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

Jonathan Becker† Chris Edmond Virgiliu Midrigan‡

Daniel Yi Xu§

Preliminary and Incomplete

This Draft: December 2025

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 across 170 US Economic Areas. We 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. We also show that the welfare costs of markups are large on average and very unevenly distributed. The welfare costs of markups for smaller, poorer, more remote locations are up to 20 times larger than the welfare costs for larger, richer, more central locations.

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, the 2024 PHBS Shenzen workshop on macroeconomics and finance, the 2025 workshop of the Australian Macroeconomics Society, and seminar participants at the University of Nottingham—Ningbo.

†Stony Brook University, jonathan.becker@stonybrook.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 productivity and in terms of the number and geographic location of their establishments. 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 as in Atkeson and Burstein (2008) determines the markups each firm charges in each of its destination markets. Equilibrium wages in each location are determined by local labor market clearing conditions.

We calibrate our model to match the operations of some 220,000 US manufacturing firms organized into 363 6-digit NAICS sectors, with an average of some 600 firms per sector. We take our geographical locations to be 170 US Economic Areas as constructed by the US Bureau of Economic Analysis. 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 ensure every firm in the model reproduces a real firm's *geographic footprint* — we place each firms' establishments exactly where they appear in the data. We choose the parameters of our model governing the distribution of productivity across firms and the correlation between a firm's establishment count and its productivity to match key facts on national sales concentration and the share of employment accounted for by multi-establishment firms. We parameterize sector-specific iceberg trade costs so that the model exactly reproduces sector-specific gravity regressions using state-to-state trade flows from the Commodity Flows Survey for 3-digit NAICS manufacturing sectors.

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 tradeable 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 0.15, 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 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 more pronounced in sectors characterized by low spatial frictions.

We find that these spatial frictions are a quantitatively significant determinant of the macroeconomic losses due to market power. Our benchmark model implies an aggregate markup of 1.26 with sector-level markups ranging from 1.62 at the 99th percentile to 1.13 at the 1st percentile. If we calibrate the model to match the same national 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 short, 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 endogenously 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 (2024), 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 such improvements in transportation particularly benefit the largest, most productive firms who can sell to even more locations than they did previously. Consistent with this, we find a modest rise in national sales concentration. Our model predicts that a 20% reduction in trade costs leads the top-4 national sales share to increase from 0.44 in our benchmark model to 0.45.

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 0.58 in our benchmark to 0.56. 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 1.26 to 1.24 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.

The welfare costs of markups in our model are large on average and very unevenly distributed across locations. We measure these welfare costs by asking how much the representative consumer in each location would gain from policies that eliminate markups. We find that the average costs are high, about 5.8% in consumption-equivalent terms, and range from 1% or less in the largest, richest, most central locations to as much as 20% in the smallest, poorest, most remote locations. An otherwise equivalent model that abstracts from spatial frictions would find a lower average cost, on the order of 3-4%, as in Edmond, Midrigan and Xu (2023). Abstracting from geography and spatial frictions leads to not just an inability to characterize the variation in the burden of market power distortions across locations but also to a sizeable understatement of the aggregate welfare costs of markups.

Finally, we ask to what extent can worker mobility mitigate the welfare costs of markups. Our benchmark model features workers that are immobile across locations. In an extension we consider a setup where workers have heterogeneous preferences for different location-specific amenities, pinning down labor supply to each location. We find that, when we parameterize the model to match the same initial allocation of labor across locations as in our benchmark, the welfare costs of markups end up being almost identical. With worker mobility, eliminating markups leads to larger changes in consumption in the most attractive

[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.

locations, but these locations also receive labor inflows so that the changes in consumption per worker are almost exactly the same as in our benchmark model.

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 cities in a model of monopolistic competition with Kimball (1995) 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 markups.

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,[2] 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.

[2] We discuss an extension with *labor mobility* across locations in Section 7 below.

Firms source from establishments in multiple locations. Each firm i selling in destination market k sources goods from establishments in multiple locations $j = 1, ..., J$. Specifically, we assume that each firm i supplies k with a CES aggregate *across establishments*

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

For simplicity we assume that the elasticity of substitution across establishments within a given firm is the same, γ, as the elasticity of substitution across firms within a given sector.

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 per efficiency unit. Aggregating the budget constraints of workers in location j gives

$$P_j C_j = W_j E_j L_j + \Pi_j \tag{4}$$

where Π_j denotes aggregate profits paid out to workers in location j.

Distribution of profits. We assume that firm ownership is perfectly diversified across locations with profits paid out in proportion to labor income

$$\Pi_j = \bar{\pi} W_j E_j L_j \tag{5}$$

This implies that every location has the same labor and profit income shares

$$\left(\frac{1}{1+\bar{\pi}}, \frac{\bar{\pi}}{1+\bar{\pi}} \right) \tag{6}$$

for some constant $\bar{\pi} \geq 0$ to be determined in equilibrium.

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

$$c_{ijk}(s) = \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{-\gamma} \underbrace{\left(\frac{P_{ik}(s)}{P_k(s)} \right)^{-\gamma} \left(\frac{P_k(s)}{P_k} \right)^{-\theta} C_k}_{=C_{ik}(s)} \tag{7}$$

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

$$P_k = \left(\int_0^1 P_k(s)^{1-\theta} \, ds \right)^{\frac{1}{1-\theta}}, \qquad P_k(s) = \left(\sum_{i=1}^{N(s)} P_{ik}(s)^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \tag{8}$$

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

$$P_{ik}(s) = \left(\sum_{j=1}^{J} (\tau_{jk}(s) p_{ijk}(s))^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \tag{9}$$

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 produced at location j. For simplicity we assume that the output shipped by firm i from source j to destination k is linear in labor

$$y_{ijk}(s) = z_{ij}(s) l_{ijk}(s) \tag{10}$$

Resource constraints. 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

$$y_{ijk}(s) = \tau_{jk}(s) c_{ijk}(s) \tag{11}$$

Marginal cost. Taking the wage rate W_j as given, firm i can source goods from j for any destination k at marginal cost

$$\frac{W_j}{z_{ij}(s)} \tag{12}$$

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

$$\Pi_{ik}(s) = \sum_{j=1}^{J} \left(p_{ijk}(s) - \frac{W_j}{z_{ij}(s)} \right) y_{ijk}(s) \tag{13}$$

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

$$\sum_{j=1}^{J} \frac{W_j}{z_{ij}(s)} y_{ijk}(s) = \sum_{j=1}^{J} \frac{\tau_{jk}(s) W_j}{z_{ij}(s)} \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{-\gamma} \underbrace{C_{ik}(s)}_{=1} \tag{14}$$

subject to the firm-level price index (9). The Lagrangian for this problem can be written

$$\mathcal{L} = \sum_{j=1}^{J} \frac{\tau_{jk}(s) W_j}{z_{ij}(s)} \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{-\gamma} - \lambda_{ik}(s) \sum_{j=1}^{J} \left(\left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right)^{1-\gamma} - 1 \right) \tag{15}$$

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

$$\gamma \frac{\tau_{jk}(s) W_j}{z_{ij}(s)} = (\gamma - 1) \lambda_{ik}(s) \left(\frac{\tau_{jk}(s) p_{ijk}(s)}{P_{ik}(s)} \right) \tag{16}$$

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

$$p_{ijk}(s) = \mu_{ik}(s) \frac{W_j}{z_{ij}(s)}, \qquad \mu_{ik}(s) = \frac{\gamma}{\gamma - 1} \left(\frac{P_{ik}(s)}{\lambda_{ik}(s)} \right) \tag{17}$$

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 (9) and eliminating the multiplier gives

$$P_{ik}(s) = \mu_{ik}(s) \cdot \mathsf{MC}_{ik}(s) \tag{18}$$

where

$$\mathsf{MC}_{ik}(s) = \left(\sum_{j=1}^{J} \left(\frac{\tau_{jk}(s) W_j}{z_{ij}(s)} \right)^{1-\gamma} \right)^{\frac{1}{1-\gamma}} \tag{19}$$

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 (7), (13) and (18) to write the firm's profits from destination k

$$\begin{aligned}\Pi_{ik}(s) &= (P_{ik}(s) - \mathsf{MC}_{ik}(s))\ C_{ik}(s) \\ &= (P_{ik}(s) - \mathsf{MC}_{ik}(s)) \left(\frac{P_{ik}(s)}{P_k(s)}\right)^{-\gamma} \left(\frac{P_k(s)}{P_k}\right)^{-\theta} C_k \end{aligned} \tag{20}$$

with each firm internalizing the effect of their price $P_{ik}(s)$ on the sector-level price index $P_k(s)$ in (8). 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

$$\mu_{ik}(s) = \frac{\varepsilon_{ik}(s)}{\varepsilon_{ik}(s) - 1} \tag{21}$$

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

$$\varepsilon_{ik}(s) = \left(\omega_{ik}(s)\frac{1}{\theta} + (1 - \omega_{ik}(s))\frac{1}{\gamma}\right)^{-1} \tag{22}$$

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

$$\omega_{ik}(s) := \frac{P_{ik}(s)C_{ik}(s)}{\sum_{i=1}^{N(s)} P_{ik}(s)C_{ik}(s)} = \frac{P_{ik}(s)^{1-\gamma}}{\sum_{i=1}^{N(s)} P_{ik}(s)^{1-\gamma}} \tag{23}$$

Since the elasticity of substitution across firms is larger than 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 (21) and (22) to write the inverse markup as a linear function of the sales share

$$\frac{1}{\mu_{ik}(s)} = 1 - \frac{1}{\varepsilon_{ik}(s)} = \frac{\gamma - 1}{\gamma} - \left(\frac{1}{\theta} - \frac{1}{\gamma}\right) \omega_{ik}(s) \tag{24}$$

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 $P_{ik}(s) = \mu_{ik}(s)\mathsf{MC}_{ik}(s)$ into (23) to get

$$\omega_{ik}(s) = \frac{(\mu_{ik}(s)\mathsf{MC}_{ik}(s))^{1-\gamma}}{\sum_{i=1}^{N(s)}(\mu_{ik}(s)\mathsf{MC}_{ik}(s))^{1-\gamma}} \tag{25}$$

Here we see that, conditional on other firms' markups, each firm's market share is *strictly decreasing* in its markup. Together, equations (24) and (25) 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 (25) 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

$$\frac{1}{\mu_{ik}(s)} = \frac{\gamma - 1}{\gamma} - \left(\frac{1}{\theta} - \frac{1}{\gamma}\right) \frac{(\mu_{ik}(s)\mathsf{MC}_{ik}(s))^{1-\gamma}}{\sum_{i=1}^{N(s)}(\mu_{ik}(s)\mathsf{MC}_{ik}(s))^{1-\gamma}} \tag{26}$$

This condition implicitly determines the distribution of markups $\mu_{ik}(s)$ within and across locations as a function of the distribution of marginal costs $\mathsf{MC}_{ik}(s)$ within and across locations. The marginal costs $\mathsf{MC}_{ik}(s)$ are exogenous to each firm but, because they depend on the wages W_j, 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. The labor market in each location j clears when the total supply of efficiency units of labor $E_j L_j$ equals the total labor demand in that location

$$E_j L_j = \int_0^1 \sum_{k=1}^{J} \sum_{i=1}^{N(s)} l_{ijk}(s)\, ds \tag{27}$$

Solving the model. We solve the model as follows. We first solve the fixed point problem (26) for the function that maps marginal costs $\mathsf{MC}_{ik}(s)$ into markups $\mu_{ik}(s)$. We then guess a vector of wages W_j, with one wage normalized to 1 as numeraire. This vector of wages implies marginal costs $\mathsf{MC}_{ik}(s)$ and hence markups $\mu_{ik}(s)$, prices $P_{ik}(s)$ and quantities $C_{ik}(s)$, etc. We can then update the wage guess efficiently by exploiting the fact that, conditional on markups, budget constraints are linear in wages. The full details of our computational procedure are given in the Appendix.

We next briefly outline how the spatial distribution of markups $\mu_{ik}(s)$ affects aggregate productivity within and across locations.

2.2 Aggregation

Underlying all our aggregation results is an endogenous bilateral *productivity network*, a collection of productivity levels $\bar{z}_{jk}(s)$ for each sector s that forms a graph on the nodes j, k with directed edges from origins j to destinations k.

Productivity network. To derive this productivity network, first let $\bar{c}_{jk}(s)$ denote the composite formed from goods shipped from j to k within a given sector

$$\bar{c}_{jk}(s) = \left(\sum_{i=1}^{N(s)} c_{ijk}(s)^{\frac{1-\gamma}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}} \tag{28}$$

Similarly, let $\bar{l}_{jk}(s) = \sum_i l_{ijk}(s)$ denote the labor used to produce this composite and let $\bar{y}_{jk}(s) = \tau_{jk}(s)\bar{c}_{jk}(s)$ denote the amount of this composite that has to be produced at j for $\bar{c}_{jk}(s)$ to arrive at k. The productivity network is then given by the collection of $\bar{z}_{jk}(s) := \bar{y}_{jk}(s)/\bar{l}_{jk}(s)$. In keeping with this notation, let $\bar{\mu}_{jk}(s)$ denote the implied markup, satisfying $\bar{p}_{jk}(s) = \bar{\mu}_{jk}(s)W_j/\bar{z}_{jk}(s)$ where

$$\bar{p}_{jk}(s) = \left(\sum_{i=1}^{N(s)} p_{ijk}(s)^{1-\gamma}\right)^{\frac{1}{1-\gamma}} = \left(\sum_{i=1}^{N(s)} \left(\frac{\mu_{ik}(s)}{z_{ij}(s)}\right)^{1-\gamma}\right)^{\frac{1}{1-\gamma}} W_j \tag{29}$$

is the price index for the composite good. Then following standard arguments, as in Edmond, Midrigan and Xu (2015, 2023), we get productivity

$$\bar{z}_{jk}(s) = \left(\sum_{i=1}^{N(s)} \left(\frac{\mu_{ik}(s)}{\bar{\mu}_{jk}(s)}\right)^{-\gamma} z_{ij}(s)^{\gamma-1}\right)^{\frac{1}{\gamma-1}} \tag{30}$$

Notice that in the special case of no dispersion in markups this reduces to a standard CES productivity index that depends only on the exogenous firm-level productivity $z_{ij}(s)$. More generally, markup dispersion reduces $\bar{z}_{jk}(s)$ below this benchmark. Notice also that $\bar{z}_{jk}(s)$ does not depend on the trade costs $\tau_{jk}(s)$. This is because, within a given sector s, the trade costs $\tau_{jk}(s)$ apply to all firms shipping from j to k equally.

With this expression for $\bar{z}_{jk}(s)$ in hand, the markup $\bar{\mu}_{jk}(s)$ on the composite good is

$$\bar{\mu}_{jk}(s) = \left(\sum_{i=1}^{N(s)} \frac{1}{\mu_{ik}(s)} \, \omega_{ijk}(s) \right)^{-1} = \frac{\sum_{i=1}^{N(s)} \mu_{ik}(s)^{1-\gamma} \, z_{ij}(s)^{\gamma-1}}{\sum_{i=1}^{N(s)} \mu_{ik}(s)^{-\gamma} \, z_{ij}(s)^{1-\gamma}} \tag{31}$$

where

$$\omega_{ijk}(s) = \left(\frac{p_{ijk}(s)}{\bar{p}_{jk}(s)} \right)^{1-\gamma} = \left(\frac{\mu_{ik}(s)}{z_{ij}(s)} \frac{\bar{z}_{jk}(s)}{\bar{\mu}_{jk}(s)} \right)^{1-\gamma} \tag{32}$$

denotes the sales share of firm i in shipments from j to k. In short, as in Edmond, Midrigan and Xu (2015, 2023), the markup on the composite good is a sales-weighted harmonic average of the firm-level markups $\mu_{ik}(s)$ for destination k.

Location-specific productivity and markups. With the bilateral productivity network $\bar{z}_{jk}(s)$ and associated markups $\bar{\mu}_{jk}(s)$ determined, we can aggregate further to get location-specific productivity and markups. For example, let $\bar{Z}_k(s) := C_k(s)/L_k(s)$ denote real consumption per worker in location k. Following the same steps, this can be written

$$\bar{Z}_k(s) = \left(\sum_{j=1}^{J} \left(\frac{\bar{\mu}_{jk}(s)}{\bar{\mu}_k(s)} \right)^{-\gamma} \left(\frac{\bar{z}_{jk}(s)}{\tau_{jk}(s)} \right)^{\gamma-1} \left(\frac{W_j}{\bar{W}_k(s)} \right)^{-\gamma} \right)^{\frac{1}{\gamma-1}} \tag{33}$$

where $\bar{\mu}_k(s)$ denotes the location-specific markup, which again can be written as sales-weighted harmonic average of the underlying $\bar{\mu}_{jk}(s)$, and where the wage index $\bar{W}_k(s)$ is implicitly defined by $P_k(s) = \bar{\mu}_k(s)\bar{W}_k(s)/\bar{Z}_k(s)$. This expression for $\bar{Z}_k(s)$ is similar to that given for the productivity nodes $\bar{z}_{jk}(s)$ in equation (30) above, but differs in two ways. First, the expression for $\bar{Z}_k(s)$ also depends on trade costs $\tau_{jk}(s)$, since trade costs reduce the contributions that high-productivity sources j make to destinations k that are costly to ship to. Second, the expression for $\bar{Z}_k(s)$ also depends on the relative wage $W_j/\bar{W}_k(s)$, which is absent from the expression for $\bar{z}_{jk}(s)$ since those productivity nodes refer to firms who are all paying the same wage W_j to produce in j.[3]

Overall we see that markup dispersion reduces aggregate productivity, both because markup dispersion *across firms* within a given destination k directly reduces productivity $\bar{z}_{jk}(s)$ at each node in the productivity network, as in equation (30), and because for any

[3] This relative wage is also missing from the equivalent expression in Edmond, Midrigan and Xu (2015). They study a two-location model with a form of *aggregate symmetry* so that $W_j = \bar{W}_k(s)$ for $j, k = 1, 2$.

given network of $\bar{z}_{jk}(s)$, markup dispersion *across locations*, reduces aggregate productivity $\bar{Z}_k(s)$ at each destination k, as in equation (33). In this sense, the spatial dispersion in markups creates *endogenous misallocation*, both across firms and across locations.

3 Quantifying the model

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

3.1 Benchmark parameterization

Our geographical locations are Economic Areas (EAs) constructed by the US Bureau of Economic Analysis. There are $J = 170$ EAs in our data. Each EA is built around an urban core, either larger Metropolitan Statistical Areas (MSAs) or smaller Micropolitan Statistical Areas, along with adjacent counties with strong commuting ties. Importantly, these EAs are built to reflect the fact that economic activity spans administrative/jurisdictional boundaries. For example, the Chicago EA includes both the Chicago metropolitan area along with other counties in Illinois, Indiana, and Wisconsin where workers have strong connections to Chicago. EAs are very diverse in size, ranging from the greater Los Angeles area which accounts for around 14.6% of total employment, all the way down to Scottsbluff, Nebraska (near the Wyoming border) which accounts for 0.0012% of total employment:

1	Los Angeles-Riverside-Orange County	14.6%	of total employment
2	New York-North New Jersey-Long Island	7.2%	
3	Chicago-Gary-Kenosha	6.9%	
	⋮		
169	San Angelo, TX	0.0013%	
170	Scottsbluff, NE-WY	0.0012%	

Sectors. We calibrate our model to match the operations of firms in 363 NAICS 6-digit manufacturing sectors, examples of which include *breakfast cereal* (sector 311230), *ready-mix concrete* (327320), *aircraft engine & engine parts* (336412), *optical instrument & lens manufacturing* (333314), and *wood kitchen cabinet & countertop manufacturing* (337110).

Labor supply. For each $j = 1, ..., J$ we measure the number of workers L_j as manufacturing employment from the County Business Patterns (CBP) aggregated to the EA level. We likewise measure the wage bill from the CBP aggregated to the EA level and choose the efficiency units E_j for each location so that the wage bill in our model for each location $W_j E_j L_j$ matches the wage bill measured in the data for that location.

Firms and establishments. On average there are about 600 firms nationally per NAICS 6-digit sector, most of which are small. But there is considerable variation in the number of firms across sectors, ranging from a low of $N(s) = 8$ firms nationally for *custom roll forming* (sector 332114), a specialized manufacturing activity focused on the metal forming process, to $N(s) = 11,492$ for *machine shops* (332710) and $N(s) = 14,279$ for *commercial printing* (323111), which covers printing of stationery, advertising materials, etc. All together we have $N = \sum_s N(s) = 219,365$ firms. We set the number of establishments for each firm by counting the number of EA locations where firm i has at least one establishment in National Establishment Time Series (NETS) county-level data aggregated to the EA level. So, for example, if a given firm has an establishment in each of two counties that belong to the same EA we call that 'one' establishment for the purposes of our model.

Firms and establishment locations. We then populate the economy by placing each firm's establishments exactly where they appear in the data, so every model firm mirrors a real firm's geographic footprint. Let $\mathcal{J}_i(s) \subseteq \{1, ..., J\}$ denote the set of locations where firm i has establishments and let $n_i(s) = |\mathcal{J}_i(s)| = \sum_j \mathbb{1}\{j \in \mathcal{J}_i(s)\}$ denote the number of locations where firm i has establishments. We take this set $\mathcal{J}_i(s)$ exactly from the data.

Firm-level productivity fixed effects. We abstract from location-specific firm productivity effects and assume that firm-level productivity can be written

$$z_{ij}(s) = z_i(s) \cdot \mathbb{1}\{j \in \mathcal{J}_i(s)\} \tag{34}$$

In other words, a firm's productivity is equal to the firm-level productivity fixed effect $z_i(s)$ in all locations where it has an establishment and zero in all locations where it has no establishments. Since we have taken the set of locations $\mathcal{J}_i(s)$ for each firm directly from the data, the only thing left to do is to assign these firm-level productivity fixed effects.

To assign the firm-level productivity fixed effects, we simulate a large number of paired uniform ranks (u, v) using a Gumbel copula

$$\mathcal{C}(u, v) = \exp\left(-\left[(-\ln u)^{\frac{1}{1-\rho}} + (-\ln v)^{\frac{1}{1-\rho}}\right]^{1-\rho}\right) = \text{Gumbel Copula}(\rho) \tag{35}$$

We then transform the first rank into a Pareto productivity draw $z = F^{-1}(u)$ where

$$F(z) = 1 - z^{-\xi} = \text{Pareto}(\xi)$$

and transform the second rank into an empirical draw of establishment counts. For each establishment-count category, e.g., all firms with n establishments, we take the simulated pool of productivity-count pairs and condition on that count to obtain candidate productivity draws. We then randomly sample the exact number of real firms in that category and assign these productivity levels to the corresponding firms in their real geographic locations.

Example. Suppose in the data there are 1123 firms that have 10 establishments. In the model, we can place the establishments of each of these 1123 firms into the exact EA that they are in the data. But what we do not observe is their firm-level productivity. From the Gumbel simulation, however, we have an extremely large set of (productivity, establishment-count) pairs. We drop all pairs whose establishment count is not 10, which leaves us with a still very large sample of productivity draws from the conditional distribution of firm-level productivity (conditional on an establishment count of 10). We then randomly sample 1123 productivity draws with replacement from this conditional distribution and assign them to the 1123 firms we have already placed in their real geographic locations.

With this procedure in place, we are left with two parameters to pin down, the the Pareto tail ξ and the Gumbel rank correlation parameter ρ, as discussed further below.[4]

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}

$$\ln \tau_{jk}(s) = \delta(s) \ln d_{jk} \tag{36}$$

This gives us a further set of parameters to pin down, the trade cost coefficients $\delta(s)$.

3.2 Calibration

Calibration strategy. We assign values to two 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

[4]For the Gumbel copula in (35), the parameter $\rho \in [0, 1]$ corresponds to the robust rank correlation coefficient known as 'Kendall's tau' commonly used to summarize dependence in heavy-tailed distributions (Nelsen, 2006). If $\rho = 0$ the copula simplifies to $\mathcal{C}(u, v) = uv$ so that the ranks are independent, If $\rho \to 1$ the copula approaches $\mathcal{C}(u, v) = \min[u, v]$ so that the ranks are perfectly dependent. For simplicity we refer to ρ as the Gumbel correlation parameter.

Table 1: Parameterization

Parameter		Value	Target
Assigned Values			
Elas. substitution across sectors	θ	1.25	Sector-level markups and concentration
Elas. substitution within sectors	γ	10	(Edmond, Midrigan and Xu, 2023)
Method of Moments			
Pareto tail firm productivity	ξ	10.35	National concentration
Gumbel rank correlation	ρ	0.81	Employment share multi-estab firms
Trade cost wrt distance	$\delta(s)$		Gravity coeff. 3-digit NAICS

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 to $\gamma = 10$ so that the model matches the sector-level relationship between inverse markups and sales concentration observed in US manufacturing data.[5] This relationship can be obtained by multiplying both sides of equation (24) by market shares $\omega_{ik}(s)$ and summing across firms and locations to get

$$\frac{1}{\mu(s)} = \frac{\gamma - 1}{\gamma} - \left(\frac{1}{\theta} - \frac{1}{\gamma}\right) \text{HHI}(s) \tag{37}$$

where HHI(s) denotes the sector's Herfindahl-Hirschman index (HHI) of sales concentration.[6]

[5] Edmond, Midrigan and Xu (2023) calibrate their parameters to match the slope coefficient of this relationship, estimated in differences over time, jointly with their other parameters which target measures of concentration in 4-digit US Census of Manufactures data. For their preferred specifications, they obtain estimates of the across-sector elasticity of substitution θ between 1.15 and 1.35 and estimates of the within-sector elasticity γ between 7 and 13. We set $\theta = 1.25$ and $\gamma = 10$ as the rough midpoints of these ranges.

[6] That is, if we let ω_i denote the sales share of firm i, then the HHI $:= \sum_i \omega_i^2$. For example, if there are N firms with $\omega_i = 1/N$ then the HHI $= 1/N$.

Calibrated parameters. We are left with the need to calibrate the Pareto tail, the Gumbel correlation, and the trade cost coefficients for each sector

$$\xi \,,\, \rho \,,\, \delta(s)$$

We calibrate these parameters internally using the simulated method of moments. We jointly target (i) measures of national sales concentration, to pin down the Pareto tail parameter ξ, (ii) measures of the employment share of multi-establishment firms to pin down the Gumbel correlation parameter ρ, and (iii) sector-level gravity regressions to pin down the trade cost coefficients, $\delta(s)$. Importantly, we calibrate the model using measures of *national* sales concentration, not local concentration.

(i) NATIONAL SALES CONCENTRATION. We target the average top-4 national sales shares and national sales HHI. In the US Census of Manufactures, the average top-4 national sales share for 6-digit NAICS sectors is 42% 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 54% of employment.

(iii) GRAVITY REGRESSIONS. Recall that $\bar{p}_{jk}(s)\bar{y}_{jk}(s)$ denotes the value of shipments from j to k. We estimate sector-specific gravity regressions of the form

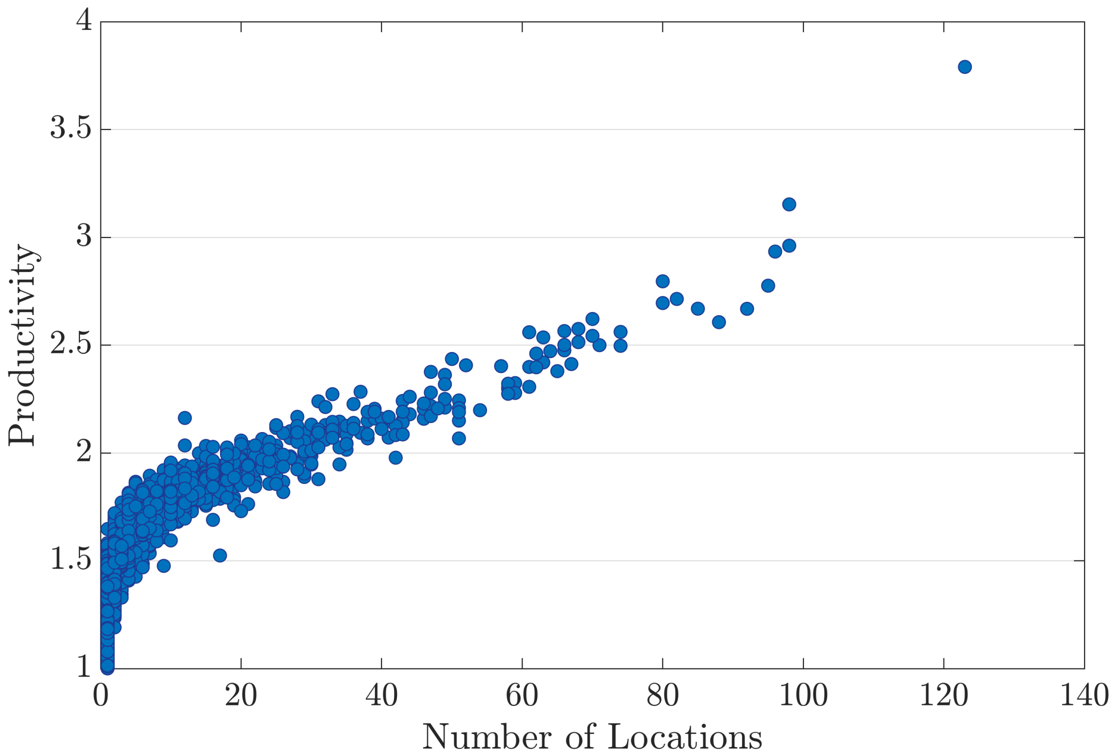
$$\ln(\bar{p}_{jk}(s)\bar{y}_{jk}(s)) = \gamma_j(s) + \gamma_k(s) + \beta(s)\ln d_{jk} + \epsilon_{jk}(s) \tag{38}$$

where $\gamma_j(s)$,$\gamma_k(s)$ denote sector-specific source and destination fixed effects. We estimate these gravity regressions using county-to-county trade flows from the Commodity Flows Survey aggregated to the EA level 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 tradeable, such as *computers & electronics* (sector 334) and *electric equipment & appliances* (335), have estimated $\beta(s)$ that are small in magnitude. Goods that are less easily tradeable, such as *wood* (321), *petroleum, asphalt and coal* (324) and *non-metallic minerals* (327) have large negative estimated $\beta(s)$.

In the model we simulate data for each of our 363 6-digit sectors and, for each sector s, calculate the total value of shipments from j to k as

$$\bar{p}_{jk}(s)\bar{y}_{jk}(s) = \sum_{i=1}^{N(s)} p_{ijk}(s)y_{ijk}(s). \tag{39}$$

Figure 1: Firm-Level Productivity $z_i(s)$ and Establishment-Count $n_i(s)$



> To be consistent with our empirical gravity regressions from the Commodity Flows Survey, we aggregate these simulated shipment flows to a 3-digit cluster of sectors and choose the parameters $\delta(s)$ in our specification (36) so that the estimated $\beta(s)$ in the model gravity regressions match their empirical counterparts from (38).

We report our internally calibrated parameters governing the productivity distribution and the operations of multi-establishment firms across locations in Table 1. Jointly with our other parameters, our model matches the data on national sales concentration with a Pareto tail $\xi = 10.35$, implying considerably thinner tails than the model of oligopolistic competition in Edmond, Midrigan and Xu (2023), which abstracts from spatial frictions. Our model matches the 54% employment share of multi-establishment firms with a Gumbel correlation of $\rho = 0.81$ between a firm's productivity fixed effect $z_i(s)$ and its number of establishments $n_i(s)$, as illustrated in Figure 1.

Model fit. We report key moments in the data and their model counterparts in Table 2, highlighting in red the moments targeted by our calibration procedure. The model does a

Table 2: Model Fit

Moments [targeted]	Data	Model
National Concentration		
Top 4 sales share	0.42	0.44
Top 20 sales share	0.73	0.65
HHI sales	0.10	0.10
Local Concentration		
HHI production	0.36	0.37
Multi-Establishment Firms		
Fraction multi-establishment firms	0.03	0.03
Employment share of multi-establishment firms	0.54	0.53
Sales share of multi-establishment firms	0.62	0.55

good job of reproducing the average amount of national sales concentration, matching the national sales HHI exactly and slightly overshooting the national top-4 sales share. The model also reproduces the employment share of multi-establishment firms almost exactly. We report the 3-digit gravity coefficients $\beta(s)$ we estimate from the CFS and their model counterparts in Figure 2. Importantly, our model exactly reproduces the sector-level gravity effects that pin down our spatial trade frictions.

Model validation. In 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. That said, the model undershoots the national top-20 share and the sales share of multi-establishment firms.

Sales concentration in local 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

Figure 2: Gravity Coefficients $\beta(s)$ in Data and Model

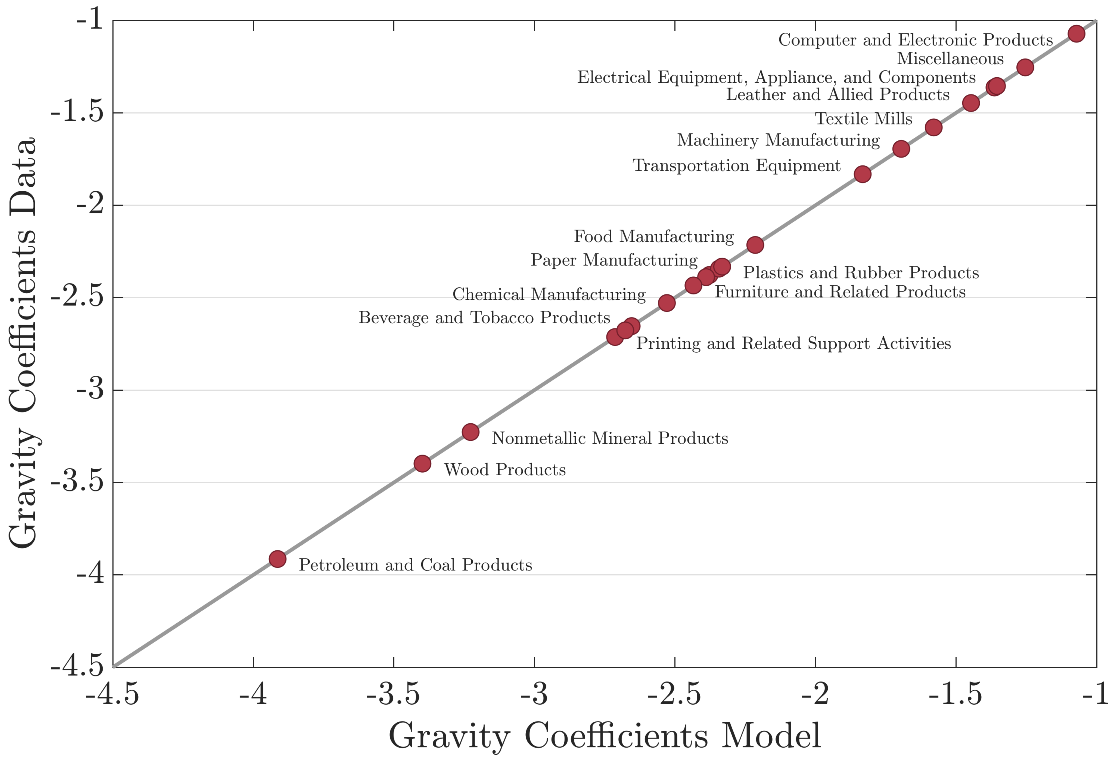


Table 3: Sales Concentration in Local Markets

Moment	Model
Local Sales Concentration	
Top 1 sales share	0.27
Top 4 sales share	0.58
Top 20 sales share	0.85
HHI sales	0.15

concentration. We report our benchmark model's implication for local sales concentration in Table 3. Intuitively, we find that:

National Sales		**Local Sales**		**Local Production**
HHI = 0.10	<	HHI = 0.15	<	HHI = 0.37
least concentrated				*most concentrated*

Since the inverse of the HHI corresponds to the number of equally-sized firms, to interpret these concentration statistics more intuitively, it is as if there are on average 10 equally-sized firms nationally, about 6-7 equally-sized firms selling locally, and just under 3 equally-sized firms producing locally. The fact that local sales concentration is less than local production concentration reflects the fact that most goods are at least somewhat tradeable. While the production of goods may be quite concentrated at specific source locations, destination markets generally receive goods from a range of sources, pushing local sales concentration lower than production concentration. That said, the fact that local sales concentration is greater than national sales concentration reflects the fact that goods cannot be traded *frictionlessly*, i.e., the economy is genuinely geographically segmented. In the next section we document these results more systematically and show that the aggregate implications of this geographic segmentation for productivity and consumer welfare can be large.

4 Quantitative importance of spatial frictions

In this section we present two results highlighting the importance of spatial frictions for sector-level outcomes across locations, with a special emphasis on patterns of concentration and measures of competition and market power. First, we show that measures of local production concentration, of the kind readily computed from data on shipments, provides a poor guide to the amount of competition firms face in the locations where people live and consume. Second, we go on to show that intranational spatial frictions matter in the aggregate. In particular, an otherwise equivalent model that abstracts from spatial frictions leads to a quantitatively significant understatement of both the aggregate markup and the aggregate productivity losses due to markup dispersion.

4.1 Production concentration does not explain sales concentration

In our model, spatial frictions shape the amount of local competition. Goods that are easily tradeable can be shipped from the most productive source locations to almost any destination market, increasing the amount of competition amongst producers of tradeable goods in those markets. Goods that are less easily tradeable will be shipped to a more limited set of destinations, inhibiting the amount of competition amongst producers of less-tradeable 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 3 reports average local production concentration (as measured by HHIs) across the 170 EAs in our model. These range from around 0.16 to nearly 1, indicating areas where many goods are produced by a single firm. Figure 4 reports average local sales concentration (again measured by HHIs) in our model. These range from around 0.12 to 0.17. In short, sales concentration is both much lower on average and much less variable than production concentration. Moreover production concentration and sales concentration are not strongly correlated across locations. For example, the greater New York City area has both low production and sales concentration, while the greater Seattle area has only moderate production concentration but some of the highest sales concentration in our model.

To reinforce this point, Figure 5 reports the scatter of local production concentration and local sales concentration HHIs for the 170 EAs in our model. If local production concentration was a good predictor of local sales concentration, we would expect this scatter to be clustered tightly around the 45°-line. But instead the scatter is close to a horizontal line — local production concentration is simply not very informative about local sales concentration. If anything, the slope coefficient is slightly negative, indicating that having higher production concentration predicts that a location will have *lower* sales concentration. In short, the usual kind of local production concentration that we can readily measure with data on shipments is simply not very informative about the local sales concentration that matters for competition and measures of market power.

Spatial frictions and local sales concentration. To further highlight the importance of tradeability, Figure 6 reports the scatter of local production HHIs against local sales HHIs when we split 3-digit sectors into 'high gravity' sectors with high spatial frictions that are relatively costly to ship across locations, e.g., *petroleum & coal* products or *wood* products and 'low gravity' sectors with low spatial frictions that are much less costly to ship across locations, e.g., *computer & electronic* products. On average, high gravity sectors have higher levels of sales concentration than low gravity sectors. In high gravity sectors, trade costs can substantially limit the amount of competition — especially for smaller and more geographically remote locations. High gravity sectors also have sales concentration that is considerably more variable across locations than low gravity sectors. Indeed we see that for low gravity sectors, the local sales HHIs are tightly clustered at just above 0.1, i.e., just above the national sales HHI. For these low gravity sectors, goods are traded in something much closer to a single national market.

Figure 3: Local Production Concentration (HHI)

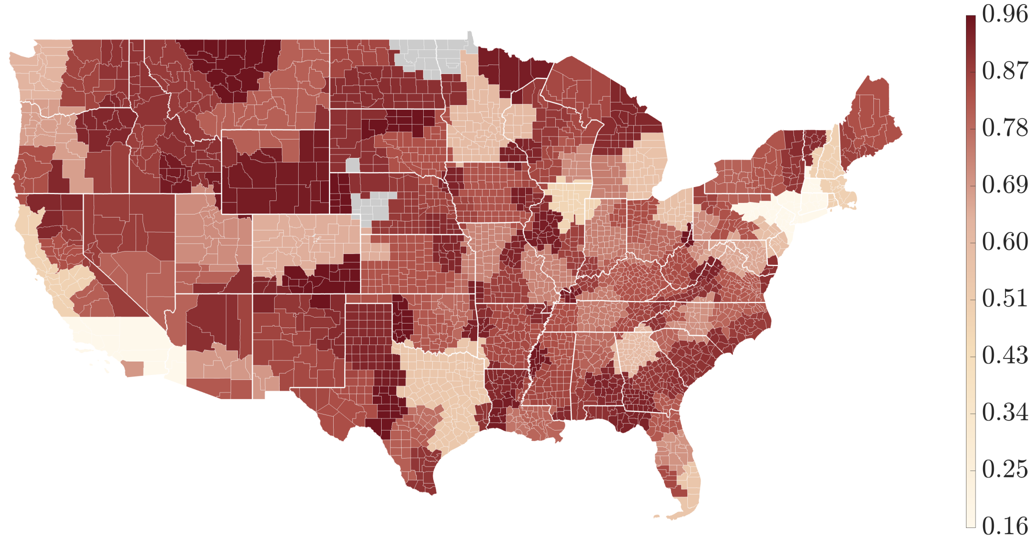


Figure 4: Local Sales Concentration (HHI)

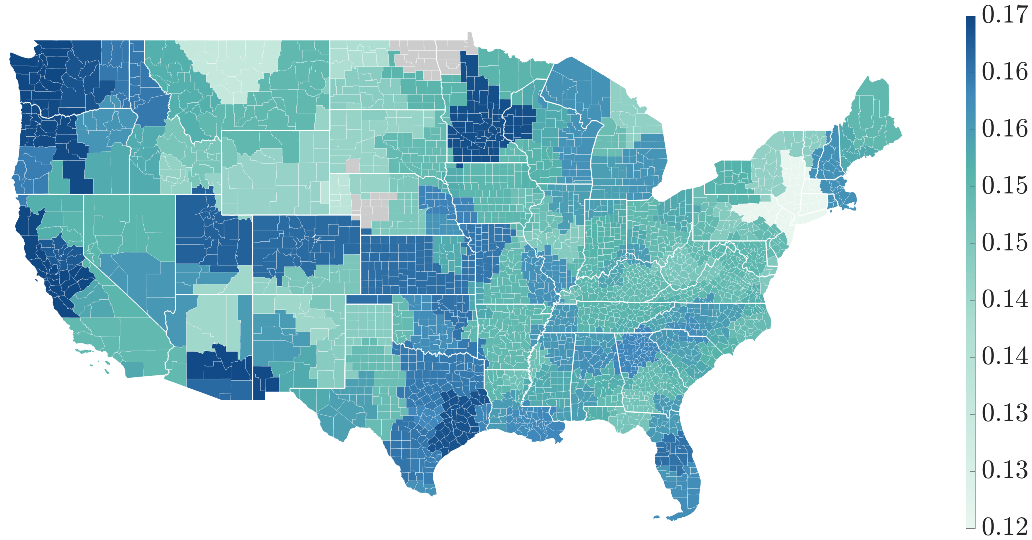


Figure 5: Production HHI $\neq$ Sales HHI

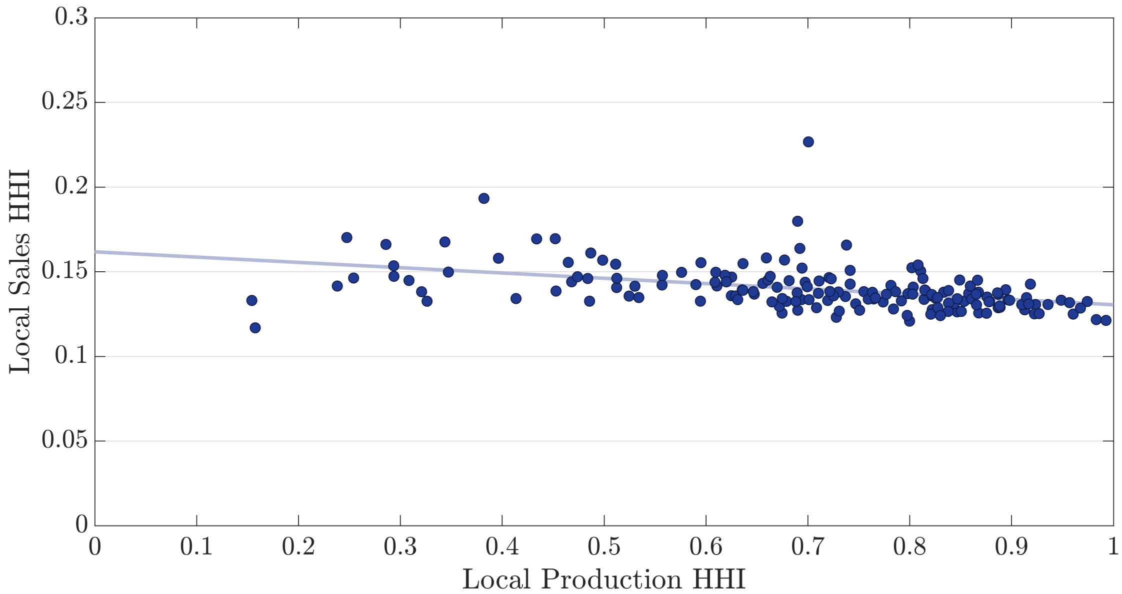


Figure 6: High Gravity Sectors Have Higher Sales HHI

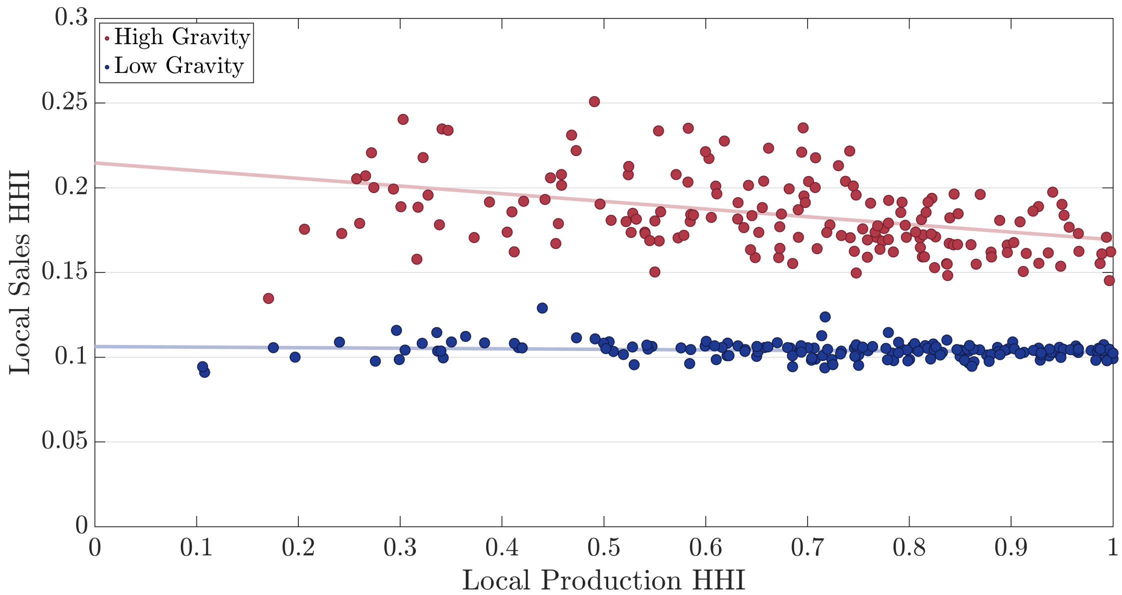
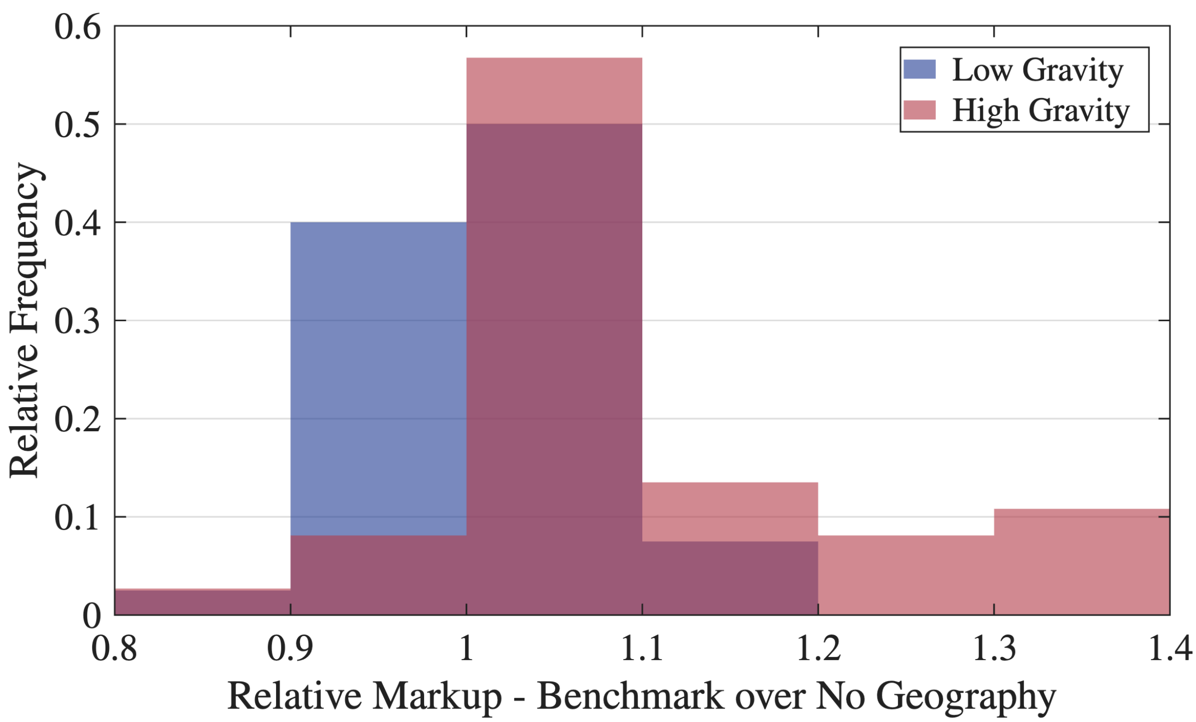


Table 4: Markup Distribution, No Geography

Percentile	Benchmark Model	No Geography
p01	1.13	1.13
p10	1.15	1.14
p25	1.18	1.16
p50	1.23	1.18
p75	1.30	1.22
p90	1.41	1.29
p99	1.62	1.50
Aggregate Markup	1.26	1.19

Figure 7: Role of Gravity



4.2 Abstracting from spatial frictions understates market power

We next quantify the significance of geography and spatial frictions for aggregate outcomes. To do this, we use an otherwise equivalent model that abstracts from geography and spatial frictions but matches the same facts on national sales concentration as our benchmark model. The results of this exercise are given in Table 4, which reports the distribution of sector-level markups and the aggregate markup. Across all sectors, markups in the model without geography are both lower and less dispersed than in our benchmark model with geography. For example, the median markup falls from 1.23 in our benchmark model with geography to 1.18 in the model without geography. In terms of markup dispersion, for our benchmark model with geography, the log p90/p50 markup ratio is about $\ln(1.41/1.23) = 0.137$, which falls to about $\ln(1.29/1.18) = 0.089$, i.e., about two-thirds as much dispersion, in the model without geography. Overall the model without geography implies both a lower aggregate markup, down from 1.26 to 1.19, and lower productivity losses due to misallocation. In this sense, abstracting from spatial frictions leads to a quantitatively significant *understatement* of the macroeconomic losses associated with market power.

To reinforce this point, we again split 3-digit 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, is 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.

5 Divergence between local and national 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.

5.1 Reduction in trade costs and diverging trends in concentration

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 (2024) 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 5 reports the effects of such changes in trade costs on concentration. Moving forward from 1963 to the present, the model predicts that in response to a reduction in trade costs the average national top-4 sales share increases modestly from 0.43 to 0.44 while the average local top-4 sales share decreases from 0.61 to 0.58. In short, the model predicts that a reduction in trade costs endogenously drives national sales concentration and local sales concentration in opposite directions. Intuitively, in response to lower trade costs, the most productive firms expand by accessing more distant markets. This increases national sales concentration and increases local production concentration. But this pattern of expansion also leads to more competition in destination markets and hence lower local sales concentration.

Divergence or convergence? It is conventional in this literature to refer to these opposite movements in national and local sales concentration as a form of *divergence* — and we have used this language too. But from the perspective of our model, the opposite-signed changes in national and local sales concentration in response to a change in trade costs are in fact a kind of *convergence*, with national sales concentration increasing from low initial levels and local sales concentration decreasing from high initial levels. In the limit as intranational trade costs disappear, the distinction between national and local sales concentration also disappears. This can be seen in the last column of Table 5, which shows that in this 'free trade' limit (while maintaining labor immobility), the national concentration moments exactly equal the local concentration moments. An observer of this process would see a stark pattern of increasing national concentration and decreasing local concentration, e.g., with the average national top-4 sales share starting at 0.46 in our benchmark economy and increasing from below to its limit of 0.57 while the average local top-4 sales share starts at 0.69 in our benchmark economy and decreases from above to the same limit. In this limit, there is effectively a single national market for each good.

5.2 Implications for markups and consumption

This exercise makes clear that a reduction in intranational trade costs can lead to an increase in national sales concentration. We now show that nonetheless this reduction in trade costs makes markets more competitive and increases consumption per worker in most locations, despite the increase in national sales concentration. The increase in national sales concentration is simply a byproduct of the most productive firms being able to sell in more locations.

But what really matters for competitive conditions is how much competition these firms face in the markets where they sell their goods.

Changes in markup distribution. To understand these changes in competitive conditions, Table 6 reports the effects of changes in intranational 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. For example, starting in 1963 with 20% higher trade costs and moving forward to our benchmark economy, we find that the aggregate markup decreases from 1.27 to 1.26 while dispersion, as measured by the log p90/p50 markup ratio decreases from about $\ln(1.44/1.24) = 0.150$ to about $\ln(1.41/1.23) = 0.137$. Further decreases in trade costs in turn lead to further decreases in the aggregate markup and markup dispersion. That said, while the qualitative direction of these changes is clear, the changes are modest — e.g., with the aggregate markup decreasing by about 0.8% in response to a 20% reduction in trade costs from 1963 to our benchmark economy. But as we will now see, these modest changes in the sector-level markup distribution mask considerable heterogeneity in markup changes across locations.

Spatial heterogeneity in markup changes. To see the heterogeneity in markup changes across locations, Figure 8 reports the percentage decrease in average markups across our 170 EAs in response to a 20% reduction in intranational trade costs. There is considerable spatial variation in markup changes, ranging from around a 0.5% decrease in the greater metropolitan area centered on New York City and around a 0.6% decrease in the greater Los Angeles area to around a 1.8% or 1.9% decrease in rural Utah, Colorado and Kansas. In other words, the markup changes in more remote locations are up to around 4 times as large as the markup changes in more central locations. Intuitively, the pro-competitive effects of a given reduction in trade costs are strongest for the most 'closed' locations and weaker for more 'open' locations where producers already face substantial competition (see Edmond, Midrigan and Xu, 2015, for further discussion of these pro-competitive effects).

Table 5: Changes in Trade Costs

	20% Increase	Benchmark	20% Decrease	Free Trade
	Increasing National Sales Concentration ⇑			
Top 4 share	0.43	0.44	0.45	0.49
HHI sales	0.10	0.10	0.10	0.11
	Increasing Local Production Concentration ⇑			
HHI production	0.36	0.37	0.38	0.40
	Decreasing Local Sales Concentration ⇓			
Top 4 share	0.61	0.58	0.56	0.49
HHI sales	0.16	0.15	0.13	0.11

Table 6: Effects of Changes in Trade Costs on Markup Distribution

Percentile	20% Increase	Benchmark	20% Decrease	Free Trade
p01	1.13	1.13	1.12	1.12
p10	1.16	1.15	1.14	1.13
p25	1.19	1.18	1.17	1.15
p50	1.24	1.23	1.21	1.19
p75	1.32	1.30	1.28	1.26
p90	1.44	1.41	1.39	1.34
p99	1.66	1.62	1.58	1.52
Aggregate Markup	1.27	1.26	1.24	1.21

Figure 8: Decrease in Markups

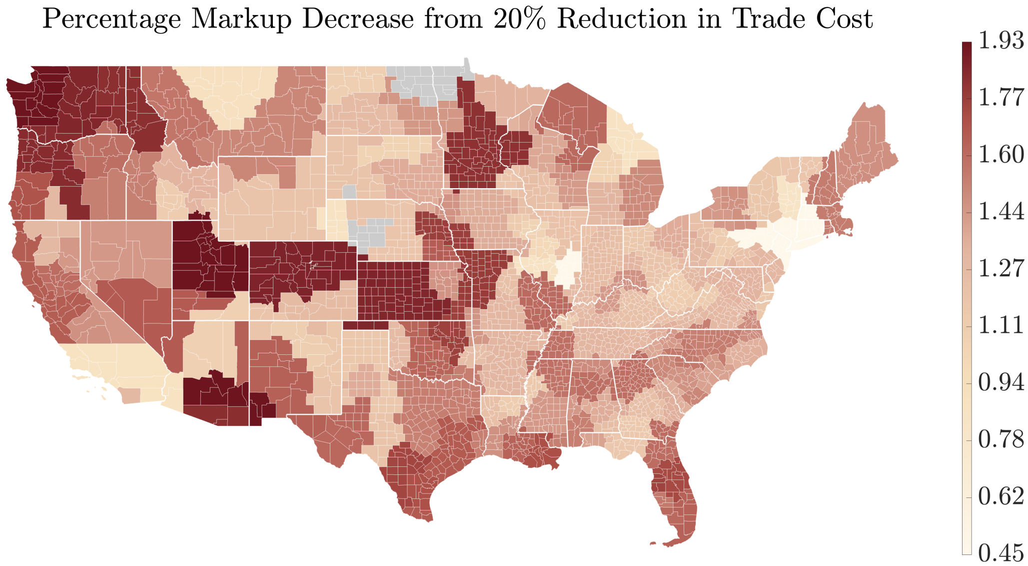
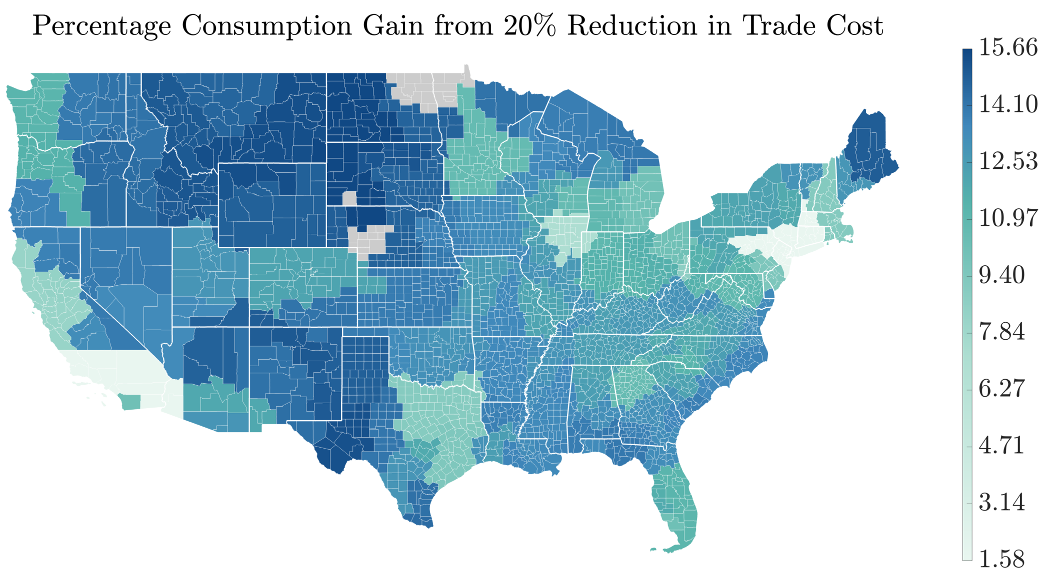


Figure 9: Total Consumption Gains



Spatial heterogeneity in consumption gains. Figure 9 reports the percentage increase in final consumption C_k for each of our 170 EAs in response to a 20% reduction in intranational trade costs.[7] There is considerably more spatial variation in consumption gains, ranging from around a 1.6% increase in the greater metropolitan area centered on New York City to increases of nearly 16% in remote parts of North and South Dakota. That is, there are nearly 10-times differences in consumption gains across locations as compared to the 4-times differences in markup decreases across locations. The largest consumption gains are generally in remote locations with more modest gains in coastal and large metro areas.

To be clear, these consumption gains from the reduction in intranational trade costs are driven by both the pro-competitive effects — with reduced markup dispersion reducing the productivity losses due to misallocation — and the standard Ricardian gains from trade that we would have in an otherwise equivalent model with constant markups. In the next section we isolate the markup channel by calculating the consumption-equivalent welfare gains from eliminating markups holding trade costs fixed.

6 Welfare costs of markups

In this section we quantify the welfare costs of markups in each location. We measure these welfare costs by asking how much the representative consumer in each location would gain in consumption from policies that eliminate markups. Since our benchmark model features inelastic factor supply, all of the gains from eliminating markups are due to eliminating markup *dispersion*, i.e., to reducing the productivity losses due to misallocation. We find that the average welfare costs of markups are large, about 5.8% in consumption-equivalent terms and vary considerably across locations, from 1% or less in the largest, richest, most central locations to more than 20% in the smallest, poorest, most remote locations.

Eliminating markups. A simple policy that eliminates markups is to pay each firm i in sector s a destination-specific sales subsidy $\chi_{ik}(s) \geq 1$ to induce the firm to set establishment-level prices equal to establishment-level marginal cost. From (17) we get

$$\chi_{ik}(s) \cdot p_{ijk}(s) = \mu_{ik}(s) \cdot \frac{W_j}{z_{ij}(s)} \tag{40}$$

which evidently induces marginal-cost pricing when the subsidy equals the markup, $\chi_{ik}(s) = \mu_{ik}(s)$ for all j. The subsidy is independent of establishment location j because the firm's

[7] Since labor L_k for each location is fixed in our benchmark model, this is equivalent to the percentage increase in consumption per worker C_k/L_k.

optimal markup is independent of establishment location. To isolate the welfare costs of markup distortions we suppose that these subsidies are funded by lump-sum taxes.

Costs of markups. Table 7 reports our main results for the welfare costs of markups in our benchmark model. On average, the representative consumer would gain 5.8% in consumption-equivalent terms from eliminating markups. Since factor supply is inelastic in our benchmark model, this is entirely due to the location-specific productivity gains from the reduced misallocation that results from eliminating markup dispersion so that relative prices are aligned with relative marginal costs everywhere. The median is similar, 5.6%. When we compute the costs of markups in our model without geography, as discussed in Section 4.2 above, we find that the representative consumer would gain 3.7% in consumption-equivalent terms from eliminating markups.[8] So again we see that abstracting from spatial frictions leads to a quantitatively significant understatement of the welfare costs of markups. Just as importantly, there is also substantial variation in the welfare costs of markups across locations — the welfare costs of markups are generally large and *very unevenly distributed.*

Spatial heterogeneity in the welfare costs of markups. To see this unevenness, Figure 10 reports the full geographic variation in the consumption gains from eliminating markups across our 170 EAs. Naturally, the welfare costs of markups are larger in locations where markups are larger to begin with. Controlling for initial markups, we also find that (i) locations which have lower initial levels of consumption per worker experience larger gains, and (ii) locations which have higher initial trade shares experience larger gains from eliminating markups.

In this sense, policies that eliminate markups do not just have important aggregate welfare benefits, they have stark implications for the relative gains experienced by different geographic areas, with smaller, poorer, more remote locations benefitting from the elimination of markups to a considerably greater extent than larger, richer, more central locations. Overall, we find that in the smallest, poorest, most remote locations the welfare costs of markups can be as large as 20% in consumption-equivalent terms — say 3-4 times as large as the average — but can be small, or even occasionally negative, in the largest, richest, most central locations.[9]

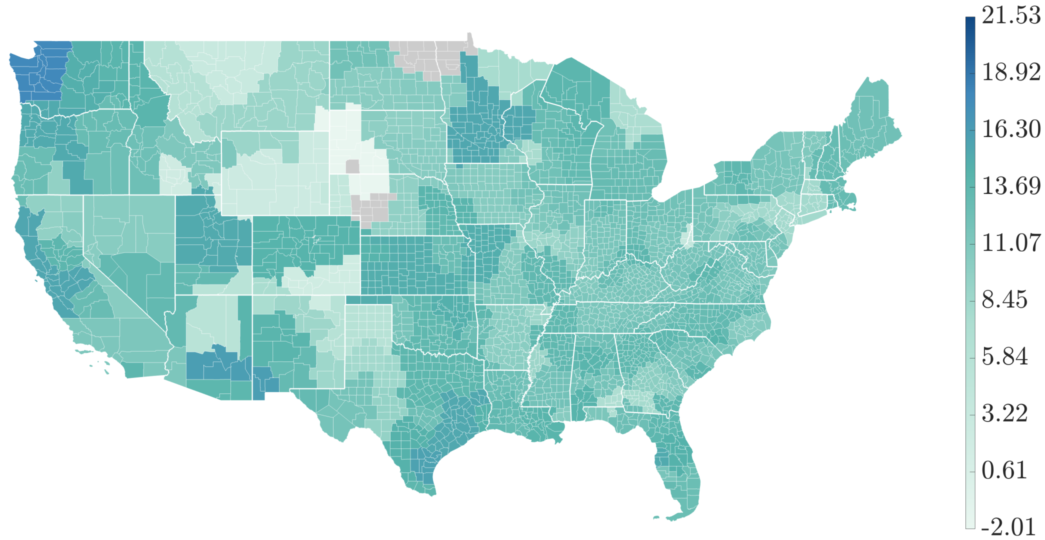
[8] Likewise Edmond, Midrigan and Xu (2023) find that the aggregate productivity losses due to misallocation in their oligopolistically competitive model that abstracts from spatial frictions are about 3-4%.

[9] Recall that we assume that profit income is distributed in proportion to labor income, i.e., that profit income is relatively high in high income locations. But large, highly competitive locations will have low markups. This means that when markups are eliminated, such locations experience a relatively small benefit from increased competition that can be overwhelmed by the loss of profit income resulting in a consumption loss. As can be seen from Figure 10 this is a quite rare occurrence .

Table 7: Eliminating Markups: Benchmark Model

Percentile	Consumption Gain, %
p01	0.9
p10	3.6
p25	3.9
p50	5.6
p75	6.9
p90	9.1
p99	14.5
Average	5.8

Figure 10: Consumption Gains From Eliminating Markups



7 Labor mobility

In this section we consider an extension of our benchmark model to allow for workers to move across locations. We use a setup in the spirit of Lagakos and Waugh (2013) and Kline and Moretti (2014) where workers have different preferences for different location-specific amenities, characterized by idiosyncratic differences in Fréchet draws, to pin down labor supply to each location. One might reasonably guess that labor mobility acts to *mitigate* the welfare costs of markups — workers can now move to locations that provide higher wages and/or lower prices. But quantitatively we find that labor mobility does little to mitigate the welfare costs of markups once we parameterize the model with labor mobility to match the same initial allocation of labor across locations as in our benchmark calibration.

Location-specific amenities. Each location $j = 1, ..., J$ is characterized by a location-specific amenity value $A_j > 0$ that is common to all workers (e.g., the location's climate). Each worker is characterized by a vector of idiosyncratic amenity draws, one for each location

$$\boldsymbol{v} = (v_1, v_2, ..., v_J) \tag{41}$$

Specifically, we assume that each v_j is drawn IID from a standard Fréchet distribution

$$\text{Prob}[v_j \leq v] = \exp(-v^{-\sigma}), \qquad \sigma > 0 \tag{42}$$

with tail parameter σ. As in our benchmark model, a worker that supplies labor in location j provides E_j efficiency units of labor.

Location choice. Let $c_j(\boldsymbol{v})$ denote the consumption of a worker of type $\boldsymbol{v}$ if they choose location j and let $u_j(\boldsymbol{v})$ denote their payoff from this choice. Consumption satisfies the individual worker's budget constraint

$$P_j c_j(\boldsymbol{v}) = (1 + \bar{\pi}) W_j E_j \tag{43}$$

where as in the benchmark model we continue to assume that profit income is paid out in proportion to labor income for some constant $\bar{\pi} \geq 0$ to be determined in equilibrium. A worker's payoff from choosing location j is then given by

$$u_j(\boldsymbol{v}) = A_j v_j(\boldsymbol{v}) c_j(\boldsymbol{v}) = A_j v_j(\boldsymbol{v})(1 + \bar{\pi}) \frac{W_j}{P_j} E_j \tag{44}$$

where $v_j(\boldsymbol{v})$ denotes the specific amenity draw for location j of a worker of type $\boldsymbol{v}$. The problem of an individual worker is to choose a location j that maximizes their payoff

$$u(\boldsymbol{v}) = \max_{j=1,...,J} u_j(\boldsymbol{v}) \tag{45}$$

Labor supply. Let $\bar{L} > 0$ denote the total mass of workers. Following standard Fréchet calculations, the mass of workers supplying labor to location j is given by

$$L_j = \frac{\left(A_j E_j \frac{W_j}{P_j}\right)^{\sigma}}{\sum_j \left(A_j E_j \frac{W_j}{P_j}\right)^{\sigma}} \bar{L} \tag{46}$$

Profit income per location cancels out because of our assumption that locations receive profit income in proportion to labor income. In short, we have a labor supply curve, increasing in the real wage W_j/P_j for each location, with elasticity given by the Fréchet tail parameter σ. A higher σ reduces the dispersion in idiosyncratic amenity draws and so makes relative labor supply across locations more responsive to relative differences in real wages.

Labor market clearing. The labor market in location j clears when the total supply of efficiency units of labor $E_j L_j$ equals the total labor demand in that location

$$E_j L_j = \frac{\left(A_j E_j \frac{W_j}{P_j}\right)^{\sigma}}{\sum_j \left(A_j E_j \frac{W_j}{P_j}\right)^{\sigma}} \bar{L} = \int_0^1 \sum_{k=1}^{J} \sum_{i=1}^{N(s)} l_{ijk}(s)\, ds \tag{47}$$

Parameterization. To solve this version of the model we fix the efficiency units of labor E_j at their benchmark values and assign a labor supply elasticity of $\sigma = 2$, in line with the range of values discussed by Fajgelbaum, Morales, Serrato and Zidar (2019). We choose the common location-specific amenity values A_j so that the model replicates manufacturing employment L_j from the County Business Patterns aggregated to the EA level. In other words, we choose A_j so that the model with labor mobility rationalizes the allocation of employment across locations in our benchmark calibration.

Quantitative results. Table 8 reports the effects of eliminating markups, across key percentiles of the distribution, for the version of the model with mobile labor and our benchmark model with immobile labor. For the model with mobile labor we report both the percentage changes in consumption per worker, C_j/L_j, and in aggregate consumption, C_j. In terms of consumption per worker, we find that labor mobility makes little difference quantitatively. For example, the average gain in consumption per worker from eliminating markups is 5.8% in both the model with mobile labor and the benchmark. Only in the upper tail of the distribution of consumption gains do we find noticeable differences, with the p90 gain in consumption per worker falling from 9.1% in the benchmark model to 8.9% in the model with labor mobility. In this sense, labor mobility mitigates the losses due to markups in those locations that are most severely impacted by markup distortions — but even in this

Table 8: Eliminating Markups: Labor Mobility

	Immobile	Mobile	
Percentile	C_j/L_j	C_j/L_j	C_j
p10	3.6	3.7	-0.4
p25	3.9	4.1	0.9
p50	5.6	5.6	5.4
p75	6.9	6.9	9.3
p90	9.1	8.9	15.6
Average	5.8	5.8	6.0

upper tail the size of the mitigating effect is small. For most locations, labor mobility has has even smaller effects on the welfare costs of markups.

That said, the model with labor mobility implies more substantial differences in the changes in aggregate consumption and employment across locations. For example, the p75 gain in consumption per worker is the same, 6.9%, in both models. But with mobile labor, that 6.9% gain can be decomposed into a 9.3% gain in aggregate consumption and an approximately 2% increase in employment. Likewise, the p90 gain in consumption per worker of 8.9% with mobile labor can be decomposed into a large 15.6% gain in aggregate consumption and a 6-7% increase in employment. The locations which gain the most from the elimination of markups see large increases in employment. Intuitively, there are large labor inflows to locations which are growing substantially. At the other end of the spectrum, the p10 gain of consumption per worker of 3.7% with mobile labor masks a 0.4% *decrease* in aggregate consumption and an approximately 4% decrease in employment. These locations still gain from the elimination of markups in consumption per worker terms, but they are shrinking both in absolute terms and relative to the rest of the economy. These implications for the reallocation of labor across locations are of course absent from our benchmark model.

8 Conclusion

We study the spatial distribution of economic activity 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 across 170 US Economic Areas. We show spatial frictions can have large aggregate effects, increasing both the aggregate markup and the productivity losses due to misallocation. We 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. We also show that the welfare costs of markups are large on average and very unevenly distributed. Smaller, poorer, more remote locations have costs some 20 times the costs of larger, richer, more central locations.

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, "The Long-Run Effects of Transportation Productivity on the US Economy," December 2024. NBER Working Paper 33248.

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.

Fajgelbaum, Pablo D., Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar, "State Taxes and Spatial Misallocation," *Review of Economic Studies*, January 2019, *86* (1), 333–376.

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.

Kline, Patrick and Enrico Moretti, "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs," *Annual Review of Economics*, August 2014, *6*, 629–662.

Lagakos, David and Michael E. Waugh, "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, April 2013, *103* (2), 948–980.

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

Nelsen, Roger B., *An Introduction to Copulas*, 2nd ed., Springer, 2006.

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.*