j \end{aligned} \tag{A4}$$

We want to derive P_j as the expenditure to buy one unit of C_j, which is $Z_j|_{C_j=1}$, and it naturally follows that

$$C_j = \left(\sum_i \left(\frac{1}{I}\right)^{\frac{1}{\eta}} (\frac{Z_j P_{inj}^{-\eta}}{\sum_{i,n} P_{inj}^{1-\eta}})^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} = Z_j \left(\frac{1}{I}\right)^{\frac{1}{\eta-1}} (\sum_{i,n} P_{inj}^{1-\eta})^{\frac{1}{\eta-1}}$$

$$\Rightarrow P_j = \left(\sum_{i,n} \left(\frac{1}{I}\right) P_{inj}^{1-\eta} \right)^{\frac{1}{1-\eta}}, \forall j$$

From (A2) we know that $\sum_{i,n} P_{inj} C_{inj} = Z_j$ and $Z_j = P_j C_j$ from the definition of P_j. We can write $\sum_{i,n} P_{inj} = C_{inj} = P_j C_j$. Thus we can do similar algebra to P_j

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

and the first order condition is

$$C_j = (\frac{P_j}{P_{j'}})^{-\theta} C_{j'}. \tag{A5}$$

We have

$$\begin{aligned} Z &= \int_j P_j C_j dj = \int_j P_j (\frac{P_j}{P_{j'}})^{-\theta} C_{j'} dj = P_{j'}^{\theta} C_{j'} \int_j P_j^{1-\theta} dj \\ \Rightarrow C_j &= \frac{Z P_j^{-\theta}}{\int_j P_j^{1-\theta} dj}, \forall j \end{aligned} \tag{A6}$$

Similarly, we want to derive P as the expenditure to buy one unit of C, which is $Z|_{C=1}$, and it naturally follows that

$$C = \left(\int_j \left(\frac{1}{J}\right)^{\frac{1}{\theta}} C_j^{\frac{\theta-1}{\theta}} dj \right)^{\frac{\theta}{\theta-1}} = \left(\int_j \left(\frac{1}{J}\right)^{\frac{1}{\theta}} (\frac{Z P_{j'}^{-\theta}}{\int_j P_j^{1-\theta} dj} dj')^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}} = Z \left(\frac{1}{J}\right)^{\frac{1}{\theta-1}} \int_j P_j^{1-\theta} dj^{\frac{1}{\theta-1}}$$

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

With C_{inj} in (A4), C_j in (A6), and $Z = PC$, we can get

$$C_{inj} = \frac{Z_j P_{inj}^{-\eta}}{\sum_{i,n} P_{inj}^{1-\eta}} = \frac{1}{J}\frac{1}{I} P_{inj}^{-\eta} P_j^{\eta-\theta} P^{\theta} C,$$

To get the inverse demand function, we start by within-market demand

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

Inverting this expression, we get

$$P_{inj} = I^{-\frac{1}{\eta}} \left(\frac{C_{inj}}{C_j} \right)^{-\frac{1}{\eta}} P_j \tag{A7}$$

Next, we invert the between-market demand function to express P_j as a function of market aggregates

$$P_j = J^{-\frac{1}{\theta}} \left(\frac{C_j}{C} \right)^{-\frac{1}{\theta}} P \tag{A8}$$

Replacing P_j in equation (A7) by (A8) gives us

$$P_{inj} = \left(\frac{1}{J} \right)^{\frac{1}{\theta}} \left(\frac{1}{I} \right)^{\frac{1}{\eta}} \left(\frac{C_{inj}}{C_j} \right)^{-\frac{1}{\eta}} \left(\frac{C_j}{C} \right)^{-\frac{1}{\theta}} P \tag{A9}$$

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;

$$\text{Min}_{L_j} \left(\int_j \left(\frac{1}{J} \right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj \right)^{\frac{\hat{\theta}}{\hat{\theta}+1}} \text{ s.t} \int_J W_j L_j \geq Z \tag{A10}$$

Then the optimal allocation is given by ;

$$\frac{\hat{\theta}}{\hat{\theta}+1} \left(\int_j \left(\frac{1}{J} \right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj \right)^{\frac{\hat{\theta}}{\hat{\theta}+1}-1} \left(\frac{1}{J} \right)^{\frac{-1}{\hat{\theta}}} \frac{\hat{\theta}+1}{\hat{\theta}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}-1} = \lambda W_j \tag{A11}$$

$$\underbrace{\left(\int_j \left(\frac{1}{J} \right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}} dj \right)^{\frac{\hat{\theta}}{\hat{\theta}+1}-1}}_{L^{\frac{-1}{\hat{\theta}}}} \left(\frac{1}{J} \right)^{\frac{-1}{\hat{\theta}}} L_j^{\frac{\hat{\theta}+1}{\hat{\theta}}-1} = \lambda W_j \tag{A12}$$

$$\frac{1}{J}^{\frac{-1}{\hat{\theta}}} L^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1}{\hat{\theta}}} = \lambda W_j \tag{A13}$$

Next multiply each side by L_j and integrate across J

$$\frac{1}{J}^{\frac{-1}{\hat{\theta}}} L^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1+\hat{\theta}}{\hat{\theta}}} = \lambda W_j L_j \tag{A14}$$

$$L^{\frac{-1}{\hat{\theta}}} \underbrace{\int_j \frac{1}{J}^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1+\hat{\theta}}{\hat{\theta}}} dj}_{L^{\frac{\hat{\theta}+1}{\hat{\theta}}}} = \lambda \int_j W_j L_j dj \tag{A15}$$

$$L^{\frac{-1}{\hat{\theta}}} L^{\frac{\hat{\theta}+1}{\hat{\theta}}} = \lambda \int_j W_j L_j dj \tag{A16}$$

$$\tag{A17}$$

which is equivalent to;

$$L = \lambda \int_j W_j L_j dj \tag{A18}$$

We define the market 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.

$$\frac{1}{J}^{\frac{-1}{\hat{\theta}}} L^{\frac{-1}{\hat{\theta}}} L_j^{\frac{1}{\hat{\theta}}} = \frac{W_j}{W} \tag{A19}$$

$$L_j^{\frac{1}{\hat{\theta}}} = \frac{1}{J}^{\frac{1}{\hat{\theta}}} \frac{W_j}{W} L^{\frac{1}{\hat{\theta}}} \tag{A20}$$

$$L_j = \left(\frac{1}{J}\right) \left(\frac{W_j}{W}\right)^{\hat{\theta}} L \tag{A21}$$

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

$$\int_J W_j L_j dj = \left(\frac{1}{J}\right) \left(\frac{1}{W}\right)^{\hat{\theta}} L \int_J W_j^{1+\hat{\theta}} dj \tag{A22}$$

$$WL = \left(\frac{1}{J}\right) \left(\frac{1}{W}\right)^{\hat{\theta}} L \int_J W_j^{1+\hat{\theta}} dj \tag{A23}$$

$$W^{1+\hat{\theta}} = \left(\frac{1}{J}\right) \int_J W_j^{1+\hat{\theta}} dj \tag{A24}$$

$$W = \left(\left(\frac{1}{J}\right) \int_J W_j^{1+\hat{\theta}} dj\right)^{\frac{1}{1+\hat{\theta}}} \tag{A25}$$

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

$$L_{inj} = \left(\frac{1}{I}\right)\left(\frac{W_{inj}}{W_j}\right)^{\hat{\eta}} L_j \tag{A26}$$

The market wage index is;.

$$W_j = \left(\left(\frac{1}{I}\right)\sum_{i,n} W_{inj}^{1+\hat{\eta}}\right)^{\frac{1}{1+\hat{\eta}}} \tag{A27}$$

Then the firm level labor supply curve is given by;

$$L_{inj} = \left(\frac{1}{J}\right)\left(\frac{1}{I}\right)\left(\frac{W_{inj}}{W_j}\right)^{\hat{\eta}}\left(\frac{W_j}{W}\right)^{\hat{\theta}} L \tag{A28}$$

To derive the inverse labor supply function we can write;

$$\left(\frac{1}{J}\right)^{-1}\frac{L_j}{L} = \left(\frac{W_j}{W}\right)^{\hat{\theta}} \tag{A29}$$

$$W_j = \left(\frac{1}{J}\right)^{\frac{-1}{\hat{\theta}}}\left(\frac{L_j}{L}\right)^{\frac{1}{\hat{\theta}}} W \tag{A30}$$

similarly at the firm level;

$$W_{inj} = \left(\frac{1}{I}\right)^{\frac{-1}{\hat{\eta}}}\left(\frac{L_{inj}}{L_j}\right)^{\frac{1}{\hat{\eta}}} W_j \tag{A31}$$

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

$$W_{inj} = \frac{1}{J}^{\frac{-1}{\hat{\theta}}}\frac{1}{I}^{\frac{-1}{\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} W \tag{A32}$$

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 $s_{nj} = \sum_i s_{inj}$ and $e_{nj} = \sum_i e_{inj}$, respectively. Firm's problem here is to choose an output level Y_{inj} for each establishment i simultaneously to maximize its profit:

$$\Pi_{nj} = \max_{Y_{inj}} \sum_i \left(P_{inj}Y_{inj} - \frac{W_{inj}}{A_{inj}}Y_{inj}\right)$$

The FOC gives:

$$P_{inj} + \frac{\partial P_{inj}}{\partial Y_{inj}} Y_{inj} + \sum_{i' \neq i} \left(\frac{\partial P_{i'nj}}{\partial Y_{inj}} Y_{i'nj} \right) = \frac{1}{A_{inj}} \left[W_{inj} + \frac{\partial W_{inj}}{\partial L^d_{inj}} L^d_{inj} + \sum_{i' \neq i} \left(\frac{\partial W_{i'nj}}{\partial L^d_{inj}} L^d_{i'nj} \right) \right]$$

Note that:

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

and similarly,

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

The FOC can be rewritten into:

$$\left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta} (1 - s_{nj}) \right] P_{inj} = \left[1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}} (1 - e_{nj}) \right] \frac{W_{inj}}{A_{inj}}, \tag{A33}$$

where markup and markdown are relatively defined as:

$$\begin{aligned}
\mu_{inj} &\equiv \frac{P_{inj}}{MC_{inj}} = \left(1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta} \left(1 - s_{nj} \right) \right)^{-1} \\
\delta_{inj} &\equiv \frac{MRPL_{inj}}{W_{inj}} = \left(1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}} \left(1 - e_{nj} \right) \right).
\end{aligned} \tag{A34}$$

A.3 Solving the equilibrium

The firm's FOC (A33) 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 (A33), we derive:

$$P_{inj} = \frac{\left[1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}} (1 - e_{nj}) \right]}{\left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta} (1 - s_{nj}) \right]} \frac{W_{inj}}{A_{inj}}. \tag{A35}$$

Plug in the inverse labor supply function (A32):

$$P_{inj} = \frac{\left[1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj})\right]}{\left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right]} \frac{J^{\frac{1}{\hat{\theta}}} I^{\frac{1}{\hat{\eta}}} \left(\frac{L_{ikj}}{L_j}\right)^{\frac{1}{\hat{\eta}}} \left(\frac{L_j}{L}\right)^{\frac{1}{\hat{\theta}}} W}{A_{inj}}. \tag{A36}$$

Finally, using the CES property $e_{inj} = I^{\frac{1}{\hat{\eta}}} \left(\frac{L_{inj}}{L_j}\right)^{1+\frac{1}{\hat{\eta}}}$, we can write P_{inj} in terms of s_{inj}, e_{inj}, A_{inj}, and other market or economy-level variables:[24]

$$P_{inj} = \frac{\left[1 + \frac{1}{\hat{\theta}} e_{nj} + \frac{1}{\hat{\eta}}(1 - e_{nj})\right]}{\left[1 - \frac{1}{\theta} s_{nj} - \frac{1}{\eta}(1 - s_{nj})\right]} \frac{J^{\frac{1}{\hat{\theta}}} I^{\frac{1}{1+\hat{\eta}}} e_{inj}^{\frac{1}{1+\hat{\eta}}} \left(\frac{L_j}{L}\right)^{\frac{1}{\hat{\theta}}} W}{A_{inj}}. \tag{A37}$$

STEP 2 :MAPPING BETWEEN SALES AND WAGE BILL We begin by using the definition of wage bill share; in the third equation, we use the demand function:

$$\begin{aligned}
e_{inj} &= M^{\frac{1}{\hat{\eta}}} \left(\frac{L_{inj}}{L_j}\right)^{1+\frac{1}{\hat{\eta}}} \\
&= M^{\frac{1}{\hat{\eta}}} \left(\frac{Y_{inj}/A_{inj}}{\left(\sum_{i',n'} M^{\frac{1}{\hat{\eta}}} \left(Y_{i'n'j}/A_{i'n'j}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}\right)^{\frac{\hat{\eta}}{\hat{\eta}+1}}}\right)^{1+\frac{1}{\hat{\eta}}} \\
&= \left(\frac{\left(P_{inj}\right)^{-\eta}/A_{inj}}{\left(\sum_{i',n'} \left(\left(P_{i'n'j}\right)^{-\eta}/A_{i'n'j}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}\right)^{\frac{\hat{\eta}}{\hat{\eta}+1}}}\right)^{1+\frac{1}{\hat{\eta}}} \\
&= \left[\sum_{i',n'} \left(\frac{\left(P_{i'n'j}/P_{inj}\right)^{-\eta}}{A_{i'n'j}/A_{inj}}\right)^{\frac{\hat{\eta}+1}{\hat{\eta}}}\right]^{-1}.
\end{aligned} \tag{A38}$$

On the other hand, we have:

$$s_{inj} = \frac{1}{M}\left(\frac{P_{inj}}{P_j}\right)^{1-\eta} \quad \Leftrightarrow \quad \frac{P_{i'n'j}}{P_{inj}} = \left(\frac{s_{i'n'j}}{s_{inj}}\right)^{\frac{1}{1-\eta}}. \tag{A39}$$

[24] *Proof*: Using the inverse labor supply function, we have

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

Therefore, we successfully link two kinds of shares:

$$e_{inj} = \left[\sum_{i',n'} \left(\left(\frac{s_{i'n'j}}{s_{inj}} \right)^{\frac{\eta}{\eta-1}} \frac{A_{inj}}{A_{i'n'j}} \right)^{\frac{\hat{\eta}+1}{\hat{\eta}}} \right]^{-1}. \tag{A40}$$

STEP 3 : EQUATION IN SHARES We are able to solve the problem in sales shares by using equality:

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

where the second equality uses equation (A37). Therefore, we can solve s_{inj} from following equation system:

$$s_{inj} = \frac{\left[\frac{1+\frac{1}{\hat{\theta}}e_{nj}+\frac{1}{\hat{\eta}}(1-e_{nj})}{1-\frac{1}{\theta}s_{nj}-\frac{1}{\eta}(1-s_{nj})} \frac{e_{inj}^{\frac{1}{1+\hat{\eta}}}}{A_{inj}} \right]^{1-\eta}}{\sum_{i',n'} \left[\frac{1+\frac{1}{\hat{\theta}}e_{n'j}+\frac{1}{\hat{\eta}}(1-e_{n'j})}{1-\frac{1}{\theta}s_{n'j}-\frac{1}{\eta}(1-s_{n'j})} \frac{e_{i'n'j}^{\frac{1}{1+\hat{\eta}}}}{A_{i'n'j}} \right]^{1-\eta}}, \tag{A41}$$

where

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

STEP 4 :SOLVING FOR THE LEVELS IN THE ECONOMY

- The equilibrium system is:

$$\text{FOC:} \quad A_{inj}P_{inj} = \mu_{inj}\delta_{inj}W_{inj}$$

$$\text{Firm-level LS:} \quad W_{inj} = J^{\frac{1}{\hat{\theta}}} I^{\frac{1}{\hat{\eta}}} \left(\frac{L_{inj}}{L_j} \right)^{\frac{1}{\hat{\eta}}} \left(\frac{L_j}{L} \right)^{\frac{1}{\hat{\theta}}} W$$

$$\text{Aggregate LS:} \quad L = (\overline{\varphi}W)^{\varphi}$$

$$\text{Firm-level demand:} \quad Y_{inj} = \frac{1}{inj} \left(\frac{P_{inj}}{P_j} \right)^{-\eta} \left(\frac{P_j}{P} \right)^{-\theta} Y$$

$$\text{Firm-level inverse demand:} \quad P_{inj} = J^{-\frac{1}{\theta}} I^{-\frac{1}{\eta}} \left(\frac{Y_{inj}}{Y_j} \right)^{-\frac{1}{\eta}} \left(\frac{Y_j}{Y} \right)^{-\frac{1}{\theta}} P$$

- Besides, we have the relationship in share:

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

- Hence, we can write FOC as:

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

- 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:

$$Y = \left[\int_j \frac{1}{J} \left(\alpha_j\right)^{\frac{\theta-1}{\theta}} dj \right]^{\frac{\widehat{\theta}+\theta}{\theta-1}} \left(\frac{P}{W} \right)^{\widehat{\theta}} L \tag{A42}$$

- 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

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

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

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

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

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

$$\left(\frac{W}{P}\right)^{\widehat{\theta}} = I^{\frac{1}{\eta-1}} \left(\frac{1}{J}\right) X \tag{A43}$$

Finally, we normalize $P = 1$ and use the 3 aggregation equations for the goods market clearing (A42), labor market clearing (A43) 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:

$$e_{inj} = \frac{1}{I}\left(\frac{W_{inj}}{W_j}\right)^{(1+\hat{\eta})} = \left[\frac{W_{inj}}{\left(\sum_{i,n} W_{inj}^{1+\hat{\eta}}\right)^{\frac{1}{1+\hat{\eta}}}}\right]^{(1+\hat{\eta})} \tag{A44}$$

$$s_{inj} = \frac{1}{I}\left(\frac{P_{inj}}{P_j}\right)^{1-\eta} = \left[\frac{P_{inj}}{\left(\sum_{i,n} P_{inj}^{1-\eta}\right)^{\frac{1}{1-\eta}}}\right]^{(1-\eta)} \tag{A45}$$

$$e_{nj} = \frac{1}{I}\sum_i \left(\frac{W_{inj}}{W_j}\right)^{(1+\hat{\eta})} = \frac{\sum_i W_{inj}^{1+\hat{\eta}}}{\sum_{i,n} W_{inj}^{1+\hat{\eta}}} \tag{A46}$$

$$s_{nj} = \frac{1}{I}\sum_i \left(\frac{P_{inj}}{P_j}\right)^{1-\eta} = \frac{\sum_i P_{inj}^{1-\eta}}{\sum_{i,n} P_{inj}^{1-\eta}} \tag{A47}$$

Substituting these expressions into (12), we can now express the first order condition as:

$$\begin{aligned} P_{inj}\left[1 - \frac{1}{\theta}\left[\frac{\sum_i P_{inj}^{1-\eta}}{\sum_{i,n} P_{inj}^{1-\eta}}\right] - \frac{1}{\eta}\left(1 - \left[\frac{\sum_i P_{inj}^{1-\eta}}{\sum_{i,n} P_{inj}^{1-\eta}}\right]\right)\right] & \\ = \frac{W_{inj}}{A_{inj}}\left(1 + \frac{1}{\hat{\theta}}\left[\frac{\sum_i W_{inj}^{1+\hat{\eta}}}{\sum_{i,n} W_{inj}^{1+\hat{\eta}}}\right] + \frac{1}{\hat{\eta}}\left(1 - \left[\frac{\sum_i W_{inj}^{1+\hat{\eta}}}{\sum_{i,n} W_{inj}^{1+\hat{\eta}}}\right]\right)\right) & \end{aligned} \tag{A48}$$

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

$$\begin{aligned} P_{inj} &= \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} Y_{inj}^{-\frac{1}{\eta}} Y_j^{\frac{1}{\eta}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P \\ &= \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} Y_{inj}^{-\frac{1}{\eta}} \left[\left(\sum_{i,n} Y_{inj}^{\frac{\eta}{\eta-1}}\right)^{\frac{\eta-1}{\eta}}\right]^{\frac{1}{\eta}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P \\ &= \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} (A_{inj}L_{inj})^{-\frac{1}{\eta}} \left[\left(\sum_{i,n} (A_{inj}L_{inj})^{\frac{\eta}{\eta-1}}\right)^{\frac{\eta-1}{\eta}}\right]^{\frac{1}{\eta}-\frac{1}{\theta}} Y^{\frac{1}{\theta}} P \end{aligned} \tag{A49}$$

$$W_{inj} = \frac{1}{J}^{\frac{1}{-\hat{\theta}}} \frac{1}{I}^{\frac{1}{-\hat{\eta}}} L_{inj}^{\frac{1}{\hat{\eta}}} L_j^{\frac{1}{\hat{\theta}}-\frac{1}{\hat{\eta}}} L^{-\frac{1}{\hat{\theta}}} W \tag{A50}$$

Plugging equation (A49) and equation (A50) in (A48), gives us each firm's first order condition only in terms of A_{inj} and L_{inj}.

$$\begin{aligned} \frac{1}{J}^{\frac{1}{\theta}} \frac{1}{I}^{\frac{1}{\eta}} (A_{inj}L_{inj})^{-\frac{1}{\eta}} \left[\left(\frac{1}{I}^{\frac{1}{\eta}} \sum_{i,n} (A_{inj}L_{inj})^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}\frac{(\theta-\eta)}{\eta\theta}}\right] \left[1 - \frac{1}{\theta}\frac{\sum_i (A_{inj}L_{inj})^{\frac{\eta-1}{\eta}}}{\sum_{i,n}(A_{inj}L_{inj})^{\frac{\eta-1}{\hat{\eta}}}} - \frac{1}{\eta}\left(1 - \frac{\sum_i (A_{inj}L_{inj})^{\frac{\eta-1}{\eta}}}{\sum_{i,n}(A_{inj}L_{inj})^{\frac{\eta-1}{\eta}}}\right)\right] A & \\ = \frac{1}{J}^{\frac{-1}{\hat{\theta}}} \frac{1}{I}^{\frac{-1}{\hat{\eta}}} \frac{(L_{inj})^{\frac{1}{\hat{\eta}}}}{A_{inj}} \left[\left(\frac{1}{I}^{\frac{-1}{\hat{\eta}}} \sum_{i,k} (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}\right)^{\frac{\hat{\eta}}{\hat{\eta}+1}\frac{(\hat{\eta}-\hat{\theta})}{\hat{\eta}\hat{\theta}}}\right] \left[1 + \frac{1}{\hat{\theta}}\frac{\sum_i (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}}{\sum_{i,n}(L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}} + \frac{1}{\hat{\eta}}\left(1 - \frac{\sum_i (L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}}{\sum_{i,n}(L_{inj})^{\frac{\hat{\eta}+1}{\hat{\eta}}}}\right)\right] & \end{aligned} \tag{A51}$$

where $A = W^{-1}L^{1/\hat{\theta}}Y^{1/\theta}$ 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}).

A.5 Backing out productivity distribution in shares

In this section we show how to back out the productivity distribution using data on employment L_{inj} for any given N and elasticities.

1. Calculate wage bill shares, e_{inj}, by definition

$$e_{inj} = \frac{W_{inj}L_{inj}}{\sum_{i'n'}\left(W_{i'n'j}L_{i'n'j}\right)}$$

   which also gives us the markdown

2. In each market, guess sales shares s_{inj}

   (a) compute markup from shares

   (b) compute relative TFP, $\alpha_{inj} := A_{inj}/A_{11j}$, by

$$\alpha_{inj} = \left(\frac{s_{inj}}{s_{11j}}\right)^{\frac{\eta}{\eta-1}} \frac{L_{11j}}{L_{inj}}$$

   (c) **Method 1.** update new sales shares, s'_{inj}, by euqation zero

$$s'_{inj} = \frac{\left(\frac{\mu_{ij}\delta_{ij}W_{inj}}{\alpha_{inj}}\right)^{1-\eta}}{\sum_{i',n'}\left(\frac{\mu_{i'j}\delta_{i'j}W_{i'n'j}}{\alpha_{i'n'j}}\right)^{1-\eta}}$$

   (d) Solve the true sales shares by fixed point theorem.

3. Get economy-wide relative TFP, $\gamma_{inj} := A_{inj}/A_{111}$, by:

$$\gamma_{inj} = \left\{\frac{\mu_{ij}\delta_{ij}W_{inj}}{\mu_{11}\delta_{11}W_{111}} \frac{(L_{111})^{-\frac{1}{\eta}}\left[\left(\sum_{i',n'}\left[(\alpha_{i'n'1}L_{i'n'1})\right]^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}\right]^{\frac{1}{\eta}-\frac{1}{\theta}}}{(L_{inj})^{-\frac{1}{\eta}}\alpha_{inj}^{\frac{1}{\theta}-\frac{1}{\eta}}\left[\left(\sum_{i',n'}\left[\left(\alpha_{i'n'j}L_{i'n'j}\right)\right]^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}\right]^{\frac{1}{\eta}-\frac{1}{\theta}}}\right\}^{\frac{\theta}{\theta-1}}$$

4. Finally, we can pin down the level by:

$$A_{inj} = \frac{\gamma_{inj}}{P}\left\{\sum_j \frac{1}{J}\left[\sum_{i,n}\frac{1}{I}\left(\frac{\mu_{ij}\delta_{ij}W_{inj}}{\gamma_{inj}}\right)^{1-\eta}\right]^{\frac{1-\theta}{1-\eta}}\right\}^{\frac{1}{1-\theta}}$$

A.6 Labor Market Elasticity Estimation

Results for Tradeables and Non-Tradeables. In Table (A1) and Table (A2), we replicate the same exercise as before except that we estimate the model on both the tradeables and the non-tradeables sector. 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.

Table A1: **Estimates of reduced-form parameters: Tradeables and Non-Tradeables**

A. OLS and Second-Stage IV Estimates					
	OLS	IV		OLS	IV
	(1)	(2)		(3)	(4)
$\frac{1}{\hat{\eta}}$	-0.254	0.305	$\frac{1}{\hat{\theta}} - \frac{1}{\hat{\eta}}$	0.080	0.012
	(1.3e-4)	(0.018)		(4.36e-5)	(2.8e-4)
Sector x Year FE	Yes	Yes	Sector FE	Yes	Yes
Establishment FE	Yes	Yes	Year FE	Yes	Yes
B. First-Stage Regressions for the IV					
$\tau_{X(i)t}$	-	-0.004	$\bar{\tau}_{jt}$	-	-0.237
		(6.32e-5)			(2.2e-4)
Sector x Year FE	-	Yes	Sector FE	-	Yes
Establishment FE	-	Yes	Year FE	-	Yes
No. of obs.	47,030,000	47,030,000	No. of obs.	47,030,000	47,030,000

Notes: Standard errors are reported in parenthesis. Estimates of γ in columns 3 and 4 are conditional on the estimates of columns 1 and 2, respectively. We define a market as all the establishments in the tradeable and the non-tradeable sectors within NAICS-6. Number of observations are common for both the first and the second-stage.

Table A2: **Estimates of structural parameters: Tradeables and Non-Tradeables**

Parameter	Description	IV Estimate
$\hat{\eta}$	Within-market elasticity	3.28
$\hat{\theta}$	Between-market elasticity	3.16

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 $\hat{\eta} > \hat{\theta} > 0$. However, we do find that the gap between $\hat{\eta}$ and $\hat{\theta}$ is lower as compared to the sample where we account only for the tradeables.