Trade Liberalization, Labor Market Power, and Misallocation across Firms: Quantitative Evidence from China's WTO Accession [*]

Enze Xie
Peking University

Mingzhi Xu
Peking University

Miaojie Yu
Liaoning University

November 19, 2022

Abstract

We study the effect of heterogeneous monopsony power across Chinese firms on the equilibrium allocation of labor. The analysis finds that heterogeneity in monopsony power results from the heterogeneity in workers' preferences for the jobs provided by an individual firm. We build a model to derive that the variance in logarithmic markdown serves as a *sufficient statistic* that can be used to infer the impact of heterogeneous monopsony power on overall production efficiency at the national level. We further quantify the welfare effect associated with labor reallocation after China joined the World Trade Organization in 2001. The empirical results suggest that lower input tariffs decrease the level of misallocation in the labor market, which is captured by the decreased variance in logarithmic markdown. In contrast, a reduction in output tariffs has no significant effect. Overall, the improved allocation of labor accounts for 5% of the total production efficiency gains from China's trade liberalization.

Keywords: heterogeneous monopsony power, misallocation, trade liberalization, markdown distribution

JEL Codes: F12, F14, F16, J42

[*]Miaojie Yu is thankful for financial support from the National Social Science Foundation of China (Major Program, Grant No. 20ZDA050), National Fund for Distinguished Young Scholars (Grant No. 71625007) and Beijing Outstanding Young Scientist Program. Mingzhi Xu is thankful for financial support from the Peking University Initiative Scientific Research Program (Grant No. 7101302576) and Peking University International Strategic Cooperation Research Program (Grant No. 7101702318). Enze Xie, National School of Development at Peking University, Beijing, China, email: ezxie2019@nsd.pku.edu.cn; Mingzhi Xu, Institute of New Structural Economics at Peking University, Beijing, China, email: mingzhixu@nsd.pku.edu.cn; Miaojie Yu, Liaoning University, Shenyang, China, email: mjyu@lnu.edu.cn

1 Introduction

Countrywide aggregate total factor productivity (TFP) is determined not only by the average level of firm TFP, but also by whether the factors of production are efficiently distributed across firms. An inefficient case of the latter has been referred to as factor misallocation, which is an important source of welfare loss that has captured much attention in the academic arena since the work by Hsieh & Klenow (2009). Previous studies have primarily focused on markups and misallocation, whereby only the dispersion in markups matters for aggregate output (Edmond et al., 2015; Peters, 2020). In contrast, the misallocation of inputs, such as labor, has been relatively understudied. In this paper, we aim to fill the gap by studying the effect of heterogeneous monopsony power in the labor market across Chinese firms on the equilibrium misallocation of labor, and quantifying the welfare effect associated with labor reallocation after China joined the World Trade Organization (WTO) in 2001.

In the analysis, heterogeneous monopsony power in the labor market across firms results from the fact that workers' preferences among the occupations and jobs provided by an individual firm are heterogeneous, and such heterogeneity has been found to account for a sizable portion of firm size heterogeneity (Haltiwanger et al. (1999; 2007). To demonstrate how heterogeneous monopsony power at the firm level leads to welfare consequences at the aggregate level, we offer a model with monopolistic competition in the product markets and monopsonistic competition in the labor markets. The key model feature is that workers' preference for a job is governed by a generalized extreme value distribution (GEV) that has three layers, inspired by Fajgelbaum et al. (2011) and Card et al. (2018). Most importantly, we show that given the estimated parameters, the first and second moments of the industry distribution of monopsony power are sufficient statistics of the effect of imperfect competition in labor markets on the aggregate economy.

At the micro level, the theoretical results show that the firm-level monopsony power leads firms to produce less output, charge a higher price, use less input, substitute non-labor input for labor, and increses the marginal revenue product of labor and revenue-based TFP (TFPR). These micro evidence is consistent with the previous finding of Hsieh & Klenow (2009) and Restuccia & Rogerson (2008). Notably, we also find that an improvement in Hicks-neutral TFP can result in biased technological change towards intermediate inputs. The intuition lies in the fact that firms' monopsony power gives rise to the increasing marginal labor costs, which makes intermediate inputs relatively preferable when firms expand due to a productivity improvement. At the macro level, we find that heterogeneity in firm monopsony power leads to the misallocation of the factors of production and causes aggregate efficiency loss. These findings are similar to those of Hsieh & Klenow (2009), who focused on distortions and misallocation. Our model derives that the variance in logarithmic markdown serves as a *sufficient statistic*, which we use to infer the overall efficiency of production at the national level.

Next, we quantify the welfare effect associated with labor reallocation after China joined the WTO in 2001. Our empirical specification is a difference-in-difference (DID) regression, where we explore the impact of trade liberalization on the variance in logarithmic markdown. Markdowns are estimated following Brooks et al. (2021b), who define markdown as the ratio of the marginal revenue product of labor to the wage. Following Amiti & Konings (2007), we use industry-level input and output tariffs as proxies for trade liberalization. The empirical evidence shows that lower input tariffs reduce the variance in logarithmic markdown on average, and the reductions in firm markdowns are larger among firms with higher monopsony power. In contrast, reductions in output tariffs have no significant effect. The results suggest that input trade liberalization helps to mitigate misallocation in the labor market and improve aggregate production efficiency. Our empirical results are robust to alternative measures of markdown and different specifications.

The quantitative results show that the total production efficiency gain from input trade liberalization is 6%, of which 5.7% is from the direct impact of reductions in input tariffs on firms' production efficiency, and 0.3% originates from its impact on the variance in the logarithmic markdown. Our paper has important implications. First, we show that heterogeneous monopsony power across firms gives rise to misallocation and hurts economic efficiency. Second, trade liberalization, especially in intermediate inputs, significantly mitigates the misallocation of allocating labor across firms and improves welfare. To this end, domestic policies, such as those that protect labor rights, would be complementary to trade policies.

This paper contributes to the literature in mainly three aspects. First, our work complements the literature on misallocation by studying an additional source of misallocation, which is heterogeneous monopsony power in the labor market. Previous work mainly analyzed misallocation in the capital market and the product market (Edmond et al., 2021; Epifani & Gancia, 2011; Hsu et al., 2020; Liu & Ma, 2021; Lu & Yu, 2015; Midrigan & Xu, 2014; Peters, 2020). The labor market has received considerably less attention. In the spirit of Hsieh & Klenow (2009), our model shows that firm heterogeneity in labor market power generates differentiation in the marginal revenue product of labor, leads to misallocation across firms, and results in an efficiency loss.[1]

Second, our work contributes to the literature on detecting the impact of trade liberalization on labor market power.[2] With the revival of interest in imperfect competition in the labor market, dozens of papers have focused on the interaction between trade and imperfectly competitive labor markets.[3] Ahsan & Mitra (2014); Dobbelaere & Wiersma (2020); Kondo et al. (2021); MacKenzie (2021) and Macedoni & Tyazhelnikov (2019) investigate the impact of trade liberalization's impact on firms' monopsony power. Several other papers focus on the impact of import competition (Caselli et al., 2021; Mertens, 2020), international status (Dobbelaere & Kiyota, 2018) and foreign direct investment (FDI) (Lu et al., 2019) on labor market power,

[1]Efficient allocation of production factors across firms can be achieved only if the marginal products of different firms are equivalent in a static model of production and demand. Any deviation from this setting incurs misallocation (Hsieh & Klenow, 2009; Restuccia & Rogerson, 2008).

[2]Previous work mainly concentrates on studying trade liberalization's effect on product market power, i.e., markup (Brandt et al., 2017; De Loecker et al., 2016; Edmond et al., 2015; Fan et al., 2018; Levinsohn, 1993; Liu & Ma, 2021).

[3]The persistent fall of the labor share across countries (Dorn et al., 2017; Karabarbounis & Neiman, 2014), the stagnation of wage growth (Gould, 2014), and high inequality within a country (Card et al., 2013) have made imperfect competition in the labor market to be the focus of a burgeoning line of research over the past two decades (Boal & Ransom, 1997; Manning, 2003a;b).

respectively. However, the predominant focus of the previous literature is the determinants of labor market power, while the role of heterogeneity in firms' labor market power has received less attention. Our work demonstrates that input trade liberalization relieves the misallocation, serving as an additional channel for gains from trade.

Third, our work provides a model that characterizes firms' heterogeneous monopsony power in the labor market with good tractability. As summarized by Manning (2021), in the labor economics context, scholars often impose imperfect competition on the labor market by taking search frictions (Burdett & Mortensen, 1998; Wu, 2020) or preference heterogeneity (or job idiosyncrasy) (Card et al., 2018; Manning, 2003b) into account. In the international trade literature, in particular, there are also two types of models to characterize labor market power. One is the oligopsonistic model (Macedoni & Tyazhelnikov, 2019; MacKenzie, 2021), and the other is the monopsonistic model (Egger et al., 2021; Jha & Rodriguez-Lopez, 2021; Macedoni, 2021). Within the monopsonistic model, the small-firm setting implies a constant monopsony power that does not vary across firms (Jha & Rodriguez-Lopez, 2021), while the big-firm setting implies the opposite (Brooks et al., 2021b).[4] Apparently, the oligopsonistic model is also a big-firm setting, so it induces oligopsony power that is positively correlated with firm size or relative size but at the expense of losing tractability. Inspired by Fajgelbaum et al. (2011) and Card et al. (2018), we build a three-tier nested demand structure following the small firm setting of the monopsonistic model to characterize firm-level variable monopsony power with good tractability.

The remainder of the paper proceeds as follows: Section 2 presents the model. Section 3 offers an overview of the data, the background of China's WTO accession, and the estimation strategy for the key variables in our regressions. Section 4 displays the empirical specification and results. Section 5 quantifies the welfare impact associated with labor reallocation after China's WTO accession. Section 6 concludes.

2 Theoretical Model

In this section, we build a model to introduce heterogeneous monopsony power and characterize its impact on firms' behavior and the aggregate economy. We begin with firm-level results and then move to aggregate-level implications. The economy is populated by L units of consumers-workers; each consumer is endowed with one unit of labor, and the supply of labor is inelastic.[5] Here we consider an economy in autarky.

2.1 Consumer Preferences and Labor Supply

We use i to denote consumer, where the consumer has heterogeneous preferences across different job offers (horizontal job differentiation). The indirect utility function of consumer i working in manufacturing

[4] Macedoni (2021) and Berger et al. (2019) introduce variable firm-level labor market power through the substitution between consumption and leisure, with the former in the monopsonistic labor market and the latter in the oligopsonistic labor market.

[5] Throughout the paper, we use the terms consumer and worker as synonyms.

industry s, firm j and occupation o is:[6].

$$u_{isjo} = \ln(w_{sjo}) + \epsilon_{io} \tag{1}$$

where w_{sjo} denotes wages. The idiosyncratic additive component of utility, ϵ_{io}, captures the other attributes of the job (such as workers' relationship with the boss, working environment, cost of commuting cost, and so forth), which consumers evaluate differently. We take the ϵ term to be distributed independently across the population of consumers according to a GEV distribution, which we denote by G(ϵ):

$$G_{\epsilon}(\boldsymbol{\epsilon}) = \exp\left\{ - \left(\sum_{s\in S} \left[\sum_{j\in M_s} \left(\sum_{o\in M_j} \exp\left(-\epsilon_o/\theta_j\right) \right)^{\theta_j/\theta_s} \right]^{\theta_s/\theta} \right)^{\theta} \right\} \tag{2}$$

with $\theta_s \in (0,1)$ for all $s \in S$ and $\theta_j \in (0,1)$ for all $j \in M_s$, where S indexes the finite set of industries, and M_s and M_j denotes the total number of firms in industry s and the total number of occupations in firm j, respectively. Consumers choose the industry, firm, and occupation in which they will work, in sequence, to obtain the highest utility. Because ϵ is distributed according to a GEV distribution, the probability of consumer i working in occupation o, firm j, industry s is given as follows:

$$Pr(\underset{s\in S, j\in M_s, o\in M_j}{\arg\max} \; u_{isjo} = s,j,o) = \eta_{o|j}\cdot\eta_{j|s}\cdot\eta_s, \;\forall o\in M_j, j\in M_s, s\in S \tag{3}$$

where

$$\eta_{o|j} = \frac{e^{ln(w_{sjo})/\theta_j}}{\sum_{k\in M_j} e^{ln(w_{sjk})/\theta_j}}, \;\forall o \in M_j \tag{4}$$

is the probability of consumer i choosing to work in occupation o conditional on working in firm j.

$$\eta_{j|s} = \frac{\left[\sum_{k\in M_j} e^{ln(w_{sjk})/\theta_j}\right]^{\theta_j/\theta_s}}{\sum_{\omega\in M_s}\left[\sum_{k\in M_\omega} e^{ln(w_{s\omega k})/\theta_\omega}\right]^{\theta_\omega/\theta_s}}, \;\forall j\in M_s \tag{5}$$

is the probability of consumer i choosing to work in firm j conditional on working in industry s, and

$$\eta_s = \frac{\left\{\sum_{\omega\in M_s}\left[\sum_{k\in M_\omega} e^{ln(w_{s\omega k})/\theta_\omega}\right]^{\theta_\omega/\theta_s}\right\}^{\theta_s/\theta}}{\sum_{g\in S}\left\{\sum_{\omega\in M_g}\left[\sum_{k\in M_\omega} e^{ln(w_{s\omega k})/\theta_\omega}\right]^{\theta_\omega/\theta_g}\right\}^{\theta_g/\theta}}, \;\forall s\in S \tag{6}$$

is the probability of consumer i choosing to work in industry s. As a result, the probability of a consumer

[6]The indirect utility function is a simplified version of Card et al. (2018). We assume away firms' heterogeneous amenity and reservation wage, to obtain analytical results, as noted by Jha & Rodriguez-Lopez (2021)

i choosing to work in occupation o, firm j, and industry s is:

$$Pr(\underset{s\in S, j\in M_s, o\in M_j}{\arg\max} u_{isjo} = s, j, o) = \lambda_{sj} e^{ln(w_{sjo})/\theta_j} \tag{7}$$

where $\lambda_{sj} = \eta_s \cdot \eta_{j|s} \cdot \frac{1}{\sum_{k\in M_j} e^{ln(w_{sjk})/\theta_j}}$, λ_{sj} is common to all occupations within industry s and firm j. We can derive the labor supply curve for each occupation by multiplying the probability above by total labor supply, that is

$$l_{sjo} = (L\lambda_{sj}) e^{ln(w_{sjo})/\theta_j} \tag{8}$$

We have already introduced monopsony power into our model. Equation (8) shows that for each occupation, firms face an upward-sloping labor supply curve. Following Manning (2003a), we use markdown ψ to measure firms' monopsony power, which is defined as the ratio of the marginal revenue product of labor to the wage. Using the labor supply function in equation (8), the occupation-level markdown is

$$\psi_{sjo} = 1 + \frac{\partial w_{sjo}}{\partial l_{sjo}} \frac{l_{sjo}}{w_{sjo}} = 1 + \theta_j = \psi_{sj} \tag{9}$$

We obtain the firm-level variable markdown from equation (9). As noted by Fajgelbaum et al. (2011), θ_j is known as the dissimilarity parameter, and it measures the degree of heterogeneity in workers' preference over different occupations provided by individual firm j.[7] As a result, equation (9) shows that greater heterogeneity in preferences for different occupations within a firm gives the firm greater monopsony power. The intuition is as follows.

First, from the perspective of the worker, the perceived heterogeneity in the nonpecuniary rewards of different occupations within the firm increases with θ_j. As a result, workers are more likely to apply for the occupation with the highest value of the idiosyncratic term ϵ_o, which in turn gives the firm greater monopsony power over each occupation. On average, firms' monopsony power increases with the degree of heterogeneity of workers' preferences.[8]

Second, from the viewpoint of firms, Booth et al. (2000) emphasize that firms play an important role in fostering collaboration among colleagues to draw on each other's strengths and increase efficiency. Worker heterogeneity affects the range of tasks that firms conduct. Firms with more heterogeneous occupations

[7]Equation (8) also shows that θ_j is the inverse elasticity of labor supply. Hence, higher θ_j implies more inelastic labor supply and larger monopsony power.

[8]As aforementioned, there are two main sources of monopsony power, preference heterogeneity and search frictions. From the perspective of search frictions, a greater value of θ_j could be viewed as larger directed job search costs, which gives firms more monopsony power (Wu, 2020). Moreover, once the worker has already been allocated to a specific occupation, a greater value of θ_j implies that the switching costs of changing jobs within the firm are higher as well, leading to more monopsony power (Ransom, 2021). Alternatively, a greater value of θ_j could be interpreted as more occupation-specific human capital, as Bachmann et al. (2022) point out that job-specific human capital gives rise to firm monopsony power because job change leads to a loss of human capital. In addition, a greater value of θ_j could be associated with fewer outside options due to the larger dissimilarity between the workers current occupation and alternative occupations, which also increases the firm's monopsony power (Schubert et al., 2019).

conduct more complex tasks and avoid fierce competition, as they have greated monopsony power over their workers.[9] Fox (2010) stresses that firms should only hire the workers who have a great non-pecuniary desire to work and then compress the wage below their marginal revenue product of labor to achieve higher profits. In our scenario, the idea is the same. Workers' idiosyncratic preferences are unobservable for firms, but when workers choose a specific occupation, their preferences are revealed. Hence, firms have incentives to expand the scope of occupations to take advantage of heterogeneous workers and compress wages to the largest extent.[10]

2.2 Production

The final product Q is produced by a representative firm in competitive final goods markets. This firm combines the output Q_s of S manufacturing industries using Cobb-Douglas production technology:

$$Q = \prod_{s=1}^{S} \mathrm{Q}_s^{\alpha_s}, \text{where } \alpha_s \in (0,1),\ \forall s \in S, \sum_{s=1}^{S} \alpha_s = 1 \tag{10}$$

Cost minimization implies $\mathrm{P_s Q_s} = \alpha_s \mathrm{PQ}$ ($\forall s \in S$). We make the final product Q the numeraire, and thus $P \equiv 1$. Furthermore, within each industry s, there exists a number of firms, denoted by M_s.[11] Each firm produces a variety, and varieties are combined together to yield the industry-level output, according to a constant elasticity of substitution aggregation:

$$\mathrm{Q_s} = \Big(\sum_{j \in M_s} q_{sj}^{\rho_s}\Big)^{\frac{1}{\rho_s}}, \rho_s \in (0,1) \tag{11}$$

where ρ_s governs the elasticity of substitution between different varieties, and we allow it to vary across industries. The cost minimization of output production implies that:

$$q_{sj} = \left(\frac{p_{sj}}{\mathrm{P_s}}\right)^{-\frac{1}{1-\rho_s}} \mathrm{Q_s} \tag{12}$$

where $\mathrm{P_s}$ is the industry level price index, i.e., $\mathrm{P_s} = \left(\sum_{j \in M_s} p_{sj}^{-\frac{\rho_s}{1-\rho_s}}\right)^{-\frac{1-\rho_s}{\rho_s}}$. Finally, we assume a Cobb-Douglas production function for each individual firm within the industry:

$$q_{sj} = \varphi_{sj} m_{sj}^{\beta_s} l_{sj}^{1-\beta_s}, \text{where } \beta_s \in (0,1) \tag{13}$$

where φ_{sj} is the Hicks-neutral TFP. l_{sj} indexes the composite labor, and m_{sj} denotes other production

[9] Firms differ in employee composition even within narrowly defined industries (Haltiwanger et al., 1999; 2007).

[10] There is a similar story in the product market. Macedoni et al. (2020) show that in a model with heterogeneous consumers, consumer heterogeneity increases markups and makes markups differ across products. Our model has heterogeneous workers, whose preferences leads to markdowns that differ across firms.

[11] We shut down firm dynamic here, and hence M_s is exogenously given.

factors, such as intermediate inputs and capital. For the convenience, we write m_{sj} for the intermediate input. We assume that m_{sj} comprises the full sets of final products combined according to a Cobb-Douglas aggregation such as equation (10). As a result, the overall price index, P, becomes the price w^m of the intermediate input m_{sj}, that is, $w^m = P \equiv 1$. Firms have different occupations, so l_{sj} is composed of different types of labor, which, combined according to Cobb-Douglas aggregation by assumption, yields

$$l_{sj} = \prod_{o=1}^{M_j} (l_{sjo})^{\gamma_o}, \text{where } \gamma_o \in (0,1),\ \forall o \in M_j, \sum_{o=1}^{M_j} \gamma_o = 1 \tag{14}$$

The labor structure implies that firm-level markdown is the Cobb-Douglas aggregation of occupation-level markdown, which is in line with equation (9):

$$\psi_{sj} = \prod_{o=1}^{M_j} (\psi_{sjo})^{\gamma_o} = \prod_{o=1}^{M_j} (1+\theta_j)^{\gamma_o} = (1+\theta_j)^{\sum_{o=1}^{M_j}\gamma_o} = 1+\theta_j = \psi_{sjo} \tag{15}$$

2.3 Firm-Level Equilibrium

The optimization problem of each firm is to choose the quantities of the input and the output to maximize profits, subject to three constraints: labor supply (equation 8), output demand (equation 12), and production function (equation 13). Formally, firms' profit maximization problem is defined as follows:

$$\max_{q_{sj}, m_{sj}, l_{sjo}\ \forall o \in M_j} p_{sj}(q_{sj})q_{sj} - \left[\sum_{o=1}^{M_j} w_{sjo}(l_{sjo})l_{sjo}\right] - w^m m_{sj} \tag{16}$$

The Cobb-Douglas production function of an individual firm implies that the price of output can be expressed as the product of the markup and marginal cost. Thus, we have the following:

$$p_{sj} = \frac{1}{\rho_s}\left\{\frac{1}{\varphi_{sj}}\left[\frac{\prod_{o=1}^{M_j}\left(\frac{w_{sjo}\psi_{sjo}}{\gamma_o}\right)^{\gamma_o}}{1-\beta_s}\right]^{1-\beta_s}\left(\frac{w^m}{\beta_s}\right)^{\beta_s}\right\} \tag{17}$$

Solving firms' profit maximization problem, l_{sj} can be expressed as a function of productivity φ_{sj} and firm-level markdown ψ_{sj}, which is the following:[12]

$$l_{sj} = \kappa_{sj}\psi_{sj}^{\frac{1-\beta_s\rho_s}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}}\varphi_{sj}^{\frac{-\rho_s}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} \tag{18}$$

where $\kappa_{sj} = \left\{\left(\frac{1}{P_s\rho_s}\right)\left(\frac{w^m}{\beta_s}\right)^{\beta_s\rho_s}\left[\frac{1}{(1-\beta_s)\gamma_o}\left(\frac{1}{L\lambda_{sj}}\right)^{\theta_j}\right]^{1-\beta_s\rho_s} Q_s^{\rho_s-1}\right\}^{\frac{1}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} > 0$. With $\beta_s \in (0,1)$, $\rho_s \in (0,1)$ and $\theta_j \in (0,1)$; thus, we can easily obtain $\frac{1-\beta_s\rho_s}{\rho_s-1+\theta_j(\beta_s\rho_s-1)} < 0$ and $\frac{-\rho_s}{\rho_s-1+\theta_j(\beta_s\rho_s-1)} > 0$. Hence, we have $\frac{\partial l_{sj}}{\partial\psi_{sj}} < 0$ and $\frac{\partial l_{sj}}{\partial\varphi_{sj}} > 0$. Given productivity, firms with more monopsony power employ fewer

[12] Appendix B shows the details of the derivation of the firm profit maximization problem.

workers; and given monopsony power, more productive firms employ more workers. Similarly, we can also obtain the expression for m_{sj} as a function of productivity and monopsony power, which is the following:

$$m_{sj} = \chi_{sj}\kappa_{sj}\psi_{sj}^{\frac{\rho_s(1-\beta_s)}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}}\varphi_{sj}^{\frac{-\rho_s(\theta_j+1)}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} \tag{19}$$

where $\frac{\partial m_{sj}}{\partial \psi_{sj}} < 0$ and $\frac{\partial m_{sj}}{\partial \varphi_{sj}} > 0$. Given productivity, firms with monopsony power use less intermediate input. Given monopsony power, firms with higher productivity use more intermediate input.

Proposition 1. *Firms with monopsony power in the labor market have an unproportionally higher intermediate input-labor ratio. The improvement in Hicks-neutral TFP results in an intermediate input-biased production technology change.*

Next, we consider the factor ratio of the firm. We can derive the intermediate input-labor ratio as a function of productivity and monopsony power, that is:

$$\frac{m_{sj}}{l_{sj}} = \chi_{sj}\psi_{sj}^{\frac{\rho_s-1}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}}\varphi_{sj}^{\frac{-\rho_s\theta_j}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} \tag{20}$$

where $\chi_{sj} = \left(\frac{1}{1-\beta_s}\right)\left(\frac{w^m\gamma_o}{\beta_s}\right)^{-1}\left(\frac{1}{L\lambda_{sj}}\right)^{\theta_j}\kappa_{sj}^{\theta_j} > 0$, $\frac{\partial(m_{sj}/l_{sj})}{\partial\psi_{sj}} > 0$ and $\frac{\partial(m_{sj}/l_{sj})}{\partial\varphi_{sj}} > 0$. Given productivity, firms with greater monopsony power use more of the intermediate input relative to labor. Notably, given monopsony power, the improvement in Hicks-neutral productivity induces an intermediate input-biased production technology change. The intuition is that monopsonistic competition implies increasing marginal costs of labor. When productivity increases, the firm expands, the relative price of labor with respect to the price of the intermediate input increases, which makes the latter preferable.

Here, we investigate the impact of monopsony power on the firm's output. Similarly, the output can also be expressed as a function of productivity and monopsony power:

$$q_{sj} = \Delta_{sj}\psi_{sj}^{\frac{1-\beta_s}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}}\varphi_{sj}^{-\frac{1+\theta_j}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} \tag{21}$$

where $\Delta_{sj} = \left[\left(\frac{1}{1-\beta_s}\right)\left(\frac{w^m\gamma_o}{\beta_s}\right)^{-1}\left(\frac{1}{L\lambda_{sj}}\right)^{\theta_j}\right]^{\beta_s}\kappa_{sj}^{\beta_s\theta_j+1} > 0$, $\frac{\partial q_{sj}}{\partial\psi_{sj}} < 0$ and $\frac{\partial q_{sj}}{\partial\varphi_{sj}} > 0$. Equation (21) implies that, given productivity, firms with monopsony power produce less and charge higher prices. Given monopsony power, a more productive firm produces more.

Proposition 2. *Firms' monopsony power in the labor market increases the marginal revenue product of labor and incurs a higher TFPR in comparison with a competitive labor market.*

The first order conditions of the firm's profit maximization problem can be rearranged to obtain the following expression in terms of the marginal revenue product of the factors of production:

$$\text{MRPL}_{sj} = (1-\beta_s)\rho_s\frac{p_{sj}q_{sj}}{l_{sj}} = w_{sj}\psi_{sj} = \Lambda_{sj}\psi_{sj}^{\frac{\rho_s-1}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}}\varphi_{sj}^{-\frac{\rho_s\theta_j}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} \tag{22}$$

$$\text{MRPM}_{sj} = \beta_s \rho_s \frac{p_{sj} q_{sj}}{m_{sj}} = w^m \tag{23}$$

where $\Lambda_{sj} = \frac{1}{\gamma_o}\left(\frac{1}{L\lambda_{sj}}\right)^{\theta_j} {\kappa_{sj}}^{\theta_j} > 0$. By definition, $\text{TFPR}_{sj} = p_{sj}$, $\text{TFPQ}_{sj} = p_{sj}\varphi_{sj}$.[13] Plugging equations (13), (22) and (23) into the expression for TFPR_{sj}, we have the following:

$$\text{TFPR}_{sj} \propto {\psi_{sj}}^{\frac{(\rho_s-1)(1-\beta_s)}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} {\varphi_{sj}}^{-\frac{\rho_s\theta_j(1-\beta_s)}{\rho_s-1+\theta_j(\beta_s\rho_s-1)}} \tag{24}$$

Equation (24) implies that, when productivity or monopsony power is held fixed, change in the other one — higher productivity or larger monopsony power — results in a higher TFPR_{sj}.[14] To sum up, the firm-level analysis shows that markdown ψ_{sj} serves as a *sufficient statistics* of the effect of labor market power on firm-level variables. Specifically, monopsony power leads a firm to use less of the input, produce less output, and charge higher prices. More interestingly, an increase in monopsony power and Hicks-neutral productivity results in intermediate input-biased production technology change. In a competitive labor market, resource allocation is uniquely determined by productivity. In contrast, in a monopsonistic labor market, resource allocation is distorted by heterogeneous monopsony power, and misallocation shows up.

2.4 Misallocation and Efficiency Loss

First, we can express aggregate output as a function of l_s, m_s, and industry-level TFP:

$$Q = \prod_{s=1}^{S} {Q_s}^{\alpha_s} = \prod_{s=1}^{S} \left(\text{TFP}_s {m_s}^{\beta_s} {l_s}^{1-\beta_s}\right)^{\alpha_s} \tag{25}$$

As a result, we can express industry-level TFP as follows:

$$\text{TFP}_s = \frac{Q_s}{{m_s}^{\beta_s} {l_s}^{1-\beta_s}} \tag{26}$$

Following the derivation of Hsieh & Klenow (2009), industry-level TFP is given by:[15]

$$\text{TFP}_s = \left[\sum_{j\in M_s} {\varphi_{sj}}^{\frac{\rho_s}{1-\rho_s}} \left(\frac{\overline{\text{MRPL}}_s}{\text{MRPL}_{sj}}\right)^{\frac{(1-\beta_s)\rho_s}{1-\rho_s}}\right]^{\frac{1-\rho_s}{\rho_s}} \tag{27}$$

[13] TFPQ and TFPR refer to quantity-based and revenue-based total factor productivity respectively (Foster et al., 2008).

[14] If we assume away labor market power, i.e. $\forall s \in S, \forall j \in M_s$, $\theta_j = 0$, $\psi_{sj} = 1$. The expression of TFRP_{sj} becomes $\text{TFPR}_{sj} = \frac{1}{\rho_s}(\frac{w^m}{\beta_s})^{\beta_s}[\frac{1}{(1-\beta_s)\gamma_o}]^{1-\beta_s} = \text{TFPR}_s$, which means there is no TFPR dispersion within industries, which is similar to Hsieh & Klenow (2009). In Hsieh & Klenow (2009), the distortion is exogenously given, and firms are price takers in the input market. In our model, the distortion is endogenously derived from workers' heterogeneous preferences and firms have wage-setting power in the labor market.

[15] Since the industrial outputs combine with each other according to a Cobb-Douglas aggregation, $\log \text{TFP}$ is the weighted average of the $\log \text{TFP}_s$, i.e., $\log \text{TFP} = \sum_{s=1}^{S} \alpha_s \log \text{TFP}_s$. Hence, proposition 3 applies to national-level productivity as well.

Equation (27) reveals that industry-level TFP is homogeneous of degree zero in monopsony power. The average level of monopsony power has no impact on industry-level TFP.

Proposition 3. *Firms' heterogeneous monopsony power in the labor market introduces an intra-industry misallocation. It results in a TFP loss, where firms with below-average monopsony power (less distorted) overproduce and firms with above-average monopsony power (more distorted) underproduce. The variance in logarithmic markdown serves as a sufficient statistic of the negative impact of the heterogeneity of firms' monopsony power on total productivity.*

Following Hsieh & Klenow (2009), we assume that φ_{sj}, ψ_{sj}, and w_{sj} are jointly log-normally distributed. The following is a simple closed-form expression for industry-level aggregate TFP:[16]

$$\log \mathrm{TFP}_s = \log\left(\sum_{j\in M_s} {\varphi_{sj}}^{\frac{\rho_s}{1-\rho_s}}\right)^{\frac{1-\rho_s}{\rho_s}} - \Gamma_{1s} \mathrm{var} \log \psi_{sj} \tag{28}$$

where:

$$\Gamma_{1s} = \frac{(\beta_s\rho_s - 1)(\beta_s - 1)}{2(1-\rho_s)} > 0$$

The negative effect of monopsony power on industry-level TFP can be summarized as the variance in $\log \psi_{sj}$. In short, the heterogeneity of firms' markdowns incurs an efficiency loss. To verify this more clearly, we consider the ratio of labor allocation between two firms, j and k:

$$\frac{l_{sj}}{l_{sk}} = \left(\frac{\varphi_{sj}}{\varphi_{sk}}\right)^{\frac{-\rho_s}{\rho_s - 1 + \theta_j(\beta_s\rho_s - 1)}} \left(\frac{\psi_{sj}}{\psi_{sk}}\right)^{\frac{1-\beta_s\rho_s}{\rho_s - 1 + \theta_j(\beta_s\rho_s - 1)}} \tag{29}$$

When $\psi_{sj} = \psi_{sk} = \psi_s$, there is no dispersion in firm-level markdown within the same industry. Then the allocation of labor across different firms is completely determined by productivity. By contrast, when we introduce heterogeneous monopsony power, labor is inefficiently reallocated from the more to the less distorted firm: given productivity, φ, $l_{sj} > l_{sk} \iff \psi_{sj} < \psi_{sk}$. Clearly, the allocation of the factors of production is not only determined by productivity, but also distorted by firms' monopsony power, which gives rise to misallocation and loss of production efficiency.

3 Data, Background and Measurement of Key Variables

3.1 Data and Processing

To quantify the welfare impact associated with labor reallocation after China joined the WTO, we rely on the following two large panel data sets: tariff data and firm-level production data. Tariff data can be accessed directly from Brandt et al. (2017), who provides input and output tariffs at the 4-digit level of

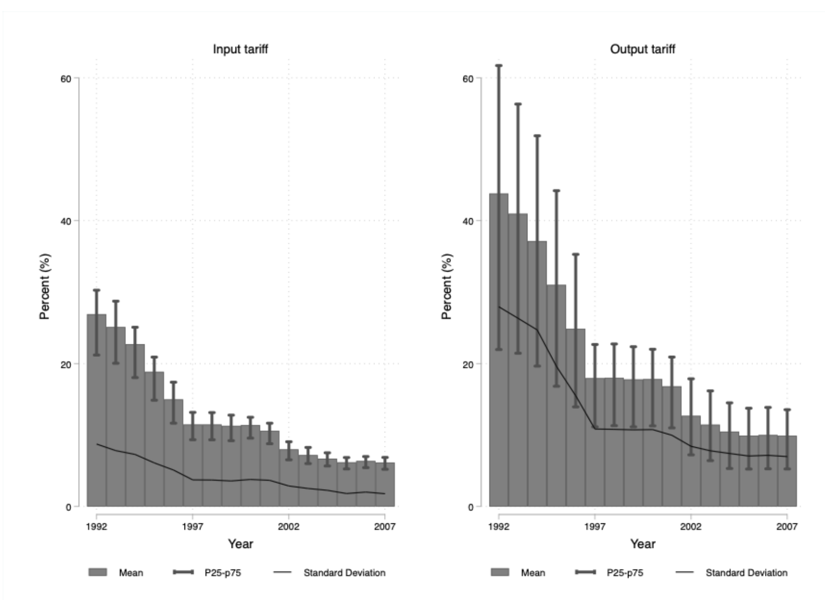
[16] For the details of the derivation, please refer to Appendix D and E.

the China Industrial Classification (CIC-4).[17] The firm-level production data is from Annual Survey of Industrial Firms (ASIF) collected by China's National Bureau of Statistics. The ASIF data has been widely used in academic research, for instance, Brandt et al. (2017; 2012) and Yu (2015). This data set covers all state-owned enterprises (SOEs) and non-state-owned enterprises (non-SOE) with annual sales greater than RMB 5 million ($ 770,000). Complete information on the three major accounting statements (i.e., balance sheet, profit and loss account, and cash flow statement) is also available. Although the data set contains rich information, some of the samples are still noisy and therefore misleading (Brandt et al., 2014). Following Ahn et al. (2011), Cai & Liu (2009), Brandt et al. (2017; 2012; 2014) and Yu (2015), we omit outliers and trade intermediaries and only retain manufacturing industry firms. Appendix A shows the details.

3.2 Background: China's WTO Accession

In 1986, China applied to the General Agreement on Tariffs and Trade to restore its status as a contracting party. After 15 years of long negotiations, China finally rejoined the WTO on December 11, 2001. To join the WTO, China had to carry out substantial unilateral trade liberalization reforms. Figure 1 shows the arithmetic averages and 25th and 75th percentiles of the input and output tariffs at the CIC-4 industry level from 1992 to 2007. From 1992 to 1997, China's input and output tariffs both showed a downward trend. The average input tariff fell from 26.9% to 11.4%, and the average output tariff fell from 43.7% to 17.9%. From 1997 to 2001, the input and output tariffs remained stable. After joining the WTO in 2001, both tariffs showed a downward trend. After 2005, both tariffs remained stable. Figure 1 also shows that the input tariff is lower than the output tariff, but the gap between the two has decreased over time.

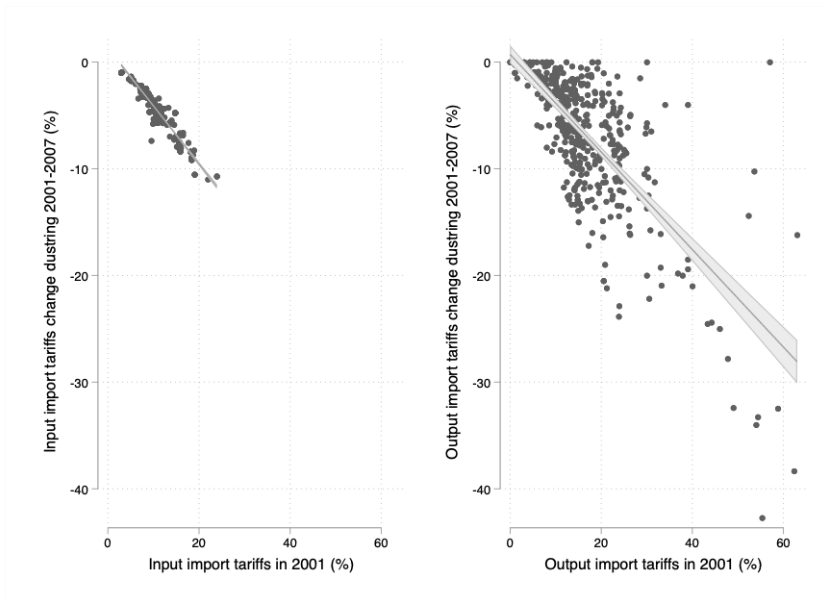
Figure 1: Input Tariff and Output Tariff Evolution during 1992-2007 (CIC-4)



[17]Both the input tariff and the output tariff are import tariffs. For illustration convenience, we omit the "import".

In addition, when China entered the WTO, there was large industry heterogeneity in import tariffs. The reduction in tariffs after China's accession to the WTO also showed great heterogeneity. Figure 2 shows the relationship between the level of input and output tariffs in 2001 and the corresponding changes in tariffs from 2001 to 2007. Clearly, the industry's initial tariff level at the time of WTO entry was significantly positively correlated with its tariff reduction. The higher the initial tariff level was, the greater was the subsequent tariff reduction. This phenomenon holds for both input and output tariffs. This study uses this feature of industrial tariffs in a DID empirical method to explore the impact of trade liberalization.

Figure 2: Correlation between Tariffs in 2001 and Changes in Tariffs over 2001-2007



3.3 Measurement of Key Variables

There are several key variables in the empirical regression specification mentioned later: firms' monopsony power as the explained variable and input and output tariffs as the core explanatory variables.

3.3.1 Monopsony Power

Our methods for estimating firm monopsony power follows the work of Brooks et al. (2021a;b).[18] By constructing a structural model with monopolistic competition in the product market and monopsonistic competition in the input market, the key point of their paper can be summarized as follows:

$$\mu_m^{\mathrm{DLW}} = \mu \times \psi_m \tag{30}$$

[18]Caselli et al. (2021) and Morlacco (2019) apply a similar algorithm. Moreover, our model is consistent with Brooks et al. (2021a;b), and their algorithm can be directly applied to our estimation. Appendix E shows the details.

where μ_m^{DLW} is the markup formula from De Loecker & Warzynski (2012) (DLW). The subscript m denotes different inputs, such as capital (K), labor (L) and the intermediate input (M). The variable μ represents the firms' true markup in the product market and does not vary across the inputs. The variable ψ_m represents the firms markdown in input market m. Equation (30) shows that when the input market is not perfectly competitive, the DLW formula for the markup is the product of the true markup and the input-specific markdown. Since μ does not vary with the inputs, we can take the ratio of equation (30) for different inputs to eliminate μ, that is:

$$\frac{\mu_m^{\mathrm{DLW}}}{\mu_{m'}^{\mathrm{DLW}}} = \frac{\psi_m}{\psi_{m'}} \tag{31}$$

Brooks et al. (2021a;b) further assume that, there exists a factor (empirically, we use the intermediate input) for which all firms are price takers, that is, $\psi_M \equiv 1$. With this assumption and focusing on the labor market, equation (31) can be further expressed as following:

$$\frac{\mu_L^{\mathrm{DLW}}}{\mu_M^{\mathrm{DLW}}} = \frac{\psi_L \times \mu}{\psi_M \times \mu} = \psi_L$$

Equation (31) is at the core of our methodology to estimate firms' monopsony power. It shows that, labor markdown can be obtained by dividing the two DLW markups. We estimate the DLW markup according to De Loecker & Warzynski (2012), which is:

$$\mu_m^{\mathrm{DLW}} = \frac{\theta_m}{\alpha_m} \tag{32}$$

where θ_m refers to the output elasticity of input m, which can be obtained by production function estimation, and α_m denotes the firm-specific expenditure share of input m, which is directly observable in the data.[19] However, ordinary least squares (OLS) estimation of the production function suffers from two endogeneity problems, simultaneity bias and sample selection. To solve these problems, we adopt the control function approach to estimate the production function. As aforementioned, intermediate input is important for estimating the markdown. Naturally, the production function must incorporate the intermediate input and hence obey the gross output formula. We use the Ackerberg et al. (2015) (ACF) method as the baseline for the empirical analysis since it has been widely accepted in many influential pieces of research. Moreover, Gandhi et al. (2020) (GNR) point out that the estimations of gross output production function and the value-added production function are not interchangeable theoretically. The previous method may confront a lack of identification when estimating the output elasticity of intermediate input. Nevertheless, Gandhi

[19]Since we can not observe quantity in the ASIF database, we adjust the factor share by using the exponential of the first stage regression residual from the production function estimation, according to De Loecker & Warzynski (2012). Note that, We can conduct this adjustment only if the production function is estimated using the method of Ackerberg et al. (2015); Gandhi et al. (2020); Levinsohn & Petrin (2003) or Olley & Pakes (1992), instead of simple OLS regression or OLS regression with firm fixed effect. For the detail of implementation, please refer to De Loecker & Warzynski (2012)

et al. (2020) identify the output elasticity of the intermediate input by using the cross-equation constraint between the production function and the first order condition with respect to the intermediate input. This allows the output elasticity of the input to differ across firms within the same industry, which is superior to the Cobb-Douglas production function setup. Therefore, we use Gandhi et al. (2020) and De Loecker & Warzynski (2012) to estimate markup and markdown, which we use for robustness analysis. We use other production function estimation methods (Levinsohn & Petrin, 2003; Olley & Pakes, 1992) (LP and OP, respectively) to estimate the markup and markdown as an additional robustness check since they differ from each other in the timing assumption of labor determination or proxy variables.[20]

3.3.2 Input and Output Tariffs

The output tariff is obtained by aggregating the Harmonized System (HS) 8-digit level product tariff by using the correspondence table between the CIC coding system and the HS coding system, which is the following:

$$\text{OutputTariff}_{st} = (\sum_{p \in s} \tau_{pt})/n_{st} \tag{33}$$

where s and p denote industry (CIC-4) and product (HS-8), respectively; t denotes year; τ denotes product tariff; and n_s refers to the total number of products in industry s. The aggregation is an unweighted average, to avoid bias caused by the negative correlation between trade volume and tariff level (Amiti & Konings, 2007; Brandt et al., 2017). Then, following the method proposed by Amiti & Konings (2007), contrary to the construction of output tariffs, input tariffs are a weighted average of output tariffs, with the weights given by the input share from the 2002 Chinese Input-Output Table.

$$\text{InputTariff}_{st} = \sum_{s'} w_{s's}^{2002} \times \text{OutputTariff}_{s't}, \text{where } w_{s's}^{2002} = \frac{\text{input}_{s's}^{2002}}{\sum_{s''} \text{input}_{s''s}^{2002}} \tag{34}$$

4 Empirical Application

4.1 Empirical Specification

As Figure 2 indicates, the initial levels of input and output tariffs in 2001 are positively correlated with the magnitude of the tariff reduction over 2001-2007. As a result, we can use the DID method to explore trade liberalization. The initial levels of the input and output tariffs act as group variables, and industries with a high initial tariff (treatment group) experience a larger extent of tariff reduction, while industries with

[20]Appendix B displays the relevant summary statistics and figures of markdown and the estimated average output elasticity of different production factors for different industries .

a low initial tariff (control group) experience a smaller extent of tariff reduction. The empirical specification is defined as follows:

$$y_{st} = \beta_0 + (\beta_1 \text{InputTariff}_s^{2001} + \beta_2 \text{OutputTariff}_s^{2001}) \cdot \text{Post02}_t + \mathcal{X}_{st}'\gamma + \lambda_s + \lambda_t + \epsilon_{st} \tag{35}$$

where s, t denote industry (CIC-4) and year respectively. y_{st} refers to the variance of log markdown (the second term in equation 28) or the weighted average of TFP (the first term in equation 28).[21] Moreover, $\text{InputTariff}_s^{2001}$ denotes the input tariff at the CIC-4 industry level in 2001 while $\text{OutputTariff}_s^{2001}$ denotes the output tariff at the CIC-4 industry level in 2001. Post02_t indicates the WTO accession dummy variable, and it takes the value of 1 in 2002 and thereafter, otherwise 0. λ_s is CIC-4 industry fixed effect and λ_t is year fixed effect, with a purpose to control factors that don't vary with industry and time respectively. Standard errors are clustered at the CIC-4 industry level according to Lu & Yu (2015). To eliminate the potential threat of omitted variable bias, we also control dozens of control variables at the CIC-4 industry level, denoted by $\mathcal{X}_{st}'$ which includes the mean value of the fixed asset and the number of firms. The parameters of interest are β_1 and β_2. If $\beta_1(\beta_2)$ is positive, it indicates that the input tariff (output tariff) reduction increases the variance in logarithmic markdown within the industry, *vice versa*.

4.2 Baseline Results

Table 1 reports the baseline results of the regression equation (35). In column (1), we only include the regressor of interest, industry, and year fixed effects. Column (1) indicates that, the coefficient of $\text{InputTariff}_s^{2001} \cdot \text{Post02}_t$ is statistically significant and negative. Since the tariff level in 2001 is positively correlated with tariff reduction between 2001 and 2007, the higher the tariff level in 2001, the larger trade liberalization is realized. Hence, the results demonstrate that input tariff reduction reduces the variance in logarithmic markdown. In contrast, the coefficient of $\text{OutputTariff}_s^{2001} \cdot \text{Post02}_t$ is negative but statistically insignificant, which means that the reduction in output tariff has no impact. In column (2), we take time-varying industry-level attributes into account. Following Lu & Yu (2015), we control the mean value of fixed assets and the number of firms in each industry to account for the entry barrier. The results are not affected. The DID specification requires that the tariff in 2001 be randomly determined. However, this may not be the truth. Following Lu & Yu (2015), we identify variables that had a significant impact on the tariff

[21] Since production function estimation is prone to measurement error, there exists some negative value of markdown. If we simply take the logarithm of the original value of markdown, the negative values are dropped, which significantly changes the distribution (hence variance) of the logarithm of markdown within the manufacturing industry. In light of this, we adopt a Hyperbolic Sine Transformation to take care of this concern, which is:

$$\text{var} \log \psi_s \triangleq \text{var} \log(\psi_{sj} + \sqrt{{\psi_{sj}}^2 + 1})$$

Bellemare & Wichman (2020) points out that this transformation has the following good properties: (1) it is similar to a logarithm; (2) it allows retaining zero-valued (and even negative-valued) observations. This transformation is widely used in economic literature (e.g., Liu & Qiu (2016)).

in 2001. As shown in Appendix A Table B.2 and B.3, four determinants stand out: (1) the output share of SOEs; (2) the output share of domestic firms; (3) export intensity; and (4) the average wage per worker. Taking the interaction terms between the tariff determinants and Post02$_t$ into account, we show in column (3) that our results still remain.

Table 1: Baseline Results

Dependent variables: Variance in log(markdown) (ACF)	(1)	(2)	(3)
Input Tariff$_{01}$ $\times$ Post$_{02}$	-0.475***	-0.475***	-0.447***
	(0.158)	(0.158)	(0.146)
Output Tariff$_{01}$ $\times$ Post$_{02}$	-0.065	-0.064	-0.040
	(0.054)	(0.055)	(0.045)
Average fixed assets (log)		-0.016	-0.015
		(0.015)	(0.016)
Number of firms (log)		-0.006	0.004
		(0.009)	(0.010)
Output share of SOEs$_{01}$ $\times$ Post$_{02}$			0.100***
			(0.028)
Output share of domestic$_{01}$ $\times$ Post$_{02}$			0.028
			(0.041)
Average wage per worker$_{01}$ $\times$ Post$_{02}$			-0.057***
			(0.019)
Export Intensity$_{01}$ $\times$ Post$_{02}$			-0.079***
			(0.019)
Observations	4,179	4,179	4,140
R-squared	0.796	0.797	0.819

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels. In each column, we control both year fixed effect and CIC-4 industry fixed effect.

4.3 Robustness Check

4.3.1 Identifying Assumption

First, to control for other contemporary policy reforms (SOE reform and the relaxation of FDI regulation), which may serve as confounding factors for our estimation, we take two variables into account: the share of SOEs among domestic firms and the number of foreign-invested firms. The result is reported in column (1) in Table 2. These additional controls leave the main result unaffected.

Table 2: Checks on Identifying Assumptions

Dependent variables: Variance in log(markdown) (ACF)	(1)	(2)	(3)
Input Tariff$_{01}$ $\times$ Post$_{02}$	-0.407***	-0.422***	
	(0.128)	(0.127)	
Output Tariff$_{01}$ $\times$ Post$_{02}$	-0.050	-0.041	
	(0.041)	(0.040)	
SOE share	-0.050*	-0.052**	0.018
	(0.027)	(0.025)	(0.047)
FDI (log)	-0.005	-0.005	0.010
	(0.005)	(0.005)	(0.007)
Total exports (log)		0.001	0.004
		(0.003)	(0.005)
Input Tariff			0.220
			(0.227)
Output Tariff			-0.112
			(0.101)
Observations	3,923	3,903	1,510
R-squared	0.859	0.864	0.925

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at 1%, 5%, 10% levels. In each column, we control year fixed effect, CIC-4 industry fixed effect, control variables, and the interaction terms between Post02$_t$ and tariff determinant.

On the one hand, China's accession to the WTO required import tariff reduction. On the other hand, WTO On the one hand, China's accession to WTO requires import tariff reduction. On the other hand, WTO accession is also associated with export tariff reduction and export market expansion. To take this into account, we add CIC-4 level total exports as an additional control regressor. The result is shown in column (2) in Table 2. Qualitatively, the signs and significance of the coefficient of the variable of interest are confirmed. In addition, as a placebo test, we investigate the impact of trade liberalization on the variance in logarithmic markdown in the pre-WTO period (i.e., 1998-2001), following Topalova (2010). As Figure 2 shows, the reduction of the import tariff during this period was subtle. Hence, it is expected that the coefficients of the regressor of interest would be insignificant. Column (3) in Table 2 verifies our prediction. Finally, another prerequisite for using the DID estimation strategy is that the treatment and control groups should satisfy the parallel trends assumption before the policy shock. We use the event study approach to check whether this assumption is satisfied:

$$
\begin{aligned}
y_{st} = &\beta_0 + \sum_{m=1998}^{m=2007} \beta_m \text{InputTariff}_s^{2001} \cdot \text{m}_t + \sum_{n=1998}^{n=2007} \beta_n \text{OutputTariff}_s^{2001} \cdot \text{n}_t \\
&+ \mathcal{X}_{st}' \gamma + \lambda_s + \lambda_t + \epsilon_{st}
\end{aligned} \tag{36}
$$

where we set the last year before China accession to the WTO (i.e., 2001) as the base year; m_t and n_t are year dummy variables; and other things are held equal.

Figure 3: The Dynamic Effect of Trade Liberalization

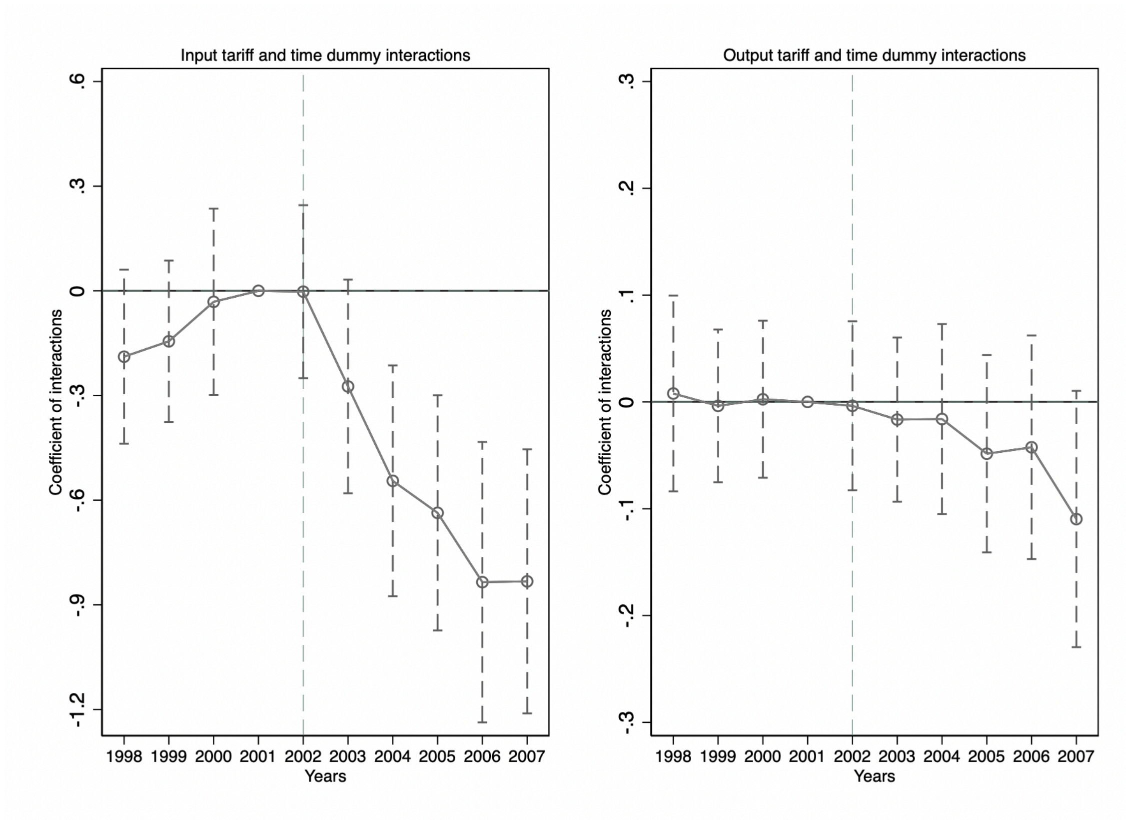


Figure 3 exhibits the estimates of β_m and β_n and the 95% confidence intervals. The estimates of β_m and β_n are insignificant between 1998 and 2001, which suggests that there is no ex-ante difference between the treatment and control groups. Moreover, the impact of input trade liberalization on the variance in logarithmic markdown is negative, and the magnitude is increasing, while the output trade liberalization has no significant effect.

4.3.2 Other Robustness Checks

Alternative measure of markdown. As shown in section 3, the markdown is obtained from production function estimation. As aforementioned, we use the ACF method for benchmark analysis. In this section, we estimate the production function using the GNR method, OP method, LP method, Simple OLS method, and OLS with fixed effects. Table 3 shows that these distinctive production function estimation strategies yield results that are comparable to our benchmark estimates.

Table 3: Alternative Measure of Markdown (GNR & Cobb-Douglas)

Dependent variables: Variance in log(markdown)	GNR (1)	OP (2)	LP (3)	OLS (4)	OLS FE (5)
Input Tariff$_{01}$ $\times$ Post$_{02}$	-0.480***	-0.493***	-0.506***	-0.503***	-0.461***
	(0.163)	(0.131)	(0.124)	(0.126)	(0.125)
Output Tariff$_{01}$ $\times$ Post$_{02}$	0.032	0.022	0.013	0.024	-0.020
	(0.052)	(0.040)	(0.035)	(0.036)	(0.037)
Observations	3,923	3,902	3,903	3,923	3,923
R-squared	0.802	0.910	0.941	0.938	0.852

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels. In each column, we control year fixed effect, CIC-4 industry fixed effect, control variables, the interaction terms between Post02$_t$ and tariff determinant, other contemporary policy reforms and total exports.

As pointed out in section 3.3.1, the Cobb-Douglas production function specification implies that there is a common constant output elasticity within the industry, which might underestimate the heterogeneity of firms' monopsony power, although the robustness of the results using the GNR method mitigates this concern to some extent. In this section, we move a step further and estimate the production function with a translog specification:

$$\begin{aligned}\tilde{q}_{it} = \beta_l \tilde{l}_{it} + \beta_k \tilde{k}_{it} + \beta_m \tilde{m}_{it} + \beta_{ll} \tilde{l}_{it}^2 + \beta_{kk} \tilde{k}_{it}^2 + \beta_{mm} \tilde{m}_{it}^2 + \\ \beta_{lk} \tilde{l}_{it} \tilde{k}_{it} + \beta_{lm} \tilde{l}_{it} \tilde{m}_{it} + \beta_{km} \tilde{k}_{it} \tilde{m}_{it} + \beta_{lkm} \tilde{l}_{it} \tilde{k}_{it} \tilde{m}_{it} + \omega_{it} + \varepsilon_{it}\end{aligned} \tag{37}$$

where the variables with a tilde refers to the logarithm of the level; for instance, $\tilde{q}_{it}$ denotes the log of the output of the firm. ω_{it} indicates firm-specific productivity, and ε_{it} is an independent and identically distributed error. Similarly, we estimate the translog production function for each CIC-2 industry separately. After we obtain the estimation of the coefficient in equation (37), we can obtain the output elasticity of the production function by taking partial derivatives. Then, following De Loecker & Warzynski (2012) and Brooks et al. (2021b), we can estimate the markup and markdown. The regression results using markdown in terms of the translog production function specification are displayed in Table 4. All the coefficients show consistent patterns.

Table 4: Alternative Measure of Markdown (Translog)

Dependent variables: Variance in log(markdown)	OP (1)	LP (2)	ACF (3)	OLS (4)	OLS FE (5)
Input Tariff$_{01}$ $\times$ Post$_{02}$	-0.365	-2.116**	-1.132*	-2.279***	-1.849***
	(0.693)	(1.067)	(0.671)	(1.144)	(0.603)
Output Tariff$_{01}$ $\times$ Post$_{02}$	0.002	0.135	-0.228	0.103	0.198
	(0.290)	(0.289)	(0.142)	(0.311)	(0.142)
Observations	3,902	3,903	3,903	3,923	3,923
R-squared	0.861	0.852	0.876	0.864	0.822

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels. In each column, we control year fixed effect, CIC-4 industry fixed effect, control variables, the interaction terms between Post02$_t$ and tariff determinant, other contemporary policy reforms and total exports.

Alternative regression specification. Table 5 summarizes the results. In column (1), we follow Amiti & Konings (2007) and use input and output tariffs as key explanatory variables. In column (2), we follow Brandt et al. (2017) and use the one-period lag of the input and output tariffs as key explanatory variables to address endogeneity. In column (3), we follow Bertrand et al. (2004) and compress the data into a two-period panel to address the serial correlation problem. All the results are in line with our baseline results.

Table 5: Alternative Empirical Specification

Dependent variables: Variance in log(markdown) (ACF)	Tariff (1)	Tariff (lag) (2)	Two-period (3)
Input Tariff	0.991***	1.133***	
	(0.212)	(0.226)	
Output Tariff	0.083	0.108**	
	(0.052)	(0.053)	
Input Tariff$_{01}$ $\times$Post$_{02}$			-0.430***
			(0.133)
Output Tariff$_{01}$ $\times$Post$_{02}$			-0.049
			(0.042)
Observations	3,903	3,520	790
R-squared	0.866	0.876	0.938

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels. In each column, we control year fixed effect, CIC-4 industry fixed effect, control variables, the interaction terms between Post02$_t$ and tariff determinant, other contemporary policy reforms and total exports.

4.3.3 Quantile Regression

To dive into the details of how trade liberalization affects the variance in logarithmic markdown, we turn to the heterogeneous response of the logarithm of the markdown at different quantiles by industry and year. In particular, we regress p5, p25, p50, p75, p95, and the mean value of the logarithm of the markdown on the aforementioned regressor. The results are summarized in Table 6. Input trade liberalization reduces firms' monopsony power, which is consistent with Kondo et al. (2021). This *"pro-competitive"* effect is most obvious at large quantiles, thereby compressing the distribution of the logarithm of the markdown. As a result, the variance in logarithmic markdown decreases. Output trade liberalization also has a subtle negative influence on firms' monopsony power, but only at a relatively large quantiles. As a result, the variance in total does not change much in response to output trade liberalization.

Table 6: Quantile Regression

Dependent variables:	(1)	(2)	(3)	(4)	(5)	(6)
Quantile of log(markdown) (ACF)	Mean	p5	p25	p50	p75	p95
Input $\text{Tariff}_{01} \times \text{Post}_{02}$	-0.408**	-0.079	-0.268*	-0.299	-0.445	-1.209***
	(0.195)	(0.079)	(0.148)	(0.204)	(0.296)	(0.410)
Output $\text{Tariff}_{01} \times \text{Post}_{02}$	-0.090	0.013	0.005	-0.112	-0.209**	-0.094
	(0.067)	(0.028)	(0.061)	(0.073)	(0.098)	(0.138)
Observations	3,903	3,903	3,903	3,903	3,903	3,903
R-squared	0.930	0.787	0.904	0.925	0.907	0.847

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels. In each column, we control both year fixed effect, CIC-4 industry fixed effect, control variables, the interaction terms between Post02_t and tariff determinant, other contemporary policy reforms and total exports.

4.4 Heterogeneity Analysis

First, we divide the firms into two groups: incumbents and exits/entrants. As Melitz (2003) points out, these firms differ from each other in productivity. Moreover, the incumbent firms may have greater compared with the exits/entrants. The first two columns in Table 7 exhibit the results, which indicate that the effect of tariff reduction on the variance in logarithmic markdown is remarkably large for incumbent firms. Second, the last two columns in Table 7 show the subgroup regressions for SOEs and non-SOEs. Chen et al. (2019) propose that the competition between SOEs is lower compared to that between non-SOEs, and SOEs have priority over access to capital, land, and other production factors. Meanwhile, SOEs undertake the responsibility of stabilizing employment and other social objectives. The regression results show that the effect of input trade liberalization is larger for non-SOEs than SOEs.

Table 7: Heterogeneity Analysis (Firm Characteristic)

Dependent variables: Variance in log(markdown)	(1) Incumbent	(2) Exit/Entrant	(3) SOEs	(4) Non-SOEs
Input Tariff$_{01}$ × Post$_{02}$	-0.562***	-0.406**	-0.400*	-0.454***
	(0.181)	(0.168)	(0.228)	(0.149)
Output Tariff$_{01}$ × Post$_{02}$	0.082	-0.007	-0.075	-0.034
	(0.052)	(0.051)	(0.070)	(0.041)
Observations	3,892	3,896	3,875	3,902
R-squared	0.696	0.679	0.621	0.826

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels. In each column, we control year fixed effect, CIC-4 industry fixed effect, control variables, the interaction terms between Post02$_t$ and tariff determinant, other contemporary policy reforms and total exports.

Next, we consider firm heterogeneity in terms of location. China's opening to the outside world was not achieved overnight but through a gradual process. There exists huge differences between coastal areas and inland areas in China in terms of the opening-up policy, governing mode, and degree of competition. Columns 1 and 2 in Table 8 report the estimation results using the coastal province and inland province subsamples, respectively.[22] The results show that input trade liberalization has had a larger impact on firms in inland areas than coastal areas.

Table 8: Heterogeneity Analysis (Firm Location)

Dependent variables: Variance in log(markdown)	(1) Coastal	(2) Inland	(3) High Mobility	(4) Low Mobility
Input Tariff$_{01}$ × Post$_{02}$	-0.378**	-0.495**	-0.358**	-0.666***
	(0.165)	(0.210)	(0.148)	(0.170)
Output Tariff$_{01}$ × Post$_{02}$	-0.079*	0.066	-0.050	0.046
	(0.046)	(0.066)	(0.042)	(0.053)
Observations	3,903	3,871	3,901	3,897
R-squared	0.829	0.623	0.759	0.733

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels. In each column, we control year fixed effect, CIC-4 industry fixed effect, control variables, the interaction terms between Post02$_t$ and tariff determinant, other contemporary policy reforms and total exports.

[22] According to the *"Notice of the State Planning Commission and the National Bureau of Statistics on the Division of Coastal and Inland Areas"*, the coastal provinces include Liaoning, Hebei, Beijing, Tianjin, Shandong, Shanghai, Zhejiang, Fujian, Guangdong, and Guangxi, and the rest are collectively referred to as inland areas.

Brooks et al. (2021b) and Manning (2003b) propose that labor supply elasticity faced by individual firms is another important determinant of firms' labor market power. The more elastic the labor supply is, the less is firms' labor market power. Hence, we split up the sample into two groups: firms located in provinces with high labor mobility and firms located in provinces with low labor mobility.[23] The regression results using these two subgroups are displayed in columns 3 and 4 in Table 8. We find that the influence of input trade liberalization is significantly larger in the provinces with low labor mobility.

5 Welfare Implication

First, we detect the impact of both input and output trade liberalizations on the industry-level weighted average TFP. The regression results are reported in Table 9. Column 3 shows that input tariff reduction is associated with higher TFP, which is consistent with the findings of Amiti & Konings (2007) and Brandt et al. (2017). Output tariff reduction has a negative but insignificant impact on average TFP.

Table 9: Trade Liberalization and its Impact on Average TFP

Dependent variables: Weighted Average TFP (ACF)	(1)	(2)	(3)
Input Tariff$_{01}$ $\times$ Post$_{02}$	0.515	0.807	1.118*
	(0.632)	(0.600)	(0.624)
Output Tariff$_{01}$ $\times$ Post$_{02}$	-0.226	-0.223	-0.122
	(0.251)	(0.238)	(0.243)
Observations	4,041	4,041	4,001
R-squared	0.615	0.624	0.623

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% level. In each column, we control both year fixed effect and CIC-4 industry fixed effect. In column (2), we further add control variables. In column (3), we take tariff determinants into account. To save place, we only report results of variables of interest.

Second, with the estimation of the markup and coefficient of the interaction term in Table 1 and 9, respectively, we can back out the parameter ρ_s and then infer Γ_s. Together with the output share α_s, which can be obtained directly from the data, we can finally calculate the aggregate impact of input trade liberalization on nationwide TFP. We denote the TFP gains from trade as:

[23] The data on labor mobility in different provinces comes from Fan Gang, Wang Xiaolu, Zhu Hengpeng. *"China's Marketization Index - A Report on the Relative Progress of Marketization in Various Regions"* Beijing: Economic Science Press, 2010. They provide annual data on labor mobility at the provincial level. Provinces with labor mobility above the mean labor mobility of all provinces are classified into the high labor mobility subgroup, and the rest of the provinces are grouped into the low labor mobility subgroup.

$$\widehat{\text{TFP}} = \log\text{TFP}^{\text{trade}} - \log\text{TFP}^{\text{counterfactual}} \tag{38}$$

where *trade* denotes trade liberalization, and *counterfactual* denotes absence of trade liberalization. In sum, the TFP gains from trade are 6%, of which 5.7% is from the direct impact of trade liberalization on weighted TFP, and 0.3% stems from the indirect impact of trade liberalization or the reallocation effect on the variance in logarithmic markdown. Our results show that, with an imperfectly competitive labor market, the impact of trade liberalization on heterogeneous firms' labor market power serves as an important channel of gains from trade, and it can improve production efficiency by means of mitigating misallocation.

6 Conclusion

In this paper, we have studied the impact of firms' heterogeneous monopsony power in the labor market on production allocation efficiency. We built a model with monopolistic competition in the product market and monopsonistic competition in the labor market. Our model shows that workers' heterogeneous preferences for occupations and firms' differentiated occupational composition serve as new sources of firms' heterogeneous monopsony power. From the micro point of view, firms with monopsony power produce less, use less input, charge higher prices, and use more non-labor input, compared with other firms. From the macro point of view, firms' heterogeneous monopsony power gives rise to misallocation and results in efficiency loss. The paper has shown that the variance in logarithmic markdown serves as a *sufficient statistic* of the negative impact of firms' heterogeneous monopsony power on total production efficiency. To the best of our knowledge, this is the first study to demonstrate that firms' heterogeneous monopsony power leads to misallocation and production efficiency loss.

Using China's accession to the WTO as a semi-natural experiment, we quantified the welfare impact of labor reallocation related to trade liberalization, with detailed data on Chinese firms' production and tariffs. The empirical results support that input trade liberalization decreases the variance in logarithmic markdown and relieves the misallocation of the factors of production across firms, while output trade liberalization has no significant effect. Overall, improvement in the allocation of labor accounts for 5% of the total production efficiency gains from China's WTO accession, which suggests that the *"reallocation"* effect of trade in an imperfectly competitive labor market serves as an important channel for gains from trade.

Our paper has important policy implications. First, firms' relative market power in the labor market is also important. Policy design should pay more attention to the distributional impacts on the market participants. Second, trade liberalization can restrict firms with greater labor market power and improve allocation efficiency. In this sense, trade policy and domestic policy, such as labor protection, are complementary.

References

Ackerberg, Daniel A, Kevin Caves & Garth Frazer. 2015. Identification properties of recent production function estimators. *Econometrica* 83(6). 2411–2451.

Ahn, JaeBin, Amit K Khandelwal & Shang-Jin Wei. 2011. The role of intermediaries in facilitating trade. *Journal of International Economics* 84(1). 73–85.

Ahsan, Reshad N & Devashish Mitra. 2014. Trade liberalization and labor's slice of the pie: Evidence from indian firms. *Journal of Development Economics* 108. 1–16.

Amiti, Mary & Jozef Konings. 2007. Trade liberalization, intermediate inputs, and productivity: Evidence from indonesia. *American Economic Review* 97(5). 1611–1638.

Bachmann, Ronald, Gökay Demir & Hanna Frings. 2022. Labor market polarization, job tasks, and monopsony power. *Journal of Human Resources* 57(S). S11–S49.

Bellemare, Marc F & Casey J Wichman. 2020. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics* 82(1). 50–61.

Berger, David W, Kyle F Herkenhoff & Simon Mongey. 2019. Labor market power. *Technical Report, National Bureau of Economic Research* .

Bertrand, Marianne, Esther Duflo & Sendhil Mullainathan. 2004. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119(1). 249–275.

Boal, William M & Michael R Ransom. 1997. Monopsony in the labor market. *Journal of Economic Literature* 35(1). 86–112.

Booth, Alison L, Gylfi Zoega et al. 2000. *Why do firms invest in general training?:'good'firms and'bad'firms as a source of monopsony power*, vol. 2536. Centre for Economic Policy Research.

Brandt, Loren, Johannes Van Biesebroeck, Luhang Wang & Yifan Zhang. 2017. Wto accession and performance of chinese manufacturing firms. *American Economic Review* 107(9). 2784–2820.

Brandt, Loren, Johannes Van Biesebroeck & Yifan Zhang. 2012. Creative accounting or creative destruction? firm-level productivity growth in chinese manufacturing. *Journal of Development Economics* 97(2). 339–351.

Brandt, Loren, Johannes Van Biesebroeck & Yifan Zhang. 2014. Challenges of working with the chinese nbs firm-level data. *China Economic Review* 30. 339–352.

Brooks, Wyatt J, Joseph P Kaboski, Illenin O Kondo, Yao Amber Li & Wei Qian. 2021a. Infrastructure investment and labor monopsony power. *IMF Economic Review* 69(3). 470–504.

Brooks, Wyatt J, Joseph P Kaboski, Yao Amber Li & Wei Qian. 2021b. Exploitation of labor? classical monopsony power and labor's share. *Journal of Development Economics* 150. 102627.

Burdett, Kenneth & Dale T Mortensen. 1998. Wage differentials, employer size, and unemployment. *International Economic Review* 257–273.

Cai, Hongbin & Qiao Liu. 2009. Competition and corporate tax avoidance: Evidence from chinese industrial firms. *The Economic Journal* 119(537). 764–795.

Card, David, Ana Rute Cardoso, Joerg Heining & Patrick Kline. 2018. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics* 36(S1). S13–S70.

Card, David, Jörg Heining & Patrick Kline. 2013. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics* 128(3). 967–1015.

Caselli, Mauro, Lionel Nesta & Stefano Schiavo. 2021. Imports and labour market imperfections: Firm-level evidence from france. *European Economic Review* 131. 103632.

Chen, Cheng, Wei Tian & Miaojie Yu. 2019. Outward fdi and domestic input distortions: Evidence from chinese firms. *The Economic Journal* 129(624). 3025–3057.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal & Nina Pavcnik. 2016. Prices, markups, and trade reform. *Econometrica* 84(2). 445–510.

De Loecker, Jan & Frederic Warzynski. 2012. Markups and firm-level export status. *American Economic Review* 102(6). 2437–2471.

Dobbelaere, Sabien & Kozo Kiyota. 2018. Labor market imperfections, markups and productivity in multinationals and exporters. *Labour Economics* 53. 198–212.

Dobbelaere, Sabien & Quint Wiersma. 2020. The impact of trade liberalization on firms' product and labor market power. *IZA Discussion Paper No. 12951, Institute of Labor Economics, Bonn, Germany* .

Dorn, David, Lawrence F Katz, Christina Patterson, John Van Reenen et al. 2017. Concentrating on the fall of the labor share. *American Economic Review* 107(5). 180–85.

Edmond, Chris, Virgiliu Midrigan & Daniel Yi Xu. 2015. Competition, markups, and the gains from international trade. *American Economic Review* 105(10). 3183–3221.

Edmond, Chris, Virgiliu Midrigan & Daniel Yi Xu. 2021. How costly are markups? *Technical Report, National Bureau of Economic Research* .

Egger, Hartmut, Udo Kreickemeier, Christoph Moser & Jens Wrona. 2021. Exporting and offshoring with monopsonistic competition. *The Economic Journal* 132(644). 1449–1488.

Epifani, Paolo & Gino Gancia. 2011. Trade, markup heterogeneity and misallocations. *Journal of International Economics* 83(1). 1–13.

Fajgelbaum, Pablo, Gene M Grossman & Elhanan Helpman. 2011. Income distribution, product quality, and international trade. *Journal of Political Economy* 119(4). 721–765.

Fan, Haichao, Xiang Gao, Yao Amber Li & Tuan Anh Luong. 2018. Trade liberalization and markups: Micro evidence from china. *Journal of Comparative Economics* 46(1). 103–130.

Foster, Lucia, John Haltiwanger & Chad Syverson. 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review* 98(1). 394–425.

Fox, Jeremy T. 2010. Estimating the employer switching costs and wage responses of forward-looking engineers. *Journal of Labor Economics* 28(2). 357–412.

Gandhi, Amit, Salvador Navarro & David A Rivers. 2020. On the identification of gross output production functions. *Journal of Political Economy* 128(8). 2973–3016.

Gould, Elise. 2014. Why america's workers need faster wage growth—and what we can do about it. *Economic Policy Institute Briefing Paper* 382.

Haltiwanger, John C, Julia I Lane & James Spletzer. 1999. Productivity differences across employers: The roles of employer size, age, and human capital. *American Economic Review* 89(2). 94–98.

Haltiwanger, John C, Julia I Lane & James R Spletzer. 2007. Wages, productivity, and the dynamic interaction of businesses and workers. *Labour Economics* 14(3). 575–602.

Hsieh, Chang-Tai & Peter J Klenow. 2009. Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics* 124(4). 1403–1448.

Hsu, Wen-Tai, Yi Lu & Guiying Laura Wu. 2020. Competition, markups, and gains from trade: A quantitative analysis of china between 1995 and 2004. *Journal of International Economics* 122. 103266.

Jha, Priyaranjan & Antonio Rodriguez-Lopez. 2021. Monopsonistic labor markets and international trade. *European Economic Review* 140. 103939.

Karabarbounis, Loukas & Brent Neiman. 2014. The global decline of the labor share. *The Quarterly Journal of Economics* 129(1). 61–103.

Kondo, Illenin, Yao Amber Li & Wei Qian. 2021. Trade liberalization and labor monopsony: Evidence from chinese firms. *Available at SSRN 4088005* .

Levinsohn, James. 1993. Testing the imports-as-market-discipline hypothesis. *Journal of international Economics* 35(1-2). 1–22.

Levinsohn, James & Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70(2). 317–341.

Liu, Qing & Larry D Qiu. 2016. Intermediate input imports and innovations: Evidence from chinese firms' patent filings. *Journal of International Economics* 103. 166–183.

Liu, Zhengwen & Hong Ma. 2021. Input trade liberalization and markup distribution: Evidence from china. *Economic Inquiry* 59(1). 344–360.

Lu, Yi, Yoichi Sugita, Lianming Zhu et al. 2019. Wage markdowns and fdi liberalization. *Hitotsubashi Institute for Advanced Study Discussion Paper HIES-*$E_{8}3, HitotsubashiUnversity, Tokyo.$

Lu, Yi & Linhui Yu. 2015. Trade liberalization and markup dispersion: evidence from china's wto accession. *American Economic Journal: Applied Economics* 7(4). 221–53.

Macedoni, Luca. 2021. Monopsonistic competition, trade, and the profit share. *The Scandinavian Journal of Economics* 124(2). 488–515.

Macedoni, Luca & Vladimir Tyazhelnikov. 2019. Oligopoly and oligopsony in international trade .

Macedoni, Luca, Mingzhi Xu & Robert C Feenstra. 2020. Large firms in retail markets: Multiple products for heterogeneous consumers. *Available at SSRN 3475289* .

MacKenzie, Gaelan. 2021. Trade and market power in product and labor markets. *Technical Report, Bank of Canada* .

Manning, Alan. 2003a. *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press.

Manning, Alan. 2003b. The real thin theory: monopsony in modern labour markets. *Labour Economics* 10(2). 105–131.

Manning, Alan. 2021. Monopsony in labor markets: A review. *ILR Review* 74(1). 3–26.

Melitz, Marc J. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6). 1695–1725.

Mertens, Matthias. 2020. Labor market power and the distorting effects of international trade. *International Journal of Industrial Organization* 68. 102562.

Midrigan, Virgiliu & Daniel Yi Xu. 2014. Finance and misallocation: Evidence from plant-level data. *American Economic Review* 104(2). 422–58.

Morlacco, Monica. 2019. Market power in input markets: Theory and evidence from french manufacturing. *Unpublished, March* 20. 2019.

Olley, Steven & Ariel Pakes. 1992. The dynamics of productivity in the telecommunications equipment industry.

Peters, Michael. 2020. Heterogeneous markups, growth, and endogenous misallocation. *Econometrica* 88(5). 2037–2073.

Ransom, Tyler. 2021. Labor market frictions and moving costs of the employed and unemployed. *Journal of Human Resources* 0219–10013R2.

Restuccia, Diego & Richard Rogerson. 2008. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics* 11(4). 707–720.

Schubert, Gregor, Anna Stansbury & Bledi Taska. 2019. Mitigating monopsony: Occupational mobility and outside options. *Working Paper, Washington Center for Equitable Growth, Washington, DC* .

Topalova, Petia. 2010. Factor immobility and regional impacts of trade liberalization: Evidence on poverty from india. *American Economic Journal: Applied Economics* 2(4). 1–41.

Wu, Liangjie. 2020. *Partially directed search in the labor market*: The University of Chicago dissertation.

Yu, Miaojie. 2015. Processing trade, tariff reductions and firm productivity: Evidence from chinese firms. *The Economic Journal* 125(585). 943–988.

Appendix A Details of the Data Processing

Following Cai & Liu (2009), Brandt et al. (2017; 2012; 2014) and Yu (2015), we conduct the following data cleaning process:

- Bbservations with missing key financial variables (such as total assets, net value of fixed assets, sales and gross value of the firm's output) are excluded
- Firms with workers fewer that eight are dropped from the sample
- Following the basic rules of the Generally Accepted Accounting Principles (GAAP), we eliminate the observations if any of the following criterias are met.
  - Liquid assets are greater than total assets
  - Total fixed assets are greater than total assets
  - The net value of fixed assets is greater than total assets
  - The firm's identification number is missing
  - An invalid established time exists (e.g., the opening month is later than December or earlier than January)

In the Annual Survey of Industrial Firms (ASIF) data, there exists some trading companies that do not produce themselves (Ahn et al., 2011). Following Brandt et al. (2017), we delete these trading companies by identifying key words in their firm name. Moreover, the ASIF database includes mining industries; manufacturing industries; and electricity, gas, and water production and supply industries. We only retain firms in the manufacturing industry and omit the other two types of firms.

Since the ASIF data does not report the actual capital stock of the company, we use the method of Brandt et al. (2012) to convert the book value of capital into the comparable actual capital stock. Meanwhile, the China Industrial Classification (CIC) 4-digit code is adjusted to be consistent over time and the nominal variables, such as output value, sales value, and intermediate input value, are converted into real variables using the deflator provided by Brandt et al. (2012).

We omit observations from Tobacco (CIC2, 16) and other manufacturing (CIC2, 43) industries due to the lack of observations.

Appendix B Figures and Tables

Figure B1 shows the mean markdowns Ackerberg et al. (2015) (ACF, Cobb-Douglas specification) for the CIC-2 industry level during 1998-2007. There exists huge heterogeneity in terms of the average markdown across different CIC-2 industries. The average markdowns fall in the range from 0.29 to 2.71, and the mean value of markdowns is 1.04. Non-ferrous Metals, Communications and Computers, and Food from Agricultural Products are the industries with the first three largest markdowns. By contrast, Printing and Recording Media, Beverages, and Wood and Products are the industries with the smallest three markdowns.

Figure B1: Estimated Markdown for CIC-2 Industry Level

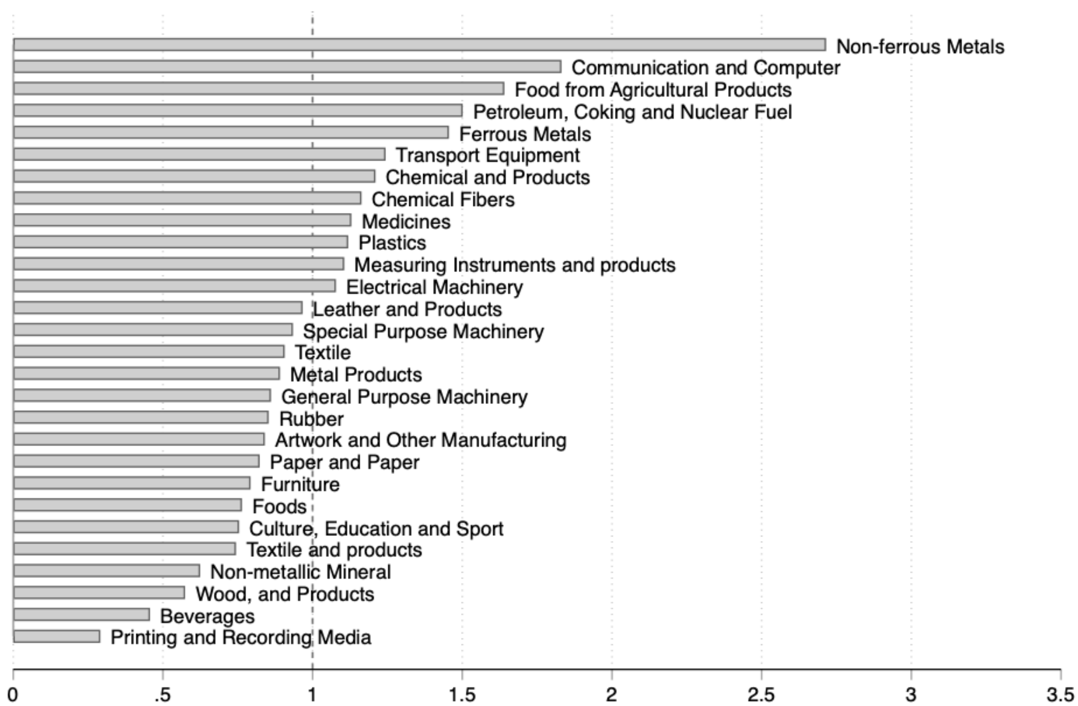


Table B.1 reports the average of estimated output elasticity of the production function at the CIC-2 industry level calculated by ACF method (Cobb-Douglas specification). The mean values of the average output elasticities of labor, capital and intermediate inputs are around 0.07, 0.04, and 0.85, respectively.

Table B.1: Average Output Elasticity by CIC-2 Sector (ACF, Cobb-Douglas Specification)

Industry	No.obs.	β_L	β_K	β_M	RTS
13 Food from Agricultural Products	117,337	0.06	0.04	0.85	0.95
14 Foods	47,219	0.06	0.04	0.88	0.98
15 Beverages	32,793	0.03	0.02	0.89	0.94
17 Textile	162,311	0.07	0.03	0.86	0.96
18 Textile and products	92,868	0.09	0.04	0.80	0.93
19 Leather and Products	46,210	0.08	0.02	0.83	0.93
20 Wood, and Products	41,812	0.04	0.02	0.89	0.95
21 Furniture	22,315	0.07	0.03	0.85	0.95
22 Paper and Paper	57,176	0.06	0.03	0.87	0.96
23 Printing and Recording Media	40,234	0.04	0.02	0.83	0.89
24 Culture, Education and Sport	25,481	0.08	0.04	0.81	0.93
25 Petroleum, Coking and Nuclear Fuel	16,827	0.05	0.03	0.87	0.96
26 Chemical and Products	140,435	0.07	0.04	0.86	0.97
27 Medicines	40,905	0.09	0.06	0.82	0.97
28 Chemical Fibers	9,779	0.06	0.03	0.92	1.01
29 Rubber	23,021	0.07	0.05	0.84	0.96
30 Plastics	89,596	0.08	0.05	0.83	0.96
31 Non-metallic Mineral	165,781	0.06	0.04	0.88	0.98
32 Ferrous Metals	46,040	0.06	0.03	0.91	1.00
33 Non-ferrous Metals	34,267	0.08	0.03	0.88	0.99
34 Metal Products	103,756	0.07	0.04	0.86	0.97
35 General Purpose Machinery	146,404	0.08	0.05	0.85	0.99
36 Special Purpose Machinery	81,070	0.09	0.07	0.85	1.01
37 Transport Equipment	92,192	0.11	0.07	0.84	1.03
39 Electrical Machinery	114,855	0.08	0.04	0.86	0.98
40 Communication and Computer	64,512	0.17	0.08	0.80	1.05
41 Measuring Instruments and products	27,326	0.12	0.04	0.81	0.97
42 Artwork and Other Manufacturing	37,422	0.08	0.03	0.84	0.95

Following the method of Lu & Yu (2015), we detect the determinants of the input and output tariffs in 2001 at the CIC-4 level, including *political factors* (output shares of state-owned enterprisess (SOEs), output share of other domestic firms, total employment in log, and employment growth rate over the past several years), *economic factors* (average wage per worker in log, capital-labor ratio in log, value-added ratio, and industry age), and *industrial policy* (export intensity). Tables B.2 and B.3 display the estimation results for input and output tariffs, respectively. Four variables have significant impacts on both input and output tariffs in 2001: (1) output share of SOEs, (2) output share of other domestic firms, (3) average wage per worker in log, and (4) export intensity.

Table B.2: Determinants of Input Tariffs in 2001

Dependent variables: Input tariff (2001)	Political (1)	Economic (2)	Industrial policy (3)
Output shares of SOEs (2001)	-0.033***	-0.052***	-0.061***
	(0.010)	(0.013)	(0.014)
Output shares of other domestic firms (2001)	-0.049***	-0.076***	-0.090***
	(0.010)	(0.011)	(0.013)
Log total employment (2001)	-0.002*	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)
Employment growth rate (98-01)	-0.027*	-0.012	-0.013
	(0.014)	(0.013)	(0.013)
Log average wage per worker (2001)		-0.042***	-0.041***
		(0.007)	(0.007)
Log capital-labor ratio (2001)		0.010**	0.006
		(0.004)	(0.004)
Value-added ratio (2001)		-0.014	-0.024
		(0.031)	(0.031)
Industry age (2001)		0.001	0.001
		(0.000)	(0.000)
Export intensity (2001)			-0.017***
			(0.006)
Observations	521	521	521
R-squared	0.049	0.115	0.123

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels.

Recall that, Figure 2 in the main text shows that the reduction in tariffs between 2001 and 2007 has a positive correlation with the level of tariffs in 2001. Hence, the regression results in Tables B.2 and B.3 are reasonable, for instance., industries with higher SOEs output as a share of domestic firms output experienced less tariff reduction after China's WTO accession.

Table B.3: Determinants of Output Tariffs in 2001

Dependent variables:	Political	Economic	Industrial policy
Output Tariff (2001)	(1)	(2)	(3)
Output shares of SOEs (2001)	-0.084***	-0.122***	-0.152***
	(0.029)	(0.039)	(0.041)
Output shares of other domestic firms (2001)	-0.121***	-0.164***	-0.208***
	(0.027)	(0.027)	(0.034)
Log total employment (2001)	-0.001	-0.005	-0.005
	(0.003)	(0.003)	(0.003)
Employment growth rate (98-01)	-0.073**	-0.046	-0.049
	(0.032)	(0.033)	(0.032)
Log average wage per worker (2001)		-0.096***	-0.094***
		(0.019)	(0.019)
Log capital-labor ratio (2001)		0.029***	0.019*
		(0.011)	(0.011)
Value-added ratio (2001)		0.088	0.059
		(0.094)	(0.094)
Industry age (2001)		0.001	0.000
		(0.001)	(0.001)
Export intensity (2001)			-0.052***
			(0.017)
Observations	521	521	521
R-squared	0.039	0.089	0.099

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, * denote significance at the 1%, 5%, 10% levels.

Appendix C Mathematical Derivation of Firm Profit Maximization

Solving the profit maximization problem, we can obtain the following two first order conditions:

$$p_{sj}q_{sj}(1-\beta_s)\gamma_o = \frac{1}{\rho_s}w_{sjo}\psi_{sjo}l_{sjo},\ \forall o \in M_j \tag{C.1}$$

$$p_{sj}q_{sj}\beta_s = \frac{1}{\rho_s}w^m m_{sj} \tag{C.2}$$

Dividing equation (C.2) by equation (C.1), we can get:

$$m_{sj} = \left(\frac{w_{sjo}\psi_{sjo}}{1-\beta_s}\right)\left(\frac{w^m\gamma_o}{\beta_s}\right)^{-1} l_{sjo} \tag{C.3}$$

We can also aggregate m_{sj} according to a Cobb-Douglas aggregation analogous to equation (14). Together with equation (14), we have:

$$m_{sj} = \prod_{o=1}^{M_j}(m_{sj})^{\gamma_o} = \left(\frac{\psi_{sj}}{1-\beta_s}\right)\left(\frac{w^m\gamma_o}{\beta_s}\right)^{-1}\left[\prod_{o=1}^{M_j}(w_{sjo})^{\gamma_o}\right] l_{sj} \tag{C.4}$$

From equation (8), we obtain the occupation level inverse supply of labor function:

$$w_{sjo} = \left(\frac{l_{sjo}}{L\lambda_{sj}}\right)^{\theta_j} \tag{C.5}$$

Plugging equation (C.4) into equation (C.5), we finally derive the expression for m_{sj} as the following:

$$m_{sj} = \left(\frac{\psi_{sj}}{1-\beta_s}\right)\left(\frac{w^m\gamma_o}{\beta_s}\right)^{-1}\left(\frac{1}{L\lambda_{sj}}\right)^{\theta_j} {l_{sj}}^{\theta_j+1} \tag{C.6}$$

Plug the expression for m_{sj} into the production function of the individual firm, the output can be written as follows:

$$q_{sj} = \varphi_{sj}\left[\left(\frac{\psi_{sj}}{1-\beta_s}\right)\left(\frac{w^m\gamma_o}{\beta_s}\right)^{-1}\left(\frac{1}{L\lambda_{sj}}\right)^{\theta_j}\right]^{\beta_s} {l_{sj}}^{\beta_s\theta_j+1} \tag{C.7}$$

Together with equation (C.5), we can obtain the p_{sj} as a function of l_{sj}, that is:

$$p_{sj} = \frac{1}{\rho_s}\frac{1}{\varphi_{sj}}\left(\frac{w^m}{\beta_s}\right)^{\beta_s}\left(\frac{1}{1-\beta_s}\right)^{1-\beta_s}\left(\frac{\psi_{sjo}}{\gamma_o}\right)^{1-\beta_s}\left(\frac{l_{sj}}{L\lambda_{sj}}\right)^{\theta_j(1-\beta_s)} \tag{C.8}$$

Solving equations (12), (15), (C.7) and (C.8) simultaneously, we can finally express l_{sj} as a function of productivity $varphi_{sj}$ and firm-level wage markdown ψ_{sj}, which is:

$$l_{sj} = \kappa_{sj} \psi_{sj}^{\frac{1-\beta_s \rho_s}{\rho_s - 1 + \theta_j(\beta_s \rho_s - 1)}} \varphi_{sj}^{\frac{-\rho_s}{\rho_s - 1 + \theta_j(\beta_s \rho_s - 1)}} \tag{C.9}$$

Plugging equation (C.9) into equation (C.6), we can obtain equations (19) and (20), which are the expression of the intermediate input and the intermediate input-labor ratio respectively. Solving equation (C.7) and equation (C.9) simultaneously, we can obtain the expression for the firms' output as the function of productivity and wage markdown as well, which is exactly equation (21) in the main text of this paper.

The first order condition of firms profit maximization problem can be rearranged to get the following expression in terms of the marginal product revenue of the production factors:

$$\text{MPRL}_{sj} = (1 - \beta_s)\rho_s \frac{p_{sj} q_{sj}}{l_{sj}} = w_{sj} \psi_{sj} = \Lambda_{sj} \psi_{sj}^{\frac{\rho_s - 1}{\rho_s - 1 + \theta_j(\beta_s \rho_s - 1)}} \varphi_{sj}^{-\frac{\rho_s \theta_j}{\rho_s - 1 + \theta_j(\beta_s \rho_s - 1)}} \tag{C.10}$$

$$\text{MPRM}_{sj} = \beta_s \rho_s \frac{p_{sj} q_{sj}}{m_{sj}} = w^m \tag{C.11}$$

where $\Lambda_{sj} = \frac{1}{\gamma_o} \left(\frac{1}{L \lambda_{sj}} \right)^{\theta_j} \kappa_{sj}^{\theta_j} > 0$. By definition, $\text{TFPR}_{sj} = p_{sj} \text{TFPQ}_{sj} = p_{sj} \varphi_{sj}$. Plugging equations (13), (C.10) and (C.11) into the expression for TFPR_{sj}, we have:

$$\begin{aligned} \text{TFPR}_{sj} =& \frac{1}{\rho_s} \left(\frac{\text{MRPM}_{sj}}{\beta_s} \right)^{\beta_s} \left(\frac{\text{MRPL}_{sj}}{1 - \beta_s} \right)^{1-\beta_s} \propto \text{MRPL}_{sj}^{1-\beta_s} \\ =& \psi_{sj}^{\frac{(\rho_s - 1)(1-\beta_s)}{\rho_s - 1 + \theta_j(\beta_s \rho_s - 1)}} \varphi_{sj}^{-\frac{\rho_s \theta_j (1-\beta_s)}{\rho_s - 1 + \theta_j(\beta_s \rho_s - 1)}} \end{aligned} \tag{C.12}$$

Appendix D Mathematical Derivation of Aggregate TFP

Equations (C.10) and (C.11) can be rearranged to obtain l_{sj} and m_{sj} as the function of MRPL_{sj} and MRPM_{sj} respectively, which are:

$$l_{sj} = \frac{(1-\beta_s)\rho_s p_{sj} q_{sj}}{\text{MRPL}_{sj}} = (1-\beta_s)\rho_s \frac{1}{\text{MRPL}_{sj}} \frac{p_{sj} q_{sj}}{P_s Q_s} P_s Q_s \tag{D.1}$$

$$m_{sj} = \frac{\beta_s \rho_s p_{sj} q_{sj}}{\text{MRPM}_{sj}} = \beta_s \rho_s \frac{1}{\text{MRPM}_{sj}} \frac{p_{sj} q_{sj}}{P_s Q_s} P_s Q_s \tag{D.2}$$

Hence, we can obtain the industry level usage of intermediate inputs and labor as the following:

$$l_s = \sum_{j \in M_s} l_{sj} = (1-\beta_s)\rho_s P_s Q_s / \overline{\text{MRPL}}_s \tag{D.3}$$

$$m_s = \sum_{j \in M_s} m_{sj} = \beta_s \rho_s P_s Q_s / \overline{\text{MRPM}}_s \tag{D.4}$$

where

$$1/\overline{\text{MRPL}}_s = \sum_{j \in M_s} \frac{1}{\text{MRPL}_{sj}} \frac{p_{sj} q_{sj}}{P_s Q_s} \tag{D.5}$$

$$1/\overline{\text{MRPM}}_s = \sum_{j \in M_s} \frac{1}{\text{MRPM}_{sj}} \frac{p_{sj} q_{sj}}{P_s Q_s} \tag{D.6}$$

denote the reciprocal of the weighted average of the value of the marginal revenue product of labor and intermediate inputs within an industry, respectively.

Then we can express aggregate output as a function of l_s, m_s, and industry-level TFP:

$$Q = \prod_{s=1}^{S} {Q_s}^{\alpha_s} = \prod_{s=1}^{S} \left(\text{TFP}_s {m_s}^{\beta_s} {l_s}^{1-\beta_s} \right)^{\alpha_s} \tag{D.7}$$

As a result, we can express the industry level TFP as the following:

$$\text{TFP}_s = \frac{Q_s}{{m_s}^{\beta_s} {l_s}^{1-\beta_s}} \tag{D.8}$$

Plugging equations (D.3) and (D.4) into equation (D.8), we can express industry-level TFP as:

$$\text{TFP}_s = \overline{\text{TFPR}}_s \frac{1}{P_s} \tag{D.9}$$

where

$$\overline{\text{TFPR}}_s = \frac{1}{\rho_s}\left(\frac{\overline{\text{MRPM}}_s}{\beta_s}\right)^{\beta_s}\left(\frac{\overline{\text{MRPL}}_s}{1-\beta_s}\right)^{1-\beta_s} \tag{D.10}$$

is a geometric average of the average marginal revenue product of intermediate input and labor in the industry.

By definition, we have $p_{sj} = \text{TFPR}_{sj}/\text{TFRQ}_{sj} = \text{TFPR}_{sj}/\varphi_{sj}$, together with the expression for the manufacturing industry-level price index; thus, we have

$$\frac{1}{P_s} = \left(\sum_{j\in M_s} p_{sj}^{\frac{\rho_s}{\rho_s-1}}\right)^{\frac{1-\rho_s}{\rho_s}} = \left[\sum_{j\in M_s}\left(\frac{\text{TFPR}_{sj}}{\varphi_{sj}}\right)^{\frac{\rho_s}{\rho_s-1}}\right]^{\frac{1-\rho_s}{\rho_s}} \tag{D.11}$$

Equation (D.9) together with equation (D.11) imply that:

$$\text{TFP}_s = \left[\sum_{j\in M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}}\left(\frac{\overline{\text{TFPR}}_s}{\text{TFPR}_{sj}}\right)^{\frac{\rho_s}{1-\rho_s}}\right]^{\frac{1-\rho_s}{\rho_s}} \tag{D.12}$$

Equations (C.12) and (D.10) imply that:

$$\frac{\overline{\text{TFPR}}_s}{\text{TFPR}_{sj}} = \left(\frac{\overline{\text{MRPM}}_s}{\text{MRPM}_{sj}}\right)^{\beta_s}\left(\frac{\overline{\text{MRPL}}_s}{\text{MRPL}_{sj}}\right)^{1-\beta_s} \tag{D.13}$$

$\text{MRPM}_{sj} = w^m$ does not vary with industry and firm; hence, we can simplify equation (D.6) and obtain that $\overline{\text{MRPM}}_s = \text{MRPM}_{sj} = w^m \equiv 1$.

Following Hsieh & Klenow (2009), we assume that φ_{sj}, ψ_{sj}, and w_{sj} are jointly log-normally distributed, there is a simple closed-form expression for industry-level aggregate TFP:

$$\begin{aligned}\log \text{TFP}_s = \log\left(\sum_{j\in M_s}\varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}}\right)^{\frac{1-\rho_s}{\rho_s}} - \Gamma_{1s}\left[\text{var}\log w_{sj} + \text{var}\log\psi_{sj}\right] \\ - \Gamma_{2s}\,\text{cov}(\log w_{sj}, \log\psi_{sj})\end{aligned} \tag{D.14}$$

where:

$$\Gamma_{1s} = \frac{(\beta_s\rho_s-1)(\beta_s-1)}{2(1-\rho_s)}, \Gamma_{2s} = \frac{(\beta_s\rho_s-1)(\beta_s-1)}{1-\rho_s}$$

If we assume there is the only variation in $\log\psi_{sj}$, equation (D.14) can be further simplified as:

$$\log \text{TFP}_s = \log\left(\sum_{j\in M_s}\varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}}\right)^{\frac{1-\rho_s}{\rho_s}} - \Gamma_{1s}\text{var}\log\psi_{sj} \tag{D.15}$$

In this case, the negative effect of monopsony power on industry level TFP can be summarized as the variance in $\log\psi_{sj}$. In short, labor market markdown dispersion incurs an efficiency loss.

Appendix E Mathematical Derivation of the Logarithm of Aggregate TFP

The problem of firm maximization can be expressed alternatively as the following:

$$\max_{m_{sj}, l_{sj}} p_{sj}(q_{sj})q_{sj} - w_{sj}(l_{sj})l_{sj} - w^m m_{sj} \tag{E.1}$$

where $w_{sj} = \prod_{o=1}^{M_j} \left(\frac{w_{sjo}}{\gamma_o}\right)^{\gamma_o} = \frac{1}{\gamma_o}\left(\frac{1}{L\lambda_{sj}}\right)^{\theta_j} {l_{sj}}^{\theta_j}$. Solving the profit maximization problem, we can obtain the following two first order conditions:

$$\beta_s \varphi_{sj} \left(\frac{m_{sj}}{l_{sj}}\right)^{\beta_s - 1} p_{sj}\rho_s = w^m \tag{E.2}$$

$$(1-\beta_s)\varphi_{sj} \left(\frac{m_{sj}}{l_{sj}}\right)^{\beta_s} p_{sj}\rho_s = w^{sj}\psi_{sj} \tag{E.3}$$

Taking the ratio of equations (E.2) and (E.3), we can express the intermediate-labor ratio as:

$$\frac{m_{sj}}{l_{sj}} = \left(\frac{\beta_s}{1-\beta_s}\right)\left(\frac{w_{sj}\psi_{sj}}{w^m}\right) \tag{E.4}$$

Moreover, we can rearrange the demand function (equation 12) to get the following expression for price:

$$p_{sj} = \left(\frac{q_{sj}}{Q_s}\right)^{\rho_s - 1} P_s \tag{E.5}$$

Plugging the production function (equation 13) into equation (E.5), we can obtain the following two equations:

$$p_{sj} = \left[\frac{\varphi_{sj}(m_{sj}/l_{sj})^{\beta_s} l_{sj}}{Q_s}\right]^{\rho_s - 1} P_s \tag{E.6}$$

$$p_{sj} = \left[\frac{\varphi_{sj}(m_{sj}/l_{sj})^{\beta_s - 1} m_{sj}}{Q_s}\right]^{\rho_s - 1} P_s \tag{E.7}$$

Solving equations (E.3,E.5, and E.7) simultaneously, we can obtain the expression for l_{sj} as the following:

$$l_{sj} = {\varphi_{sj}}^{\frac{\rho_s}{1-\rho_s}} \left(\frac{w_{sj}\psi_{sj}}{1-\beta_s}\right)^{\frac{\beta_s\rho_s - 1}{1-\rho_s}} \left(\frac{w^m}{\beta_s}\right)^{\frac{-\beta_s\rho_s}{1-\rho_s}} (\rho_s P_s)^{\frac{1}{1-\rho_s}} Q_s \tag{E.8}$$

Similarly, solving equations (E.2), (E.4) and (E.6) simultaneously, we can obtain the expression for m_{sj} as the following:

$$m_{sj} = {\varphi_{sj}}^{\frac{\rho_s}{1-\rho_s}} \left(\frac{w_{sj}\psi_{sj}}{1-\beta_s}\right)^{\frac{\beta_s\rho_s - \rho_s}{1-\rho_s}} \left(\frac{w^m}{\beta_s}\right)^{\frac{-(\beta_s\rho_s + 1 - \rho_s)}{1-\rho_s}} (\rho_s P_s)^{\frac{1}{1-\rho_s}} Q_s \tag{E.9}$$

We can further aggregate l_{sj} and m_{sj} across different firms within the same manufacturing industry to obtain the industry-level intermediate inputs and labor:

$$l_s = \sum_{j=1}^{M_s} l_{sj} = (\rho_s P_s)^{\frac{1}{1-\rho_s}} Q_s \left(\frac{w^m}{\beta_s}\right)^{\frac{-\beta_s\rho_s}{1-\rho_s}} \sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} \left(\frac{w_{sj}\psi_{sj}}{1-\beta_s}\right)^{\frac{\beta_s\rho_s-1}{1-\rho_s}} \tag{E.10}$$

$$m_s = \sum_{j=1}^{M_s} m_{sj} = (\rho_s P_s)^{\frac{1}{1-\rho_s}} Q_s \left(\frac{w^m}{\beta_s}\right)^{\frac{-(\beta_s\rho_s+1-\rho_s)}{1-\rho_s}} \sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} \left(\frac{w_{sj}\psi_{sj}}{1-\beta_s}\right)^{\frac{\beta_s\rho_s-\rho_s}{1-\rho_s}} \tag{E.11}$$

Using the definition of TFP_s, we have:

$$\text{TFP}_s = \frac{Q_s}{m_s{}^{\beta_s} l_s{}^{1-\beta_s}} = \left(\frac{Q_s}{m_s}\right)^{\beta_s} \left(\frac{Q_s}{l_s}\right)^{1-\beta_s} \tag{E.12}$$

Solving equations (E.10), (E.11), and (E.12), we can further express TFP_s as:

$$\text{TFP}_s = \frac{\left[\frac{1}{\rho_s P_s}\left(\frac{w^m}{\beta_s}\right)^{\beta_s}\left(\frac{1}{1-\beta_s}\right)^{1-\beta_s}\right]^{\frac{1}{1-\rho_s}}}{\left[\sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s-1}{1-\rho_s}}\right]^{1-\beta_s} \left[\sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s-\rho_s}{1-\rho_s}}\right]^{\beta_s}} \tag{E.13}$$

We can also get the expression for q_{sj} as a function of φ_{sj}, ϕ_{sj} by integrating equations (13) and (E.4):

$$q_{sj} = \varphi_{sj}{}^{\frac{1}{1-\rho_s}} \left(\frac{w_{sj}\psi_{sj}}{1-\beta_s}\right)^{\frac{\beta_s-1}{1-\rho_s}} \left(\frac{w^m}{\beta_s}\right)^{\frac{-\beta_s}{1-\rho_s}} (\rho_s P_s)^{\frac{1}{1-\rho_s}} Q_s \tag{E.14}$$

Plug equation (E.14) into equation (E.11), we can obtain the industry-level output as:

$$Q_s = (\rho_s P_s)^{\frac{1}{1-\rho_s}} Q_s \left(\frac{w^m}{\beta_s}\right)^{\frac{-\beta_s}{1-\rho_s}} \left(\frac{1}{1-\beta_s}\right)^{\frac{\beta_s-1}{1-\rho_s}} \left\{\sum_{j=1}^{M_s} \left[\varphi_{sj}{}^{\frac{1}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s-1}{1-\rho_s}}\right]^{\rho_s}\right\}^{\frac{1}{\rho_s}} \tag{E.15}$$

Rearranging equation (E.15), we have the following expression:

$$\left[\frac{1}{\rho_s P_s}\left(\frac{w^m}{\beta_s}\right)^{\beta_s}\left(\frac{1}{1-\beta_s}\right)^{1-\beta_s}\right]^{\frac{1}{1-\rho_s}} = \left\{\sum_{j=1}^{M_s} \left[\varphi_{sj}{}^{\frac{1}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s-1}{1-\rho_s}}\right]^{\rho_s}\right\}^{\frac{1}{\rho_s}} \tag{E.16}$$

The left hand side of equation (E.16) is exactly the numerator of TFP_s in equation (E.13). Finally, we can obtain the expression for industry-level aggregate TFP as:

$$\mathrm{TFP}_s = \frac{\left\{ \sum_{j=1}^{M_s} \left[\varphi_{sj} (w_{sj}\psi_{sj})^{\beta_s - 1} \right]^{\frac{\rho_s}{1-\rho_s}} \right\}^{\frac{1}{\rho_s}}}{\left[\sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s - 1}{1-\rho_s}} \right]^{1-\beta_s} \left[\sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s - \rho_s}{1-\rho_s}} \right]^{\beta_s}} \tag{E.17}$$

Taking the log of equation (E.17), we have:

$$\begin{aligned} \log \mathrm{TFP}_s =& \frac{1}{\rho_s} \log \left\{ \sum_{j=1}^{M_s} \left[\varphi_{sj} (w_{sj}\psi_{sj})^{\beta_s - 1} \right]^{\frac{\rho_s}{1-\rho_s}} \right\} - \beta_s \log \left[\sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s - \rho_s}{1-\rho_s}} \right] \\ & (1-\beta_s) \log \left[\sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s - 1}{1-\rho_s}} \right] \end{aligned} \tag{E.18}$$

Following Hsieh & Klenow (2009), we assume that w_{sj}, φ_{sj} and ψ_{sj} jointly obey a log-normal distribution. Hence, we can simplify the composition term of $\log \mathrm{TFP}_s$ one by one, that is:

$$\begin{aligned} \log & \left\{ \sum_{j=1}^{M_s} \left[\varphi_{sj} (w_{sj}\psi_{sj})^{\beta_s - 1} \right]^{\frac{\rho_s}{1-\rho_s}} \right\} = \log M_s + \frac{\rho_s}{1-\rho_s} \Big[E(\log \varphi_{sj}) + (\beta_s - 1) E(\log w_{sj}) \\ & + (\beta_s - 1) E(\log \psi_{sj}) \Big] + \frac{1}{2} \left(\frac{\rho_s}{1-\rho_s} \right)^2 \operatorname{var} \log \varphi_{sj} + \frac{1}{2} \left[\left(\frac{\rho_s}{1-\rho_s} \right) (1-\beta_s) \right]^2 \operatorname{var} \log w_{sj} \\ & + \frac{1}{2} \left[\left(\frac{\rho_s}{1-\rho_s} \right) (1-\beta_s) \right]^2 \operatorname{var} \log \psi_{sj} + \left(\frac{\rho_s}{1-\rho_s} \right)^2 (\beta_s - 1) \operatorname{cov}(\log \varphi_{sj}, \log w_{sj}) \\ & + \left(\frac{\rho_s}{1-\rho_s} \right)^2 (\beta_s - 1) \operatorname{cov}(\log \varphi_{sj}, \log \psi_{sj}) + \left[\left(\frac{\rho_s}{1-\rho_s} \right) (1-\beta_s) \right]^2 \operatorname{cov}(\log \psi_{sj}, \log w_{sj}) \end{aligned} \tag{E.19}$$

$$\begin{aligned} \log & \left[\sum_{j=1}^{M_s} \varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}} (w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s - \rho_s}{1-\rho_s}} \right] = \log M_s + \Big[\frac{\rho_s}{1-\rho_s} E(\log \varphi_{sj}) + \frac{\beta_s\rho_s - \rho_s}{1-\rho_s} E(\log w_{sj}) \\ & + \frac{\beta_s\rho_s - \rho_s}{1-\rho_s} E(\log \psi_{sj}) \Big] + \frac{1}{2} \left(\frac{\rho_s}{1-\rho_s} \right)^2 \operatorname{var} \log \varphi_{sj} + \frac{1}{2} \left(\frac{\beta_s\rho_s - \rho_s}{1-\rho_s} \right)^2 \operatorname{var} \log w_{sj} \\ & + \frac{1}{2} \left(\frac{\beta_s\rho_s - \rho_s}{1-\rho_s} \right)^2 \operatorname{var} \log \psi_{sj} + \left(\frac{\rho_s}{1-\rho_s} \right) \left(\frac{\beta_s\rho_s - \rho_s}{1-\rho_s} \right) \operatorname{cov}(\log \varphi_{sj}, \log w_{sj}) \\ & + \left(\frac{\rho_s}{1-\rho_s} \right) \left(\frac{\beta_s\rho_s - \rho_s}{1-\rho_s} \right) \operatorname{cov}(\log \varphi_{sj}, \log \psi_{sj}) + \left(\frac{\beta_s\rho_s - \rho_s}{1-\rho_s} \right)^2 \operatorname{cov}(\log \psi_{sj}, \log w_{sj}) \end{aligned} \tag{E.20}$$

$$\begin{aligned}
\log\Big[\sum_{j=1}^{M_s}\varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}}(w_{sj}\psi_{sj})^{\frac{\beta_s\rho_s-1}{1-\rho_s}}\Big] &= \log M_s + \Big[\frac{\rho_s}{1-\rho_s}E(\log\varphi_{sj}) + \frac{\beta_s\rho_s-1}{1-\rho_s}E(\log w_{sj}) \\
&+ \frac{\beta_s\rho_s-1}{1-\rho_s}E(\log\psi_{sj})\Big] + \frac{1}{2}\Big(\frac{\rho_s}{1-\rho_s}\Big)^2\text{var}\log\varphi_{sj} + \frac{1}{2}\Big(\frac{\beta_s\rho_s-1}{1-\rho_s}\Big)^2\text{var}\log w_{sj} \\
&+ \frac{1}{2}\Big(\frac{\beta_s\rho_s-1}{1-\rho_s}\Big)^2\text{var}\log\psi_{sj} + \Big(\frac{\rho_s}{1-\rho_s}\Big)\Big(\frac{\beta_s\rho_s-1}{1-\rho_s}\Big)\text{cov}(\log\varphi_{sj},\log w_{sj}) \\
&+ \Big(\frac{\rho_s}{1-\rho_s}\Big)\Big(\frac{\beta_s\rho_s-1}{1-\rho_s}\Big)\text{cov}(\log\varphi_{sj},\log\psi_{sj}) + \Big(\frac{\beta_s\rho_s-1}{1-\rho_s}\Big)^2\text{cov}(\log\psi_{sj},\log w_{sj})
\end{aligned}\tag{E.21}$$

Plugging equations (E.19), (E.20), and (E.21) back into equation (E.18), we can obtain:

$$\begin{aligned}
\log\text{TFP}_s = \log\Big(\sum_{j\in M_s}\varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}}\Big)^{\frac{1-\rho_s}{\rho_s}} &- \frac{(\beta_s\rho_s-1)(\beta_s-1)}{2(1-\rho_s)}[\text{var}\log w_{sj} + \text{var}\log\psi_{sj}] \\
&- \frac{(\beta_s\rho_s-1)(\beta_s-1)}{1-\rho_s}\text{cov}(\log w_{sj},\log\psi_{sj})
\end{aligned}\tag{E.22}$$

which is exactly the same as equation (D.14).

If we assume that there is only variation in $\log\psi_{sj}$, equation (E.22) can be further simplified as equation (E.23), which is exactly equation (28):

$$\log\text{TFP}_s = \log\Big(\sum_{j\in M_s}\varphi_{sj}{}^{\frac{\rho_s}{1-\rho_s}}\Big)^{\frac{1-\rho_s}{\rho_s}} - \Gamma_{1s}\text{var}\log\psi_{sj}\tag{E.23}$$

Appendix F Applicability of the Brooks et al. (2021a;b) Algorithm to Our Model

The first order condition of the firm profit maximization problem can be expressed alternatively as the following:

$$\beta_s \varphi_{sj} \left(\frac{m_{sj}}{l_{sj}} \right)^{\beta_s - 1} p_{sj} \rho_s = w^m \tag{F.1}$$

$$(1 - \beta_s) \varphi_{sj} \left(\frac{m_{sj}}{l_{sj}} \right)^{\beta_s} p_{sj} \rho_s = w^{sj} \psi_{sj} \tag{F.2}$$

We can divide equation (F.2) by equation (F.1) to obtain:

$$\psi_{sj} = \frac{\frac{1-\beta_s}{w_{sj} l_{sj} / p_{sj} q_{sj}}}{\frac{\beta_s}{w^m m_{sj} / p_{sj} q_{sj}}} = \frac{\mu_L^{\text{DLW}}}{\mu_M^{\text{DLW}}} \tag{F.3}$$

Equation (F.3) indicates that, we can use Brooks et al. (2021a;b) method to estimate firms' monopsony power in the labor market.