Local Concentration, National Concentration, and the Spatial Distribution of Markups*

Jonathan Becker† Chris Edmond Virgiliu Midrigan‡

Daniel Yi Xu§

Preliminary and Incomplete

This Draft: August 2024

Abstract

We study the spatial distribution of production and consumption in a quantitative model with multi-establishment firms, oligopolistic competition, and endogenously variable markups. We calibrate our model to match US Census of Manufactures firm and establishment data and intranational trade flows from the Commodity Flows Survey. We show that spatial frictions can have large aggregate effects, increasing both the aggregate markup and the productivity losses due to misallocation. We then show that a reduction in intranational trade costs, calibrated to match long-run trends in US manufacturing, will increase national sales concentration but decrease local sales concentration. Local markets become more competitive, markups fall, and aggregate productivity rises, despite the increase in national concentration.

Keywords: competition, misallocation, multi-establishment firms, trade flows, gravity.

JEL classifications: D4, E2, F1, L1, O4.

*We particularly thank our discussants Mayara Felix and Jake Zhao for their detailed comments and thank audience participants at the 2023 FRB Dallas conference on international economics, the 2023 OzMac workshop, and the 2024 PHBS Shenzen workshop on macroeconomics and finance.

†New York University, jb7026@nyu.edu.

‡New York University and National Bureau of Economic Research, virgiliu.midrigan@nyu.edu.

§Duke University and National Bureau of Economic Research, daniel.xu@duke.edu.

1 Introduction

Does rising national concentration mean large firms have more market power? Do technological changes that benefit large firms mean goods markets are becoming less competitive? We answer these questions using a quantitative spatial model with oligopolistic competition and endogenously variable markups. In our model, markups are determined by the amount of competition firms face in the locations where people live and consume. Technological changes that allow large firms to service more markets can generate a pattern of simultaneously rising national concentration and falling local concentration. Because of this, such technological changes can increase local competition, reduce markups, and increase aggregate productivity even while national concentration is rising.

Our model features many geographically segmented locations and heterogeneous firms that can, in general, source goods from multiple locations and sell in multiple locations. Firms compete *oligopolistically* in their destination markets. Firms are heterogeneous both in terms of their overall productivity and in terms of the number and geographic location of their establishments. Workers are geographically immobile. Taking wages in each location as given, each firm chooses an optimal, comprehensive production plan for their set of establishments that determines that firm's effective marginal cost of producing for each possible destination market. Given these firm-and-destination-specific marginal costs, oligopolistic competition in the spirt of Atkeson and Burstein (2008) then determines the markups each firm charges in each of its destination markets. In equilibrium, wages in each location are determined by local labor market clearing conditions.

We calibrate our model to match the operations of some 270,000 US manufacturing firms organised into 364 6-digit NAICS sectors, with an average of some 730 firms per sector. We take our geographical locations to be US states and Washington DC. While most firms are small and have only one establishment, larger firms tend to have multiple-establishments and multiple-establishment firms produce in a broad range of locations. We take the number of firms in each sector straight from the US Census of Manufactures. We choose the parameters of our model governing the distribution of productivity across firms and the number and locations of establishments to match key facts on national sales concentration and the distribution of firm establishments across states. We parameterize sector-specific iceberg trade costs so that the model reproduces sector-specific gravity regressions using state-to-state trade flows from the Commodity Flows Survey for 3-digit NAICS manufacturing sectors.

Our model is relatively parsimonious: there are three elasticities of substitution — across establishments, across firms, and across sectors — plus three parameters that control the distribution of productivity and the number and location of establishments, and, for each 3-digit NAICS manufacturing sector, there is one additional parameter that controls sector-specific trade costs. Despite its parsimony, the model does a good job of matching not just national concentration but also matching the diversity of multi-establishment operations across states. The model also matches, essentially perfectly, the sector-specific state-to-state gravity effects that we measure using the Commodity Flows Survey.

The fact that we match these sector-specific gravity effects is important. These gravity effects determine the quantitative significance of the spatial frictions in each sector, i.e., they determine which sectors produce goods that are intrinsically less tradable and, within a given sector, which locations are more central and which are more remote. In equilibrium, these spatial frictions play a crucial role in determining both the spatial distribution of *production* and the spatial distribution of *consumption*.

The spatial distribution of consumption is key because it's how much competition firms face in their destination markets that determines how much market power firms really have. Regardless of how concentrated production is, if firms are shipping goods to destination markets where they have to compete with many rival firms, markups will be lower and consumers will be better off. In other words, if we are interested in how much market power firms really have, we need to know how concentrated these local destination markets are.

But local sales concentration can not be directly observed in the Census of Manufactures. One of our contributions is a set of model-based measurements of local sales concentration. That is, we can use our model, which is calibrated to match national sales concentration and local production concentration, to draw inferences about local sales concentration. Intuitively, we find that local sales concentration is higher than national sales concentration but not as high as local production concentration. For example, in our benchmark model the local sales Herfindahl-Hirschman Index (HHI) is, on average, about 0.20, higher than the national sales HHI of 0.10 but lower than the local production HHI of 0.36 reported by Autor, Patterson and Van Reenen (2023). Reassuringly, the ordering of concentration implied by our model is also consistent with the findings of Benkard, Yurukoglu and Zhang (2023) who study concentration in very finely disaggregated consumer survey data.

More generally, we find that production concentration is larger and more dispersed than sales concentration and that this effect is considerably more pronounced in sectors characterized by low spatial trade frictions. Sectors with weak gravity effects, that produce more easily traded goods — such as computer and electronics manufacturing — feature weaker spatial correlation between production and consumption. Sectors with stronger gravity effects — such as wood, petroleum and coal manufacturing — feature stronger spatial correlation between production and consumption.

We find that these spatial frictions are quantitatively significant determinants of the economy-wide, macroeconomic losses due to market power. Our benchmark model with spatial frictions implies an aggregate, economy-wide markup of 1.31 with sector-level markups ranging from 1.62 at the 99th percentile of sectors down to 1.17 at the 1st percentile. If we calibrate the model to match the same national sales concentration but abstract from geography and spatial frictions we find a much lower aggregate markup of 1.18 and much less markup dispersion, and hence lower productivity losses due to misallocation, with sector-level markups ranging from 1.40 at the 99th percentile to 1.12 at the 1st percentile. In this sense, abstracting from geography and spatial frictions leads to a quantitatively significant *understatement* of the macroeconomic losses associated with market power.

We then show that technological changes that allow firms to service more markets can generate a pattern of simultaneously rising national concentration and falling local concentration. Specifically, we consider an exogenous 20% reduction in trade costs, chosen to match the findings of Coşar, Osotimehin and Popov (2022), who find a long-run decrease of 15-20% for US manufacturing over the years 1963-2017. This change in trade costs, largely due to technological improvements in transportation services, makes it easier for all firms to service more markets. And one might suspect that this trend of improvement in transportation particularly benefits the largest, most productive firms who may then be well-positioned to sell to even more locations than they did previously. Consistent with this, we find that national sales concentration does rise. Our model predicts that a 20% reduction in trade costs leads the top-4 national sales share to increase from 0.46 in our benchmark model to 0.49.

But this 20% reduction in trade costs also leads to a reduction in local sales concentration, with the top-4 local sales share decreasing from 0.69 in our benchmark to 0.66. The increase in national concentration masks the fact that local markets are becoming more competitive. This increase in competition leads the aggregate, economy-wide markup to fall from 1.31 to 1.29 and leads to lower misallocation from markup dispersion.[1] In this sense, rising national concentration provides a misleading guide to changes in market power. In this scenario, national sales concentration is rising, as is local production concentration, but markets are becoming more competitive, markups are falling, and aggregate productivity is rising.

[1]While this change in the aggregate markup seems small, it is in fact a surprisingly large change for this class of models. As discussed at length in Edmond, Midrigan and Xu (2023), for this class of models, composition effects mean that even large changes in the number of competitors within a market generate tiny effects on the aggregate markup. In their benchmark model, a *tripling* of the number of competitors leads to an insignificant third-decimal place change in the aggregate markup. See also Bernard, Eaton, Jensen and Kortum (2003) and Arkolakis, Costinot, Donaldson and Rodríguez-Clare (2019) for further examples in an international trade context.

Trends in concentration. This paper contributes to the extensive literature on the causes and consequences of the rise in concentration in the US since the early 1980s, following Grullon, Larkin and Michaely (2019), Autor, Dorn, Katz, Patterson and Van Reenen (2020), Amiti and Heise (2021), Ganapati (2021), and many others.

Diverging trends in national and local concentration? It is widely agreed that changes in national concentration may provide a misleading guide to changes in market power and that local concentration may provide a better guide. In an influential paper using National Establishment Time Series (NETS) data, Rossi-Hansberg, Sarte and Trachter (2020) argue that local concentration has been declining even while national concentration has risen. As discussed by Decker (2020), NETS data suffers from issues with coverage and accuracy. Further work using the US Census of Retail Trade by Smith and Ocampo (2024) argues that *both* national and local sales concentration have been rising since the early 1990s. Similarly, using the US Economic Census more broadly, Autor, Patterson and Van Reenen (2023) find that local sales concentration has risen but that local employment concentration has fallen. But, as argued by Benkard, Yurukoglu and Zhang (2023), measures of concentration using Census data focus on the classification of economic activity by *production*, not by *consumption* and it is the availability of good substitutes for consumers that ultimately determines how much market power producers have. Benkard et al. find decreasing local sales concentration in finely disaggregated consumer survey data. Neiman and Vavra (2023) report a similar decrease in sales concentration which they interpret as arising due to increasingly 'niche' consumption patterns.

Multi-establishment production. This paper also contributes to the recent literature on the increasing importance of multi-establishment firms, following Jia (2008), Holmes (2011), Basker, Klimek and Van (2012), Foster, Haltiwanger, Klimek, Krizan and Ohlmacher (2016), Cao, Hyatt, Mukoyama and Sager (2022), and many others. While this literature originally focused on retail trade, this phenomenon has become increasingly important for services too, as in Hsieh and Rossi-Hansberg (2023).

Spatial misallocation. Our work is closely related to two recent papers on the spatial distribution of markups. Like us, Asturias, García-Santana and Ramos (2019) develop an Atkeson and Burstein (2008) model with many locations — which they use to assess the importance of improvements in transportation infrastructure in India — but unlike us they abstract from multi-establishment firms and do not develop the implications of their model for trends in local concentration. Similarly, Franco (2023) studies spatial misallocation across

US cities in a model of monopolistic competition with Kimball demand. He emphasizes the endogenous sorting of firms across locations, which we abstract from, but does not consider the implications of multi-establishment firms for the spatial distribution of market power.

2 Model

The economy consists of many heterogeneous locations. Across the economy there are many heterogeneous firms that, in general, can source goods from multiple locations and sell in multiple locations. Firms compete *oligopolistically* in their destination markets. The economy is *geographically segmented* in two ways: (i) labor is immobile across locations, with location-specific wages pinned down by local labor market clearing conditions, and (ii) goods shipments are subject to iceberg trade costs.

2.1 Environment

There are J locations indexed by $j, k = 1, ..., J$. There is a continuum of sectors indexed by $s \in [0, 1]$. Within each sector there is a finite $n(s)$ firms indexed by $i = 1, ..., n(s)$ that compete oligopolistically in their destination markets. Trade in goods is subject to sector-specific iceberg trade costs $\tau_{jk}(s) \geq 1$ with $\tau_{jj}(s) = 1$. A notational convention that we maintain throughout is that location j refers to the *source* of a good and location k refers to a *destination* so that $\tau_{jk}(s)$, say, refers to the sector-specific cost of shipping from j to k.

Location-specific final good. In each destination market k there is a non-tradeable final good produced under perfectly competitive conditions. This location-specific final good is given by a CES aggregate *across sectors*

$$C_k = \left(\int_0^1 C_k(s)^{\frac{\theta-1}{\theta}} \, ds \right)^{\frac{\theta}{\theta-1}}, \qquad \theta > 1 \tag{1}$$

Then *within sectors*, output is given by a CES aggregate across the $n(s)$ firms in sector s

$$C_k(s) = \left(\sum_{i=1}^{n(s)} C_{ik}(s)^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}}, \qquad \gamma > \theta \tag{2}$$

We assume $\gamma > \theta$ so that goods are more substitutable within sectors than across sectors.

Firms source from establishments in multiple locations. Each firm i selling in destination market k sources goods from establishments in multiple locations $j = 1, ..., J$. Within firm i, output is given by a CES aggregate *across establishments*

$$C_{ik}(s) = \left(\sum_{j=1}^{J} c_{ijk}(s)^{\frac{\lambda-1}{\lambda}} \right)^{\frac{\lambda}{\lambda-1}}, \qquad \lambda \geq \gamma \tag{3}$$

Two special cases are worth noting: (i) the case $\lambda = \gamma$ where goods are equally substitutable within and across firms (within a given sector), and (ii) the case $\lambda = +\infty$ where a given firm will source all of its output from its least-cost establishment.

Location-specific representative consumer. Each location j is populated by L_j identical workers each endowed with e_j efficiency units of labor. Labor is immobile across locations. Each worker inelastically supplies their e_j units of labor to the local labor market and receives location-specific wage w_j. Firm ownership is perfectly diversified across locations with profits $\bar{\pi} L_j$ paid out in location j, i.e., with constant profits per worker $\bar{\pi}$. Aggregating the budget constraints of workers in location j gives

$$P_j C_j = \left(w_j e_j + \bar{\pi} \right) L_j \tag{4}$$

Demand system. This nested-CES setup implies that the demand for goods sourced from location j to be sold at destination k by firm i in sector s is given by

$$c_{ijk}(s) = \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{-\lambda} \underbrace{\left(\frac{P_{ik}(s)}{P_k(s)} \right)^{-\gamma} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k}_{=C_{ik}(s)} \tag{5}$$

As usual, the location-specific final good and sector-level price indexes are given by

$$P_k = \left(\int_0^1 P_k(s)^{1-\theta}\, ds \right)^{\frac{1}{1-\theta}}, \qquad P_k(s) = \left(\sum_{i=1}^{n(s)} P_{ik}(s)^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \tag{6}$$

But in this setup there is now also an index for aggregating prices across establishments within each firm. This firm-level price index is given by

$$P_{ik}(s) = \left(\sum_{j=1}^{J} \left(\tau_{jk}(s) p_{ijk}(s) \right)^{1-\lambda} \right)^{\frac{1}{1-\lambda}} \tag{7}$$

We implicitly set $p_{ijk}(s) = +\infty$ for any firm i that does not sell in location k.

Production. Firm i in sector s is endowed with productivity $z_{ij}(s) \geq 0$ for goods sourced from location j. We assume that goods sourced from location j but sold in destination k require labor input from both places. Let $l_{ijk}(s)$ denote labor used to produce goods in location j that are shipped to k and let $m_{ijk}(s)$ denote labor used in destination k to process goods that are shipped from j. The flow of output $y_{ijk}(s)$ from source j to destination k is given by the Cobb-Douglas technology

$$y_{ijk}(s) = z_{ij}(s)\, l_{ijk}(s)^{\alpha}\, m_{ijk}(s)^{1-\alpha} \qquad\qquad 0 < \alpha \leq 1 \tag{8}$$

We think of $m_{ijk}(s)$ as representing marketing, distribution, and other administrative services required to make goods sourced from j viable products in destination k. If $\alpha = 1$, this reduces to the usual setup where variable inputs are required only in the source location.

Given the sector-specific iceberg trade costs $\tau_{jk}(s) \geq 1$, the resource constraints on the flow of output from j to k are simply

$$y_{ijk}(s) = \tau_{jk}(s)\, c_{ijk}(s) \tag{9}$$

Labor demand. Firms take wages in each location as given. Given the Cobb-Douglas technology (8), the labor demands in source j and destination k are given by

$$w_j l_{ijk}(s) = \alpha \frac{W_{jk}}{z_{ij}(s)}\, y_{ijk}(s) \tag{10}$$

and

$$w_k m_{ijk}(s) = (1-\alpha) \frac{W_{jk}}{z_{ij}(s)}\, y_{ijk}(s) \tag{11}$$

where W_{jk} denotes the wage index implied by the Cobb-Douglas technology

$$W_{jk} = \left(\frac{w_j}{\alpha}\right)^{\alpha} \left(\frac{w_k}{1-\alpha}\right)^{1-\alpha} \tag{12}$$

Marginal cost. Hence a firm can source goods from j for destination k at marginal cost

$$\frac{W_{jk}}{z_{ij}(s)} \tag{13}$$

Profits. Since a firm can supply destination k with goods sourced from establishments at any location j, the firm's profits from sales at k are given by

$$\Pi_{ik}(s) = \sum_{j=1}^{J} \left(p_{ijk}(s) - \frac{W_{jk}}{z_{ij}(s)}\right) y_{ijk}(s) \tag{14}$$

A firm's total profits are then given by $\Pi_i(s) = \sum_{k=1}^{J} \Pi_{ik}(s)$. This objective is separable across destinations k and hence the firm maximizes total profits by maximizing profits in each destination k separately.

We characterize the firm's profit maximizing strategy in each destination k in two steps: (i) taking as given the firm's composite price for its destination market, $P_{ik}(s)$, we determine the least-cost way of servicing that destination with one unit of the firm's composite good, $C_{ik}(s) = 1$, then (ii) we characterize how the firm's price $P_{ik}(s)$ is determined through oligopolistic competition with the other firms servicing destination k. The first step implicitly gives us a characterization of the allocation of production across locations within a given firm. Given the first step, the second step is a nested-CES oligopoly problem familiar from Atkeson and Burstein (2008) and Edmond, Midrigan and Xu (2015).

Within-firm allocation. Taking as given $P_{ik}(s)$ and $C_{ik}(s) = 1$, for step (i) firm i chooses prices $p_{ijk}(s)$ for $j = 1, ..., J$ to minimize the total cost of servicing destination k

$$\sum_{j=1}^{J} \frac{W_{jk}}{z_{ij}(s)} y_{ijk}(s) = \sum_{j=1}^{J} \frac{\tau_{jk}(s) W_{jk}}{z_{ij}(s)} \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{-\lambda} \underbrace{C_{ik}(s)}_{=1} \tag{15}$$

subject to the firm-level price index (7). The Lagrangian for this problem can be written

$$\mathcal{L} = \sum_{j=1}^{J} \frac{\tau_{jk}(s) W_{jk}}{z_{ij}(s)} \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{-\lambda} - \xi_{ik}(s) \sum_{j=1}^{J} \left(\left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{1-\lambda} - 1 \right) \tag{16}$$

where $\xi_{ik}(s) \geq 0$ denotes the multiplier on the firm's constraint. The first order conditions for interior solutions simplify to

$$\lambda \frac{\tau_{jk}(s) W_{jk}}{z_{ij}(s)} = (\lambda - 1) \xi_{ik}(s) \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right) \tag{17}$$

Rearranging this we see that, at the optimum, source prices satisfy

$$p_{ijk}(s) = \mu_{ik}(s) \frac{W_{jk}}{z_{ij}(s)}, \qquad \mu_{ik}(s) = \frac{\lambda}{\lambda - 1} \left(\frac{P_{ik}(s)}{\xi_{ik}(s)} \right) \tag{18}$$

Hence the least-cost way to service destination k is to set a *destination-specific* markup $\mu_{ik}(s)$ that applies uniformly regardless of the source location j. The firm 'prices to market' in a way that reflects the demand and competitive conditions specific to market k. But by making the markup independent of j the firm *avoids distorting allocations within the firm.*

Plugging this expression for the source prices $p_{ijk}(s)$ back into the firm-level price index (7) and eliminating the multiplier gives

$$P_{ik}(s) = \mu_{ik}(s) \Phi_{ik}(s) \tag{19}$$

where

$$\Phi_{ik}(s) = \left(\sum_{j=1}^{J} \phi_{ijk}(s)^{1-\lambda} \right)^{\frac{1}{1-\lambda}}, \qquad \phi_{ijk}(s) := \frac{\tau_{jk}(s) W_{jk}}{z_{ij}(s)} \tag{20}$$

denotes the firm's marginal cost of servicing destination k with one unit of the composite good $C_{ik}(s)$. With this characterization of the within-firm allocation in hand, we can now turn to the strategic interactions between firms in each destination k.

Oligopolistic competition. For step (ii) we then need to characterize how the firm's price $P_{ik}(s)$ is determined through oligopolistic competition with the other firms servicing destination k. Given the within-firm allocation we can use (5), (14) and (19) to write the firm's profits from destination k

$$\begin{aligned} \Pi_{ik}(s) &= (P_{ik}(s) - \Phi_{ik}(s))\; C_{ik}(s) \\ &= (P_{ik}(s) - \Phi_{ik}(s)) \left(\frac{P_{ik}(s)}{P_k(s)} \right)^{-\gamma} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k \end{aligned} \tag{21}$$

with each firm internalizing the effect of their price $P_{ik}(s)$ on the sector-level price index $P_k(s)$ in (6). Given our characterization of the within-firm allocation in the first step, this second step is a standard nested-CES oligopoly problem familiar from Atkeson and Burstein (2008) and Edmond, Midrigan and Xu (2015).

As is well known, this implies that each firm sets a markup of the form

$$\mu_{ik}(s) = \frac{\varepsilon_{ik}(s)}{\varepsilon_{ik}(s) - 1} \tag{22}$$

where the demand elasticity $\varepsilon_{ik}(s)$ facing firm i is endogenous to the firm's sales share in destination k. For our benchmark model we assume that each destination market is characterized by *Cournot competition*. With this specification, the demand elasticity works out to be a sales-weighted harmonic average of the elasticities of substitution within and across sectors

$$\varepsilon_{ik}(s) = \left(\omega_{ik}(s) \frac{1}{\theta} + (1 - \omega_{ik}(s)) \frac{1}{\gamma} \right)^{-1} \tag{23}$$

where $\omega_{ik}(s)$ denotes the market share of firm i in destination market k

$$\omega_{ik}(s) := \frac{P_{ik}(s) C_{ik}(s)}{\sum_{i=1}^{n(s)} P_{ik}(s) C_{ik}(s)} = \frac{P_{ik}(s)^{1-\gamma}}{\sum_{i=1}^{n(s)} P_{ik}(s)^{1-\gamma}} \tag{24}$$

Since the elasticity of substitution across firms is larger than the elasticity of substitution across sectors, $\gamma > \theta$, the demand elasticity $\varepsilon_{ik}(s)$ facing a firm is lower for firms with larger market shares in destination k. Intuitively, firms that are small within a given market are mostly competing with other firms within the same sector and so face a relatively high demand elasticity, approaching the within-sector elasticity γ as $\omega_{ik}(s) \to 0$. At the other extreme, firms that are large within a given market are mostly competing with firms in other sectors and so face a relatively low demand elasticity, approaching the across-sector elasticity θ as $\omega_{ik}(s) \to 1$.

While intuitive, this discussion is incomplete. It simply takes market shares $\omega_{ik}(s)$ as exogenous and traces out the implications of those market shares for markups $\mu_{ik}(s)$. But in this model, markups and market shares are *jointly determined* as part of a larger fixed-point problem. To solve this problem, it turns out to be convenient to first combine (22) and (23) to write the inverse markup as a linear function of the sales share

$$\frac{1}{\mu_{ik}(s)} = 1 - \frac{1}{\varepsilon_{ik}(s)} = \frac{\gamma - 1}{\gamma} - \left(\frac{1}{\theta} - \frac{1}{\gamma}\right)\omega_{ik}(s) \tag{25}$$

From which we see that indeed a firm's markup is *strictly increasing* in its market share.

To obtain the second condition we need, we substitute prices $P_{ik}(s) = \mu_{ik}(s)\Phi_{ik}(s)$ into (24) to get

$$\omega_{ik}(s) = \frac{(\mu_{ik}(s)\Phi_{ik}(s))^{1-\gamma}}{\sum_{i=1}^{n(s)}(\mu_{ik}(s)\Phi_{ik}(s))^{1-\gamma}} \tag{26}$$

Here we see that, conditional on other firms' markups, each firm's market share is *strictly decreasing* in its markup. Together, equations (25) and (26) are two equations in two unknowns that jointly determine the markups $\mu_{ik}(s)$ and market shares $\omega_{ik}(s)$ for each i, k and s. Notice that the interactions between firms within a given market enter only through the denominator in (26) and that market shares are homogenous of degree zero in the markups.

Eliminating the market shares between these we have a single fixed point condition

$$\frac{1}{\mu_{ik}(s)} = \frac{\gamma - 1}{\gamma} - \left(\frac{1}{\theta} - \frac{1}{\gamma}\right)\frac{(\mu_{ik}(s)\Phi_{ik}(s))^{1-\gamma}}{\sum_{i=1}^{n(s)}(\mu_{ik}(s)\Phi_{ik}(s))^{1-\gamma}} \tag{27}$$

This condition implicitly determines the distribution of markups $\mu_{ik}(s)$ within and across locations as a function of the distribution of marginal costs $\Phi_{ik}(s)$ within and across locations. The marginal costs $\Phi_{ik}(s)$ are exogenous to each firm but, because they depend on the wage indexes W_{jk}, still need to be determined in equilibrium.

General equilibrium. The equilibrium of the model is pinned down by labor market clearing in each local labor market. To derive these labor market clearing conditions we first sum equations (10) and (11) over all firms within a given sector to get the amount of production labor $l_{jk}(s)$ used to produce goods shipping from j to k

$$l_{jk}(s) := \sum_{i=1}^{n(s)} l_{ijk}(s) = \alpha \frac{\tau_{jk}(s) W_{jk}}{w_j} \frac{1}{z_{jk}(s)} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k \tag{28}$$

and the amount of labor $m_{kj}(s)$ used in j to process goods shipped from k

$$m_{kj}(s) := \sum_{i=1}^{n(s)} m_{ikj}(s) = (1-\alpha) \frac{\tau_{kj}(s) W_{kj}}{w_j} \frac{1}{z_{kj}(s)} \left(\frac{P_j\,(s)}{P_j} \right)^{-\theta} C_j \tag{29}$$

where $z_{jk}(s)$ is the aggregate productivity of firms shipping from j to k. This aggregate productivity is given by a harmonic index of the productivities $z_{ij}(s)$ of each firm at source location j, specifically

$$z_{jk}(s) = \left(\sum_{i=1}^{n(s)} \frac{1}{z_{ij}(s)} \left(\frac{\phi_{ijk}(s)}{\Phi_{ik}(s)} \right)^{-\lambda} \left(\frac{P_{ik}(s)}{P_k(s)} \right)^{-\gamma} \right)^{-1} \tag{30}$$

These expressions for labor demand depend implicitly on the distribution of markups $\mu_{ik}(s)$ within and across locations through the prices $P_{ik}(s)$ that enter $l_{jk}(s), m_{kj}(s)$ and $z_{jk}(s)$.

Labor demand in location j is then found by aggregating $l_{jk}(s) + m_{kj}(s)$ over all k and sectors s. Labor supply in location j is given by $e_j L_j$, i.e., each location j is endowed with L_j workers each with e_j efficiency units of labor. The labor market in location j clears when

$$e_j L_j = \int_0^1 \sum_{k=1}^{J} \left(l_{jk}(s) + m_{kj}(s) \right) ds \tag{31}$$

Solving the model. We solve the model as follows. We first solve the fixed point problem (27) for the function that maps marginal costs $\Phi_{ik}(s)$ into markups $\mu_{ik}(s)$. We then guess a vector of wages w_j, with one wage normalized to 1 as numeraire. This guess implies wage indexes W_{jk}, marginal costs $\Phi_{ik}(s)$ and hence markups $\mu_{ik}(s)$, prices $P_{ik}(s)$ and quantities $C_{ik}(s)$, etc. As shown in the Appendix, we can then update the wage guess efficiently by exploiting the fact that, conditional on markups, budget constraints are linear in wages.

3 Quantifying the model

In this section we outline our benchmark parameterization and calibration strategy and present our model's implications for local sales concentration.

3.1 Benchmark parameterization

Our geographical locations are US states and Washington DC. Our benchmark model is set up to match the operations of some 270,000 manufacturing firms in 364 6-digit NAICS sectors across these locations.

Productivity distribution. We assume that firm-level productivity in source location j can be written

$$z_{ij}(s) = \bar{z}_i(s)\, \hat{z}_{ij}(s) \tag{32}$$

where $\bar{z}_i(s)$ is a firm-specific fixed effect and $\hat{z}_{ij}(s)$ is a location-specific effect. For parsimony, and as is standard in the literature, we assume that the firm fixed effect $\bar{z}_i(s)$ is drawn from a Pareto distribution on $[1, \infty)$ with tail parameter ξ. Conditional on firm i operating in location j, the location-specific effect $\hat{z}_{ij}(s)$ is a binary $\{0, 1\}$ outcome with distribution that depends on the firm-specific fixed effect $\bar{z}_i(s)$, as discussed below.

Operations of multi-establishment firms across locations. We parameterize the locations where a firm operates as follows. We first assign every firm a 'home' location where it operates with probability 1. For any other location j, we assume that the probability that a firm with productivity $\bar{z}_i(s)$ has an establishment at j is given by

$$\text{prob}\big[\, \hat{z}_{ij}(s) = 1 \,|\, \bar{z}_i(s) \,\big] = \frac{1}{1 + \alpha_0 \bar{z}_i(s)^{-\alpha_1}} \tag{33}$$

where α_0 determines the probability that the least-productive firm, with $\bar{z}_i(s) = 1$, has an establishment at j and where α_1 determines how quickly this probability increases for more productive firms.

Trade costs. Following Caliendo, Parro, Rossi-Hansberg and Sarte (2018), we parameterize the sector-specific iceberg trade costs $\tau_{jk}(s)$ by assuming a sector-specific log-linear relationship between trade costs and physical distance d_{jk}

$$\ln \tau_{jk}(s) = \delta(s) \ln d_{jk} \tag{34}$$

3.2 Calibration

Calibration strategy. We assign values to a small number of conventional parameters that are held constant throughout all our quantitative exercises. We calibrate the remaining parameters internally using the simulated method of moments. We calibrate the parameters governing the distribution of firm-level productivity and the operations of multi-establishment firms to match establishment-level data from the US Census of Manufactures.

We calibrate the parameters governing spatial trade frictions by requiring that the model reproduce gravity regressions based on the Commodity Flow Survey (CFS).

Assigned parameters. Following Edmond, Midrigan and Xu (2023), we set the across-sector elasticity of substitution to $\theta = 1.25$ and the within-sector elasticity of substitution to $\gamma = 10$. Intuitively, θ pins down the aggregate markup, while the difference between θ and γ determines the slope coefficient in a regression of inverse markups on market shares, as in (25). For parsimony we assume that the elasticity of substitution across goods within a firm equals the elasticity of substitution across firms within a given sector, $\lambda = \gamma$, i.e., goods produced by different establishments under the umbrella of the same firm are just as substitutable for one another as are goods sold by other firms within the same sector. For our benchmark model we assume that $\alpha = 1$ so that labor is required only in the source location. We take the number of firms $n(s)$ in each of the 364 6-digit NAICS sectors straight from the data, giving us an average of 733 firms per sector. Most of these firms are small and produce only locally. We report these parameter choices in Panel A of Table 1.

Table 1: Parameterization

Parameter		Value	Target
A. Assigned Parameters			
Elas. subs. across sectors	θ	1.25	Edmond, Midrigan, Xu (2023)
Elas. subs. within sectors	$\gamma = \lambda$	10	Edmond, Midrigan, Xu (2023)
Labor in source location	α	1	
B. Calibrated Parameters			
Pareto tail firm productivity	ξ	7.35	National sales concentration
Trade cost wrt distance	$\delta(s)$		Gravity coeff. 3-digit NAICS
Multi-establishment	$\text{prob}[\hat{z}_{ij} \mid \bar{z}_i = 1]$	0.0003	Share of multi-estab. firms
	$\text{prob}[\hat{z}_{ij} \mid \bar{z}_i = 2]$	0.082	

Calibrated parameters. We calibrate the remaining parameters internally using the simulated method of moments. We determine these parameters jointly, targeting (i) measures of national sales concentration, to pin down the Pareto tail parameter ξ, (ii) measures of the average frequency of multi-establishment and multi-state firms to pin down α_0 and α_1, and (iii) sector-level gravity regressions to pin down the elasticities of trade costs with respect to distance, $\delta(s)$.

(i) National Sales Concentration. We target the average top-4 and top-20 national sales shares and average national sales Herfindahl-Hirschman Index (HHI). In the US Census of Manufactures, the average top-4 and top-20 national sales shares for 6-digit NAICS sectors are 42% and 73% respectively while the average national HHI is 0.10.

(ii) Operations of Multi-Establishment Firms. In the US Census of Manufactures, about 4% of firms are multi-establishment and these multi-establishment firms account for about 79% of sales.

(iii) Gravity Regressions. We estimate sector-specific gravity regressions of the form

$$\ln(\text{shipments})_{jk}(s) = \gamma_j(s) + \gamma_k(s) + \beta(s) \ln d_{jk} + \epsilon_{jk}(s) \tag{35}$$

where $\gamma_j(s)$,$\gamma_k(s)$ denote sector-specific source and destination fixed effects. We estimate these gravity regressions using state-to-state trade flows from the Commodity Flows Survey for each 3-digit NAICS manufacturing sector. Our estimated slope coefficients $\beta(s)$, reported in the Appendix, measure how sensitive trade flows are to geographical distance. Goods that are more easily tradable, such as *computers & electronics* (sector 334) and *electric equipment & appliances* (sector 335), have estimated $\beta(s)$ that are small in magnitude. Goods that are less easily tradable, such as *wood* (sector 321), *petroleum, asphalt and coal* (sector 324) and *non-metallic minerals* (sector 327) have large negative estimated $\beta(s)$.

In the model we simulate data for each of our 364 6-digit sectors and, for each sector s, calculate the total value of shipments from j to k as

$$\text{shipments}_{jk}(s) = \sum_{i=1}^{n(s)} p_{ijk}(s)\, y_{ijk}(s). \tag{36}$$

To be consistent with our empirical gravity regressions from the Commodity Flows Survey we aggregate these shipment flows to a 3-digit cluster of sectors and choose the parameters $\delta(s)$ in our specification (34) so that the estimated $\beta(s)$ in the model gravity regressions match their empirical counterparts from (35).

We report our internally calibrated parameters governing the productivity distribution and the operations of multi-establishment firms across locations in Panel B of Table 1. Jointly with our other parameters, our model matches the data on national sales concentration with a productivity distribution with Pareto tail $\xi = 7.35$, an estimate which implies slightly thinner tails than the model of oligopolistic competition in Edmond, Midrigan and Xu (2023), which abstracts from spatial frictions. Our model matches the facts on the operations of multi-establishment firms across locations by setting α_0 and α_1 such that

$$\text{prob}\big[\, \hat{z}_{ij}(s) = 1 \,|\, \bar{z}_i(s) = 1 \,\big] = \frac{1}{1+\alpha_0} = 0.0003 \tag{37}$$

and

$$\text{prob}\big[\, \hat{z}_{ij}(s) = 1 \,|\, \bar{z}_i(s) = 2 \,\big] = \frac{1}{1+\alpha_0 2^{-\alpha_1}} = 0.082 \tag{38}$$

That is, for any location other than their home location, the least productive firms, with $\bar{z}_i(s) = 1$, have a tiny 0.03% chance of having an establishment there. A firm that is twice as productive as the least productive, with $\bar{z}_i(s) = 2$, has a much greater 8.2% probability of having an establishment in any location other than their home location.

Model fit. We report the moments we target in the data and their model counterparts in Panel A of Table 2. The model does a good job of reproducing the average amount of national sales concentration. Despite its parsimony, the model also captures the diversity of multi-establishment operations across states, though it somewhat overestimates the fraction of firms that are multi-establishment and under-estimates the relative size of multi-establishment firms. We report the 3-digit gravity coefficients $\beta(s)$ we estimate from the CFS and their model counterparts in Figure 1. Our model very closely reproduces the sector-level gravity effects that pin down our spatial trade frictions.

Model validation. In Panel B of Table 2 we also report some key moments that were not targeted in our calibration exercise. In the data, local production is much more concentrated than national sales, the local production HHI is 0.36 compared to the national sales HHI of 0.10. Our model reproduces this fact almost exactly, with local production HHI 0.35 and national sales HHI 0.11. Similarly, our model does a good job of reproducing the amount of concentration across states, despite the fact that we did not target this moment.

Table 2: Model Fit

Moments	Data	Model
Panel A: Targeted Moments		
Top 4 national sales share	0.42	0.46
Top 20 national sales share	0.73	0.70
HHI national sales	0.10	0.11
Fraction multi-establishment firms	0.04	0.09
Sales accounted for by multi-establishment firms	0.79	0.63
Panel B: Validation		
HHI local production	0.36	0.35
HHI across states	0.08	0.05

Figure 1: Gravity Coefficients $\beta(s)$ in Data and Model

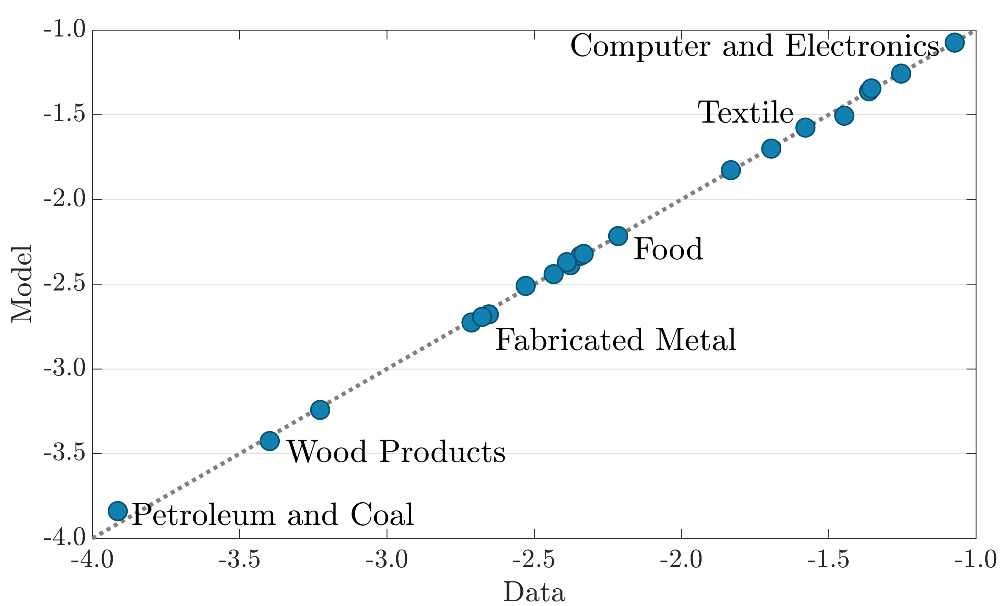


Table 3: Sales Concentration in Local Markets

Local Sales Concentration	Model
Top 1 local sales share	0.33
Top 4 local sales share	0.69
Top 20 local sales share	0.94
HHI local sales	0.20

Concentration in destination markets. Of key interest in our framework is how much competition firms face in the destination markets that they sell to. In less competitive markets, dominant firms will be able to charge high markups. Sales concentration in local markets is not something we can directly observe with the Census data. But given that our model does a good job of reproducing national sales concentration and local production concentration, it seems natural to use the model to infer the amount of local sales concentration. We report our benchmark model's implication for local sales concentration in Table 3. Intuitively, we find that local sales concentration is lower than local production concentration but not as low as national sales concentration.

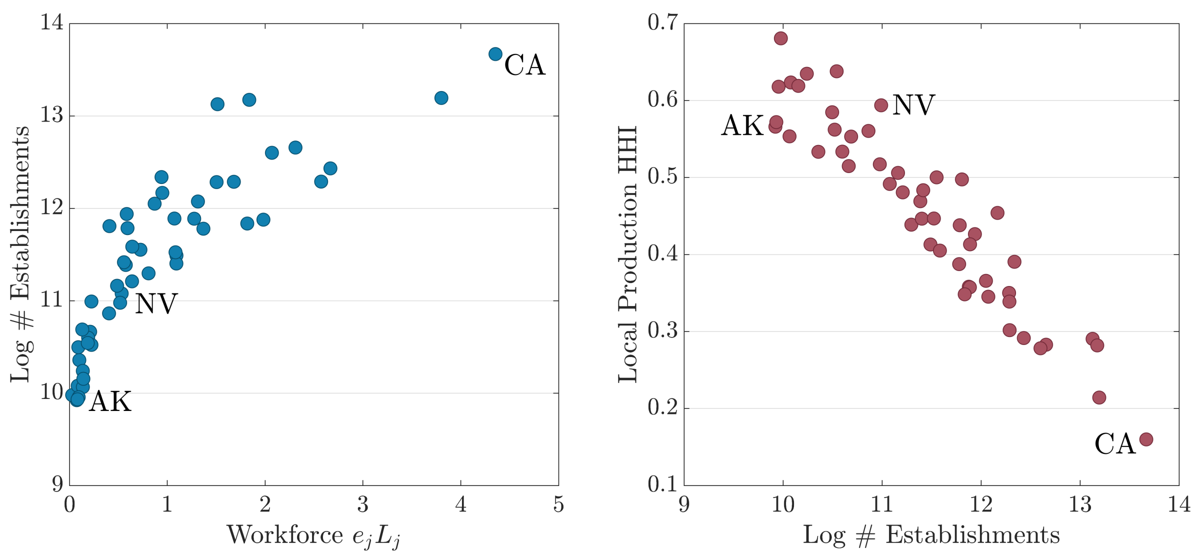
4 Quantitative results

In this section we present three quantitatve results. First, we show how spatial trade frictions influence the spatial distribution of production concentration and sales concentration. Second, we show that these spatial frictions matter in the aggregate. In particular, an otherwise equivalent model that abstracts from geography and spatial frictions leads to a quantitatively significant understatement of both the aggregate markup and the aggregate productivity losses due to markup dispersion. Third, we show that a reduction in trade costs — calibrated to match estimates of the effects of improvements in transportation technology in manufacturing — leads to increases in national sales concentration but decreases in local sales concentration. And because of this decrease in local sales concentration, such decreases in trade costs also lead to decreases in markups and markup dispersion, despite the increase in national concentration.

4.1 Spatial distribution of concentration

In our model, intranational trade costs shape the amount of local competition. Goods that are easily tradable can be shipped from the most productive source locations to almost any destination market, increasing the amount of competition amongst producers of tradable goods in those markets. Goods that are less easily tradable will be shipped to a more limited set of destinations, inhibiting the amount of competition amongst producers of less-tradable goods in those markets. Because of these effects, our model predicts that local production concentration is both higher and more dispersed than local sales concentration.

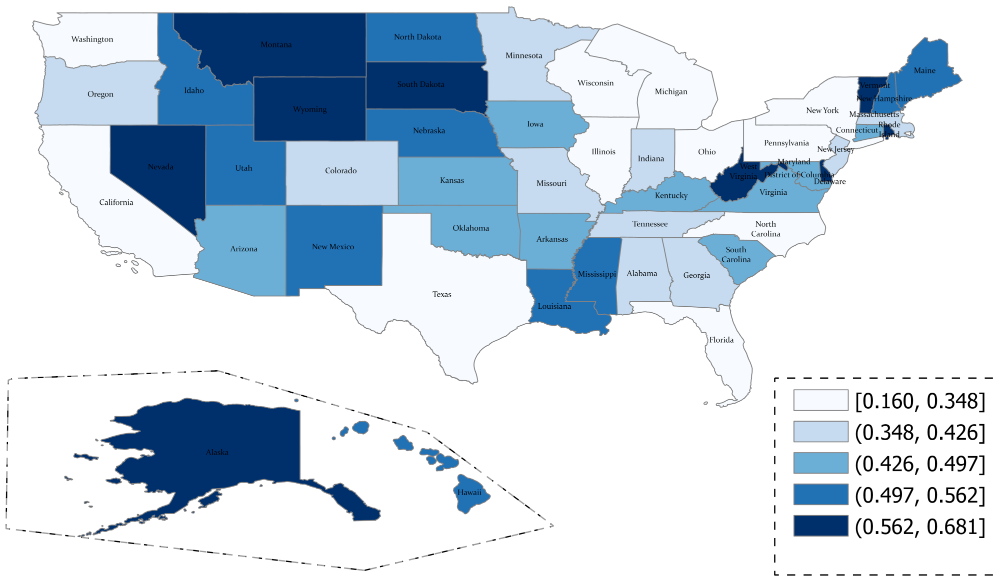
Figure 2: Location Characteristics and Local Production Concentration



Location characteristics and production concentration. Overall, the number of establishments in any given location is strongly influenced by the basic characteristics of that location. For example, locations that are endowed with large and productive labor forces, i.e., high effective labor $e_j L_j$, attract a larger number of establishments across all sectors and have lower production concentration. Figure 2 illustrates, with the left panel showing the positive spatial correlation between effective labor $e_j L_j$ and the number of establishments in each location j and the right panel showing the negative spatial correlation between the number of establishments in location j and local production concentration.

Intuitively, states like California, with large endowments of effect labor, have higher establishment density and lower production concentration compared to states with smaller endowments of effective labor and hence fewer establishments, such as Alaska and Nevada.

Figure 3: Spatial Distribution of Local Production Concentration



In short, the bulk of the dispersion in local production concentration across states is, driven by heterogeneity in location characteristics. Figure 3 shows the spatial distribution of local production concentration for our benchmark model.

Trade and sales concentration. We find that intranational trade plays a key role in determining the amount of local competition. Despite the high production concentration in states like Nevada, geographic proximity to states with many productive establishments, such as California, increases trade flows that increase local competition. Conversely, relatively isolated states, like Alaska, experience limited trade flows and hence their local production plays a more direct role in determining the amount of local competition. Figure 4 illustrates the spatial distribution of local sales concentration for our benchmark model.

The differences between the distribution of production concentration in Figure 3 and the distribution of sales concentration in Figure 4 illustrate how the presence of many productive establishments in a location reduces sales concentration not only in their home location but also in surrounding locations. Figure 5 reports the spatial correlation between production concentration and sales concentration. The spatial correlation is positive, but noisy.

Figure 4: Spatial Distribution of Local Sales Concentration

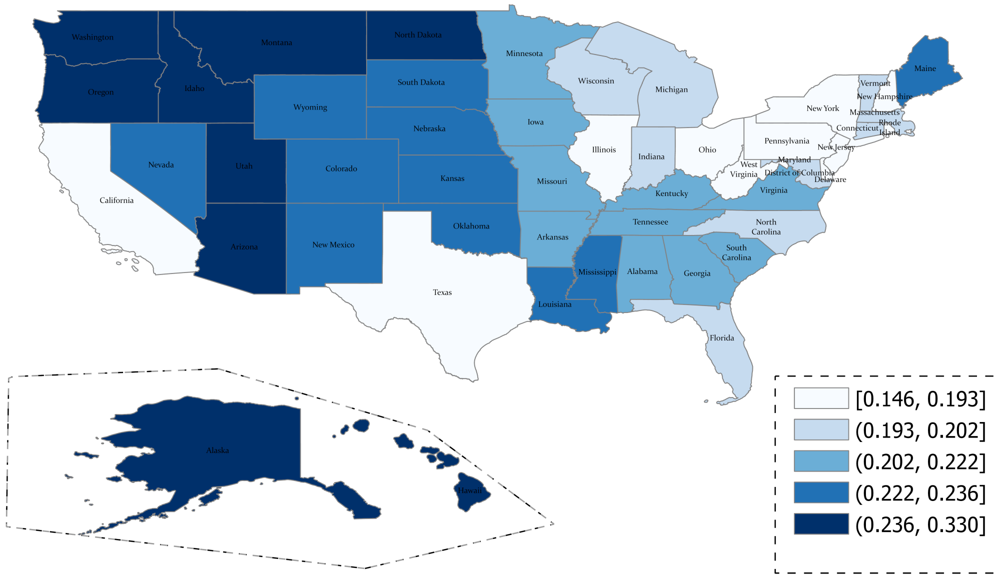


Figure 5: Spatial Correlation Production and Sales HHI

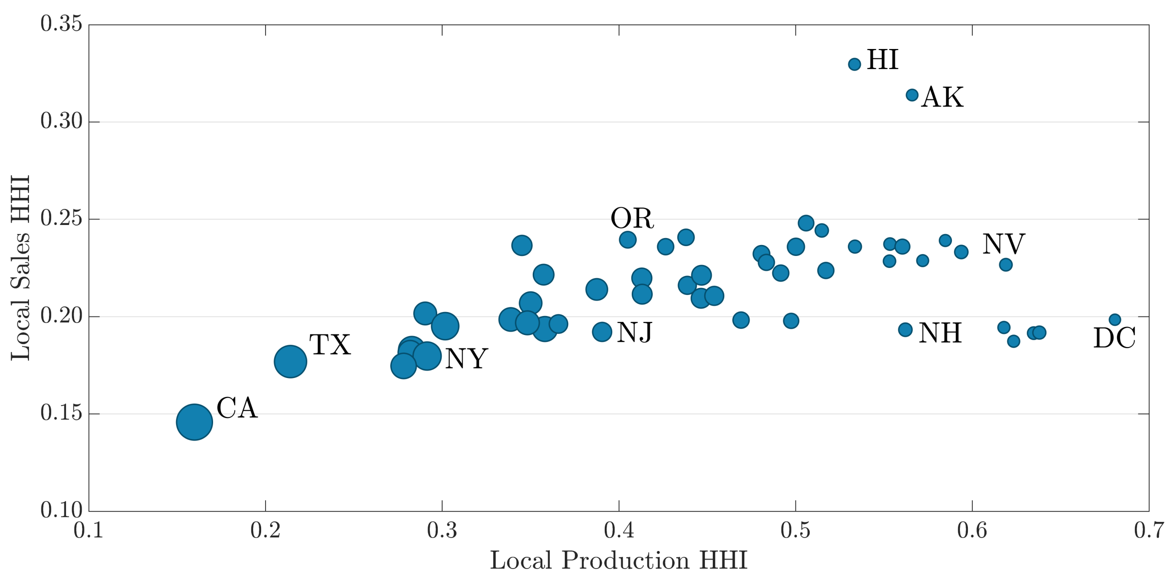
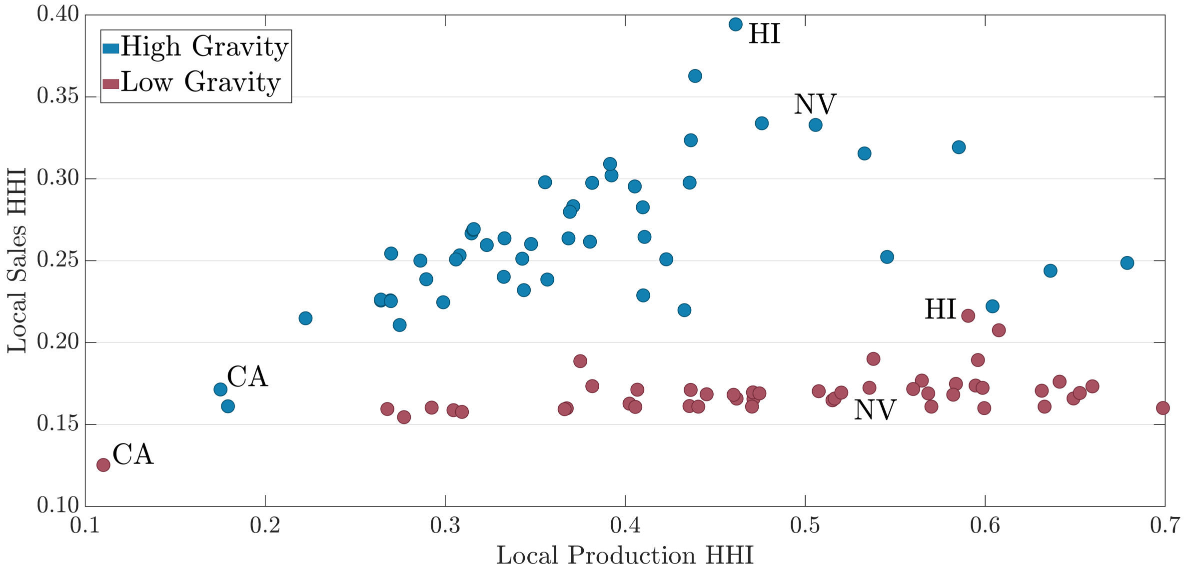


Figure 6: Spatial Correlation by Gravity



Role of gravity. The spatial correlation between production concentration and sales concentration becomes much more clear when we look at specific sectors and condition on the extent of spatial frictions. In particular, for highly tradable sectors with low spatial frictions (low gravity effects), such as *computer* and *electronics* manufacturing, there is essentially zero spatial corelation between production concentration and sales concentration, as shown by the red dots in Figure 6. By contrast, for less tradable sectors with high spatial frictions (high gravity effects), such as *wood, petroleum,* and *coal* manufacturing, there is a strong positive spatial correlation, as shown by the blue dots in Figure 6.

4.2 Geography and aggregate markups

We now quantify the significance of geography and these spatial frictions for aggregate outcomes. To achieve this, we use an otherwise equivalent model that abstracts from geography and spatial frictions. We keep our assigned parameters unchanged and calibrate this version of the model to match national sales concentration. We report the fit of this model in Table 4. The model without geography fits the national sales concentration facts about as well as our benchmark model.

Table 4: Model Fit, No Geography

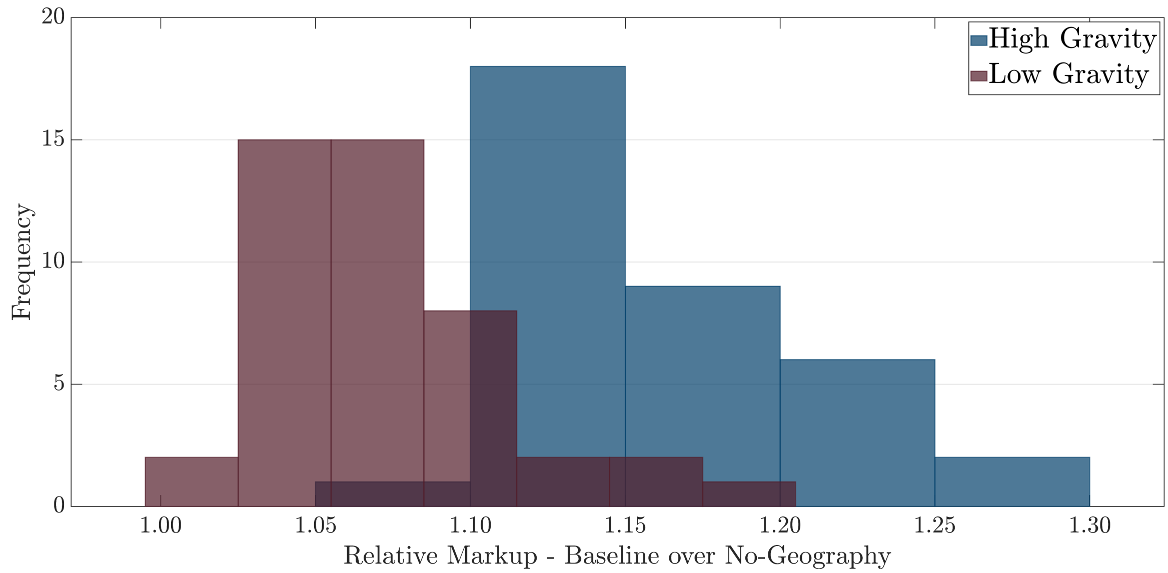
Moments	Data	Benchmark	No Geography
Top 4 sales share	0.42	0.46	0.47
Top 20 sales share	0.73	0.70	0.72
HHI sales	0.10	0.11	0.08

But the model without geography implies lower markups and less markup dispersion than our benchmark model, as shown in Table 5. Across all sectors, markups in the model without geography are both lower and less dispersed than in our benchmark model with geography. In other words, the model without geography implies both a lower aggregate markup and lower productivity losses due to misallocation. In this sense, abstracting from geography and spatial frictions leads to a quantitatively significant *understatement* of the macroeconomic losses associated with market power.

Table 5: Markup Distribution, No Geography

Percentile	Benchmark	No Geography
p01	1.17	1.12
p10	1.21	1.14
p25	1.25	1.15
p50	1.29	1.17
p75	1.36	1.20
p90	1.46	1.25
p99	1.62	1.40
Aggregate Markup	1.31	1.18

Figure 7: Sector-Level Relative Markups, by Gravity



To reinforce this point, we partition sectors into *high gravity* sectors, facing strong spatial frictions, and *low gravity* sectors, facing weak spatial frictions, and then compute the ratio of the markup in each sector in our benchmark model to its counterpart markup in the model without geography. The distribution of these relative markups for the two categories, high gravity and low gravity, are shown in Figure 7. In the benchmark model with geography, these sector-level markups are larger and more dispersed and this effect is indeed much stronger for high gravity sectors.

4.3 Reduction in trade costs and diverging trends in concentration

Recent empirical research has documented significantly different trends in national sales concentration and local sales concentration. National sales concentration, along with local production concentration, has been on the rise since the early 1980s. But local sales concentration has been on the decline. We now show that these divergent trends in national and local sales concentration emerge naturally in our model when intranational trade costs are falling over time.

Reduction in trade costs. Over time, improvements in transportation technology and infrastructure should decrease trade costs, i.e., gravity effects should be becoming weaker. Consistent with this, using interregional trade data, Coşar, Osotimehin and Popov (2022) find a 15 to 20% decrease in manufacturing distance elasticities from 1963 to 2017. Since our benchmark model is calibrated to current data, to replicate the conditions of 1963 we increase trade cost elasticities $\delta(s)$ uniformly by 20% for all 3-digit NAICS manufacturing sectors. Table 6 reports the effects of such changes in trade costs on concentration. Moving forward from 1963 to the present, a reduction in trade costs increase national sales concentration and local production concentration while decreasing local sales concentration — i.e., a reduction in trade costs implies endogenously divergent trends in national and local sales concentration.

Table 6: Effects of Changes in Trade Cost on Concentration

	20% Increase	Benchmark	20% Decrease	Free Trade
National Concentration				
Top 4 national sales share	0.44	0.46	0.49	0.57
Top 20 national sales share	0.67	0.70	0.73	0.83
HHI national sales	0.10	0.11	0.11	0.13
Local Production Concentration				
HHI local production	0.34	0.35	0.37	0.42
Local Sales Concentration				
Top 4 local sales share	0.72	0.69	0.66	0.57
Top 20 local sales share	0.95	0.94	0.92	0.83
HHI local sales	0.21	0.20	0.18	0.13

Extrapolating such reduction in trade costs forward, the model predicts that the most productive firms expand by accessing more distant markets. This results in further increases in national sales concentration and increases in local production concentration. But this pattern of expansion also leads to more competition in destination markets and hence lower local sales concentration. In the limit as trade costs disappear, while maintaining the assumption of immobile labor, we find that national and local sales concentration *converge*. In this 'free trade' limit, there is no meaningful distinction between national and local sales.

Table 7: Effects of Changes in Trade Costs on Markup Distribution

Percentile	20% Increase	Benchmark	20% Decrease	Free Trade
p01	1.18	1.17	1.16	1.13
p10	1.23	1.21	1.20	1.16
p25	1.27	1.25	1.23	1.18
p50	1.32	1.29	1.27	1.22
p75	1.38	1.36	1.34	1.28
p90	1.49	1.46	1.43	1.39
p99	1.66	1.62	1.60	1.55
Aggregate Markup	1.33	1.31	1.29	1.24

Markups. Table 7 reports the effects of such changes in trade costs on the sector-level markup distribution and the aggregate markup. Since markups are determined in large part by the amount of competition in local markets, and local sales concentration is declining as trade costs decrease from 1963 to the present, it is not surprising that we find that the reduction in trade costs leads to both lower markups and lower markup dispersion — and hence lower productivity losses due to misallocation.

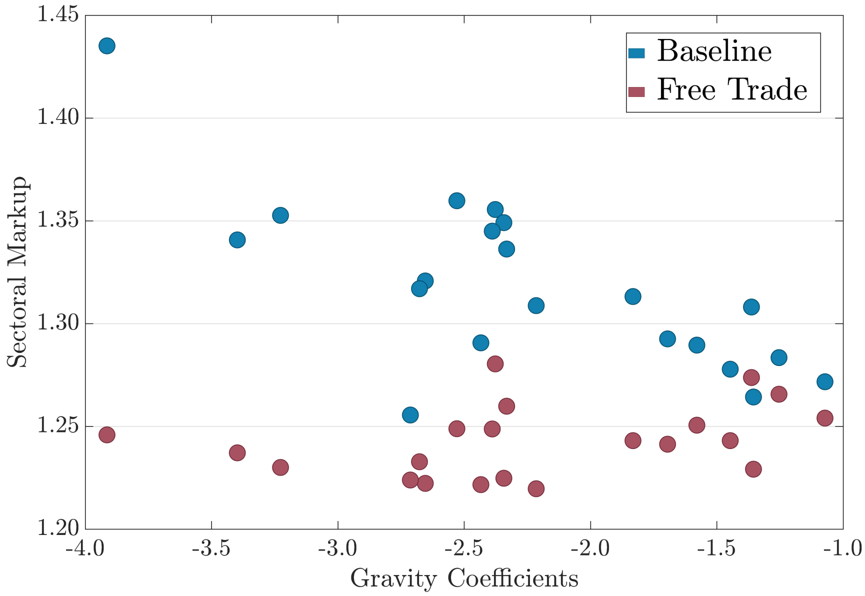
Implications for markups, by gravity. Table 8 partitions sectors into *high gravity* sectors, facing strong spatial frictions, and *low gravity* sectors, facing weak spatial frictions. We already know that markups are larger and more dispersed in high gravity sectors. Now we see that a reduction in trade costs also has larger effects on the markups of firms in high-gravity sectors, especially in the upper-tail of the markup distribution. Overall, for high-gravity sectors the reduction in the aggregate markup in response to a 20% reduction in trade costs is about twice as large as the reduction in the aggregate markup for low-gravity sectors. This can also be seen in Figure 8 where we plot sector-level markups against sector-level gravity coefficients $\beta(s)$ for our benchmark calibration, in blue, and the 'free trade' version of the model without trade costs, in red. Clearly the reduction in markups is larger for high gravity sectors, i.e., sectors with large negative $\beta(s)$.

Table 8: Implications for Markups, Low- vs. High-Gravity Sectors

Percentile	20% Increase	Benchmark	20% Decrease	Free Trade
Panel A: Low Gravity Sectors				
p10	1.20	1.19	1.18	1.16
p25	1.22	1.20	1.19	1.18
p50	1.26	1.24	1.23	1.22
p75	1.30	1.30	1.29	1.28
p90	1.45	1.44	1.44	1.42
Aggregate Markup	1.28	1.27	1.26	1.24
Panel B: High Gravity Sectors				
p10	1.28	1.28	1.26	1.15
p25	1.30	1.30	1.28	1.20
p50	1.37	1.36	1.33	1.24
p75	1.52	1.46	1.38	1.27
p90	1.55	1.50	1.45	1.33
Aggregate Markup	1.38	1.36	1.34	1.24

In short, while reductions in trade costs of this kind lead to an increase in national sales concentration, this increase in national concentration masks the fact local markets are becoming more competitive, markups are falling, and aggregate productivity is rising. The reductions in markups are larger in high-gravity sectors, i.e., in sectors where spatial frictions are more severe.

Figure 8: Larger Reductions in Markups for High Gravity Sectors



5 Conclusion

We study the spatial distribution of production and consumption in a quantitative model with multi-establishment firms, oligopolistic competition, and endogenously variable markups. We calibrate our model to match US Census of Manufactures firm and establishment data and intranational trade flows from the Commodity Flows Survey. We show that spatial frictions can have large aggregate effects, increasing both the aggregate markup and the productivity losses due to misallocation. We then show that a reduction in intranational trade costs, calibrated to match long-run trends in US manufacturing, will increase national sales concentration but decrease local sales concentration. Local markets become more competitive, markups fall, and aggregate productivity rises, despite the increase in national concentration

References

Amiti, Mary and Sebastian Heise, "US Market Concentration and Import Competition," May 2021. CEPR Discussion Paper 16126.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The Elusive Pro-Competitive Effects of Trade," *Review of Economic Studies*, January 2019, *86* (1), 46–80.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2019, *17* (6), 1881–1940.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 2008, *98* (5), 1998–2031.

Autor, David, Christina Patterson, and John Van Reenen, "Local and National Concentration Trends in Jobs and Sales: The Role of Structural Transformation," April 2023. NBER Working Paper 31130.

_ , David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, May 2020, *135* (2), 645–709.

Basker, Emek, Shawn D. Klimek, and Pham Hoang Van, "Supersize It: The Growth of Retail Chains and the Rise of the Big-Box Store," *Journal of Economics and Management Strategy*, 2012, *21* (3), 541–582.

Benkard, C. Lanier, Ali Yurukoglu, and Anthony Lee Zhang, "Concentration in Product Markets," September 2023. NBER Working Paper 28745.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, September 2003, *93* (4), 1268–1290.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "The Impact of Regional and Sectoral Productivity Changes on the US Economy," *Review of Economic Studies*, 2018, *85* (4), 2042–2096.

Cao, Dan, Henry R. Hyatt, Toshihiko Mukoyama, and Erick Sager, "Firm Growth Through New Establishments," 2022. Working Paper.

Coşar, A. Kerem, Sophie Osotimehin, and Latchezar Popov, "Regional and Aggregate Implications of Transportation Costs," 2022. Working Paper.

Decker, Ryan, "Discussion of: Diverging Trends in National and Local Concentration," 2020. ASSA Meetings.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, Markups, and the Gains from International Trade," *American Economic Review*, October 2015, *105* (10), 3183–3221.

_ , _ , and _ , "How Costly Are Markups?," *Journal of Political Economy*, July 2023, *131* (7), 1619–1675.

Foster, Lucia, John Haltiwanger, Shawn Klimek, C.J. Krizan, and Scott Ohlmacher, "The Evolution of National Retail Chains: How We Got Here," in Emek Basker, ed., *Handbook on the Economics of Retailing and Distribution*, Elgar, 2016.

Franco, Santiago, "Output Market Power and Spatial Misallocation," November 2023. University of Chicago Working Paper.

Ganapati, Sharat, "Growing Oligopolies, Prices, Output, and Productivity," *AEJ: Microeconomics*, August 2021, *13* (3), 309–327.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely, "Are US Industries Becoming More Concentrated?," *Review of Finance*, July 2019, *23* (4), 697–743.

Holmes, Thomas J., "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, January 2011, *79* (1), 253–302.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, "The Industrial Revolution in Services," *Journal of Political Economy: Macroeconomics*, 2023, *1* (1), 3–42.

Jia, Panle, "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, November 2008, *76* (6), 1263–1316.

Kimball, Miles S., "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit, and Banking*, 1995, *27* (4, Part 2), 1241–1277.

Neiman, Brent and Joseph Vavra, "The Rise of Niche Consumption," *AEJ: Macroeconomics*, July 2023, *15* (3), 224–264.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter, "Diverging Trends in National and Local Concentration," *NBER Macroeconomics Annual*, 2020, pp. 115–150.

Smith, Dominic A. and Sergio Ocampo, "The Evolution of US Retail Concentration," *AEJ: Macroeconomics*, 2024, *forthcoming.*

Appendix

A Solving the model: labor market clearing

In this appendix we provide further details on how we use the labor market clearing conditions in each location to update our initial guess for wages.

Labor demand. Recall that labor demand by firm i that produces in j and sells in k is given by

$$l_{ijk}(s) = \alpha \frac{W_{jk}}{w_j} \frac{y_{ijk}(s)}{z_{ij}(s)} \tag{A1}$$

and

$$m_{ijk}(s) = (1-\alpha) \frac{W_{jk}}{w_k} \frac{y_{ijk}(s)}{z_{ij}(s)} \tag{A2}$$

Plugging in the resource constraints $y_{ijk}(s) = \tau_{jk}(s) c_{ijk}(s)$ and using the demand for goods

$$\begin{aligned} l_{ijk}(s) &= \alpha \frac{\tau_{jk}(s) W_{jk}}{w_j} \frac{c_{ijk}(s)}{z_{ij}(s)} \\ &= \alpha \frac{\tau_{jk}(s) W_{jk}}{w_j} \frac{1}{z_{ij}(s)} \left(\frac{\phi_{ijk}(s)}{\Phi_{ik}(s)} \right)^{-\lambda} \left(\frac{P_{ik}(s)}{P_k(s)} \right)^{-\gamma} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k \end{aligned} \tag{A3}$$

where

$$\Phi_{ik}(s) = \left(\sum_{j=1}^{J} \phi_{ijk}(s)^{1-\lambda} \right)^{\frac{1}{1-\lambda}}, \qquad \phi_{ijk}(s) := \frac{\tau_{jk}(s) W_{jk}}{z_{ij}(s)} \tag{A4}$$

Similarly

$$m_{ijk}(s) = (1-\alpha) \frac{\tau_{jk}(s) W_{jk}}{w_k} \frac{1}{z_{ij}(s)} \left(\frac{\phi_{ijk}(s)}{\phi_{ik}(s)} \right)^{-\lambda} \left(\frac{P_{ik}(s)}{P_k(s)} \right)^{-\gamma} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k \tag{A5}$$

Summing over firms, we then have labor used to produce in j for goods sold in k

$$l_{jk}(s) := \sum_{i=1}^{n(s)} l_{ijk}(s) = \alpha \frac{\tau_{jk}(s) W_{jk}}{w_j} \frac{1}{z_{jk}(s)} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k \tag{A6}$$

and labor used in j to process goods shipped from k

$$m_{kj}(s) = \sum_{i=1}^{n(s)} m_{ikj}(s) = (1-\alpha) \frac{\tau_{kj}(s) W_{kj}}{w_j} \frac{1}{z_{kj}(s)} \left(\frac{P_j(s)}{P_j} \right)^{-\theta} C_j \tag{A7}$$

where $z_{jk}(s)$ is the index of productivity given by

$$\frac{1}{z_{jk}(s)} = \sum_{i=1}^{n(s)} \frac{1}{z_{ij}(s)} \left(\frac{\phi_{ijk}(s)}{\Phi_{ik}(s)} \right)^{-\lambda} \left(\frac{P_{ik}(s)}{P_k(s)} \right)^{-\gamma} \tag{A8}$$

Labor market clearing. The labor market clearing condition in each location j can then be written

$$e_j L_j = \int_0^1 \sum_{k=1}^{J} (l_{jk}(s) + m_{kj}(s))\, ds \tag{A9}$$

Multiplying both sides by w_j and using our expressions for $l_{jk}(s)$ and $m_{kj}(s)$ above we can write this

$$\begin{aligned} w_j e_j L_j &= \int_0^1 \sum_{k=1}^{J} \left(\alpha \frac{\tau_{jk}(s) W_{jk}}{z_{jk}(s)} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k + (1-\alpha) \frac{\tau_{kj}(s) W_{kj}}{z_{kj}(s)} \left(\frac{P_j(s)}{P_j} \right)^{-\theta} C_j \right) ds \\ &= \int_0^1 \sum_{k=1}^{J} \left(\alpha \frac{\tau_{jk}(s) W_{jk}}{z_{jk}(s) P_k(s)} \left(\frac{P_k(s)}{P_k} \right)^{1-\theta} P_k C_k + (1-\alpha) \frac{\tau_{kj}(s) W_{kj}}{z_{kj}(s) P_j(s)} \left(\frac{P_j(s)}{P_j} \right)^{1-\theta} P_j C_j \right) ds \end{aligned}$$

Interchanging the order of integration and summation we get

$$w_j e_j L_j = \sum_{k=1}^{J} (\alpha P_k C_k \Lambda_{jk} + (1-\alpha) P_j C_j \Lambda_{kj}) \tag{A10}$$

where to conserve notation we define

$$\Lambda_{jk} := \int_0^1 \frac{\tau_{jk}(s) W_{jk}}{z_{jk}(s) P_k(s)} \left(\frac{P_k(s)}{P_k} \right)^{1-\theta} ds \tag{A11}$$

Profits. Aggregating the representative consumer's budget constraint in each location

$$P_j C_j = (w_j e_j + \bar{\pi})\, L_j \tag{A12}$$

The profits of firm i in destination k are given by

$$\Pi_{ik}(s) = (P_{ik}(s) - \Phi_{ik}(s)) C_{ik}(s) = \left(1 - \frac{1}{\mu_{ik}(s)} \right) P_{ik}(s) C_{ik}(s) \tag{A13}$$

Summing over all locations, firms and sectors gives

$$\bar{\pi} L = \sum_{k=1}^{J} \int_0^1 \sum_{i=1}^{n(s)} \left(1 - \frac{1}{\mu_{ik}(s)} \right) P_{ik}(s) C_{ik}(s)\, ds \tag{A14}$$

where $L := \sum_{k=1}^{J} L_k$. Writing the sector-level markup $\mu_k(s)$ in each location k as the sales-weighted harmonic average

$$\mu_k(s) = \left(\sum_{i=1}^{n(s)} \frac{\omega_{ik}(s)}{\mu_{ik}(s)} \right)^{-1} \tag{A15}$$

where $\omega_{ik}(s)$ denotes the sales share of firm i in location k

$$\omega_{ik}(s) := \frac{P_{ik}(s)C_{ik}(s)}{\sum_{i=1}^{n(s)} P_{ik}(s)C_{ik}(s)} = \left(\frac{P_{ik}(s)}{P_k(s)}\right)^{1-\gamma} \tag{A16}$$

we can then write aggregate profits

$$\bar{\pi}L = \sum_{k=1}^{J}\int_0^1 \left(1 - \frac{1}{\mu_k(s)}\right) P_k(s)C_k(s)\,ds \tag{A17}$$

Similarly, writing the aggregate markup in each location k

$$\mu_k = \left(\int_0^1 \frac{\omega_k(s)}{\mu_k(s)}\,ds\right)^{-1} \tag{A18}$$

where

$$\omega_k(s) := \frac{P_k(s)C_k(s)}{\int_0^1 P_k(s)C_k(s)\,ds} = \left(\frac{P_k(s)}{P_k}\right)^{1-\theta} \tag{A19}$$

we can further simplify aggregate profits to

$$\bar{\pi}L = \sum_{k=1}^{J}\left(1 - \frac{1}{\mu_k}\right) P_kC_k \tag{A20}$$

Now substituting in the representative consumer's budget constraint for each location k

$$\bar{\pi}\sum_{k=1}^{J} L_k = \sum_{k=1}^{J}\left(1 - \frac{1}{\mu_k}\right)(w_ke_kL_k + \bar{\pi})L_k \tag{A21}$$

Which implies that profits per worker $\bar{\pi}$ are given by

$$\bar{\pi} = \frac{\sum_{k=1}^{J}\left(1 - \frac{1}{\mu_k}\right) w_ke_kL_k}{\sum_{k=1}^{J}\frac{1}{\mu_k}L_k} \tag{A22}$$

Back to labor market clearing. With this in place we can go back to the labor market clearing condition

$$\begin{aligned} w_je_jL_j &= \sum_{k=1}^{J}\left(\alpha P_kC_k\Lambda_{jk} + (1-\alpha)P_jC_j\Lambda_{kj}\right) \\ &= \sum_{k=1}^{J}\left(\alpha[w_ke_k + \bar{\pi}]L_k\Lambda_{jk} + (1-\alpha)[w_je_j + \bar{\pi}]L_j\Lambda_{kj}\right) \end{aligned}$$

Hence

$$
\begin{aligned}
w_j e_j L_j =& \alpha \left(\sum_{k=1}^{J} w_k e_k L_k \Lambda_{jk} + \left\{ \frac{\sum_{k=1}^{J} \left(1 - \frac{1}{\mu_k}\right) w_k e_k L_k}{\sum_{k=1}^{J} \frac{1}{\mu_k} L_k} \right\} \sum_{k=1}^{J} L_k \Lambda_{jk} \right) \\
&+ (1-\alpha) \left(\sum_{k=1}^{J} w_j e_j L_j \Lambda_{kj} + \left\{ \frac{\sum_{k=1}^{J} \left(1 - \frac{1}{\mu_k}\right) w_k e_k L_k}{\sum_{k=1}^{J} \frac{1}{\mu_k} L_k} \right\} \sum_{k=1}^{J} L_j \Lambda_{kj} \right)
\end{aligned}
$$

Collecting terms we can write this

$$
\left(1 - (1-\alpha) \sum_{k=1}^{J} \Lambda_{kj}\right) w_j e_j L_j - \alpha \sum_{k=1}^{J} \Lambda_{jk} w_k e_k L_k - \left(\alpha \gamma_j + (1-\alpha)\beta_j\right) \sum_{k=1}^{J} \left(1 - \frac{1}{\mu_k}\right) w_k e_k L_k = 0
$$

where, in slight abuse of notation, we define the coefficients

$$
\gamma_j := \frac{\sum_{k=1}^{J} \Lambda_{jk} L_k}{\sum_{k=1}^{J} \frac{1}{\mu_k} L_k} \tag{A23}
$$

and

$$
\beta_j := \frac{\sum_{k=1}^{J} \Lambda_{kj} L_j}{\sum_{k=1}^{J} \frac{1}{\mu_k} L_k} \tag{A24}
$$

Now divide both sides by $e_j L_j$ to get

$$
\left(1 - (1-\alpha) \sum_{k=1}^{J} \Lambda_{kj}\right) w_j - \alpha \sum_{k=1}^{J} \Lambda_{jk} \frac{e_k L_k}{e_j L_j} w_k - \left(\alpha \gamma_j + (1-\alpha)\beta_j\right) \sum_{k=1}^{J} \left(1 - \frac{1}{\mu_k}\right) \frac{e_k L_k}{e_j L_j} w_k = 0
$$

We can write this compactly in matrix form as

$$
\boldsymbol{B w} = 0 \tag{A25}
$$

where the off-diagonal jk elements of the $J \times J$ matrix $\boldsymbol{B}$ are given by

$$
b_{jk} := -\alpha \Lambda_{jk} \frac{e_k L_k}{e_j L_j} - \left(\alpha \gamma_j + (1-\alpha)\beta_j\right) \left(1 - \frac{1}{\mu_k}\right) \frac{e_k L_k}{e_j L_j} \tag{A26}
$$

and where the diagonal elements are given by

$$
b_{jj} := \left(1 - (1-\alpha) \sum_{k=1}^{J} \Lambda_{kj}\right) - \alpha \Lambda_{jj} - \left(\alpha \gamma_j + (1-\alpha)\beta_j\right) \left(1 - \frac{1}{\mu_j}\right) \tag{A27}
$$

By Walras' Law, the matrix $\boldsymbol{B}$ has rank $J-1$, i.e.,we can only determine relative wages.

In practice, how we solve this problem depends on whether we are computing an initial steady state or a new steady state after some shock, e.g., after a change in trade costs.

Initial steady state. For computing an initial steady state we start with an arbitrary guess at wages w_j and back out the efficiency units e_j that rationalize observed wages per worker $W_j^{data} = w_j e_j$ in each location j. With this guess we set up the matrix $\boldsymbol{B}$, compute the nullspace, normalize its first element to 1 and then update our guess for wages iteratively. Once we update the wages w_j we use

$$e_j = \frac{W_j^{data}}{w_j}$$

to update the endowments of efficiency units of labor needed to ensure wages per worker in the model are consistent with the data. If we dampen the iterations sufficiently, this algorithm converges quite nicely.

New steady state. After we compute the initial steady state, we have a vector e_j of efficiency units in each location. With this in place we can compute the response of wages to changes in the environment. To do so, we use the same algorithm as above, but no longer allow e_j to adjust as we update w_j.

B Assigning establishments to sectors and locations

In the data have 266,802 firms operating 297,221 establishments. In our calibration, we assign establishments to sectors and locations as follows:

- For each firm we draw $\bar{z}_i(s)$ from a Pareto distribution with tail ξ.
- We use location weights, the number of establishments in each state, to draw firms' home states.
- We set the probability that a firm operates in its assigned home state to 1.
- Given parameters α_0 and α_1 and firm productivity draw $\bar{z}_i(s)$, the probability that firm i operates in any other state j is given by

$$\text{prob}\left[\, \hat{z}_{ij}(s) = 1 \,|\, \bar{z}_i(s) \,\right] = \frac{1}{1 + \alpha_0 \bar{z}_i(s)^{-\alpha_1}}$$

- We draw $\hat{z}_{ij}(s)$ from $\{0, 1\}$ according to these probabilities
- If $\hat{z}_{ij}(s) = 1$, firm i operates in state j with productivity $\bar{z}_i(s)$, otherwise firm i does not have an establishment in j.