The Granular Origins of Agglomeration *

Shinnosuke Kikuchi
MIT

Daniel G. O'Connor
MIT

October 10, 2024

Abstract

A few large firms dominate many local labor markets. How does that granularity affect the geography of economic activity? And what does it mean for the efficiency of firm entry? To answer these questions, we propose a new economic geography model featuring granular firms subject to idiosyncratic shocks. We show that average wages increase in the size of the local labor market due to that granularity, and provide a sufficient statistic for the contribution of our mechanism. We further prove that too few firms enter in equilibrium. Using Japanese administrative data on manufacturing, we provide evidence consistent with our mechanism and quantify it. Our mechanism implies that markets with around 2 firms per sector have an elasticity of wages to population of 0.05 and firms capture only 85% of their contribution to production in profits. In large markets like Tokyo, the elasticity is around 0.001, and firm entry is approximately efficient. Enacting optimal place-based industrial policy would increase the number of firms in modest-sized cities by more than 30% and actually decrease the number of firms and people in Tokyo.

*We thank Daron Acemoglu, Treb Allen, Kosuke Aoki, David Atkin, David Autor, Arnaud Costinot, Dave Donaldson, Enrico Moretti, Tomoya Mori, Bob Staiger, Iván Werning, and Nathan Zorzi for their helpful comments. We also thank seminar participants at Dartmouth, Hitotsubashi, MIT, RIETI, and UTokyo. We thank Satoshi Ichikawa and Tomoko Yamaguchi for excellent research assistance. This study is a part of the project "Macroeconomy and Automation" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). Kikuchi acknowledges permission to access the microdata of the Census of Manufacture (CoM) and the Economic Census for Business Activity (ECBA), granted by the Ministry of Economy, Trade and Industry and the Ministry of Internal Affairs and Communications in Japan. We thank Kazunobu Hayakawa for providing us with a crosswalk file between the HS product code and the product code in the CoM survey used in Baek et al. (2021).

1 Introduction

A few large firms dominate many local labor markets. Kodak accounted for almost a quarter of Rochester, New York's city payroll at its peak. Toyota hires a large proportion of the workforce in its headquarter-city Toyota. And even in a large city like Seattle, software engineers are at the mercy of Microsoft. With these giants, a shock to a single firm can hurt the entire market. If Microsoft has a bad year and lays off a large proportion of its software engineers, those workers might end up unemployed or at a low-wage job in another industry. How does this exposure to firm-specific shocks shape where workers and firms are located? And is there room for place-based policies to insulate labor markets from a single firm's influence?

In this paper, we study these questions both theoretically and empirically. Our analysis builds on the basic idea that in a small market, if an individual firm has a bad shock, the workers have nowhere else to go. They are stuck at that unproductive firm. By contrast, in a large market, when a single firm becomes less productive, workers can move to another firm that is doing better and use their skills more effectively. Thus, large markets provide a "constant market for skill" as Marshall (1920) said and Krugman (1992) formalized. This labor market pooling mechanism implies that larger markets use labor more effectively than small markets, so that there are benefits to agglomeration when individual firms matter.

Central to our contribution is a new economic geography model with a continuum of sectors, each of which have a finite number of (heterogeneous) firms, and workers that are imperfectly mobile across both firms and sectors. With this general model, we can derive a sufficient statistic for the contribution of granularity to the wage premium of large cities. We can then transparently tie the theoretical mechanism to the empirical evidence and policy implications.

We start by showing that, in a granular world, there are increasing returns to scale because firms use labor more effectively in large markets. In particular, firms in large markets expand their employment more in response to productivity shocks, and so they use more workers while they are productive. Therefore, average labor productivity is higher. To see why that is, consider a sector with only one firm. To expand after a good productivity shock, that firm will need to attract workers from other sectors. Since workers have sector-specific skills, and there are costs to attracting people from other backgrounds, this will be very difficult. Microsoft in Seattle would need to offer jobs to people who do not have a strong background in software engineering, for example. By contrast, if a firm hires a small share of the market because the market is very large, it can expand by poaching workers from other firms in the same sector. We also show that the externality disappears as the region becomes very large. That is because, once a market is sufficiently large, firms no longer have any difficulty finding the workers they need. Therefore, adding more firms will not improve efficiency.

We go beyond this intuition and show that a sufficient statistic for how effectively a single firm uses labor is the covariance of that firm's log productivity and its log employment. Then the productivity of the entire market is just the average (employment-weighted) value of that covariance across firms. Thus, we can confirm our intuition of how the mechanism operates and quantify the contribution of granularity to the wage premium of large cities in a transparent, theory-consistent way.

Our last theoretical result considers the implications for policy. We show that when there are region-wide increasing returns to scale, too few firms enter in equilibrium. That is because, when firms are granular, they know their entry affects wages. Not only will their entry increase average wages in a region, but it will increase wages precisely when the firm would like to hire more workers because their own attempt to increase their employment drives up the wage. That induces a correlation between wages and idiosyncratic firm shocks, which further depresses firm profits conditional on entry since profit functions are convex in wages and productivity. Thus, there is room for place-based subsidies on firm entry, especially in locations where granularity matters. Our result also derives the implied size of the subsidy as a function of the sufficient statistic for increasing returns to scale.

Our empirical analysis focuses on Japan, where we have a panel of all manufacturing establishments with at least 4 employees every year. This data includes employment, payroll, and shipment by 6-digit product category. We define a sector as a 2-digit JSIC industry and a region as a commuting zone. Throughout the paper, we will use the term labor market to refer to a commuting zone and sectoral labor market to refer to the sector in a single commuting zone. Half of all sectoral labor markets have 13 or fewer firms, and the median HHI is 1860.91, which suggests that larger firms play an important role. There would need to be 5.5 firms of the same size in a single market to imply an HHI that large.

We start by providing evidence that granularity matters. We show that the variance of log payroll in a sectoral labor market is decreasing in the number of firms in that sector, suggesting, consistent with Gabaix (2011), that individual firms are subject to idiosyncratic shocks. And those shocks average out in larger markets. We also show that average wages have a large variance, suggesting that the extra payments from firms are not arbitraged away by workers moving in and out of the sector. Thus, moving sectors must be costly.

We then provide evidence consistent with our mechanism as suggested by the sufficient statistic. We begin by showing that the variance of log employment at a single firm in a large sectoral labor market is larger on average than a similarly situated firm in a small sectoral labor market. That suggests, if those firms are subject to similar shocks, that the firm in the larger market has an easier time expanding in response to good shocks and shrinking in response to bad ones. To provide more direct evidence, we construct revenue productivity shocks to each firm using that firm's exposure to real exchange rate movements. A great benefit of our Japanese data is that we have access to product-specific shipments at the establishment level so this is possible. Consistent with our mechanism, firms that already hire a large portion of the sectoral labor market expand their employment less in response to these shocks.

Finally, we calibrate our geography model to match the manufacturing activity in the 256 commuting zones in Japan. Ideally we would directly estimate the average covariance between log productivity and log employment by market size, but we do not observe the productivity of every firm. Instead, we use our empirical results to infer the distribution. We start by setting labor mobility costs of moving across firms and sectors to match how firm employment responds to the revenue productivity shocks. Intuitively, we find the cost of moving across firms by how much employment of a firm small relative to its sector responds to the shock. We determine the cost of moving across sectors by observing how much employment at firms large relative to their sector responds. We then calibrate the size of the idiosyncratic and sector-wide shocks to match the variance in log payroll by sectoral labor market across different regions. The size of the idiosyncratic shocks are determined by how quickly the variance of log payroll falls with the size of the region. We solve the model using the exact hat algebra of Dekle et al. (2008) to avoid explicitly calibrating location-specific parameters on entry costs, productivity, and amenities.

Granular labor market pooling implies that markets with an average of 1.9 firms per sector have the highest elasticity of wages to population at 0.051. Furthermore, firm profits are only 84.0% of the firm's marginal contribution to production. For context, Combes et al. (2011) argue that agglomeration externalities together imply an elasticity somewhere between 0.02 and 0.05, while more recent research like Kline and Moretti (2014) suggest the elasticity in manufacturing sectors could be as large as 0.4. Granularity matters less for large regions like Tokyo. There, the elasticity of wages to population is around 0.001, and firm entry is approximately efficient. We simulate our model with the optimal subsidy on firm entry and find that the number of firms in modest-sized cities increases by more than 30%. The number of firms in Tokyo actually decreases, even though there is a small positive subsidy. That is because workers leave Tokyo for smaller cities in response to the subsidies.

The rest of the paper is organized as follows. We give a short review of the literature below. In section 2, we present the model of a single location and show our theoretical results. We provide empirical evidence for the mechanism in Section 3. We embed the model of a single location in a quantitative model of economic geography in Section 4 which we then quantify in Section 5. We present the quantitative results along with the implications of implementing the optimal policy in Section 6 before concluding in Section 7.

Related Literature

The literature on the spatial agglomeration of economic activity is rich and varied. In an early contribution, Marshall (1920) proposed three reasons why firms might locate around other firms: labor market pooling, access to intermediates, and the sharing of ideas. Subsequent theory papers have formalized these ideas and offered other potential mechanisms (Miyauchi, 2018; Davis and Dingel, 2019). Duranton and Puga (2004) provide a new way to think about all of the mechanisms in their review. Many empirical studies have shown that there are benefits from agglomeration (Andersson et al., 2014; Kline and Moretti, 2014; Greenstone et al., 2010) and have also analyzed the coagglomeration patterns of sectors to infer the relative importance of different theoretical mechanisms (Ellison and Glaeser, 1997; Ellison et al., 2010). Rosenthal and Strange (2004) review the evidence.

Our paper focuses on a particular mechanism that would fall under the broad umbrella of labor market pooling. There are many microfoundations with different mechanisms of how labor market pooling can lead to agglomeration benefits (Andersson et al., 2007; Papageorgiou, 2022). Our model builds on the basic theoretical framework of Krugman (1992). Krugman considers a setting with a finite number of ex-ante identical firms where labor is perfectly mobile within a labor market but not across.[1] Overman and Puga (2010) extend Krugman's model to include multiple sectors and then test the predictions about where those sectors should be located. Other papers test these predictions in different settings (de Almeida and de Moraes Rocha, 2018; Nakajima and Okazaki, 2012). We provide a new theoretical model with ex-ante heterogeneous firms and imperfect labor mobility across firms and sectors. This allows us to derive a general sufficient statistic for the strength of this mechanism. Then we can provide evidence consistent with the

[1] Duranton and Puga (2004) also reviews and discusses the model.

mechanism and quantify its importance.

More recent work looks for direct evidence of the labor market pooling mechanism. Moretti and Yi (2024) show that workers who are laid off in large labor markets have an easier time finding work as compared to workers in small markets. This is consistent with the basic theory we lay out, though our evidence focuses on the firm response rather than the worker side. Conte et al. (2024) shows that when firms in large markets can more easily expand in response to productivity shocks, more volatile firms will sort into larger markets. We abstract from firm sorting but demonstrate how granularity could explain why firms in larger markets can expand more in response to productivity shocks.

We build on a large literature recently inspired by Gabaix (2011) that looks to quantify what the granular nature of firms means for economic activity and optimal policy. Gabaix (2011) shows that shocks to individual firms could explain nationwide fluctuations. Bernard et al. (2018) gives a framework for thinking about how a few important firms could shape the nature of international trade. Gaubert and Itskhoki (2021) discuss what granularity means for the observed comparative advantage of countries, and Gaubert et al. (2021) studies what that implies for optimal policy. We contribute to this literature by considering granularity's implications for the geography of economic activity and optimal place-based policy.

Finally, our quantitative model builds on the computational geography literature reviewed in Redding (2022). Two notable early contributions are Allen and Arkolakis (2014) and Redding (2016). We demonstrate how these models can be extended to include granular firms subject to idiosyncratic shocks in a tractable way. We also consider how the optimal policy changes in that setting (Fajgelbaum and Gaubert, 2020).

2 How Does Granularity Lead to Agglomeration?

The goal of this section is to demonstrate how granularity leads to higher wages in larger markets through the labor market pooling mechanism discussed by Marshall (1920). To focus on that, we model the sectoral labor markets of a single location with a given population. We then endogenize where people live in Section 4 to demonstrate the implications for agglomeration.

Our model in this section will allow us to identify the key mechanism through which granularity implies agglomeration benefits. This will then guide our empirics as we look for evidence consistent with that mechanism. In order to isolate the effects of granularity most transparently, we focus on an environment that is neoclassical conditional on firm entry and delay a discussion of how the results change with imperfect competition to

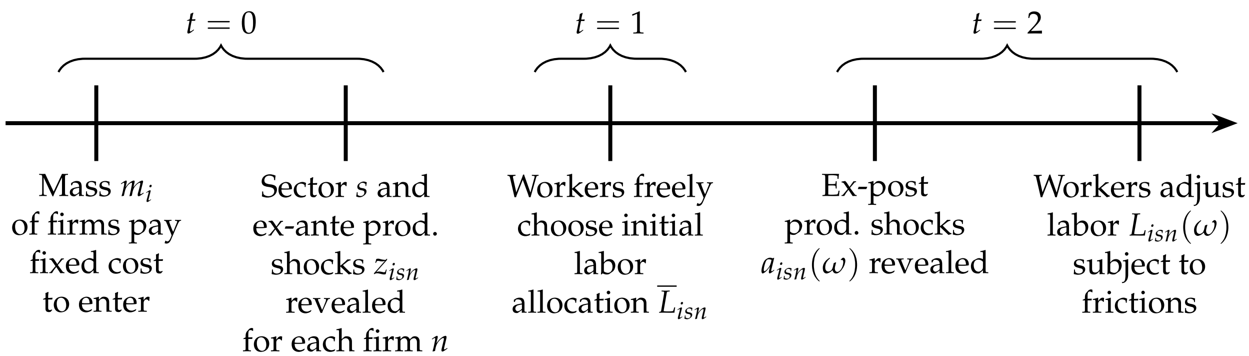


Figure 1: Timing of model

section 2.5.

2.1 Environment

We focus on a region i with a mass ℓ_i of workers and a continuum of sectors $s \in \mathcal{S}$. The sectors produce perfectly substitutable goods but hire in distinct sectoral labor markets.

Timing. There are three time periods $t \in \{0, 1, 2\}$. In period 0, a mass m_i of establishments pay a fixed cost of the traded final good in order to enter. Each firm is randomly assigned a sector s and then gets an ex-ante productivity draw z from some known distribution. Thus each sector ends up with a finite number of firms N_{is} that differ in their productivity. This captures the long run differences in firm size and will determine how exposed different markets are to short-run idiosyncratic firm shocks in period 2.

After observing those initial productivity draws, a representative worker freely allocates her labor $\overline{L}_{isn}$ across sectors s and firms n in period 1. This captures the long run decision-making of a worker free to direct her search or make training decisions to work at certain firms in certain sectors of her region.

Then, in period 2, the state of the world $\omega \in \Omega$ is revealed. This determines the short-run productivity shocks to each firm. The representative worker adjusts how much labor she supplies to each firm, $L_{isn}(\omega)$, subject to frictions of moving labor across firms and sectors on the shorter time scale. Firms then produce and sell their goods. The model timing is summarized in Figure 1.

Workers. The representative worker is risk neutral.[2] She gets utility U_i from consuming a freely traded final good, c_i, and enjoying a location-specific amenity u_i,

$$U_i = u_i c_i. \tag{1}$$

The representative worker is endowed with one unit of labor that she supplies to the market inelastically and allocates among all the active firms. In period 1, the worker can freely set her units of labor across sectors $s \in \mathcal{S}$ and firms $n \in \mathcal{N}_{is}$ taking as given each firm's ex-ante productivity z_{isn}. In particular, she chooses her vector of labor supply $\overline{\boldsymbol{L}}_{\boldsymbol{i}} \equiv \{\overline{L}_{isn}\}_{s,n}$ in the set of feasible labor allocations $\overline{\mathcal{L}}$, i.e.,

$$\overline{\boldsymbol{L}}_{\boldsymbol{i}} \in \overline{\mathcal{L}} \equiv \left\{ \overline{\boldsymbol{L}}'_{\boldsymbol{i}} | \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \overline{L}'_{isn} ds \leq 1 \right\}. \tag{2}$$

The period 1 labor decision captures the long-run labor supply decision of a worker who lives in this location. The perfect substitutability of labor across firms and sectors captures that, in the long run, the representative worker can freely make education decisions and direct her search to work at particular firms.

Then, at the beginning of period 2, the state of the world ω is revealed. This determines ex-post productivity shocks for firms. In response, the representative worker can reallocate her units of labor across firms and sectors subject to moving frictions. In particular, labor must be in some feasible set, $\mathcal{L}(\overline{\boldsymbol{L}}_{\boldsymbol{i}})$,

$$\boldsymbol{L}_{\boldsymbol{i}}(\boldsymbol{\omega}) \in \mathcal{L}(\overline{\boldsymbol{L}}_{\boldsymbol{i}}), \tag{3}$$

where $\boldsymbol{L}_{\boldsymbol{i}}(\boldsymbol{\omega}) \equiv \{L_{isn}(\omega)\}_{s,n}$ and $\mathcal{L}(\overline{\boldsymbol{L}}_{\boldsymbol{i}})$ is defined by,

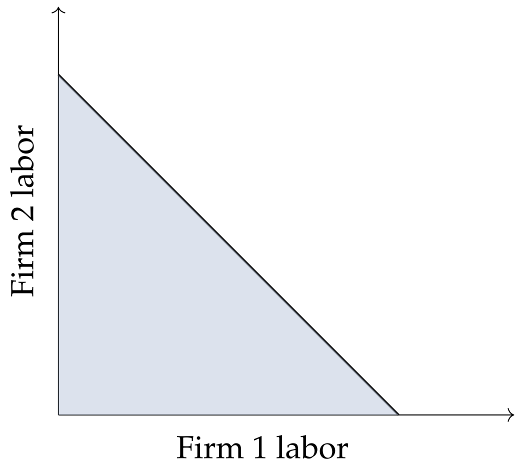
$$\begin{aligned}
\mathcal{L}(\overline{\boldsymbol{L}}_{\boldsymbol{i}}) = \Bigg\{ \boldsymbol{L}'_{\boldsymbol{i}}(\boldsymbol{\omega}) | \int_{\mathcal{S}} \overline{L}_{is} \cdot g_S \left(\frac{L'_{is}(\omega)}{\overline{L}_{is}} \right) ds \leq 1, \\
\overline{L}_{is} = \sum_{n \in \mathcal{N}_{is}} \overline{L}_{isn}, \\
L'_{is}(\omega) = \sum_{n \in \mathcal{N}_{is}} \overline{L}_{isn} \cdot g_N \left(\frac{L'_{isn}(\omega)}{\overline{L}_{isn}} \right) \Bigg\}.
\end{aligned}$$

$\overline{L}_{is}$ is how much labor the worker assigned to sector s in period 1, $L_{is}(\omega)$ is how much effective labor the worker assigns to sector s in state of the world ω, and the functions $g_N(\cdot)$

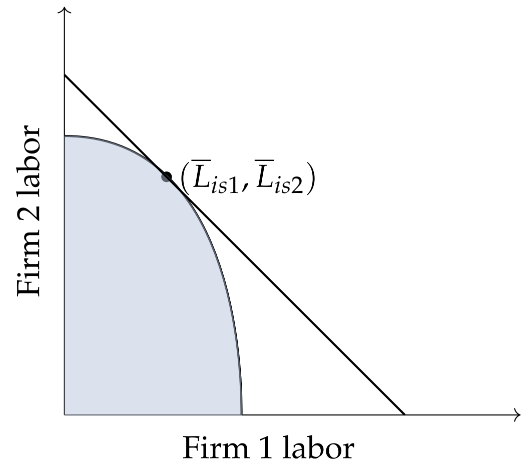
[2] We explain how this model of a representative worker can map onto a model of heterogeneous workers with comparative advantage for working at certain firms and sectors below.

Figure 2: Feasible Labor Allocations

This depicts the feasible labor set for a single sector when labor cannot move between sectors and there are 2 firms in that sector. The set of feasible labor choices in period 2 is a function of the labor choice in period 1.



(a) Possible allocations in period 1, $\overline{\mathcal{L}}$



(b) Possible allocations in period 2, $\mathcal{L}(\overline{\boldsymbol{L}}_{\boldsymbol{i}})$

and $g_S(\cdot)$ summarize the costs of moving labor across firms and sectors, respectively.

$\overline{L}_{is} \cdot g_S(L_{is}(\omega)/\overline{L}_{is})$ is the amount of total labor that the representative agent needs to supply to sector s to get $L_{is}(\omega)$ amount of sector s labor when $\overline{L}_{is}$ labor is originally supplied. Similarly, $\overline{L}_{isn} \cdot g_N(L_{isn}(\omega)/\overline{L}_{isn})$ is the amount of sector s labor that the representative agent needs to supply to firm n to get $L_{isn}(\omega)$ amount of effective labor when $\overline{L}_{isn}$ labor was originally supplied. We assume that $g_N(1) = g_S(1) = 1$ so that the initial allocation $L_{isn}(\omega) = \overline{L}_{isn}$ is feasible. g_N and g_S are increasing and convex to capture the fact that the representative agent needs to give up time in order to supply labor to each firm and that moving a lot of labor between firms becomes increasingly difficult. We illustrate an example two firm sector with no sectoral mobility in Figure 2. In period 1, the worker faces a linear trade-off between supplying labor to firms 1 and 2. Then, in period 2, the initial allocation remains feasible, but moving labor between the two firms can lead to a loss because the worker might have firm-specific skills or the worker might lose time while searching for a job.

This model nests a number of standard models from the literature. While the original model in Krugman (1992) features a single sector, we can nest a multi-sector version as a limit case. In that setting, there is perfect mobility across firms within a sector, $\overline{L}_{isn} \cdot g_N(L_{isn}(\omega)/\overline{L}_{isn}) = L_{isn}(\omega)$, and no mobility across sectors, $\overline{L}_{is} \cdot g_S(L_{is}(\omega)/\overline{L}_{is}) = \overline{L}_{is}$ if $L_{is}(\omega) \leq \overline{L}_{is}$ and infinity otherwise. Our assumption also allows for the nested constant elasticity labor supply model in Berger et al. (2022) where workers have comparative advantage working for certain firms and sectors. That comparative advantage takes a particular form to keep the model tractable. More generally, we allow for any model of the labor market as long as the deadweight loss from moving labor into a sector in the second period is proportional to the percentage change in the size of that sector from the initial allocation. The same must hold for moving across firms.

Firms. There is a continuum of potential firm entrants. To enter, a firm must pay a fixed cost $\psi_i > 0$ in terms of the freely traded final good in period 0. Those firms are then randomly assigned a sector. Thus, while a mass of firms enter, each sector only has a finite number of firms. We denote by $N_{is} \equiv |\mathcal{N}_{is}|$ the (finite) number of firms in sector s. We denote the probability mass function of N_{is} by $p_N(\cdot, m)$ when a mass of m firms enter. The average number of firms across all of the sectors needs to equal the number of firm entrants, i.e. $m_i = \int_{\mathcal{S}} N_{is} ds = \sum_{N=0}^{\infty} N p_N(N, m_i)$. We also assume that $p_N(N, m)$ is differentiable in m and for $m' > m$, $\frac{N}{m'} p_N(N, m')$ first order stochastic dominates $\frac{N}{m} p_N(N, m)$. This assumption rules out firm entry that is overly biased towards small sectors, and we will have more to say about it below. Firm n in sector s then gets an ex-ante productivity draw z_{isn} from a distribution $F_{iz}(\cdot)$ which we assume is continuous and regularly varying.[3]

In period 2, each firm n gets an idiosyncratic productivity shock $\tilde{a}_{isn}(\omega)$, a sector-wide productivity shock $\tilde{A}_{is}(\omega)$, and produces the final good $y_{isn}(\omega)$ according to,

$$y_{isn}(\omega) = a_{isn}(\omega) f(\ell_{isn}(\omega)),$$

where $a_{isn}(\omega) \equiv z_{isn} \tilde{a}_{isn}(\omega) \tilde{A}_{is}(\omega)$ is the total productivity shock to firm n and $\ell_{isn}(\omega)$ is the total amount of labor firm n hires. We assume that $f(\cdot)$ is CES with elasticity of production to labor $1 - \eta \in (0, 1)$, i.e. $f(\ell) = \ell^{1-\eta}$. Throughout we assume that the expected number of employees at a firm is finite. The idiosyncratic productivity shocks are independent and identically distributed with finite variance and $\mathbb{E}[\log \tilde{a}_{isn}(\omega)] = 0$. Similarly, sector-wide shocks are independent and identically distributed with finite variance $\mathbb{E}[\log \tilde{A}_{is}(\omega)] = 0$.

[3] Formally, $L : (0, \infty) \to (0, \infty)$ is regularly varying if $\lim_{x \to \infty} \frac{L(ax)}{L(x)} \in \mathbb{R}^+$ for all $a > 0$.

Market Clearing. Total expected production in the location is

$$Y_i = \mathbb{E}\left[\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} a_{isn}(\omega) f(\ell_{isn}(\omega)) ds\right], \tag{4}$$

where expectations are taken with respect to ω, the number of firms, and the initial productivity draws. We assume that the law of large numbers applies across this continuum of sectors so that realized production is always Y_i. Therefore, goods market clearing requires that total consumption plus the amount of the final good used as an investment must equal expected production,

$$c_i \ell_i + \psi_i = Y_i. \tag{5}$$

In the labor market, labor demanded needs to equal the individual labor supplied by each worker multiplied by the number of workers

$$\ell_{isn}(\omega) = L_{isn}(\omega)\ell_i. \tag{6}$$

2.2 Decentralized Equilibrium

Labor Supply. Workers choose their initial labor allocation $\overline{L}_{isn}$ in period 1 and their subsequent labor allocation $L_{isn}(\omega)$ in period 2 to maximize expected utility, taking wages as given. We normalize the price of the final good to 1, so workers solve the problem,

$$\begin{aligned} \overline{\boldsymbol{L}}_{\boldsymbol{i}}, \{\boldsymbol{L}_{\boldsymbol{i}}(\boldsymbol{\omega})\} \in & \underset{\overline{\boldsymbol{L}}_i' \in \overline{\mathcal{L}},\ \boldsymbol{L}_i'(\boldsymbol{\omega}) \in \mathcal{L}(\overline{\boldsymbol{L}}_i')}{\operatorname{argmax}} \quad u_i c_i \\ & \text{s.t. } c_i \leq \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} w_{isn}(\omega) L_{isn}'(\omega) ds, \end{aligned} \tag{7}$$

where $w_{isn}(\omega)$ is the equilibrium wage for firm n in sector s state of the world ω.

Labor Demand. We normalize productivity so that the price of every good is 1. Then firms maximize profits taking wages and prices as given,

$$\ell_{isn}(\omega) \in \underset{\ell_{isn}'(\omega)}{\operatorname{argmax}} \quad a_{isn}(\omega) f(\ell_{isn}'(\omega)) - w_{isn}(\omega)\ell_{isn}'(\omega). \tag{8}$$

Entry. We assume that firms enter up to the point that expected profits are equal to the fixed cost of entering. Defining $\pi_{isn}(\omega) \equiv a_{isn}(\omega) f(\ell_{isn}(\omega)) - w_{isn}(\omega)\ell_{isn}(\omega)$ as the profits firm n earns in the state of the world ω, we can write this,

$$\psi_i = \frac{\mathbb{E}\left[\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \pi_{isn}(\omega) ds\right]}{m_i}. \tag{9}$$

The numerator is the total amount of profits earned by firms in location i. The denominator is the total measure of firms that enter, so the ratio is a firm's expected profit before any firm realizes its sector or productivity shocks.

Definition 1. A *local equilibrium* consists of wages $w_{isn}(\omega)$, labor supply decisions $\overline{L}_{isn}$, $L_{isn}(\omega)$, entry decisions m_i, and labor demand decisions $\ell_{isn}(\omega)$ such that,

- Workers maximize utility taking wages as given, (7);
- Conditional on entry, firms maximize profits taking prices and wages as given, (8);
- Firms enter up to the point that expected profits are equal to the fixed cost of entering, (9); and
- Goods and labor markets clear, (5), (6).

2.3 Granular Origins of Agglomeration

We now demonstrate how granularity implies average wages increase in the size of the market. We will proceed in two steps. First, we show that wages increase in population if, in markets with more firms, the average firm can adjust its employment more easily in response to shocks. This is the key mechanism, and our empirics in section 3 provide evidence consistent with it. Throughout this paper, we will refer to this mechanism as labor market pooling.

Second, we return to granularity. We provide sufficient conditions under which our model leads to labor market pooling, and thereby agglomeration benefits. And finally, we show that the mechanism disappears in the limit as the number of firms goes to infinity, highlighting the fact that this is a fundamentally granular phenomenon.

We derive all of our results in this paper by doing a second-order approximation to expected production around the point with no ex-post productivity shocks. This keeps the model tractable and makes the mapping between theory and empirics easier and more transparent.

Labor Market Pooling and IRS. As an intermediate step, we express region-wide production as a function of the number of workers ℓ and the number of firms m,

$$Y_i(\ell, m) \equiv \mathbb{E}\left[\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} a_{isn}(\omega) f(\ell_{isn}(\omega)) ds\right].$$

This object differs from equilibrium production since firm entry is not endogenized. However, in the next lemma, we confirm that understanding this object is sufficient for understanding agglomeration both qualitatively and quantitatively.

Lemma 1. *For small ex-post shocks, if $Y_i(\ell, m)$ has increasing returns to scale, i.e. $\frac{1}{Y_i}\frac{dY_i(\alpha\ell, \alpha m)}{d\alpha}\big|_{\alpha=1} > 1$, then average wages are increasing in population, i.e. $\frac{d\log w_i}{d\log \ell_i} > 0$. In particular,*

$$\frac{d\log w_i}{d\log \ell_i} = \frac{\left(\frac{1}{Y_i}\frac{dY_i(\alpha\ell,\alpha m)}{d\alpha}\big|_{\alpha=1} - 1\right)}{1 - \eta - \left(\frac{1}{Y_i}\frac{dY_i(\alpha\ell,\alpha m)}{d\alpha}\big|_{\alpha=1} - 1\right)}.$$

Thus, we can focus on the region-wide production function. Lemma 1 clarifies how our estimates of increasing returns to scale map onto the benefits of agglomeration.

In our next proposition, we lean on the second-order approximation to find a sufficient statistic for the strength of those increasing returns to scale. We will use this statistic in three ways. First, we use it as a stepping stone in our proof that granularity leads to increasing returns to scale. Second, it elucidates the mechanism through which granularity causes increasing returns to scale. Thus, we will be able to look for evidence consistent with that mechanism. Third, we use the statistic to quantify our mechanism. We can determine how much of the urban wage premium is accounted for by granularity rather than some other externality emphasized by the literature by focusing on this statistic.

Proposition 1. *For small ex-post shocks, $Y_i(\ell, m)$ features increasing returns to scale if the (employment-weighted) average covariance between log employment and log productivity shocks is increasing in the number of firms, i.e.*

$$\frac{d}{dm}\left[\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\ell_i} \mathrm{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds\right] > 0.$$

In particular,

$$Y_i(\ell, m) \approx z_i \ell^{1-\eta} m^{\eta} \Omega(m),$$

where $z_i = \mathbb{E}[z_{isn}^{1/\eta}]^{\eta}$ and

$$\Omega(m) \equiv \mathbb{E}[a_{isn}(\omega)] + \frac{1-\eta}{2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\ell_i} \mathrm{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds. \tag{10}$$

The proof of proposition 1 is in the appendix, but we give a brief sketch here. The proof proceeds in three steps. We introduce a variable β so that the productivity of firm n is $z_{isn}(\tilde{a}_{isn}(\omega)\tilde{A}_{is}(\omega))^{\beta}$. First, we solve the model when $\beta = 0$ so that there are no ex-post shocks. Second, we do a second order approximation to regional production in β around 0. We find that $Y_i \approx z_i \ell^{1-\eta} m^{\eta} \Omega(m)$ where $\Omega(m)$ is defined

$$\begin{aligned}\Omega(m) = \mathbb{E}[a_{isn}(\omega)] + (1-\eta) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \operatorname{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds \\ - \frac{1-\eta}{2} \Psi\left(\{\operatorname{Var}(\log \ell_{isn}(\omega))\}, \{\operatorname{Var}(\log \ell_{is}(\omega))\}\right),\end{aligned} \tag{11}$$

where $\Psi(\cdot, \cdot)$ is the measure of deadweight loss that results from moving labor between firms and sectors in the second period. In the third step, we note that when wages are set competitively, they optimally trade off the deadweight loss with the productivity, leading to the simplification

$$\Psi\left(\{\operatorname{Var}(\log \ell_{isn}(\omega))\}, \{\operatorname{Var}(\log \ell_{is}(\omega))\}\right) = \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \operatorname{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds,$$

completing the proof.

Proposition 1 provides a clear interpretation of how labor market pooling leads to increasing returns to scale. Consider a firm n that never adjusts its worforce in response to idiosyncratic productivity shocks so that

$$\operatorname{Cov}(\log \ell_{isn}(\omega), \log a_{isn}(\omega)) = 0.$$

That firm would hold onto a large number of workers when it has bad productivity shocks and not expand to take advantage of good productivity shocks. Therefore, its average labor productivity would depend solely on its average productivity shock. By contrast, if the firm were to expand after a good productivity shock and shrink after a bad shock, its average labor productivity would increase because it hires more workers when its more productive. Thus, that firm could produce more goods on average while hiring the same average number of workers simply by increasing its covariance.

Proposition 1 then says that there are increasing returns to scale if increasing the size of the market improves how efficiently labor reallocates across firms in response to productivity shocks, properly weighted by the importance of each firm. Thus, if firms are better able to expand their employment in response to good productivity shocks in a larger market, then larger markets will be more productive. We will look for direct evidence of this in Section 3.

Implicit in Proposition 1 is the fact that the average covariance is only a function of the number of firms m. Thus, our empirics will use the number of firms as the relevant measure of the size of the market.

Granular Labor Market Pooling. Here, we complete the proof that there are increasing returns when firms are granular. The final assumptions are stated below.

Assumption 1. Labor is imperfectly mobile across sectors, i.e. $g_S''(\cdot) > 0$, and firms are subject to idiosyncratic shocks, i.e. $\mathrm{Var}(\log \tilde{a}_{isn}(\omega)) > 0$.

Then we have the following result.

Proposition 2. *Suppose ex-post shocks are small, and Assumption 1 holds. Then average wages are increasing in population, i.e.* $\frac{d \log w_i}{d \log \ell_i} > 0$*. And as the population goes to infinity, these agglomeration effects disappear, i.e.* $\lim_{\ell_i \to \infty} \frac{d \log w_i}{d \log \ell_i} = 0$.

We start by giving a basic intuition for why there are increasing returns to scale using Proposition 1. Consider how firm n in sector s responds to an idiosyncratic, ex-post productivity shock $\Delta \log a_{isn}(\omega)$. To first order, employment responds according to

$$\Delta \log \ell_{isn}(\omega) \approx \frac{1}{\eta_N + \eta} \left[1 - \frac{\eta_S}{\eta + \eta_N + \eta_S} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \right] \Delta \log a_{isn}(\omega), \tag{12}$$

where $\eta_N := \frac{g_N''(1)1}{g_N'(1)}$ and $\eta_S := \frac{g_S''(1)1}{g_S'(1)}$ are the inverse of the supply elasticities across firms within a sector and across sectors, respectively.

In a small region, where the mass of firms m_i is small, chances are that there are very few other firms in sector s. Therefore, firm n hires a large share of the labor force in sector s, i.e. $\overline{\ell}_{isn}/\overline{\ell}_{is}$ is large. In that case, the firm's employment does not respond much after a productivity shock because it already hires a large proportion of the sectoral labor force, and to expand it would have to attract workers from other sectors. Therefore, the firm does not effectively scale up in response to a productivity shock and ends up using labor inefficiently. In other words, in a small market, workers are stuck working at the same few firms whether or not they are productive in that state of the world. In a market with a large mass of firms m, firm n's share of the sector s labor force is smaller. Therefore, firm n's labor responds more in response to productivity shocks because it can take workers from other firms in the sector, and labor is used more efficiently.

Turning to the proof sketch, we now allow for productivity shocks to every firm. Then labor at firm n in sector s is given by

$$\Delta \log \ell_{isn}(\omega) = \frac{1}{\eta + \eta_N} \left[\Delta \log a_{isn}(\omega) - \frac{\eta_S}{\eta_N + \eta_S + \eta} \sum_{n' \in \mathcal{N}_{is}} \frac{\bar{\ell}_{isn'}}{\bar{\ell}_{is}} \Delta \log a_{isn'}(\omega) \right].$$

Then some tedious algebra, left to the appendix, shows that the employment weighted average covariance is given by

$$\int_{\mathcal{S}} \frac{\bar{\ell}_{is}}{\bar{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \operatorname{Cov}(\log a_{isn}(\omega), \log \ell_{isn}(\omega)) ds = \frac{1}{\eta + \eta_N + \eta_S} \sigma_S^2 + \frac{1}{\eta + \eta_N} \left[1 - \frac{\eta_S}{\eta + \eta_N + \eta_S} \left(\int_{\mathcal{S}} \frac{\bar{\ell}_{is}}{\bar{\ell}_i} HHI_{is} ds \right) \right] \sigma_N^2,$$

where σ_S^2 is the variance of log sector-wide shocks and σ_N^2 is the variance of the log idiosyncratic shocks. This variance depends on the number of firms in the sector only through the $HHI_{is} \equiv \sum_{n \in \mathcal{N}_{is}} \left(\bar{\ell}_{isn} / \bar{\ell}_{is} \right)^2$. Thus, the degree of increasing returns to scale is

$$\frac{d \log \Omega(m)}{d \log m} = -\frac{1}{\eta + \eta_N} \frac{\eta_S}{\eta + \eta_N + \eta_S} \frac{m \frac{d}{dm} \left[\int_{\mathcal{S}} \frac{\bar{\ell}_{is}}{\bar{\ell}_i} HHI_{is} ds \right]}{\Omega(m)} \sigma_N^2.$$

If the average HHI across sectors decreases as the number of firms m increase, Proposition 1 is satisfied, and there are increasing returns to scale. The expected HHI of a given sector s is decreasing in the number of firms N_{is}, so the last thing we need to check is that firm entry across sectors is not too weird. For example, suppose that there were only two sectors with firms in them: one with two firms and another with one firm. Then if all future firm entry goes into unoccupied sectors, the average HHI would actually increase as the average number of firms in occupied sectors decreases. Our assumption that the distribution induced by $\frac{N}{m} p(N, m)$ is increasing in first order stochastic dominance exactly rules out these strange cases.

The speed with which average HHI decreases depends on the distribution of entrants across sectors, $p_N(N, m)$, and the ex-ante productivity distribution $F_{is}(z)$. The shape of $p(N, m)$ especially matters for low m. As m becomes larger, the ex-ante productivity distribution matters more. As discussed in Gabaix (2011), if $F_{is}(z)$ has thin tails, HHI_{is} decreases approximately at the rate of N_{is}^{-1}. Then the agglomeration externality is strong when there are a small number of firms and the HHI falls quickly with new entrants. However, the HHI quickly approaches zero, at which point the average HHI cannot fall further. For example, if a sector has one firm and adds another ex-ante identical firm, the HHI drops from 1 to 0.5. If a sector already has 100 identical firms, the HHI is 0.01, and doubling the number of firms only decreases it to 0.005, not increasing productivity much. Intuitively, that is because if there are already 100 firms in a sector, it is easy for any firm to expand by attracting workers from the other 99 firms. Adding more firms does not have much effect.

If $F_{is}(z)$ has thick tails,[4] HHI_{is} decreases at a rate of $N_{is}^{-\xi}$ where $\xi \in (0,1)$. In that case, the weighted HHI does not fall as quickly, and large firms continue to be constrained in their ability to respond to productivity shocks. Thus, the externality is not as strong for small markets, but continues to matter for medium-sized, and even large, cities. Nonetheless, the externality dies out as the number of firms becomes large, and the average HHI approaches zero.

As suggested by Assumption 1, the preceding argument for increasing returns to scale can fail for two reasons. The first possibility is that firms are not subject to idiosyncratic shocks. In that case, $\sigma_N^2 = 0$, and the average variance of log marginal product does not change as the market gets larger. That is because, if all shocks are sector-wide, for a firm to expand in response to a productivity shock, it will have to attract workers from other sectors since every other firm in the sector is trying to expand as well. We can test this explicitly in the data. If firms are subject to idiosyncratic shocks, then the variance of log total payments to labor will decrease in the size of the sectoral market since the shocks to each of the firms will average out. The other possibility is that workers are freely mobile across sectors, i.e. $\eta_S = 0$. In that case, the firms are not really granular as the relevant labor market is the entire commuting zone.

2.4 Under Entry of Firms

In this section, we consider what this force for agglomeration means for policy. Conditional on firm entry, the model is competitive, so $Y_i(\ell, m)$ is not only the equilbrium level of average production for the market with ℓ workers and m firms. It is also the optimal level of production given m. Therefore, the only possible source of inefficiency is firm entry. Rewriting the goods market clearing condition (5), the first best level of entry m_i^{FB}

[4]Formally, if $F_{is}(z)$ is a pareto distribution, i.e. $F_{is}(z) = 1 - az^{-\zeta}$, then $1 < \zeta(1-\eta) < 2$. Then

$$\xi = 2\left(1 - \frac{1}{\zeta(1-\eta)}\right).$$

solves the following problem,

$$m_i^{FB} \in \underset{m'}{\operatorname{argmax}} \quad Y_i(\ell_i, m') - \psi_i m'. \tag{13}$$

The following proposition demonstrates that the optimal policy calls for a subsidy on firm entry when there are increasing returns to scale.

Proposition 3. *Suppose that shocks are small, and Assumption 1 holds. Then the optimal policy is an ad valorem subsidy on firm entry of* $\frac{1}{\eta}\frac{d\log\Omega(m)}{d\log m}\big|_{m=m_i} > 0$ *where* $\Omega(m)$ *is given by equation (10) in Proposition 1. And as the population goes to infinity, the optimal subsidy converges to 0.*

This result is the normative counterpart of Proposition 2. We sketch the proof here. Taking the first-order condition associated with the net production maximization problem gives the requirement that the marginal product of another firm needs to equal the marginal cost of entry in the first best,

$$\frac{\partial Y_i(\ell_i, m_i^{FB})}{\partial m} = \psi_i.$$

In equilibrium, on the other hand, free entry implies that total profits divided by the number of firms equals the marginal cost of entry. Total profits are simply the value of total production less payments to labor, so with a subsidy of s,

$$\frac{Y_i(\ell_i, m_i^E) - w_i\ell_i}{m_i^E} = (1-s)\psi_i,$$

where w_i is average wages and m_i^E is equilibrium entry. To find the optimal subsidy, we note that because wages are competitively set, workers are paid their marginal product so that $w_i = \frac{\partial Y_i}{\partial \ell_i}$. However, there are increasing returns to scale so that

$$Y_i(\ell_i, m_i) < \frac{\partial Y_i}{\partial m_i} m_i + \frac{\partial Y_i}{\partial \ell_i}\ell_i.$$

It follows that $\frac{Y_i(\ell_i, m_i^E) - w_i\ell_i}{m_i} < \frac{\partial Y_i}{\partial m_i} = \psi_i$ in equilibrium. Therefore, the optimal policy is a subsidy on firm entry, i.e. $s > 0$. This subsidy then converges to 0 as the market becomes infinitely large because the increasing retuns to scale of Y_i disappear in the limit.

The first welfare theorem breaks in this setting because we do not have Walrasian entry. In a competitive equilibrium, firms need to take prices and wages as given, but in our model, we allow firms to ask what their profits would be conditional on entry. That requires a firm internalizing the effect that their entry will have on wages. Firms know that by entering, wages will be higher. Not only will wages be higher on average, but they will be higher precisely when the firm would like to hire more workers because their own idiosyncratic shocks affect wages. This leads to under entry, especially in small markets where adding one more firm will have a big effect on the distribution of wages.

A skeptical reader might wonder why we depart from competitive entry. The reason is simple. An equilibrium with competitive entry does not exist. To illustrate this most transparently, consider a special case of our model where workers are freely mobile across firms within a sector s and all productivity shocks are idiosyncratic. A competitive firm would then maximize profits taking the sector-wide wages $w_{is}(\omega)$ as given. Solving for the optimal labor choice and plugging that back into the profit equation (8), we find that the profits a firm n would earn in state of the world ω is

$$\pi_{isn}(\omega) = \eta(1-\eta)^{\frac{1-\eta}{\eta}} a_{isn}(\omega)^{\frac{1}{\eta}} w_{is}(\omega)^{-\frac{1-\eta}{\eta}}.$$

Taking a second order approximation to profits around the no ex-post shock equilibrium implies that

$$\mathbb{E}[\pi_{isn}(\omega)] \approx \overline{\pi}_{isn}\left[\zeta_s - \frac{1-\eta}{\eta^2}\operatorname{Cov}(\log a_{isn}(\omega), \log w_{is}(\omega))\right]$$

where ζ_s is some sector wide constant and $\overline{\pi}_{isn}$ are firm profits when the ex-post shocks are 1. That is, profits are decreasing in the covariance between firm productivity and equilibrium wages. This is because profit functions are convex, so firms appreciate the variance that comes when productivity and wages are not correlated.

For a firm that is currently operating, wages are correlated with their shocks because their attempt to hire more workers drives up the wage. On the other hand, a potential entrant's productivity shocks are not correlated with wages. Therefore, given the current distribution of wages, the potential entrant expects strictly higher profits than the operating firm. But in any competitive equilibrium, operating firms must expect weakly positive profits while potential entrants must expect weakly negative profits. This is not possible when potential entrants expect strictly higher profits than current entrants, and therefore, no competitive equilibrium exists.

2.5 Extensions to Imperfect Labor Markets

Before turning to the empirical evidence for our mechanism, we briefly discuss how our results change if labor markets are not perfectly competitive. Since there are a finite number of firms in each sectoral labor market, and they face upward sloping labor supply curves in the short run, firms have market power that they could exploit. To shed light on the novel aspects, we have ignored that conduct up to this point. But we characterize the model when firms internalize that power and engage in Cournot competition in appendix B, as in Berger et al. (2022). We assume that firms can commit to their wage offers. Thus, firms internalize that if they were to lower their wages, the representative worker would decrease her time spent at that firm in the long run. Since labor is perfectly substitutable in the long run, this implies that the average wage markdown is zero, but the markdown could vary in response to ex-post productivity shocks.

Proposition 1 no longer holds because labor does not move between firms to efficiently trade off deadweight loss with the productivity gains of moving workers to more productive firms. In particular, a firm does not expand as much in response to productivity shocks since it increases its markdown to exploit its increased market power. However, equation (11) continues to hold. And we show that

$$\Psi\left(\{\text{Var}(\log \ell_{isn}(\omega))\}, \{\text{Var}(\log \ell_{is}(\omega))\}\right) < \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \text{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds.$$

This implies that the covariance of log employment and log productivity understates how productively firms are using workers. We further show that in the limit the equilibrium converges to the competitive one, so the covariance accurately describes how efficiently labor is used. Thus, the observed decline in the covariance between log productivity and employment will overstate the productivity benefits.

However, for given parameters, the market will use labor less efficiently if firms compete oligopsonistically. And in the limit, as the number of firms goes to infinity, they use workers just as efficiently as the competitive case. Thus, for a given set of parameters, monopsony increases the gains from being in a large market when interacted with our mechanism. We include the exact assumptions necessary to recover an adapted proposition 2 in the appendix.

Proposition 3 then becomes more complicated. The first best policy would feature subsidies on wages along with subsidies on firm entry. If the planner could not undo the distortion on wages, then workers are paid less than their marginal product and the efficiency of firm entry is ambiguous.

3 Empirical Evidence

Section 2 showed how granularity can lead to increasing returns to scale. In this section, we provide evidence that it does using administrative data from Japan. We start by showing that large firms subject to their own idiosyncratic shocks dominate hiring within a sector in many commuting zones. We also show that people rarely move across sectors, confirming the conditions of Proposition 2. We then provide evidence of our mechanism as presented in Proposition 1. In particular, we see how firm employment responds to idiosyncratic revenue productivity shocks constructed from exchange rates and trade data.

3.1 Data

In mapping the model to the data, we need to lay out what is a region and what is a sector. We interpret a region i as a commuting zone and denote by $\mathcal{I}$ the set of 256 commuting zones in Japan.[5] We map sectors $s \in \mathcal{S}$ from the theory onto 2-digit manufacturing industries in the data.[6][7] Then we interpret each establishment in the data as its own independent firm in the theory as we do not allow multi-establishmet firms. Throughout this section, we will use the term establishment.

Japanese Census of Manufactures Data. Our primary data source is the Census of Manufacture (CoM) in Japan for the manufacturing sector. The Ministry of Economy, Trade, and Industry (METI) conducts the Census of Manufacture annually to gather information on the current status of establishments in the manufacturing sector. Specifically, this census covers all manufacturing establishments in years when the last digit of the survey year is 0, 3, 5, or 8. For other years, the census covers all establishments with at least 4 employees in Japan. The CoM survey was not conducted in 2012 and 2016, and instead, another survey, the Economic Census for Business Activity (ECBA) was conducted by METI and the Ministry of Internal Affairs and Communications for data in 2011 and

[5] To construct time-consistent commuting zones from municipalities in Japan, we first follow Kondo (2023) to convert municipalities in each year into time-consistent municipality groups. Japan has 1,724 municipalities as of June 2023, including 6 municipalities in the Northern Territories. We drop these 6 municipalities as the CoM data does not cover them. We then use the converter in Adachi et al. (2020) to convert these municipality groups into commuting zones.

[6] To construct time-consistent industries, we use a crosswalk provided by RIETI to convert categories into the 2011 codes. For example, manufacturing of iron and steel industry is one of the 2-digit industries.

[7] In the theory, we assume a continuum of sectors, but in practice, we only use 23 industries. We do this because non-regular workers in Japan move freely within these broad sectors so that this is the best notion of a labor market. By assuming a continuum of sectors, we abstract away from the granularity of sectors, understating the influence of individual firms.

2015.[8] We use the ECBA survey to substitute the CoM survey in 2011 and 2015. The detailed sample construction is in Appendix E.

This data has two advantages. First, we observe panels of all the establishments with at least 4 employees. This feature allows us to compute a variety of volatility measures at the establishment level as well as commuting zone and sector-level variables.[9] Second, we observe yearly shipment values by detailed product categories for each establishment. This enables us to construct establishment-level exposure to foreign exchange rate changes using product shipment share and national, product-destination level export data.

UN Comtrade Data We supplement the CoM data with bilateral trade flows by product-year level from UN Comtrade data to compute the share of destination countries of the Japanese export at the product-year level. See the Appendix E for a detailed data construction.

Penn World Table Data We construct real exchange rates using data for the nominal exchange rate ("xr") and price level of exports ("pl_x") from Penn World Table (Feenstra et al., 2015).

3.2 Summary Statistics

Table 1 presents summary statistics for the sectoral labor markets. The average number of establishments is 53, but more than half of all sectoral labor markets have 13 or fewer establishments. The employment HHI distribution suggests individual establishments have an outsized role in many labor markets. The median HHI is 1860.91. There would need to be 5.5 establishments of the same size in a single market to get an HHI that large. Employment is similarly sparse, with half of sectoral labor markets having 346 employees or fewer. See Figure 3 for a histogram of the entire distribution of establishments and employment.

We also report the summary statistics for a number of variables we use in our subsequent analysis in Table 1. Log average wage is the average salary of workers in the sectoral labor market. The variance of log wage growth and that of log payroll growth are computed over time for each sectoral labor market. These measure the volatility of

[8]The ECBA survey covers all establishments, including establishments in non-manufacturing sectors, but we focus on establishments in the manufacturing sector to be consistent with the CoM survey.

[9]One further advantage, compared to the US LBD data, is that we can separately identify single establishments within each of the 47 prefectures.

wages and payroll of a single market, respectively. The log variance of log wage growth is smaller than that of log payroll growth, suggesting that some of the variance of log payroll growth is driven by workers moving between sectors to take advantage of economic opportunities as they arise. However, the variance of log wages is not zero. This suggests that economic opportunities are not completely arbitraged away by workers moving between sectors. Instead, workers in a sector that has a good shock see an increase in their wages.

In fact, workers are slow to move across sectors. Only 6% of full-time workers switch jobs within a year, and of those who switch jobs, 76.9% stay in the same industries.[10] [11]

Table 1: Summary Statistics: Sectoral Labor Market

	Mean	Median	Std. Dev.
Num. of Estab.	53.19	13.00	176.56
Employment HHI	3222.53	1860.91	3270.30
Employment	1651.48	346.50	4771.52
Log Avg. Wage	5.78	5.82	0.37
Log Variance of Log Wage Growth	-4.40	-4.46	1.29
Log Variance of Log Payroll Growth	-2.97	-3.05	1.38
Sample size	4,800		

Note: The tables show the summary statistics for sectoral labor markets. Each sectoral labor market is a 2-digit manufacturing industry, commuting zone pair. The variance of log wage growth, the variance of log payroll growth, and the weighted sum of log employment growth are in log units. The variance of log wage growth and log payroll growth are computed for a single sectoral labor market over time. The weighted average of log employment growth is computed by first taking the variance of log employment growth for each establishment. The variances are then averaged together weighted by each establishment's average payroll share.

3.3 Firms Are Subject to Idiosyncratic Shocks

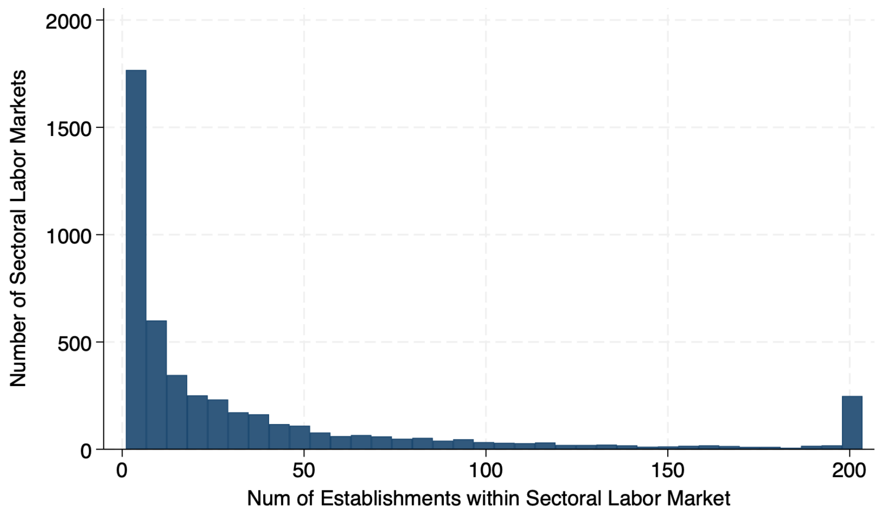
In presenting the summary statistics, we showed that large establishments dominate many sectoral labor markets and that worker movement across sectors does not arbitrage away all wage differences. Here we check the final condition of Proposition 2: establishments are subject to idiosyncratic shocks.

[10] Data is from the Japanese Panel Study of Employment Dynamics, which was provided by Recruit Works Institute. The sample is the full-time workers in 2015, and the sample size is 22,965. When computing the mobility across sectors conditional on changing jobs, we use the data for workers who have ever changed jobs in their careers.

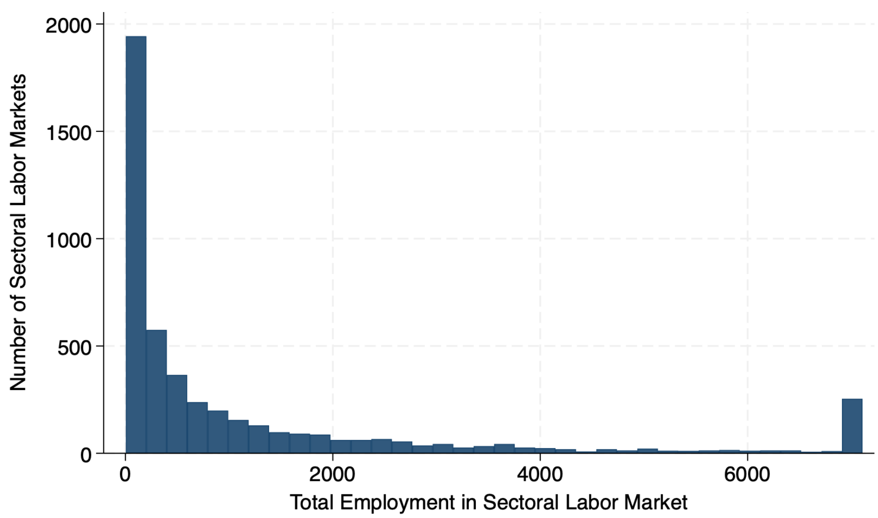
[11] The probability that a worker would stay in the same industry if she were randomly choosing a new job is 48.4%. We compute this number using the HHI of 2-digit sectoral employment within each commuting zone and take the weighted average using the total employment of each commuting zone.

Figure 3: Histogram of Number of Establishments and Employment in each Secotral Labor Market

(a) Number of Establishments



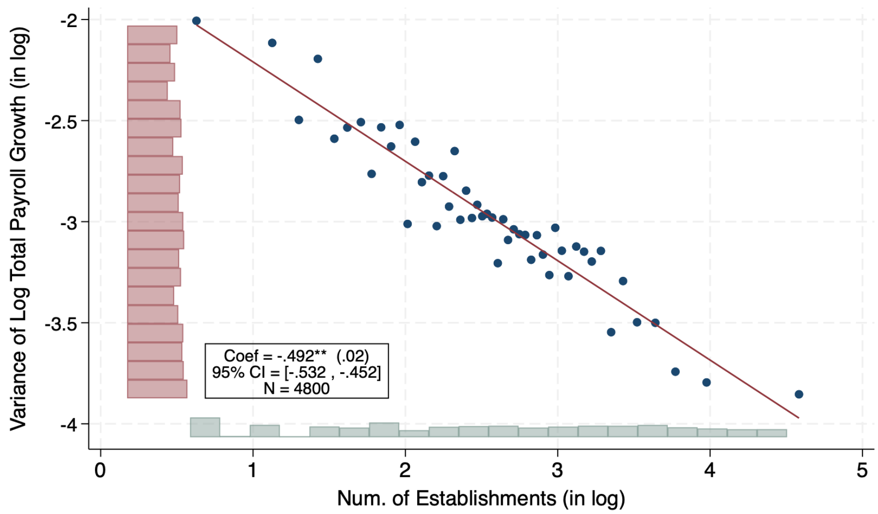
(b) Total Employment



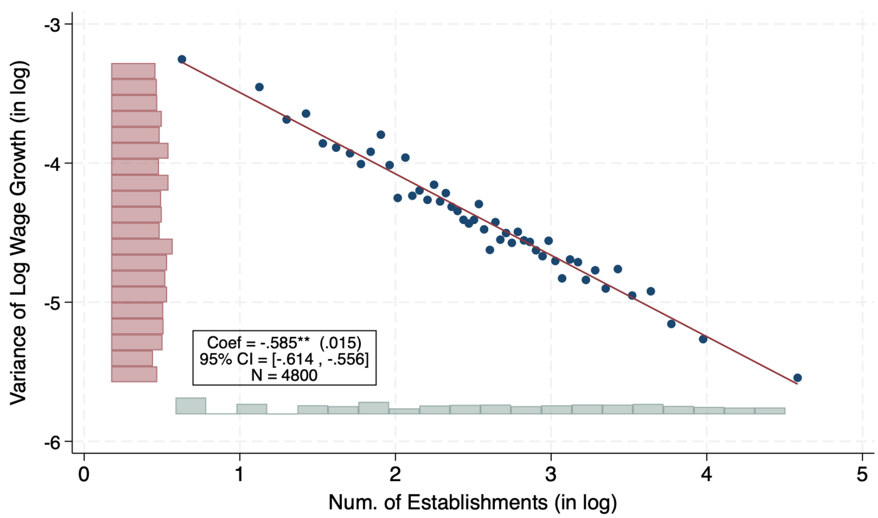
Notes: The figures show the histogram of the number of establishments and employment in each local labor market, respectively. The unit of observation is the local labor market, a pair of a JSIC 2-digit industry, and a commuting zone. To be visible, I collapse all the local labor markets with the statistics larger than the top 5% percentile to be in the same bin. The top 5% percentile is 210 for the number of establishments and 7,107 for the total employment,

Figure 4: Volatility of Payroll and Average Wage Growth and Number of Establishments

(a) Payroll



(b) Average Wage



Note: The panels show the binned scatter-plots and histograms of the relationship between volatility of log growth in total labor payment (top) and in average wage (bottom) and the number of establishments across local labor markets in Japan estimated by (15). We also show the histograms of both variables. The unit of observation is the local labor market, a pair of JSIC 2-digit industries, and a commuting zone. The vertical axis is the log variance of log growth in total labor payment (top) and average wage (bottom) over 1990-2016 in each local labor market. The horizontal axis is the number of establishments in log, averaged over 1986-2016 in each local labor market.

In order to do this, we construct a measure of volatility at the sectoral labor market level. For each sectoral labor market, we start by computing the total labor payroll in each period. We then compute one-year log growth rate and take its variance over time to get our measure. We use the variance of the growth rate rather than the variance in levels as a conservative estimate of the uncertainty in labor payments from one period to the next. This differences out persistent differences in growth rates that workers could expect.

We regress our measure of volatility in labor demand against the number of establishments in the sectoral labor market to see if larger labor markets have less volatile demand for labor. We use the specification,

$$\log\left(\mathrm{Var}\left(\log\left(\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)\right)\right)\right) = \beta\ln N_{is} + \mu_i + \mu_s + \varepsilon_{is} \tag{14}$$

where μ_i and μ_s are commuting zone fixed effects and industry fixed effects, respectively. We show the bin scatter plot in Figure 4(a) . As one can see, there is a strong negative relationship, consistent with our story that establishments are subject to idiosyncratic shocks. Small sectoral labor markets with few establishments have very volatile demand for labor as a shock to a single establishment greatly affects the overall demand for labor. By contrast, in a large market, there are many establishments, all subject to their own idiosyncratic shocks. So the idiosyncratic shocks "average out" to some extent and the demand for labor remains relatively constant.

We do a similar exercise with the variance in log wage growth at the sectoral labor market level to ensure that this variance in labor demand is not arbitraged away by workers moving between sectors. We compute the sector-specific average wage by dividing the total labor payment by the number of workers. We then find the variance of log growth in the same way we did for the total payroll. We run the regression,

$$\log\left(\mathrm{Var}\left(\log\left(\frac{\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)}{\sum_{n\in\mathcal{N}_{is}} \ell_{isn}(\omega)}\right)\right)\right) = \beta\ln N_{is} + \mu_i + \mu_s + \varepsilon_{is} \tag{15}$$

where μ_i and μ_s are commuting zone fixed effects and industry fixed effects, respectively. The bin scatter plot is in Figure 4(b). We find a similar negative relationship: the variance of log wages is higher in smaller markets because there are fewer establishments and a shock to one establishment will greatly affect labor demand and it is not arbitraged away by worker movement.

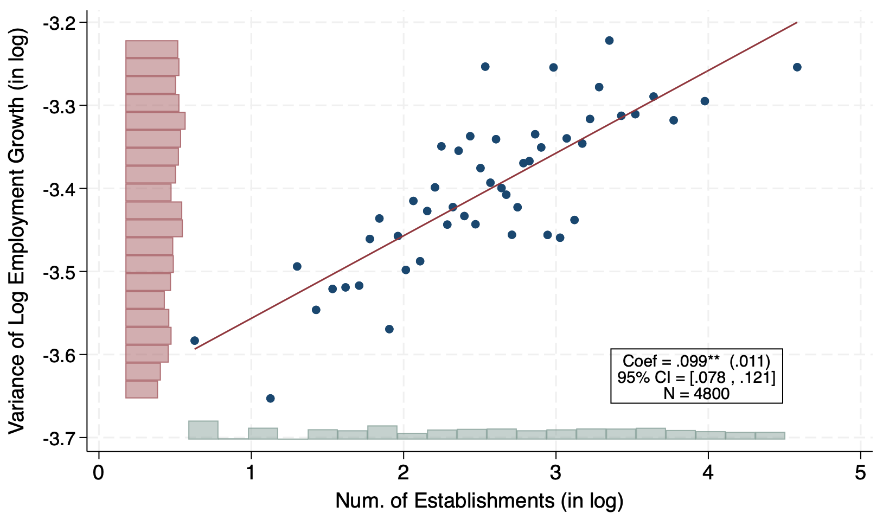
3.4 Firms in Larger Markets Expand Employment More Easily

Having confirmed all of the sufficient conditions of Proposition 2 for there to be increasing returns to scale, we now look for explicit evidence of our mechanism. We do this by looking for evidence of Proposition 1: the weighted covariance of log employment and log productivity of establishments is larger in markets with more establishments. This would provide evidence that larger markets are more productive because workers can move to where their labor is most needed.

3.4.1 Observational Evidence

We start by providing observational evidence. If establishments have an easier time finding the workers they need after a good shock in large markets, then an establishment in a larger market should have a higher variance of log employment than a similar establishment in a small market. We check this by constructing a measure of establishment employment volatility.

Figure 5: Volatility of Establishment-level Employment Growth and Number of Establishments



Note: The figure shows the binned scatterplots of the relationship between the volatility of establishment-level employment growth and the number of establishments across local labor markets in Japan. We also show the histograms of both variables. The unit of observation is the local labor market, a pair of a JSIC 2-digit industry, and a commuting zone. The vertical axis is the average of establishment-level log variance of log growth of employment over 1986-2016 in each local labor market. The horizontal axis is the number of establishments in log, averaged over 1986-2016 in each local labor market.

First, we residualize each establishment's annual employment by establishment and year-fixed effects.

$$\Delta \ln \ell_{n,t,t+1} = \alpha \ln \ell_{n,t} + \eta_t + \varepsilon^{\ell}_{n,t}$$

where $\Delta \ln \ell_{n,t,t+1}$ is the changes in log employment of establishment n from year t to $t+1$, $\ln \ell_{n,t}$ is the log employment of establishment n from year t, and η_t is a year fixed effect. We control log employment to capture the fact that large and small establishments are systematically different. Small establishments might be growing in an expected way as suggested by some papers studying establishment dynamics, such as Hopenhayn (1992).

Second, we compute yearly changes of the estimates of $\varepsilon^{\ell}_{n,t,t+1}$, $\Delta\widehat{\varepsilon}^{\ell}_{n,t,t+1}$ as follows:

$$\Delta\widehat{\varepsilon}^{\ell}_{n,t,t+1} \equiv \widehat{\varepsilon}^{\ell}_{n,t+1} - \widehat{\varepsilon}^{\ell}_{n,t}.$$

This gives a measure of unexpected growth in employment for establishment n.

Finally, we take variance of $\Delta\widehat{\varepsilon}^{\ell}_{n,t,t+1}$, $(\mathrm{Var}\, \Delta\widehat{\varepsilon}^{\ell}_{n,t,t+1})$, over time within each establishment n to get a measure of employment growth variance relative to expected growth patterns. We then pool the establishments in a sectoral labor market by taking the weighted average of $\mathrm{Var}(\Delta\widehat{\varepsilon}^{\ell}_{n,t,t+1})$, weighted by each establishment's median employment over the sample period. This gives a measure of employment volatility of establishments in the sectoral labor market.

We estimate the following log-linear model

$$\sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \cdot \mathrm{Var}(\Delta\widehat{\varepsilon}^{\ell}_{n,t,t+1}) = \delta \log N_{is} + \mu_i + \mu_s + \varepsilon_{is} \tag{16}$$

where $\log N_{is}$ is the number of establishments in commuting zone i and industry s. μ_i and μ_s are commuting zone fixed effects and industry fixed effects, respectively.

Figure 5 shows the result in binned scatter plots. One can see the clear positive relationship implying that comparing two similarly situated establishments, the establishment in the larger labor market will have a higher variance of log employment.

3.4.2 Quasi-Experimental Evidence

Finally, we look for more direct evidence that establishments in large markets can expand more easily in response to productivity shocks. In particular, we construct establishment-level demand shock from the establishment's product mix, country-level destinations, and destinations' real exchange rates. We then examine how that "adjusted real effective exchange rate shock (AREER)" affects an establishment's employment share within sectoral labor markets.

Specification Our econometric specification is an empirical version of equation (12). It is as follows:

$$\Delta \ln \ell_{n,t,t+1} = \beta_1 \Delta\mu_{n,t,t+1} + \beta_2 \left(\Delta\mu_{n,t,t+1} \cdot s_{n,t-1}\right) + \mathbf{X}'_{n,t}\Gamma + \zeta_n + \zeta_t + \varepsilon_{nt}, \tag{17}$$

where $\Delta \ln \ell_{n,t,t+1}$ is the log change in the employment of establishment n from year t to $t+1$, $\Delta\mu_{n,t,t+1}$ is the negative demand shock for establishment n, $s_{n,t-1}$ is the payroll share of establishment n in year $t-1$ within the sectoral labor market.[12] $\mathbf{X}_{n,t}$ is a vector of covariates at the establishment level, including an establishment age and its square, lagged payroll share, and lagged export ratio relative to total shipment.[13] ζ_n and ζ_t are fixed effects for establishments and year, respectively. $\varepsilon_{n,t}$ is the error term.

Establishment-Specific Negative Labor Demand Shock We proxy the establishment-specific revenue productivity shock $\Delta\mu_{n,t,t+1}$ by an establishment-level exposure to a real effective exchange rate shock, which we define as follows,[14]

$$\Delta\mu_{n,t,t+1} = \overline{\mathrm{EXP}_n} \times \left(\sum_c \overline{\omega_{n,c}} \cdot \Delta REX^{JPN}_{c,t,t+1}\right) \tag{18}$$

where $\overline{\mathrm{EXP}_n}$ is the median export share relative to the total shipment of establishment n over the period.[15] $\overline{\omega_{n,c}}$ is the median exposure of establishment n to country c over time, where the time-specific exposure $\omega_{n,c,t} \equiv \sum_p \omega_{n,p,t} \cdot \omega_{p,c,t}$ is the share of product shipment of establishment n to country c in year t. Since CoM does not report establishment-specific export destinations, we use $\omega_{n,p,t}$, the share of shipment of establishment n in 6-digit product category p in year t and $\omega_{p,c,t}$, the share of Japanese export to country c out of total shipment in 6-digit product category p from the UN Comtrade data. $\Delta REX^{JPN}_{c,t,t+1}$ is the change in the real exchange rate of Japanese Yen to the currency in the country c from t to $t+1$. Therefore, positive $\Delta REX^{JPN}_{c,t,t+1}$ means JPY appreciation against the currency

[12] We look at the interaction between the shock and the payroll share of the establishment in the period before because that is the specification suggested by equation (12). Noting that in larger markets, establishments tend to have lower payroll shares then confirms statistic (ii) of Proposition 1.

[13] Establishment ages are not surveyed in the CoM data. **[fill]**

[14] In the context of Japan, there are several papers, which study the effect of exchange fluctuations on employment responses(Hosono et al., 2015; Yokoyama et al., 2021). Some recent studies on the effect of the exchange rate on employment using firm-level exposure to trade (Nucci and Pozzolo, 2010; Ekholm et al., 2012; Yokoyama et al., 2021) using firm-level export share. Similar to ours, Dai and Xu (2017) use firm-level heterogeneity of trade partners and the heterogeneous fluctuations of exchange rates across currencies in the context of Chinese manufacturing sectors. Our specification leverages the establishment-level product mix, product-country-level export, and country-level exchange rate fluctuations.

[15] The CoM survey asked the ratio of exports in each establishment only after 2001, so we use the median, rather than the lagged value.

in country c, so that $\Delta\mu_{p,t,t+1}$ and $\Delta\mu_{n,t,t+1}$ are *negative* demand shock for product and establishment, respectively.

Definition of Employment Types We examine the effect of a JPY appreciation on employment across establishments. We define executives with compensation and permanent employees ("sei-shain") as regular workers who typically work full-time with an indefinite contract. We define non-regular workers as the sum of part-time workers and workers dispatched from temporary help agencies.

Sample Restriction We restrict the sample to 2001 to 2013 because the export share and employment by employment types are available in the CoM data since 2001, and the timing of the survey has changed from December in the previous year to June in 2014. We drop the case where the shock is JPY depreciation, when the AREER shock is negative, as we expect heterogeneity in responses of employment to positive and negative shocks. Our final sample in this analysis is an unbalanced panel of 163,121 unique establishments in the manufacturing sector from 2001 to 2013.

Table 2: Summary Statistics: Establishment

	Mean	Std. Dev.	p10	p50	p90
Employment	51.11	155.86	6.00	18.00	104.00
Payroll (in millions JPY)	223.56	1029.06	11.71	56.22	408.77
Payroll (in log, JPY)	8.75	1.42	7.07	8.63	10.62
Share of Non-Regular Workers	0.34	0.23	0.07	0.29	0.70
Log Changes in Employment	-0.01	0.19	-0.18	0.00	0.16
Log Changes in Regular Emp.	-0.01	0.27	-0.22	0.00	0.20
Log Changes in Non-Regular Emp.	-0.00	0.51	-0.51	0.00	0.51
Emp. Share within LLM	0.02	0.08	0.00	0.00	0.05
Payroll Share within LLM	0.02	0.08	0.00	0.00	0.05
Sample size	1,164,363				

Note: The tables show the summary statistics for the data used in the analysis across establishments.

Summary Statistics: Establishment Panel Table 2 shows the summary statistics for panels of the establishments in our analysis. The average and median establishments have 51 and 18 workers. The average payroll is 224 million in JPY (in 2015 value). On average, the share of non-regular workers is 34%. The changes in employment share are symmetric with the median of zero, but the volatility comes from non-regular employment with a standard deviation of 0.51, rather than regular employment with a standard deviation of 0.27. This is consistent with the findings in the previous literature in Japan.[16] The average share of employment as well as payroll within a local labor market is 2%.

Results: No Interaction Before showing our main results on the role of the market size in responses to shocks, we present evidence that our establishment-specific shocks have an impact on establishment outcomes. Table 3 shows the result without the interaction term. Column (1) uses log changes in sales, Column (2) uses log changes in total employment, Column (3) uses log changes in regular employment, and Column (4) uses log changes in non-regular employment. Column (1) shows that the 1% of the negative exchange rate shock decreases sales by 3.5%. Column (2) shows that the employment declines by 0.3%. This decline is heterogeneous across employment types: Column (3) shows that regular employment declines only by 0.3%, while Column (4) shows that establishments adjust non-regular employment by 2.6%. This is consistent with the findings in Yokoyama et al. (2021) that a JPY appreciation reduces the sales and non-regular employment of exporters.

Table 3: Effects of JPY Appreciation on Establishment Sales and Employment

	Dep. Var.: Log Changes			
			Employment by Types	
	Sales	Employment	Regular	Non-Regular
AREER Shock	-3.46	-0.25	-0.29	-2.62
	(0.17)	(0.09)	(0.12)	(0.23)
Observations	1,164,363	1,164,363	1,164,363	1,164,363
Covariates	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓
Establishment FEs	✓	✓	✓	✓

Notes: This table shows the relationship between JPY appreciation and various outcomes at the establishment level. Column (1) uses log changes in sales, Column (2) uses log changes in total employment, Column (3) uses log changes in regular employment, and Column (4) uses log changes in non-regular employment. The running variable is the adjusted real exchange rate shock at an establishment level. All columns include the following covariates: lagged share of non-regular workers, lagged payroll share within each local labor market, the establishment's age square, and the sum of the shock to other establishments within each local labor market. All columns include establishment fixed effects and year-fixed effects. Standard errors are robust against heteroscedasticity.

Results: Roles of Size in Response After confirming that our measure of shock does actually affect the establishment's sales and non-regular employment, we examine the

[16] See Morikawa (2010) and Kambayashi (2017) for the evidence that firms adjust labor more flexibly for non-regular workers.

Table 4: Effects of JPY Appreciation on Changes in Non-Regular Employment

	Dep. Var.: Log Changes in Non-Regular Emp.			
	(1)	(2)	(3)	(4)
AREER Shock	-2.62	-2.98	-0.31	-0.55
	(0.23)	(0.27)	(0.44)	(0.44)
AREER Shock $\times$ Log Payroll			-1.08	-1.12
			(0.14)	(0.15)
AREER Shock $\times$ Payroll Share		3.35	8.26	
		(1.26)	(1.41)	
AREER Shock $\times$ (Payroll Share $>$ 3%)				2.43
				(0.50)
Observations	1,164,363	1,164,363	1,164,363	1,164,363
Covariates	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓
Establishment FEs	✓	✓	✓	✓

Notes: This table shows the relationship between JPY appreciation and non-regular employment growth at the establishment level. The dependent variable is the log growth of employment share within local labor markets. The running variable is the adjusted real exchange rate shock at an establishment level. The log payroll share is normalized by subtracting the average log payroll in the entire sample when it interacts with the AREER shock. All columns include the following covariates: lagged share of non-regular workers, lagged payroll share within each local labor market, the establishment's age square, and the sum of the shock to other establishments within each local labor market. All columns include establishment fixed effects and year-fixed effects. Standard errors are robust against heteroscedasticity.

heterogeneous responses by the establishment's sizes. Table 4 shows the results. The dependent variable is the log changes in non-regular employment. The running variable is the adjusted real exchange rate shock at an establishment level. All columns include the following covariates: lagged share of non-regular workers, lagged payroll share within local labor markets, lagged log payroll, and the establishment's age square. All columns include establishment fixed effects and year-fixed effects. Regressions are weighted by the establishment's payroll. Standard errors are robust against heteroscedasticity.

Column (1) runs the regression without the interaction, which replicates Column (3) in Table 3 that a JPY appreciation decreases non-regular employment. Column (2) adds the interaction with payroll share within a local labor market. The positive estimate implies that the establishments with higher payroll share respond less to the shock, which is consistent with condition (ii) of Proposition 1. Quantitatively, if the plants differ in their payroll share by 10% pt, the elasticity decreases by 0.34, which is about 11% (=0.34/2.98) of the initial elasticity of 2.98.

However, we may find that the interaction between the shock and a establishments's payroll share is positive simply because larger establishments are different from smaller establishments–independent of their size relative to the market. For example, larger establishments may adjust labor less because they can adjust other factors more easily. Therefore, the estimate may pick up heterogeneity in responses by the absolute size, rather than relative size.

To address this concern, Column (3) adds the interaction term with the lagged log payroll of the establishment.[17] This means that we compare the elasticities of employment by payroll share within a local labor market, across establishments with the same absolute sizes. The estimate is now 8.26, which is larger.

Column (4) uses the interaction of the shock with the dummy variable, which takes one if the payroll share within a local labor market is larger than 3%. which is roughly the 80% percentile value in the sample. Again, the estimate is positive and implies that establishments with larger payroll share within a local labor market respond less to the shock. Therefore, establishments in a larger market can expand their employment more easily.

4 A Quantitative Model of Granular Economic Geography

In this section, we describe an empirical version of the model in Section 2 that allows us to estimate the strength of the mechanism using the regressions in Section 3. We

[17] We normalize the log payroll by subtracting the average log payroll share.

then embed this labor market model in a standard model of economic geography with I locations $i \in \mathcal{I} = \{1, \ldots, I\}$ with imperfect labor mobility across regions, production externalities independent of granularity, and amenity externalities. We do this by introducing a migration decision at time $t = -1$ allowing workers to choose where to live. We then assume that workers are stuck in their region as moving in response to 1 year productivity shocks is negligible in the Japanese context.

4.1 Parametric Restrictions

To fully specify the model, we need to specify the probability mass function of firm entrants $p_N(\cdot, m)$, the distribution of the ex-ante productivity shocks $F_{iz}(\cdot)$, and the costs to moving across firms and sectors, $g_N(\cdot), g_S(\cdot)$. We continue to use the second order approximation, so we do not need to fully specify the distribution of the ex-post idiosyncratic and sectoral shocks $F_a(\cdot), F_A(\cdot)$.

Firm Entry Distribution. We assume that if a mass of m_i establishments attempt to enter in location i, the number of realized entrants in any sector s is distributed Poisson with parameter m_i. The probability mass function is

$$p_N(N, m_i) = \frac{(m_i)^N e^{-m_i}}{N!}.$$

Ex-ante Productivity Shock Distribution. Following Gabaix (2011), we assume that the ex-ante productivity shocks, z_{isn} are distributed according to a Pareto distribution with shape parameter λ and scale parameter z_i.

z_i is the natural productivity of location i. It is subject to increasing returns to scale independent of our granular mechanism which take the form,

$$z_i = \bar{z}_i (\ell_i)^{\gamma_z}, \tag{19}$$

where γ_z is the agglomeration elasticity independent of our mechanism and $\bar{z}_i$ is the fundamental productivity of location i.

Labor Mobility Frictions. The labor mobility frictions take the CES form frequently used in the labor literature.[18] We have that

$$g_{\mathcal{N}}\left(\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}}\right) = \left(\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}}\right)^{\frac{1+\kappa}{\kappa}},$$

for individual firms. For moving across sectors, we have

$$g_{S}\left(\frac{L'_{is}(\omega)}{\overline{L}'_{is}}\right) = \left(\frac{L'_{is}(\omega)}{\overline{L}'_{is}}\right)^{\frac{1+\nu}{\nu}\frac{\kappa}{1+\kappa}}.$$

These imply that the elasticity of substitution across firms is κ and the elasticity of substitution across sectors is ν. It is more difficult to move across firms in different sectors than in the same sector if $\nu < \kappa$.

4.2 Migration Decision

We allow the amenities in location i to be a function of fundamentals and spillovers. In particular,

$$u_i = \overline{u}_i(\ell_i)^{\gamma_u}, \tag{20}$$

where $\overline{u}_i$ is the fundamental amenities in location i and γ_u is a measure of the externality. $\overline{u}_n$ captures fundamental differences across locations like weather and the beauty of the nearby area. γ_u captures interactions between people like traffic.

Each worker has an idiosyncratic preference for each location i, ε_i. The utility that the worker gets from living in location i is $U_i\varepsilon_i$. Each worker can move freely and so lives in her utility maximizing location. The preference shocks are distributed according to a Fréchet distribution with shape parameter θ. Therefore, the number of workers who live in location i is

$$\ell_i = \left(\frac{U_i}{U}\right)^{\theta}\ell, \tag{21}$$

where ℓ is total population, and,

$$U = \left[\sum_{i\in\mathcal{I}}(U_i)^{\theta}\right]^{\frac{1}{\theta}}. \tag{22}$$

[18] Berger et al. (2022) use the same parametric restriction and show it is equivalent to a roy model with Fréchet productivity draws across sectors and firms.

Definition 2. A *Decentralized Equilibrium* consists of population ℓ_i, average wages w_i, fundamental utility U_i, and firm entry m_i in each labor market, along with wages $w_{isn}(\omega)$, labor supply decisions $\overline{L}_{isn}$, $L_{isn}(\omega)$, and labor demand decisions $\ell_{isn}(\omega)$ such that,

- Workers choose their utility maximizing location taking average wages as given, (21), (22);
- Amenities and productivity are consistent with spillovers, (19), (20); and
- Average wages, firm entry, wages, labor supply decisions, and labor demand decisions are part of a local equilibrium.

5 A Quantification of Granular Scale Effects

In this section, we calibrate the quantitative model economic geography model introduced in Section 4 using accepted parameters from the literature and our empirical results from Section 3. Proposition 1 makes clear that we would ideally match the covariance of log productivity and log employment of individual firms. Unfortunately, we do not observe the entire distribution of productivity shocks to firms, so we infer it. In particular, we calibrate the elasticity of employment to idiosyncratic productivity shocks using our estimates of how firm employment responds to productivity shocks. We then calibrate the distribution of sector-wide shocks and idiosyncratic shocks to match the distribution of log payroll across different sectoral labor markets as described below.

5.1 Calibration

We summarize how we calibrate all of the parameters in Table 5. We discuss each parameter in more detail below.

Labor Supply and Demand Elasticities. We calibrate the labor supply and demand elasticities to match our estimates of how firm employment and sales respond to exporting productivity shocks. The details are in the appendix, but we go through the brief intuition here. To calibrate the degree of decreasing returns to scale η, we compare the coefficient on how labor responds interacted with the size of the market with the coefficient on how sales responds interacted with the size of the market using the specification of column (2) in Table 4. Intuitively, this identifies the degree of decreasing returns to scale by finding how much more a firm in a large market can produce because it can expand its employment more. We calibrate the supply elasticity across firms η_N to rationalize the

A. Labor Supply and Demand Elasticities			
Parameter	Value	Description	Source
κ	7.54	Supply elasticity across firms	
ν	0.456	Supply elasticity across sectors	Table 4
η	0.203	Decreasing returns to scale	
B. Productivity Shock Parameters			
Parameter	Value	Description	Source
λ	8.517	Ex-ante productivity Pareto tail	
σ_N^2	0.191	Variance of log idiosyncratic shocks	Variance of log sector payroll
σ_S^2	0.019	Variance of log sectoral shocks	
C. Migration and Spillover Parameters			
Parameter	Value	Description	Source
θ	3	Migration elasticity	Redding (2016)
γ_u	-0.25	Amenity spillovers	Japanese housing share
γ_z	0.03	Other productivity spillovers	Combes et al. (2011)

Table 5: Calibration summary.

increase in labor in response to the shock for a firm small relative to its market. Finally, we calibrate the elasticity of subsitution across sectors by comparing how much employment in a firm small relative to its market responds compared to in a firm that is larger. This measures how much harder it is to attract workers from other sectors. For this calibration, we use the specification in column (3) of Table 4 to control for the overall size of the firm.

The results are in Panel A of Table 5. The firm substitution parameter is $\kappa = 7.54$, suggesting the workers move relatively easily across firms within a sector. However, it is smaller than the value of 10.82 that Berger et al. (2022) estimate in the US context and much larger than the 1.02 Felix (2021) finds in the Brazilian context. We find an elasticity of substitution across sectors of $\nu = 0.456$ similar to the estimate of 0.42 that Berger et al. (2022) find and smaller than the 0.80 Felix (2021) finds. And finally, we find $\eta = 0.203$.

Productivity Shock Parameters. We jointly calibrate λ, σ_N^2, and σ_S^2 to match the region-wide average variance of log total payments to labor in sector s. This is matching Figure 4(b) on average. In the data, that is

$$\int_{\mathcal{S}} \frac{\bar{\ell}_{is}}{\ell_i} Var\left(\log\left(\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)\right)\right) ds.$$

In the model, that is

$$\left(\frac{1}{\eta_N+\eta}\left[1+\eta_N-\frac{(1-\eta)\eta_S}{\eta+\eta_N+\eta_S}\right]\right)^2\left\{\sigma_S^2+\left(\int_S\frac{\overline{\ell}_{is}}{\ell_i}HHI_{is}ds\right)\sigma_N^2\right\}.$$

We choose the parameters to minimize a quadratic loss function of the log differences weighted by population.

Migration and Spillover Parameters. We take the Fréchet parameter controlling the migration elasticity $\theta = 3$ from Redding (2016), which is consistent with the evidence from Bryan and Morten (2019) in Indonesia and Hornbeck and Moretti (2020) in the United States. We set $\gamma_u = -0.25$ to match the average housing spending in Japan.

We choose the strength of other spillovers, γ_z, to match the estimated agglomeration externality from Combes et al. (2011), 0.03.

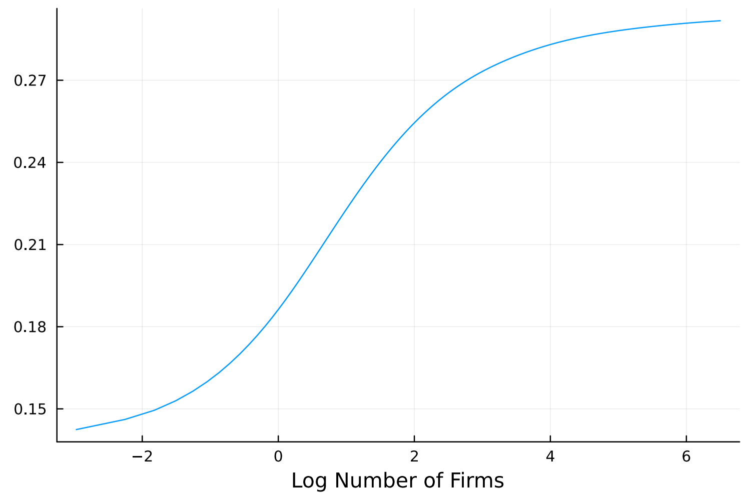
5.2 Exact Hat Algebra

We avoid explicitly calibrating the location-specific entry costs, ψ_n, fundamental productivities, $\overline{z}_n$, and fundamental amenities, $\overline{u}_n$, by solving the model in percent changes from the observed equilibrium in 2019 using the exact hat algebra method pioneered by Dekle et al. (2008). This method implicitly calibrates those parameters to exactly match the average wages, population, and number of firms in Japanese manufacturing sectors. Details are in Appendix D.

6 Quantitative Results

In Section 5, we calibrated the model from Section 4. In this section, we present the results. We start by plotting $\Omega(m)$, which summarizes the degree of increasing returns to scale due to our mechanism. We then turn to demonstrating what the mechanism means for the geography of economic activity. In particular, we calculate the implied elasticity of wages to population and the degree of under-entry for each commuting zone. Finally, we simulate how economic activity would change if the central government put in place the optimal entry subsidy, which is a place-based industrial policy.

Figure 6: $\log \Omega(m)$



6.1 Increasing Returns to Scale

We plot $\Omega(m)$ in Figure 6. As suggested by Proposition 1, the slope of the line is the degree of increasing returns to scale. And, in the data, this is determined by how quickly the average covvariance of log productivity and log employment increases as the number of firms increases. In the model, it is determined by how quickly the employment weighted average HHI across sectors decreases. As emphasized by Proposition 2, the curve levels off as the number of firms gets very large. At that point, granularity ceases to be an important driver of agglomeration.

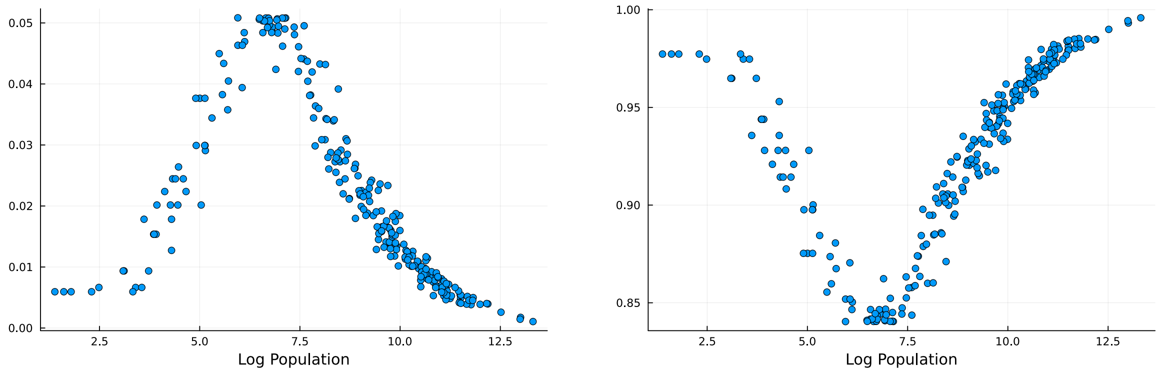
Perhaps more surprisingly, we see that the curve is relatively flat in very small regions as well, i.e. where the average number of firms is less than 1. At that point, new firm entrants are likely to enter sectors with no firms. Therefore, the employment weighted average HHI does not change much. Instead, the new sector with an HHI of 1 is simply averaged in with the other sectors with a high HHI. As the number of firms incrase, new firms are likely to enter sectors that already have firms. Therefore, those sectors see a precipitous decline in HHI, and the average HHI decreases as well.

6.2 The Geography of Economic Activity

Using this estimate of $\Omega(m)$, we can estimate the contribution of granularity to agglomeration benefits for each labor market. We provide two ways of interpreting it in Figure 7. Figure 7(b) captures the degree of under entry in each commuting zone in Japan by plotting the ratio of firm profits to a firm's marginal product. Large and small locations with very little externalities see firms mostly internalizing their impact on the productive

Figure 7: Strength of the externality

(a) Elasticity of wage to population (b) Firm Entry Wedge



capacity of the labor market. For example, firms in Tokyo capture 99.6% of their contribution to production in profits. By contrast, in modest sized cities, adding another firm will have a large impact on the distribution of wages in a sectoral labor market. Firms are therefore less likely to enter. In fact, in commuting zones with an average of 1.9 firms across sectors, profits are only 84% of the firm's marginal product.

Figure 7(a) show the elasticity of the average wage to an increase in population if the granular externality were the only reason for agglomeration (i.e. $\gamma_z = 0$). This is calculated according to lemma 1. The elasticity gets as high as 0.051 in locations with an average 1.9 firms per sector. The implied elasticity is much smaller for larger locations. For example, the elasticity is 0.001 in Tokyo. For context, Combes et al. (2011) find that most causal estimates of the urban wage premium find an elasticity between 0.02 and 0.05 when pooling across locations of all sizes. Other more recent papers like Kline and Moretti (2014) find that the elasticity could be as high as 0.4 in manufacturing industries. Thus, this mechanism can explain a significant portion of the productivity advantage of large locations, though it is not the whole story, especially for very large locations. [19]

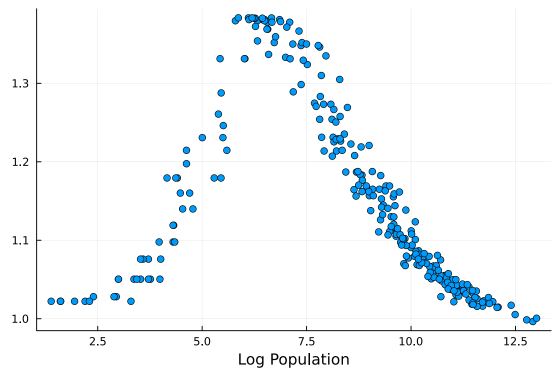
6.3 Optimal Place-Based Industrial Policy

As we know from Proposition 3, there is under entry in equilibrium. In this section, we assess the degree of under entry, and see how the geography of economic activity in Japan would change were the Japanese government to put in place the optimal entry

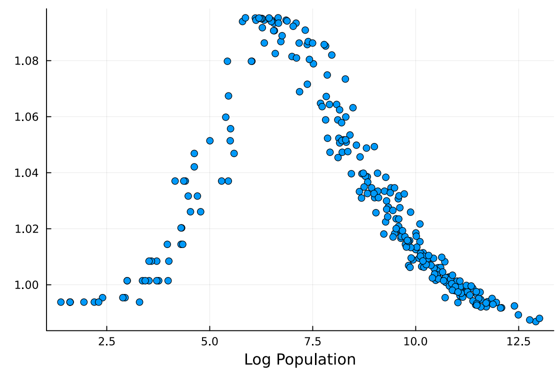
[19] We include these graphs under the assumption that firms compete oligopsonistically in Figure **??**. As suggested by the shape of $\Omega(m)$, the elasticity of wages never gets above 0.06. However, the elasticity remains above 0.05 for even medium sized cities, and the elasticity in Tokyo is around 0.01 rather than 0.001.

Figure 8: Optimal Policy

(a) Change in Number of Firms



(b) Change in Population



subsidy, paid for with a tax propotional to wages. We graph the results for the change in the number of firms and the change in population in Figure 8.

For the modest-sized commuting zones where the labor market pooling externality is strongest, the number of firms increases by 38%. This is driven in part by the direct subsidy. But it also happens because of the reallocation of economic activity. Figure 8b shows that those locations see an increase in population of 9.5%. Meanwhile, large locations like Tokyo actually see a decrease in population. That decrease in population at the top of the distribution is so large that even though there is now a subsidy on firm entry, the number of firms also declines.

7 Conclusion

When individual firms matter, there are benefits to agglomeration. Having many firms all in the same location can allow workers to move to the firm where their skills will be best used. This has important implications for understanding where economic activity takes place and optimal policy.

This paper offers simple model that allows for a transparent mapping between the data and the positive and normative implications. There is still a lot more to be said about how granularity interacts with the endogenous skill acquisition, heterogeneous firms and sectors, and sorting.

References

Adachi, D., T. Fukai, D. Kawaguchi, and Y. U. Saito (2020). Commuting zones in Japan. Discussion papers 20-E-021, Research Institute of Economy, Trade and Industry (RIETI).

Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics 129*(3), 1085–1140.

Andersson, F., S. Burgess, and J. I. Lane (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics 61*(1), 112–128.

Andersson, M., J. Klaesson, and J. P. Larsson (2014). The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? *Papers in Regional Science 93*(4), 727–747.

Baek, Y., K. Hayakawa, K. Tsubota, S. Urata, and K. Yamanouchi (2021). Tariff pass-through in wholesaling: Evidence from firm-level data in japan. *Journal of the Japanese and International Economies 62*, 101164.

Berger, D., K. Herkenhoff, and S. Mongey (2022). Labor market power. *American Economic Review 112*(4), 1147–93.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2018). Global firms. *Journal of Economic Literature 56*(2), 565–619.

Bryan, G. and M. Morten (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy 127*(5), 2229–2268.

Combes, P.-P., G. Duranton, and L. Gobillon (2011). The identification of agglomeration economies. *Journal of economic geography 11*(2), 253–266.

Conte, M., I. Méjean, T. K. Michalski, and B. Schmutz (2024). The volatility advantages of large labor markets. *Available at SSRN*.

Dai, M. and J. Xu (2017). Firm-specific exchange rate shocks and employment adjustment: Evidence from china. *Journal of International Economics 108*, 54–66.

Davis, D. R. and J. I. Dingel (2019). A spatial knowledge economy. *American Economic Review 109*(1), 153–170.

de Almeida, E. T. and R. de Moraes Rocha (2018). Labor pooling as an agglomeration factor: Evidence from the brazilian northeast in the 2002–2014 period. *EconomiA 19*(2), 236–250.

Dekle, R., J. Eaton, and S. Kortum (2008). Global rebalancing with gravity: Measuring the burden of adjustment. *IMF staff papers 55*(3), 511–540.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2063–2117. Elsevier.

Durrett, R. (2019). *Probability: theory and examples*, Volume 49. Cambridge university press.

Ekholm, K., A. Moxnes, and K. H. Ulltveit-Moe (2012). Manufacturing restructuring and the role of real exchange rate shocks. *Journal of International Economics 86*(1), 101–117.

Ellison, G. and E. L. Glaeser (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of political economy 105*(5), 889–927.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review 100*(3), 1195–1213.

Fajgelbaum, P. D. and C. Gaubert (2020). Optimal spatial policies, geography, and sorting. *The Quarterly Journal of Economics 135*(2), 959–1036.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The next generation of the penn world table. *American Economic Review 105*(10), 3150–82.

Feenstra, R. C. and J. Romalis (2014). International prices and endogenous quality. *The Quarterly Journal of Economics 129*(2), 477–527.

Felix, M. (2021). Trade, labor market concentration, and wages. *Job Market Paper*.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Gaubert, C. and O. Itskhoki (2021). Granular comparative advantage. *Journal of Political Economy 129*(3), 871–939.

Gaubert, C., O. Itskhoki, and M. Vogler (2021). Government policies in a granular global economy. *Journal of Monetary Economics 121*, 95–112.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of political economy 118*(3), 536–598.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, 1127–1150.

Hornbeck, R. and E. Moretti (2020). Estimating who benefits from productivity growth: Local and aggregate effects of city tfp shocks on wages, rents, and inequality. Technical report, Working Paper.

Hosono, K., M. Takizawa, and K. Tsuru (2015). The impact of demand shock on the employment of temporary agency workers: Evidence from japan during the global financial crisis. *Seoul Journal of Economics 28*, 265–284.

Kambayashi, R. (2017). *Standard Workers, Non-Standard Workers; Seiki no Sekai, Hiseiki no Sekai (in Japanese)*. Keio University Press.

Kline, P. and E. Moretti (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly journal of economics 129*(1), 275–331.

Kondo, K. (2023). Municipality-level Panel Data and Municipal Mergers in Japan. Technical papers 23-T-001, Research Institute of Economy, Trade and Industry (RIETI).

Krugman, P. (1992). *Geography and trade*. MIT press.

Marshall, A. (1920). *Principles of Economics*. Macmillan.

Miyauchi, Y. (2018). Matching and agglomeration: Theory and evidence from japanese firm-to-firm trade. Technical report, working paper.

Moretti, E. and M. Yi (2024). Size matters: Matching externalities and the advantages of large labor markets. Technical report, National Bureau of Economic Research.

Morikawa, M. (2010). Volatility, nonstandard employment, and productivity: An empirical analysis using firm-level data. Discussion papers 10-E-025, Research Institute of Economy, Trade and Industry (RIETI).

Nakajima, K. and T. Okazaki (2012). Labor pooling as a source of industrial agglomeration—the case of the japanese manufacturing industries—. *Economic Review 63*(3), 227–235.

Nucci, F. and A. F. Pozzolo (2010). The exchange rate, employment and hours: What firm-level data say. *Journal of International Economics 82*(2), 112–123.

Overman, H. G. and D. Puga (2010). Labor pooling as a source of agglomeration: An empirical investigation. In *Agglomeration economics*, pp. 133–150. University of Chicago Press.

Papageorgiou, T. (2022). Occupational matching and cities. *American Economic Journal: Macroeconomics 14*(3), 82–132.

Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics 101*, 148–167.

Redding, S. J. (2022). Trade and geography. *Handbook of International Economics 5*, 147–217.

Rosenthal, S. S. and W. C. Strange (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2119–2171. Elsevier.

Yokoyama, I., K. Higa, and D. Kawaguchi (2021). Employment adjustments of regular and non-regular workers to exogenous shocks: evidence from exchange-rate fluctuation. *ILR Review 74*(2), 470–510.

A Proofs for Section 2

In this appendix, we go through and prove the theoretical results in Section 2. We start by begin explicit about what we mean by second order approximation. In particular, we introduce a variable β so that the productivity shock to firm n is $z_{isn}\left(\tilde{A}_{is}(\omega)\tilde{a}_{isn}(\omega)\right)^{\beta}$. Then production is

$$Y_i(\ell, m, \beta) = \mathbb{E}\left[\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} z_{isn}\left(\overline{a}_{isn}(\omega)\right)^{\beta} f\left(\ell_{isn}(\omega, \beta)\right) ds\right],$$

where $\overline{a}_{isn}(\omega) \equiv \tilde{A}_{is}(\omega)\tilde{a}_{isn}(\omega)$. We then do a second order approximation to this in β.

Lemma 2. *To first order,*

$$\frac{\partial \log L_{isn}(\omega, \beta)}{\partial \beta} = \frac{1}{\eta + \eta_N}\left[\log \overline{a}_{isn}(\omega) - \frac{\frac{\eta_S}{\eta + \eta_N}}{1 + \frac{\eta_S}{\eta + \eta_N}}\left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \overline{a}_{isn'}(\omega)\right)\right].$$

Proof. We start by characterizing the worker's problem. It is

$$\max_{\overline{\boldsymbol{L}}_i', \boldsymbol{L}_i'(\omega)} \mathbb{E}\left[\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} w_{isn}(\omega) L'_{isn}(\omega) ds\right]$$

such that

$$\begin{aligned}
1 &= \int_{\mathcal{S}} \overline{L}'_{is} ds, \\
1 &= \int_{\mathcal{S}} \overline{L}'_{is} \cdot g_S\left(\frac{L'_{is}(\omega)}{\overline{L}'_{is}}\right) ds, \\
\overline{L}'_{is} &= \sum_{n \in \mathcal{N}_{is}} \overline{L}'_{isn}, \\
L'_{is}(\omega) &= \sum_{n \in \mathcal{N}_{is}} \overline{L}'_{isn} \cdot g_N\left(\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}}\right),
\end{aligned}$$

where we assume the inequalities bind since workers can always increase utility by working slightly more. Taking the first order conditions,

$$\begin{aligned}
w_{isn}(\omega) &= \lambda_{is}(\omega) g_N' \left(\frac{L_{isn}'(\omega)}{\overline{L}_{isn}}\right) \\
\lambda_{is}(\omega) &= \lambda_i(\omega) g_S' \left(\frac{L_{is}'(\omega)}{\overline{L}_{is}'}\right) \\
0 &= \mathbb{E}\left[\lambda_{is}(\omega)\left(g_N\left(\frac{L_{isn}'(\omega)}{\overline{L}_{isn}'}\right) - \frac{L_{isn}'(\omega)}{\overline{L}_{isn}'} g_N'\left(\frac{L_{isn}'(\omega)}{\overline{L}_{isn}'}\right)\right)\right] + \overline{\lambda}_{is} \\
\overline{\lambda}_{is} &= \mathbb{E}\left[\lambda_i(\omega)\left(g_S\left(\frac{L_{is}'(\omega)}{\overline{L}_{is}'}\right) - \frac{L_{is}'(\omega)}{\overline{L}_{is}'} g_S'\left(\frac{L_{is}'(\omega)}{\overline{L}_{is}'}\right)\right)\right] + \overline{\lambda}_i.
\end{aligned}$$

The wages are determined competitively so that

$$w_{isn}(\omega) = z_{isn}(\overline{a}_{isn}(\omega))^{\beta} \left(L_{isn}(\omega)\ell_i\right)^{-\eta}.$$

We can then start by finding the solution when $\beta = 0$. Then notice that if $L_{isn}(\omega) = \overline{L}_{isn}$ and wages are constant across all states of the world, we have

$$\begin{aligned}
w &= \lambda_{is}(\omega) g_N'(1) \\
\lambda_{is}(\omega) &= \lambda_i(\omega) g_S'(1) \\
0 &= \mathbb{E}\left[\lambda_{is}(\omega)\left(1 - g_N'(1)\right)\right] + \overline{\lambda}_{is} \\
\overline{\lambda}_{is} &= \mathbb{E}\left[\lambda_i(\omega)\left(1 - g_S'(1)\right)\right] + \overline{\lambda}_i.
\end{aligned}$$

With $g_n'(1) = g_S'(1) = 1$, we find that $w = \lambda_{is}(\omega) = \lambda_i(\omega)$, and $\overline{\lambda}_{is} = \overline{\lambda}_i = 0$. Then,

$$z_{isn} \overline{L}_{isn}^{-\eta} \ell_i^{-\eta} = w.$$

We take the derivative with respect to β and plug in for the value at $\beta = 0$. At that point, $L_{isn}'(\omega, 0) = \overline{L}_{isn}'(0)$. Then

$$\begin{aligned}
\log \overline{a}_{isn}(\omega) - \eta \frac{\partial \log L_{isn}(\omega,\beta)}{\partial \beta} &= \frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta} + \eta_N \left(\frac{\partial \log L_{isn}(\omega,\beta)}{\partial \beta} - \frac{\partial \log \overline{L}_{isn}(\beta)}{\partial \beta}\right) \\
\frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta} &= \frac{\partial \log \lambda_i(\omega)}{\partial \beta} + \eta_S \left(\frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} - \frac{\partial \log \overline{L}_{is}(\beta)}{\partial \beta}\right).
\end{aligned}$$

And for the next first order condition,

$$
\begin{aligned}
0 = &\frac{\partial \overline{\lambda}_{is}(\beta)}{\partial \beta} \\
&+ \mathbb{E}\left[\left(g_N\left(\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}}\right) - \frac{L'_{isn}(\omega)}{\overline{L}'_{isn}} g'_N\left(\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}}\right)\right)\frac{\partial \lambda_{is}(\omega,\beta)}{\partial \beta}\right] \\
&+ \mathbb{E}\left[-\lambda_{is}(\omega)\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}} g''_N\left(\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}}\right)\frac{1}{\overline{L}'_{isn}}\frac{\partial L'_{isn}(\omega,\beta)}{\partial \beta}\right] \\
&+ \mathbb{E}\left[\lambda_{is}(\omega,\beta)\frac{L'_{isn}(\omega)}{\overline{L}'_{isn}} g''_N\left(\frac{L'_{isn}(\omega)}{\overline{L}_{isn}}\right)\frac{L'_{isn}(\omega)}{\overline{L}^2_{isn}}\frac{\partial \overline{L}_{isn}(\omega,\beta)}{\partial \beta}\right] \\
\frac{\partial \overline{\lambda}_{is}(\beta)}{\partial \beta} = &\eta_N\left(\mathbb{E}\left[\frac{\partial \log L'_{isn}(\omega,\beta)}{\partial \beta}\right] - \frac{\partial \log \overline{L}'_{isn}(\beta)}{\partial \beta}\right).
\end{aligned}
$$

Similarly,

$$
\frac{\partial \overline{\lambda}_{i}(\beta)}{\partial \beta} - \frac{\partial \overline{\lambda}_{is}(\beta)}{\partial \beta} = \eta_S\left(\mathbb{E}\left[\frac{\partial \log L'_{is}(\omega,\beta)}{\partial \beta}\right] - \frac{\partial \log \overline{L}'_{is}(\beta)}{\partial \beta}\right)
$$

These go along with the constraints

$$
\begin{aligned}
0 &= \int_{\mathcal{S}} \overline{L}_{is}\frac{\partial \log \overline{L}_{is}(\beta)}{\partial \beta} ds \\
\frac{\partial \log \overline{L}_{is}(\beta)}{\partial \beta} &= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\frac{\partial \log \overline{L}_{isn}(\beta)}{\partial \beta} \\
0 &= \int_{\mathcal{S}} \overline{L}_{is}\frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} ds \\
\frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} &= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\frac{\partial \log L_{isn}(\omega)}{\partial \beta}.
\end{aligned}
$$

Then notice that if

$$
0 = \frac{\partial \overline{\lambda}_{i}(\beta)}{\partial \beta} = \frac{\partial \overline{\lambda}_{is}(\beta)}{\partial \beta} = \frac{\partial \log \overline{L}'_{is}(\beta)}{\partial \beta} = \frac{\partial \log \overline{L}'_{isn}(\beta)}{\partial \beta},
$$

the other equations can still hold. We would then have

$$\begin{aligned}
\log \bar{a}_{isn}(\omega) &= \frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta} + (\eta+\eta_N)\frac{\partial \log L_{isn}(\omega,\beta)}{\partial \beta} \\
\frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta} &= \frac{\partial \log \lambda_i(\omega)}{\partial \beta} + \eta_S \frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} \\
0 &= \mathbb{E}\left[\frac{\partial \log L'_{isn}(\omega,\beta)}{\partial \beta}\right] \\
0 &= \mathbb{E}\left[\frac{\partial \log L'_{is}(\omega,\beta)}{\partial \beta}\right] \\
0 &= \int_{\mathcal{S}} \bar{L}_{is} \frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} ds \\
\frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} &= \sum_{n\in\mathcal{N}_{is}} \frac{\bar{L}_{isn}}{\bar{L}_{is}} \frac{\partial \log L_{isn}(\omega)}{\partial \beta}.
\end{aligned}$$

So then,

$$\begin{aligned}
\frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} &= \sum_{n\in\mathcal{N}_{is}} \frac{\bar{L}_{isn}}{\bar{L}_{is}} \frac{1}{\eta+\eta_N}\left(\log \bar{a}_{isn}(\omega) - \frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta}\right) \\
&= \frac{1}{\eta+\eta_N}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\bar{L}_{isn}}{\bar{L}_{is}} \log \bar{a}_{isn}(\omega)\right) - \frac{1}{\eta+\eta_N}\frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta}.
\end{aligned}$$

Plugging in to the second equation

$$\begin{aligned}
\frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta} &= \frac{\partial \log \lambda_i(\omega)}{\partial \beta} + \eta_S \frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} \\
&= \frac{\partial \log \lambda_i(\omega)}{\partial \beta} + \frac{\eta_S}{\eta+\eta_N}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\bar{L}_{isn}}{\bar{L}_{is}} \log \bar{a}_{isn}(\omega)\right) - \frac{\eta_S}{\eta+\eta_N}\frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta} \\
&= \frac{1}{1+\frac{\eta_S}{\eta+\eta_N}}\frac{\partial \log \lambda_i(\omega)}{\partial \beta} + \frac{\frac{\eta_S}{\eta+\eta_N}}{1+\frac{\eta_S}{\eta+\eta_N}}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\bar{L}_{isn}}{\bar{L}_{is}} \log \bar{a}_{isn}(\omega)\right)
\end{aligned}$$

Therefore,

$$\frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} = -\frac{1}{\eta+\eta_N+\eta_S}\frac{\partial \log \lambda_i(\omega)}{\partial \beta} + \frac{1}{\eta+\eta_N+\eta_S}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\bar{L}_{isn}}{\bar{L}_{is}} \log \bar{a}_{isn}(\omega)\right).$$

And we also get

$$\begin{aligned}\frac{\partial \log L_{isn}(\omega,\beta)}{\partial \beta} &= \frac{1}{\eta+\eta_N}\log \overline{a}_{isn}(\omega) - \frac{1}{\eta+\eta_N}\frac{\partial \log \lambda_{is}(\omega,\beta)}{\partial \beta} \\ &= \frac{1}{\eta+\eta_N}\left[\log \overline{a}_{isn}(\omega) - \frac{\frac{\eta_S}{\eta+\eta_N}}{1+\frac{\eta_S}{\eta+\eta_N}}\left(\sum_{n'\in\mathcal{N}_{is}}\frac{\overline{L}_{isn'}}{\overline{L}_{is}}\log \overline{a}_{isn'}(\omega)\right)\right] \\ &\qquad - \frac{1}{\eta+\eta_N+\eta_S}\frac{\partial \log \lambda_i(\omega)}{\partial \beta}.\end{aligned}$$

Finally note that,

$$\begin{aligned}0 &= \int_{\mathcal{S}} \overline{L}_{is}\frac{\partial \log L_{is}(\omega,\beta)}{\partial \beta} ds \\ &= \int_{\mathcal{S}} \overline{L}_{is}\left[-\frac{1}{\eta+\eta_N+\eta_S}\frac{\partial \log \lambda_i(\omega)}{\partial \beta} + \frac{1}{\eta+\eta_N+\eta_S}\left(\sum_{n\in\mathcal{N}_{is}}\frac{\overline{L}_{isn}}{\overline{L}_{is}}\log \overline{a}_{isn}(\omega)\right)\right] ds\end{aligned}$$

$$\frac{\partial \log \lambda_i(\omega)}{\partial \beta} = \mathbb{E}[\log \overline{a}_{isn}(\omega)] = 0.$$

□

Next, we move onto the second order.

Lemma 3. *For small ex-post shocks, $Y_i(\ell, m, \beta)$ is approximately given by*

$$(\mu_z)^{\eta}(\ell_i)^{1-\eta}(m_i)^{\eta}\Omega(m_i),$$

for

$$\begin{aligned}\Omega(m_i) \equiv\ & \mathbb{E}[a_{isn}(\omega)^{\beta}] \\ & + (1-\eta)\int_{\mathcal{S}}\sum_{n\in\mathcal{N}_{is}}\frac{\ell_{isn}(0)}{\ell_i}\mathbb{E}\left[\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta}\log \overline{a}_{isn}(\omega)\beta^2\right] ds \\ & - \frac{1-\eta}{2}\eta\int_{\mathcal{S}}\sum_{n\in\mathcal{N}_{is}}\frac{\ell_{isn}(0)}{\ell_i}\mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta}\right)^2\beta^2\right] ds \\ & - \frac{1-\eta}{2}\eta_N\int_{\mathcal{S}}\sum_{n\in\mathcal{N}_{is}}\frac{\ell_{isn}(0)}{\ell_i}\mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta}\right)^2\beta^2\right] ds \\ & - \frac{1-\eta}{2}\eta_S\int_{\mathcal{S}}\frac{\ell_{is}(0)}{\ell_i}\mathbb{E}\left[\left(\frac{\partial \log \ell_{is}(\omega)}{\partial \beta}\right)^2\beta^2\right] ds.\end{aligned} \tag{23}$$

where $\mu_z = \mathbb{E}[z_{isn}^{\frac{1}{\eta}}]$.

Proof. We start by solving the model at $\beta = 0$. There is then no uncertainty so we propose a solution where $L_{isn}(\omega)$ and $w_{isn}(\omega)$ is constant across all ω. Firms maximize profits taking wages as given. Therefore

$$z_{isn}(1-\eta)\left(L_{isn}\ell_i\right)^{-\eta} = w_{isn}.$$

The worker then maximizes utility taking wages w_{isn} as given. Since labor is freely mobile in the first period, the worker will only work for firms that offer the highest wages. Thus, wages must be equalized across all firms, so that

$$L_{isn} = \frac{1}{\ell_i}\left(\frac{z_{isn}(1-\eta)}{w}\right)^{\frac{1}{\eta}}.$$

Labor market clearing requires that

$$\begin{aligned}
1 &= \int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} L_{isn} ds \\
1 &= \int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} \frac{1}{\ell_i}\left(\frac{z_{isn}(1-\eta)}{w}\right)^{\frac{1}{\eta}} ds \\
w &= \frac{1-\eta}{\ell_i^{\eta}}\left(\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}^{\frac{1}{\eta}} ds\right)^{\eta}.
\end{aligned}$$

We can then plug this into production. This gives

$$\begin{aligned}
Y_i(\ell, m, 0) &= \int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\left(L_{isn}\ell_i\right)^{1-\eta} ds \\
&= \int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\left(w^{-\frac{1}{\eta}}(z_{isn})^{\frac{1}{\eta}}(1-\eta)^{\frac{1}{\eta}}\right)^{1-\eta} ds \\
&= w^{-\frac{1-\eta}{\eta}}(1-\eta)^{\frac{1-\eta}{\eta}}\left(\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} (z_{isn})^{\frac{1}{\eta}} ds\right) \\
&= (\ell_i)^{1-\eta}\left(\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} (z_{isn})^{\frac{1}{\eta}} ds\right)^{\eta} \\
&= (\mu_z)^{\eta}(\ell_i)^{1-\eta}(m_i)^{\eta},
\end{aligned}$$

where $\mu_z = \mathbb{E}[z_{isn}^{1/\eta}]$ which we can plug in since we assume that the law of large numbers holds across the continuum of sectors, and the number of firms is m_i.

We then take the first derivative. That is

$$\frac{\partial Y_i}{\partial \beta} = \mathbb{E}\left[\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\overline{a}_{isn}(\omega)^{\beta}\ell_{isn}(\omega,\beta)^{1-\eta}\log(\overline{a}_{isn}(\omega))ds\right]$$
$$+\mathbb{E}\left[\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\overline{a}_{isn}(\omega)^{\beta}(1-\eta)\ell_{isn}(\omega,\beta)^{-\eta}\frac{\partial \ell_{isn}(\omega,\beta)}{\partial\beta}ds\right].$$

We can then take the second derivative. That is

$$\frac{\partial^2 Y_i}{\partial \beta^2} = \mathbb{E}\left[\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\overline{a}_{isn}(\omega)^{\beta}\ell_{isn}(\omega,\beta)^{1-\eta}\log(\overline{a}_{isn}(\omega))^2ds\right]$$
$$+2\mathbb{E}\left[\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\overline{a}_{isn}(\omega)^{\beta}(1-\eta)\ell_{isn}(\omega,\beta)^{-\eta}\frac{\partial \ell_{isn}(\omega,\beta)}{\partial\beta}\log(\overline{a}_{isn}(\omega))ds\right]$$
$$-\mathbb{E}\left[\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\overline{a}_{isn}(\omega)^{\beta}(1-\eta)\eta\ell_{isn}(\omega,\beta)^{-\eta-1}\left(\frac{\partial \ell_{isn}(\omega,\beta)}{\partial\beta}\right)^2ds\right]$$
$$+\mathbb{E}\left[\int_{\mathcal{S}} \sum_{n\in\mathcal{N}_{is}} z_{isn}\overline{a}_{isn}(\omega)^{\beta}(1-\eta)\ell_{isn}(\omega,\beta)^{-\eta}\frac{\partial^2 \ell_{isn}(\omega,\beta)}{\partial\beta^2}ds\right]$$

Then we plug in at $\beta = 0$. We use the fact that

$$z_{isn}\overline{a}_{isn}(\omega)^{\beta}\ell_{isn}(\omega,\beta)^{1-\eta} = \frac{1}{1-\eta}w_i(0)\ell_{isn}(\omega,0).$$

Then, to second order production is approximately given by,

$$\begin{aligned}
Y_i(\ell, m, \beta) \approx (\mu_z)^\eta (\ell)^{1-\eta} (m)^\eta \\
&+ \frac{1}{1-\eta} w_i(0) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \ell_{isn}(0) \mathbb{E}[\log(\overline{a}_{isn}(\omega))] \beta ds \\
&+ w_i(0) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \ell_{isn}(0) \mathbb{E}\left[\frac{\partial \log \ell_{isn}(\omega, \beta))}{\partial \beta}\right] \beta ds \\
&+ \frac{1}{2} \frac{1}{1-\eta} w_i(0) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \ell_{isn}(0) \mathbb{E}\left[\log(\overline{a}_{isn}(\omega))^2\right] \beta^2 ds \\
&+ w_i(0) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \ell_{isn}(0) \mathbb{E}\left[\frac{\partial \log \ell_{isn}(\omega, \beta)}{\partial \beta} \log(\overline{a}_{isn}(\omega))\right] \beta^2 ds \\
&- \eta \frac{1}{2} w_i(0) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \ell_{isn}(0) \mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega, \beta)}{\partial \beta}\right)^2\right] \beta^2 \\
&+ \frac{1}{2} w_i(0) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \mathbb{E}\left[\frac{\partial^2 \ell_{isn}(\omega, \beta)}{\partial \beta^2}\right] \beta^2 ds.
\end{aligned}$$

Then we can make a few substitutions. First, note that

$$\mathbb{E}[\overline{a}_{isn}(\omega)^\beta] \approx 1 + \mathbb{E}[\log(\overline{a}_{isn}(\omega))]\beta + \frac{1}{2}\mathbb{E}[\log(\overline{a}_{isn}(\omega))^2]\beta^2 + o(\beta^3).$$

Then we also note that

$$w_i(0)\ell = (1-\eta)(\mu_z)^\eta (\ell)^{1-\eta}(m)^\eta.$$

We also know that to first order

$$\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \ell_{isn}(0) \mathbb{E}\left[\frac{\partial \log \ell_{isn}(\omega, \beta))}{\partial \beta}\right] ds = 0.$$

Therefore, we can write,

$$
\begin{aligned}
Y_i(\ell, m, \beta) \approx (\mu_z)^\eta (\ell)^{1-\eta}(m)^\eta \Bigg\{ & 1 + \mathbb{E}[\overline{a}_{isn}(\omega)^\beta] - 1 \\
& + (1-\eta)\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\frac{\partial \log \ell_{isn}(\omega, \beta)}{\partial \beta} \log(\overline{a}_{isn}(\omega)) \right] \beta^2 ds \\
& - \eta \frac{1-\eta}{2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega, \beta)}{\partial \beta} \right)^2 \beta^2 \right] \\
& + \frac{1-\eta}{2} \frac{1}{\ell_i} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \mathbb{E}\left[\frac{\partial^2 \ell_{isn}(\omega, \beta)}{\partial \beta^2} \right] \beta^2 ds \Bigg\} \\
= (\mu_z)^\eta (\ell)^{1-\eta}(m)^\eta \Bigg\{ & \mathbb{E}[\overline{a}_{isn}(\omega)^\beta] \\
& + (1-\eta)\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\frac{\partial \log \ell_{isn}(\omega, \beta)}{\partial \beta} \left(\log(\overline{a}_{isn}(\omega)) - \frac{\eta}{2} \frac{\partial \log \ell_{isn}(\omega, \beta)}{\partial \beta} \right) \beta^2 \right] \\
& + \frac{1-\eta}{2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \mathbb{E}\left[\frac{\partial^2 L_{isn}(\omega, \beta)}{\partial \beta^2} \right] \beta^2 ds \Bigg\}.
\end{aligned}
$$

Next, we turn to transforming the last line. We do this by noting that labor needs to satisfy,

$$
\begin{aligned}
1 &= \int_{\mathcal{S}} \overline{L}_{is} ds \\
\overline{L}_{is} &= \sum_{n \in \mathcal{N}_{is}} \overline{L}_{isn} \\
1 &= \int_{\mathcal{S}} \overline{L}_{is} \cdot g_S \left(\frac{L_{is}(\omega)}{\overline{L}_{is}} \right) ds \\
L_{is}(\omega) &= \sum_{n \in \mathcal{N}_{is}} \overline{L}_{isn} \cdot g_N \left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}} \right).
\end{aligned}
$$

Then taking the derivative, we find that,

$$
\begin{aligned}
0 &= \int_{\mathcal{S}} \frac{\partial \overline{L}_{is}}{\partial \beta} ds \\
\frac{\partial \overline{L}_{is}}{\partial \beta} &= \sum_{n \in \mathcal{N}_{is}} \frac{\partial \overline{L}_{isn}}{\partial \beta} \\
0 &= \int_{\mathcal{S}} \left[g_S \left(\frac{L_{is}(\omega)}{\overline{L}_{is}} \right) \frac{\partial \overline{L}_{is}}{\partial \beta} + g_S' \left(\frac{L_{is}(\omega)}{\overline{L}_{is}} \right) \frac{\partial L_{is}(\omega)}{\partial \beta} - \frac{L_{is}(\omega)}{\overline{L}_{is}} g_S' \left(\frac{L_{is}(\omega)}{\overline{L}_{is}} \right) \frac{\partial \overline{L}_{is}}{\partial \beta} \right] ds \\
\frac{\partial L_{is}(\omega)}{\partial \beta} &= \sum_{n \in \mathcal{N}_{is}} g_N \left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}} \right) \frac{\partial \overline{L}_{isn}}{\partial \beta} + g_N' \left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}} \right) \frac{\partial L_{isn}(\omega)}{\partial \beta} - \frac{L_{isn}(\omega)}{\overline{L}_{isn}} g_N' \left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}} \right) \frac{\partial \overline{L}_{isn}}{\partial \beta}.
\end{aligned}
$$

We then take the second derivative, substituting in $L_{is}(\omega) = \overline{L}_{is}$, $\frac{\partial \overline{L}_{is}}{\partial \beta} = 0$, $g_S(1) = g_N(1) = g_S'(1) = g_N'(1) = 1$, $g_S''(1) = \eta_s$ and $g_N''(1) = \eta_N$ at $\beta = 0$. Then

$$
\begin{aligned}
0 &= \int_{\mathcal{S}} \frac{\partial^2 \overline{L}_{is}}{\partial \beta^2} ds \\
\frac{\partial^2 \overline{L}_{is}}{\partial \beta^2} &= \sum_{n \in \mathcal{N}_{is}} \frac{\partial^2 \overline{L}_{isn}}{\partial \beta^2} \\
\int_{\mathcal{S}} \frac{\partial^2 L_{is}(\omega)}{d\beta^2} ds &= -\eta_S \int_{\mathcal{S}} \overline{L}_{is} \left(\frac{\partial \log L_{is}(\omega)}{\partial \beta} \right)^2 ds \\
\sum_{n \in \mathcal{N}_{is}} \frac{\partial^2 L_{isn}(\omega)}{\partial \beta^2} &= \frac{\partial^2 L_{is}(\omega)}{\partial \beta^2} - \eta_N \sum_{n \in \mathcal{N}_{is}} \overline{L}_{isn} \left(\frac{\partial \log L_{isn}(\omega)}{\partial \beta} \right)^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{1-\eta}{2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \mathbb{E} \left[\frac{\partial^2 L_{isn}(\omega, \beta)}{\partial \beta^2} \beta^2 \right] = &-\frac{1-\eta}{2} \eta_N \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \left(\frac{\partial \log L_{isn}(\omega)}{\partial \beta} \right)^2 \beta^2 ds \\
&- \frac{1-\eta}{2} \eta_S \int_{\mathcal{S}} \overline{L}_{is} \left(\frac{\partial \log L_{is}(\omega)}{\partial \beta} \right)^2 \beta^2 ds.
\end{aligned}
$$

Plugging this back into the second order approximation to the production function gives (23). □

Then we can come back and prove lemma 1.

Lemma 1. *For small ex-post shocks, if $Y_i(\ell, m)$ has increasing returns to scale, i.e. $\frac{1}{Y_i} \frac{dY_i(\alpha\ell, \alpha m)}{d\alpha}\big|_{\alpha=1} >$*

1, *then average wages are increasing in population, i.e.* $\frac{d \log w_i}{d \log \ell_i} > 0$. *In particular,*

$$\frac{d \log w_i}{d \log \ell_i} = \frac{\frac{dY_i(\alpha\ell, \alpha m)}{d\alpha}\big|_{\alpha=1} - 1}{1 - \eta - \left(\frac{dY_i(\alpha\ell, \alpha m)}{d\alpha}\big|_{\alpha=1} - 1\right)}.$$

Proof. Production is approximately given by $Y(\ell, m) = \ell^{1-\eta} m^{\eta} \Omega(m)$. Since wages are competitively set, they are equal to the marginal product of labor. Thus, $w\ell = (1 - \eta)Y(\ell, m)$. Then free entry implies

$$\psi = \frac{Y(\ell, m) - w\ell}{m} = \eta \frac{Y(\ell, m)}{m}.$$

Thus, we can take the total differential to get

$$\begin{aligned} d \log Y &= (1 - \eta) d \log \ell + \eta d \log m + \frac{\partial \log \Omega(m)}{\partial \log m} d \log m \\ d \log w &= d \log Y - d \log \ell \\ 0 &= d \log Y - d \log m. \end{aligned}$$

Solving, we get

$$d \log w = \frac{\frac{\partial \log \Omega(m)}{\partial \log m}}{1 - \eta - \frac{\partial \log \Omega(m)}{\partial \log m}} d \log \ell.$$

We complete the proof by noting that

$$\frac{dY(\alpha\ell, \alpha m)}{d\alpha}\bigg|_{\alpha=1} - 1 = \frac{\partial \log \Omega(m)}{\partial \log m}.$$

□

Then we can return to prove Proposition 1.

Proposition 1. *For small ex-post shocks, $Y_i(\ell, m)$ features increasing returns to scale if the (employment-weighted) average covariance between log employment and log productivity shocks is increasing in the number of firms, i.e.*

$$\frac{d}{dm}\left[\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\ell_i} \operatorname{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds\right] > 0.$$

In particular,

$$Y_i(\ell, m) \approx \ell^{1-\eta} m^{\eta} \Omega(m),$$

where

$$\Omega(m) \equiv \mathbb{E}[a_{isn}(\omega)] + \frac{1-\eta}{2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \operatorname{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds.$$

Proof. We know from Lemma 3 that

$$\begin{aligned}
\Omega(m_i) \equiv\ & \mathbb{E}[a_{isn}(\omega)^{\beta}] \\
& + (1-\eta) \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta} \log \overline{a}_{isn}(\omega) \beta^2\right] ds \\
& - \frac{1-\eta}{2} \eta \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta}\right)^2 \beta^2\right] ds \\
& - \frac{1-\eta}{2} \eta_N \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta}\right)^2 \beta^2\right] ds \\
& - \frac{1-\eta}{2} \eta_S \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{is}(\omega)}{\partial \beta}\right)^2 \beta^2\right] ds.
\end{aligned}$$

And recall from Lemma 2 that

$$\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta} = \frac{1}{\eta + \eta_N} \left[\log \overline{a}_{isn}(\omega) - \frac{\frac{\eta_S}{\eta + \eta_N}}{1 + \frac{\eta_S}{\eta + \eta_N}} \left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \overline{a}_{isn'}(\omega)\right)\right].$$

We then solve for $\mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta}\right)^2 \beta^2\right]$. Defining $X_N \equiv \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega,\beta)}{\partial \beta}\right)^2 \beta^2\right] ds$, we get

$$X_N = \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{isn}(\omega, \beta)}{\partial \beta} \right)^2 \beta^2 \right] ds$$

$$= \frac{\beta^2}{(\eta + \eta_N)^2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\left[\log \bar{a}_{isn}(\omega) - \frac{\frac{\eta_S}{\eta+\eta_N}}{1 + \frac{\eta_S}{\eta+\eta_N}} \left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \bar{a}_{isn'}(\omega) \right) \right]^2 \right] ds$$

$$= \frac{\beta^2}{(\eta + \eta_N)^2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\log \bar{a}_{isn}(\omega)^2 - 2 \frac{\eta_S}{\eta + \eta_N + \eta_S} \log \bar{a}_{isn}(\omega) \left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \bar{a}_{isn'}(\omega) \right) \right]$$

$$+ \frac{\beta^2}{(\eta + \eta_N)^2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\eta_S}{\eta + \eta_N + \eta_S} \right)^2 \left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \bar{a}_{isn'}(\omega) \right)^2 \right]$$

$$= \frac{\beta^2}{(\eta + \eta_N)^2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\log \bar{a}_{isn}(\omega)^2 \right] ds$$

$$- 2 \frac{\beta^2}{(\eta + \eta_N)^2} \frac{\eta_S}{\eta_N + \eta_N + \eta_S} \left(1 - \frac{1}{2} \frac{\eta_S}{\eta_N + \eta_N + \eta_S} \right) \int_{\mathcal{S}} \frac{\ell_{is}}{\ell_i} \mathbb{E}\left[\left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \bar{a}_{isn'}(\omega) \right)^2 \right] ds.$$

Moving onto the variance in sectoral labor, we have $X_S \equiv \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{is}(\omega)}{\partial \beta} \right)^2 \beta^2 \right] ds$. Then

$$X_S = \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\frac{\partial \log \ell_{is}(\omega)}{\partial \beta} \right)^2 \beta^2 \right] ds$$

$$= \frac{\beta^2}{(\eta + \eta_N + \eta_S)^2} \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\sum_{n'} \frac{\ell_{isn'}}{\ell_{is}} \log \bar{a}_{isn'}(\omega) \right)^2 \right] ds$$

Defining $X_2 \equiv (\eta + \eta_N)X_N + \eta_S X_S$, we have

$$
\begin{aligned}
X_2 = & \frac{\beta^2}{\eta+\eta_N} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\log \overline{a}_{isn}(\omega)^2\right] ds \\
& - 2\frac{\beta^2}{\eta+\eta_N}\frac{\eta_S}{\eta+\eta_N+\eta_S}\left(1 - \frac{1}{2}\frac{\eta_S}{\eta+\eta_N+\eta_S}\right) \int_{\mathcal{S}} \frac{\ell_{is}}{\ell_i} \mathbb{E}\left[\left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \overline{a}_{isn'}(\omega)\right)^2\right] ds \\
& + \frac{\beta^2 \eta_S}{(\eta+\eta_N+\eta_S)^2} \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\sum_{n'} \frac{\ell_{isn'}}{\ell_{is}} \log \overline{a}_{isn'}(\omega)\right)^2\right] ds \\
= & \frac{\beta^2}{\eta+\eta_N} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\log \overline{a}_{isn}(\omega)^2\right] ds \\
& - \frac{\beta^2}{\eta+\eta_N}\frac{\eta_S}{\eta+\eta_N+\eta_S} \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\sum_{n'} \frac{\ell_{isn'}}{\ell_{is}} \log \overline{a}_{isn'}(\omega)\right)^2\right] ds.
\end{aligned}
$$

Next, we solve for the weighted covariance. Define $X_1 \equiv \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}}{\ell_i} \mathbb{E}\left[\frac{\partial \log \ell_{isn}}{\partial \beta} \log \overline{a}_{isn}(\omega) \beta^2\right] ds$. Then we can write

$$
\begin{aligned}
X_1 = & \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}}{\ell_i} \mathbb{E}\left[\frac{\partial \log \ell_{isn}}{\partial \beta} \log \overline{a}_{isn}(\omega) \beta^2\right] ds \\
= & \frac{\beta^2}{\eta+\eta_N} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}}{\ell_i} \mathbb{E}\left[\left(\log \overline{a}_{isn}(\omega) - \frac{\frac{\eta_S}{\eta+\eta_N}}{1+\frac{\eta_S}{\eta+\eta_N}}\left(\sum_{n' \in \mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \log \overline{a}_{isn'}(\omega)\right)\right) \log \overline{a}_{isn}(\omega)\right] ds \\
= & \frac{\beta^2}{\eta+\eta_N} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\log \overline{a}_{isn}(\omega)^2\right] ds \\
& - \frac{\beta^2}{\eta+\eta_N}\frac{\eta_S}{\eta+\eta_N+\eta_S} \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\sum_{n'} \frac{\ell_{isn'}}{\ell_{is}} \log \overline{a}_{isn'}(\omega)\right)^2\right] ds.
\end{aligned}
$$

In particular, $X_1 = X_2$. Therefore,

$$
\begin{aligned}
\Omega(m) &= \mathbb{E}[a_{isn}(\omega)^\beta] + (1-\eta)X_1 - \frac{1-\eta}{2}X_2 \\
&= \mathbb{E}[a_{isn}(\omega)^\beta] + \frac{1-\eta}{2}X_1 \\
&= \mathbb{E}[a_{isn}(\omega)^\beta] + \frac{1-\eta}{2} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \operatorname{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds,
\end{aligned}
$$

completing the proof.

$\square$

Proposition 2. *Suppose ex-post shocks are small, and Assumption 1 holds. Then average wages are increasing in population, i.e.* $\frac{d \log w_i}{d \log \ell_i} > 0$. *And as the population goes to infinity, these agglomeration effects disappear, i.e.* $\lim_{\ell_i \to \infty} \frac{d \log w_i}{d \log \ell_i} = 0$.

Proof. In order to prove this result, we need to show that $\int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\ell_i} \operatorname{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds$ is increasing in the number of firms. Notice, from the proof of the previous proposition,

$$\begin{aligned} X = & \frac{\beta^2}{\eta + \eta_N} \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}(0)}{\ell_i} \mathbb{E}\left[\log \bar{a}_{isn}(\omega)^2\right] ds \\ & - \frac{\beta^2}{\eta + \eta_N} \frac{\eta_S}{\eta + \eta_N + \eta_S} \int_{\mathcal{S}} \frac{\ell_{is}(0)}{\ell_i} \mathbb{E}\left[\left(\sum_{n'} \frac{\ell_{isn'}}{\ell_{is}} \log \bar{a}_{isn'}(\omega)\right)^2\right] ds, \end{aligned}$$

where $X \equiv \int_{\mathcal{S}} \sum_{n \in \mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\ell_i} \operatorname{Cov}\left(\log a_{isn}(\omega), \log \ell_{isn}(\omega)\right) ds$. Then we compute the variances.

$$\begin{aligned} \mathbb{E}\left[\log \bar{a}_{isn}(\omega)^2 \beta^2\right] &= \mathbb{E}\left[\log \bar{a}_{isn}(\omega)\right]^2 \beta^2 + \operatorname{Var}(\beta \log \bar{a}_{isn}(\omega)) \\ &= \mathbb{E}\left[\beta \log \tilde{a}_{isn}(\omega) + \beta \log \tilde{A}_{is}(\omega)\right]^2 + \operatorname{Var}(\beta \log \tilde{a}_{isn}(\omega) + \beta \log \tilde{A}_{is}(\omega)) \\ &= \sigma_S^2 + \sigma_N^2, \end{aligned}$$

where the last line follows from the fact that the idiosyncratic shock and the sector shock are independent and their log means are zero. Similarly,

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{n' \in \mathcal{N}_{is}} \frac{\bar{L}_{isn'}}{\bar{L}_{is}} \log \bar{a}_{isn'}(\omega) \beta\right)^2\right] &= \mathbb{E}\left[\sum_{n' \in \mathcal{N}_{is}} \frac{\bar{L}_{isn'}}{\bar{L}_{is}} \beta \log \bar{a}_{isn'}(\omega)\right]^2 + \operatorname{Var}\left(\sum_{n' \in \mathcal{N}_{is}} \frac{\bar{L}_{isn'}}{\bar{L}_{is}} \beta \log \bar{a}_{isn'}(\omega)\right) \\ &= \operatorname{Var}\left(\beta \log \tilde{A}_{is}(\omega) + \sum_{n' \in \mathcal{N}_{is}} \frac{\bar{L}_{isn'}}{\bar{L}_{is}} \beta \log \tilde{a}_{isn'}(\omega)\right) \\ &= \operatorname{Var}\left(\beta \log \tilde{A}_{is}(\omega)\right) + \sum_{n' \in \mathcal{N}_{is}} \left(\frac{\bar{L}_{isn'}}{\bar{L}_{is}}\right)^2 \operatorname{Var}(\beta \log \tilde{a}_{isn'}(\omega)) \\ &= \sigma_S^2 + \left[\sum_{n' \in \mathcal{N}_{is}} \left(\frac{\bar{L}_{isn'}}{\bar{L}_{is}}\right)^2\right] \sigma_N^2. \end{aligned}$$

Therefore, we can write

$$X = \frac{1}{\eta+\eta_N}\left(\sigma_S^2+\sigma_N^2\right) - \frac{1}{\eta+\eta_N}\frac{\eta_S}{\eta+\eta_N+\eta_S}\left(\sigma_S^2 + \int_{\mathcal{S}}\frac{\ell_{is}}{\ell_i}HHI_{is}ds\sigma_N^2\right)$$
$$= \frac{1}{\eta+\eta_N+\eta_S}\sigma_S^2 + \frac{1}{\eta+\eta_N}\left[1-\frac{\eta_S}{\eta+\eta_N+\eta_S}\int_{\mathcal{S}}\frac{\ell_{is}}{\ell_i}HHI_{is}ds\right]\sigma_N^2,$$

where $HHI_{is} \equiv \left[\sum_{n'\in\mathcal{N}_{is}}\left(\frac{\overline{L}_{isn'}}{\overline{L}_{is}}\right)^2\right]$.

Then, to complete the prooof, we need to show that $\int_{\mathcal{S}}\frac{\ell_{is}}{\ell_i}HHI_{is}ds$ is decreasing in m. Then we also need to show that its derivative converages to a constant. Recall that,

$$\overline{L}_{isn} = \frac{1}{\ell_i}z_{isn}^{\frac{1}{\eta}}\left(\frac{1-\eta}{w}\right)^{\frac{1}{\eta}},$$

where

$$w = \frac{1-\eta}{\ell_i^\eta}m^\eta\left(\mu_z\right)^\eta$$

and $\mu_z = \mathbb{E}[z_{isn}^{\frac{1}{\eta}}]$. Therefore, $\overline{L}_{isn} = z_{isn}^{\frac{1}{\eta}}/(m\mu_z)$. And we can write

$$\int_{\mathcal{S}} L_{is}HHI_{is}ds = \int_{\mathcal{S}}\overline{L}_{is}\sum_{n\in\mathcal{N}_{is}}\frac{(\overline{L}_{isn})^2}{(\overline{L}_{is})^2}ds$$
$$= \int_{\mathcal{S}}\left(\sum_{n'\in\mathcal{N}_{is}} z_{isn'}^{1/\eta}/(m\mu_z)\right)\sum_{n\in\mathcal{N}_{is}}\frac{(z_{isn}^{1/\eta}/(m\mu_z))^2}{\left(\sum_{n'\in\mathcal{N}_{is}} z_{isn'}^{1/\eta}/(m\mu_z)\right)^2}ds$$
$$= \frac{1}{m\mu_z}\int_{\mathcal{S}} N_{is}\frac{\sum_{n'\in\mathcal{N}_{is}} z_{isn'}^{1/\eta}}{N_{is}}\frac{\sum_{n\in\mathcal{N}_{is}}(z_{isn}^{1/\eta})^2}{\left(\sum_{n'\in\mathcal{N}_{is}} z_{isn'}^{1/\eta}\right)^2}ds$$
$$= \frac{1}{\mu_z}\left[\sum_{N=0}^{\infty}\frac{N}{m}\mathbb{E}\left[\frac{\sum_{n'\in\mathcal{N}} z_{isn'}^{1/\eta}}{N}\frac{\sum_{n\in\mathcal{N}}(z_{isn}^{1/\eta})^2}{\left(\sum_{n'\in\mathcal{N}} z_{isn'}^{1/\eta}\right)^2}\right]p(N,m)\right].$$

I will denote by $\psi_N \equiv \mathbb{E}\left[\frac{\sum_{n'\in\mathcal{N}} z_{isn'}^{1/\eta}}{N}\frac{\sum_{n\in\mathcal{N}}(z_{isn}^{1/\eta})^2}{\left(\sum_{n'\in\mathcal{N}} z_{isn'}^{1/\eta}\right)^2}\right]$. By assumption, ψ_N is decreasing in N. I will also define $F(N,m) \equiv \sum_{n=N}^{\infty}\frac{n}{m}p(n,m)$. Notice that because $m = \sum_{N=0}^{\infty}Np(N,m)$, this is 1 minus a CDF. Furthermore, $F(1,m) = 1$ for all m.

Then consider m' and m'' with $m' > m''$. By our assumption on FOSD, $F(N,m') \geq F(N,m'')$ for all N. Notice that this must be strict for some N. If not, then $\frac{N}{m'}p(N,m') = \frac{N}{m''}p(N,m'')$, i.e. $p(N,m') = \frac{m'}{m''}p(N,m'')$. This is contradiction for $m' > m''$ because it would be impossible for both $p(N, m')$ and $p(N, m'')$ to sum to 1 across all N.

Notice that we can write

$$\begin{aligned}\int_{\mathcal{S}} L_{is} HHI_{is} ds &= \frac{1}{\mu_z} \sum_{N=0}^{\infty} \frac{N}{m} p(N, m) \psi_N \\ &= \frac{1}{\mu_z} \left(\psi_1 F(1, m) + (\psi_2 - \psi_1) F(2, m) + (\psi_3 - \psi_2) F(3, m) \cdots \right).\end{aligned}$$

Therefore,

$$\begin{aligned}X &= \frac{1}{\mu_z} \sum_{N=0}^{\infty} \frac{N}{m'} p(N, m') \psi_N - \frac{1}{\mu_z} \sum_{N=0}^{\infty} \frac{N}{m''} p(N, m'') \psi_N \\ &= \frac{1}{\mu_z} \left(\psi_1 (F(1, m') - F(1, m'')) + (\psi_2 - \psi_1)(F(2, m') - F(2, m'')) + \cdots \right)\end{aligned}$$

Then $F(1, m') = F(1, m'')$, $\psi_n - \psi_{n-1} < 0$, and $F(n, m') - F(n, m'') \geq 0$ for $n \geq 2$, with one strict. Thus,

$$\frac{1}{\mu_z} \sum_{N=0}^{\infty} \frac{N}{m'} p(N, m') \psi_N - \frac{1}{\mu_z} \sum_{N=0}^{\infty} \frac{N}{m''} p(N, m'') \psi_N < 0.$$

It follows that average HHI is decreasing in the number of firms.

To complete the proof, we need to show that as the population goes to infinity, these agglomerations effects disappear. In particular, we need to show that

$$\frac{-\frac{d}{dm} \left[\int_{\mathcal{S}} \frac{\ell_{is}}{\ell_i} HHI_{is} ds \right] \sigma_N^2}{\mathbb{E}[a_{isn}(\omega)] + \frac{1-\eta}{2} \left[\frac{1}{\eta + \eta_N + \eta_S} \sigma_{\mathcal{S}}^2 + \frac{1}{\eta + \eta_N} \left(1 - \frac{\eta_S}{\eta + \eta_N + \eta_S} \int_{\mathcal{S}} \frac{\ell_{is}}{\ell_i} HHI_{is} ds \right) \sigma_N^2 \right]} \to 0.$$

We start by showing that $\psi_N \to 0$ as $N \to \infty$. Then we will show that that implies $\int_{\mathcal{S}} \overline{L}_{is} HHI_{is} ds \to 0$. Note that because $1 - F_{iz}$ is regularly varying, there exists α and slowly varying function L[20] such that $1 - F_{iz}(x) = x^{-\alpha} L(x)$. Starting with ψ_N, note that we can rewrite it,

$$\psi_N = \mathbb{E} \left[\frac{a_N}{N^2} \frac{\sum_{n \in \mathcal{N}} (z_{isn}^{1/\eta})^2}{a_N} \frac{1}{N^{-1} \left(\sum_n z_{isn}^{1/\eta} \right)} \right],$$

where a_N is defined so that $\mathbb{P}(z_{isn}^{2/\eta} > a_N) = N^{-1}$. Then by the strong law of large number, $N^{-1} \left(\sum_n z_{isn}^{1/\eta} \right) \to \mathbb{E}[z_{isn}^{1/\eta}]$ a.s.. If $\mathbb{E}[(z_{isn})^{2/\eta}]$ exists, then we can cancel a_N and

[20] A function is slowly varying if for every $a > 0$, $\frac{L(ax)}{L(x)} \to 1$ as $x \to \infty$.

use the strong law of large numbers to note that $N^{-1}\left(\sum_n z_{isn}^{2/\eta}\right) \to \mathbb{E}[z_{isn}^{2/\eta}]$. Then that is multiplied by N^{-1} and the part in expectations converges to 0 almost surely.

If the variance does not exist, we need to use Levy's Theorem.[21] By Lévy's theorem, $\frac{\sum_{n\in\mathcal{N}}(z_{isn}^{1/\eta})^2}{a_N}$ converges in distribution to a non-degenerate distribution. That simply leaves a_N/N^2. Note that

$$\begin{aligned}\frac{a_N}{N^2} &= a_N \mathbb{P}(z_{isn}^{2/\eta} > a_N)^2 \\ &= a_N a_N^{-\alpha} L(a_N^{1/2})^2.\end{aligned}$$

This converges to 0 as $a_N \to \infty$ if $\alpha > 1$. But note that since the mean exists, α must be greater than 1. Further note that $a_N \to \infty$ as $N \to \infty$ otherwise the variance would exist. Thus, the part in the expectations must converge to 0 in distribution.

Note that $p(N,m) \leq 1$ for all N and m. Therefore, $\frac{N}{m}p(N,m) \to 0$ as $m \to 0$. Then take $\varepsilon > 0$. Since $\psi_N \to 0$, there exists a $\overline{N}$ such that $\psi_N < \mu_z \frac{\varepsilon}{2}$ for all $N > \overline{N}$. Next note that for every $N' < \overline{N}$, there exists a $m_{N'}$ such that $\frac{N'}{m}p(N',m) \leq \mu_z \frac{\varepsilon}{2\overline{N}\psi_{N'}}$ for all $m > m_{N'}$. Take $\overline{m} = \max_{N' \leq \overline{N}} m_{N'}$. Then for $m > \overline{m}$,

$$\begin{aligned}\int_{\mathcal{S}} L_{is} HHI_{is} ds &= \frac{1}{\mu_z}\sum_{N=0}^{\infty}\frac{N}{m}p(N,m)\psi_N \\ &\leq \frac{1}{\mu_z}\sum_{N=0}^{\overline{N}}\frac{N}{m}p(N,m)\psi_N + \frac{1}{\mu_z}\psi_{\overline{N}}\sum_{\overline{N}+1}^{\infty}p(N,m) \\ &\leq \frac{1}{\mu_z}\sum_{N=0}^{\overline{N}}\frac{N}{m}p(N,m)\psi_N + \frac{1}{\mu_z}\psi_{\overline{N}} \\ &\leq \frac{1}{\mu_z}\sum_{N=0}^{\overline{N}}\frac{N}{m}p(N,m)\psi_N + \frac{\varepsilon}{2} \\ &\leq \sum_{N=0}^{\overline{N}}\frac{\varepsilon}{2\overline{N}\psi_N}\psi_N + \frac{\varepsilon}{2} \\ &= \varepsilon,\end{aligned}$$

completing the proof.

□

[21] See Durrett (2019).

B Oligopsony

B.1 Environment

Under oligopsony, we assume that firms no longer take as given the wages in every state of the world. Instead, firms maxmize profits taking as given the labor demanded by the other firms in its sector and the worker's opportunities in other sectors. For simplicity, we use the constant elasticity version of the model, i.e. $g_S(x) = x^{1+\eta_S}$ and $g_N(x) = x^{1+\eta_N}$. The problem of firm n is then

$$\max_{L_{isn}(\omega),\overline{L}_{isn},L_{is}(\omega),\overline{L}_{is},w_{isn}(\omega)} \mathbb{E}\left[a_{isn}(\omega) f\left(L_{isn}(\omega)\ell_i\right) - w_{isn}(\omega)L_{isn}(\omega)\ell_i\right]$$

where

$$w_{isn}(\omega) = \lambda_i(\omega)(1+\eta_S)(1+\eta_N)\left(\frac{L_{is}(\omega)}{\overline{L}_{is}}\right)^{\eta_S}\left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}}\right)^{\eta_N},$$

$$-\overline{\lambda}_{is} = \mathbb{E}\left[\lambda_i(\omega)(1+\eta_S)\left(\frac{L_{is}(\omega)}{\overline{L}_{is}}\right)^{\eta_S}\left(\left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}}\right)^{1+\eta_N} - \frac{L_{isn}(\omega)}{\overline{L}_{isn}}(1+\eta_N)\left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}}\right)^{\eta_N}\right)\right]$$

$$\overline{\lambda}_{is} = \mathbb{E}\left[\lambda_i(\omega)\left(\left(\frac{L_{is}(\omega)}{\overline{L}_{is}}\right)^{1+\eta_S} - \frac{L_{is}(\omega)}{\overline{L}_{is}}(1+\eta_S)\left(\frac{L_{is}(\omega)}{\overline{L}_{is}}\right)^{\eta_S}\right)\right] + \overline{\lambda}_i,$$

and

$$\begin{aligned}
\overline{L}_{is} &= \sum_{n'\in\mathcal{N}_{is}} \overline{L}_{isn'} \\
1 &= \int_{\mathcal{S}} \overline{L}_{is} ds \\
L_{is}(\omega) &= \sum_{n'\in\mathcal{N}_{is}} \overline{L}_{isn'} \cdot \left(\frac{L_{isn'}(\omega)}{\overline{L}_{isn'}}\right)^{1+\eta_N} \\
1 &= \int_{s} \overline{L}_{is} \cdot \left(\frac{L_{is}(\omega)}{\overline{L}_{is}}\right)^{1+\eta_S}.
\end{aligned}$$

B.2 Linear-Quadratic Approximation

We do a quadratic approximation to the profit function and linear approximation to the constraints around the point with no ex-post shocks just as we did in the competitive environment.

B.2.1 Profit Function

Define

$$\Pi_{isn}(\omega) = a_{isn}(\omega) f\left(L_{isn}(\omega)\ell_i\right) - w_{isn}(\omega) L_{isn}(\omega)\ell_i.$$

Then, starting with the first order, we have

$$\begin{aligned}
\frac{\partial \Pi_{isn}}{\partial \log a_{isn}(\omega)} &= a_{isn}(\omega) f\left(L_{isn}(\omega)\ell_i\right) \\
\frac{\partial \Pi_{isn}}{\partial \log L_{isn}(\omega)} &= a_{isn}(\omega) f'\left(L_{isn}(\omega)\ell_i\right) L_{isn}(\omega)\ell_i - w_{isn}(\omega) L_{isn}(\omega)\ell_i \\
\frac{\partial \Pi_{isn}}{\partial \log w_{isn}(\omega)} &= -w_{isn}(\omega) L_{isn}(\omega)\ell_i.
\end{aligned}$$

Then the second order derivatives are

$$\begin{aligned}
\frac{\partial^2 \Pi_{isn}(\omega)}{\partial \log a_{isn}(\omega)^2} &= a_{isn}(\omega) f\left(L_{isn}(\omega)\ell_i\right) \\
\frac{\partial^2 \Pi_{isn}(\omega)}{\partial \log a_{isn}(\omega) \partial \log L_{isn}(\omega)} &= a_{isn}(\omega) f'\left(L_{isn}(\omega)\ell_i\right) L_{isn}(\omega)\ell_i \\
\frac{\partial^2 \Pi_{isn}(\omega)}{\partial \log L_{isn}(\omega)^2} &= a_{isn}(\omega) f'\left(L_{isn}(\omega)\ell_i\right) L_{isn}(\omega)\ell_i + a_{isn}(\omega) f''\left(L_{isn}(\omega)\ell_i\right) \left(L_{isn}(\omega)\ell_i\right)^2 \\
&\quad - w_{isn}(\omega) L_{isn}(\omega)\ell_i \\
\frac{\partial^2 \Pi_{isn}(\omega)}{\partial \log L_{isn}(\omega) \partial \log w_{isn}(\omega)} &= -w_{isn}(\omega) L_{isn}(\omega)\ell_i \\
\frac{\partial^2 \Pi_{isn}(\omega)}{\partial \log w_{isn}(\omega)^2} &= -w_{isn}(\omega) L_{isn}(\omega)\ell_i.
\end{aligned}$$

Then the second order approximation around $\beta = 0$ is then

$$\Pi_{isn} \propto \mathbb{E}\left[-\hat{w}_{isn}(\omega) + \hat{a}_{isn}(\omega)\hat{L}_{isn}(\omega) - \frac{1}{2}\eta \hat{L}_{isn}(\omega)^2 - \hat{L}_{isn}(\omega)\hat{w}_{isn}(\omega) - \frac{1}{2}\hat{w}_{isn}(\omega)^2\right]$$

where $\hat{x}$ denotes log deviations from $\beta = 0$ value. In order to get a quadratic-linear approximation, we need to transform this function to be completely second order. Thus, we need a new form for $\hat{w}_{isn}(\omega)$. Note that the first constraint is already log-linear, so that

$$\hat{w}_{isn}(\omega) = \hat{\lambda}_i(\omega) + \eta_S\left(\hat{L}_{is}(\omega) - \hat{\bar{L}}_{is}\right) + \eta_N\left(\hat{L}_{isn}(\omega) - \hat{\bar{L}}_{isn}\right).$$

Meanwhile, the other two equations determining labor supply give

$$\overline{\lambda}_{is} = \eta_N(1+\eta_S)\mathbb{E}\left[\lambda_i(\omega)\left(\frac{L_{is}(\omega)}{\overline{L}_{is}}\right)^{\eta_S}\left(\frac{L_{isn}(\omega)}{\overline{L}_{isn}}\right)^{1+\eta_N}\right]$$

and

$$\overline{\lambda}_i = \eta_S\mathbb{E}\left[\lambda_i(\omega)\left(\frac{L_{is}(\omega)}{\overline{L}_{is}}\right)^{1+\eta_S}\right] + \overline{\lambda}_{is}.$$

Substituting out $\overline{\lambda}_{is}$ we can write

$$\begin{aligned}
&\frac{\overline{\lambda}_i}{\lambda_i(\omega)}\left(\hat{\overline{\lambda}}_i - \hat{\lambda}_i(\omega)\right) + \frac{\overline{\lambda}_i}{\lambda_i(\omega)}\frac{1}{2}\left(\hat{\overline{\lambda}}_i - \hat{\lambda}_i(\omega)\right)^2 \\
&\qquad = \mathbb{E}\left[\eta_S(1+\eta_S)\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right) + \frac{\eta_S(1+\eta_S)^2}{2}\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right)^2\right] \\
&\qquad\quad + \mathbb{E}\left[\eta_N(1+\eta_S)\left(\eta_S\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right) + (1+\eta_N)\left(\hat{L}_{isn}(\omega) - \hat{\overline{L}}_{isn}\right)\right)\right] \\
&\qquad\quad + \frac{1}{2}\mathbb{E}\left[\eta_N(1+\eta_S)\left(\eta_S\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right) + (1+\eta_N)\left(\hat{L}_{isn}(\omega) - \hat{\overline{L}}_{isn}\right)\right)^2\right] \\
&\qquad = \mathbb{E}\left[\eta_S(1+\eta_N)(1+\eta_S)\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right) + \eta_N(1+\eta_S)(1+\eta_N)\left(\hat{L}_{isn}(\omega) - \hat{\overline{L}}_{isn}\right)\right] \\
&\qquad\quad + \mathbb{E}\left[\frac{\eta_S(1+\eta_S)^2}{2}\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right)^2\right] \\
&\qquad\quad + \frac{1}{2}\mathbb{E}\left[\eta_N(1+\eta_S)\left(\eta_S\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right) + (1+\eta_N)\left(\hat{L}_{isn}(\omega) - \hat{\overline{L}}_{isn}\right)\right)^2\right]
\end{aligned}$$

Then dropping the $\lambda_i(\omega)$ and $\overline{\lambda}_i$ as the firm has no effect on it, profits are given by

$$\begin{aligned}
\Pi_{isn} = &\frac{1}{(1+\eta_N)(1+\eta_S)}\mathbb{E}\left[\frac{\eta_S(1+\eta_S)^2}{2}\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right)^2\right] \\
&+ \frac{1}{(1+\eta_N)(1+\eta_S)}\mathbb{E}\left[\frac{\eta_N(1+\eta_S)}{2}\left(\eta_S\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right) + (1+\eta_N)\left(\hat{L}_{isn}(\omega) - \hat{\overline{L}}_{isn}\right)\right)^2\right] \\
&+ \mathbb{E}\left[\hat{a}_{isn}(\omega)\hat{L}_{isn}(\omega) - \frac{1}{2}\eta\hat{L}_{isn}(\omega)^2 - \hat{L}_{isn}(\omega)\hat{w}_{isn}(\omega) - \frac{1}{2}\hat{w}_{isn}(\omega)^2\right]
\end{aligned}$$

Then we can do a first order approximation to the constraints

B.2.2 Constraints

We have

$$\hat{w}_{isn}(\omega) = \hat{\lambda}_i(\omega) + \eta_S\left(\hat{L}_{is}(\omega) - \hat{\overline{L}}_{is}\right) + \eta_N\left(\hat{L}_{isn}(\omega) - \hat{\overline{L}}_{isn}\right)$$

$$\frac{\overline{\lambda}_i}{\lambda_i(\omega)}\left(\hat{\hat{\lambda}}_i - \hat{\lambda}_i(\omega)\right) = (1+\eta_N)(1+\eta_S)\mathbb{E}\left[\eta_S(\hat{L}_{is}(\omega) - \hat{\hat{L}}_{is}) + \eta_N(\hat{L}_{isn}(\omega) - \hat{\hat{L}}_{isn})\right]$$

along with

$$\begin{aligned}
\hat{\hat{L}}_{is} &= \sum_{n'\in\mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}}\hat{\hat{L}}_{isn'} \\
0 &= \int_{\mathcal{S}} \hat{\hat{L}}_{is} ds \\
\hat{L}_{is}(\omega) &= \sum_{n'\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\hat{L}_{isn'}(\omega) \\
0 &= \int_{\mathcal{S}} \hat{L}_{is}(\omega) ds.
\end{aligned}$$

B.2.3 Combining Things

Then to first order, $\hat{\hat{\lambda}}_i = \hat{\lambda}_i(\omega) = \hat{\hat{L}}_{is} = \hat{\hat{L}}_{isn} = 0$. We can then make those substitutions and also substitute in for $\hat{w}_{isn}(\omega)$. We can then get

$$\begin{aligned}
\Pi_{isn} = &\frac{1}{2}\eta_S\frac{1+\eta_S}{1+\eta_N}\mathbb{E}\left[\hat{L}_{is}(\omega)^2\right] + \frac{1}{2}\frac{\eta_N}{1+\eta_N}\mathbb{E}\left[(\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega))^2\right] \\
&+ \mathbb{E}[\hat{a}_{isn}(\omega)\hat{L}_{isn}(\omega)] - \frac{1}{2}\eta\mathbb{E}[\hat{L}_{isn}(\omega)^2] - \mathbb{E}[\hat{L}_{isn}(\omega)\hat{w}_{isn}(\omega)] - \frac{1}{2}\mathbb{E}[\hat{w}_{isn}(\omega)^2] \\
= &\mathbb{E}[\hat{a}_{isn}(\omega)\hat{L}_{isn}(\omega)] + \frac{1}{2}\eta_S\frac{1+\eta_S}{1+\eta_N}\mathbb{E}[\hat{L}_{is}(\omega)^2] - \eta\frac{1}{2}\mathbb{E}\left[\hat{L}_{isn}(\omega)^2\right] \\
&+ \frac{1}{2}\frac{\eta_N}{1+\eta_N}\mathbb{E}\left[(\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega))^2\right] - \frac{1}{2}\mathbb{E}[(\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega))^2] + \frac{1}{2}\mathbb{E}[\hat{L}_{isn}(\omega)] \\
= &\mathbb{E}[\hat{a}_{isn}(\omega)\hat{L}_{isn}(\omega)] + \frac{1}{2}\eta_S\frac{1+\eta_S}{1+\eta_N}\mathbb{E}[\hat{L}_{is}(\omega)^2] + (1-\eta)\frac{1}{2}\mathbb{E}\left[\hat{L}_{isn}(\omega)^2\right] \\
&- \frac{1}{2}\frac{1}{1+\eta_N}\mathbb{E}\left[(\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega))^2\right]
\end{aligned}$$

subject to the constraint

$$\hat{L}_{is}(\omega) = \sum_{n'\in\mathcal{N}_{is}} \frac{\overline{L}_{isn'}}{\overline{L}_{is}}\hat{L}_{isn'}(\omega),$$

and

$$\hat{w}_{isn}(\omega) = \eta_S\hat{L}_{is}(\omega) + \eta_N\hat{L}_{isn}.$$

Taking the first order conditions, gives

$$0 = \hat{a}_{isn}(\omega) + (1-\eta)\,\hat{L}_{isn}(\omega) - \frac{1}{1+\eta_N}(\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega)) + \lambda_L(\omega)\frac{\overline{L}_{isn}}{\overline{L}_{is}} + \lambda_w(\omega)\eta_N$$

$$0 = \eta_S \frac{1+\eta_S}{1+\eta_N} \hat{L}_{is}(\omega) - \lambda_L(\omega) + \lambda_w(\omega)\eta_S$$

$$0 = -\frac{1}{1+\eta_N} (\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega)) - \lambda^w(\omega)$$

Substituting in

$$\begin{aligned}
0 &= \hat{a}_{isn}(\omega) + (1-\eta)\,\hat{L}_{isn}(\omega) - \frac{1}{1+\eta_N}(\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega)) \\
&\quad + \eta_S \frac{1+\eta_S}{1+\eta_N} \hat{L}_{is}(\omega) \frac{\overline{L}_{isn}}{\overline{L}_{is}} + \lambda_w(\omega)\eta_S \frac{\overline{L}_{isn}}{\overline{L}_{is}} + \lambda_w(\omega)\eta_N \\
&= \hat{a}_{isn}(\omega) + (1-\eta)\,\hat{L}_{isn}(\omega) - \frac{1}{1+\eta_N}(\hat{w}_{isn}(\omega) + \hat{L}_{isn}(\omega)) \left(1 + \eta_N + \eta_S \frac{\overline{L}_{isn}}{\overline{L}_{is}}\right) \\
&\quad + \eta_S \frac{1+\eta_S}{1+\eta_N} \hat{L}_{is}(\omega) \frac{\overline{L}_{isn}}{\overline{L}_{is}}
\end{aligned}$$

Then we will substitute in for $\hat{w}_{isn}(\omega) = \eta_N \hat{L}_{isn} + \eta_S \hat{L}_{is}$. This gives

$$\begin{aligned}
0 &= \hat{a}_{isn}(\omega) + (1-\eta)\,\hat{L}_{isn}(\omega) - \frac{1}{1+\eta_N}(\eta_S \hat{L}_{is}(\omega) + (1+\eta_N)\hat{L}_{isn}(\omega)) \left(1 + \eta_N + \eta_S \frac{\overline{L}_{isn}}{\overline{L}_{is}}\right) \\
&\quad + \eta_S \frac{1+\eta_S}{1+\eta_N} \hat{L}_{is}(\omega) \frac{\overline{L}_{isn}}{\overline{L}_{is}} \\
&= \hat{a}_{isn}(\omega) - \left(\eta + \eta_N + \eta_S \frac{\overline{L}_{isn}}{\overline{L}_{is}}\right) \hat{L}_{isn}(\omega) \\
&\quad + \left(-\eta_S - \frac{\eta_S^2}{1+\eta_N} \frac{\overline{L}_{isn}}{\overline{L}_{is}} + \eta_S \frac{1+\eta_S}{1+\eta_N} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\right) \hat{L}_{is}(\omega) \\
0 &= \hat{a}_{isn}(\omega) - \left(\eta + \eta_N + \eta_S \frac{\overline{L}_{isn}}{\overline{L}_{is}}\right) \hat{L}_{isn}(\omega) - \eta_S \left(1 - \frac{1}{1+\eta_N} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\right) \hat{L}_{is}(\omega)
\end{aligned}$$

Then we define

$$\Psi_{isn} \equiv \frac{1}{\eta + \eta_N + \eta_S \frac{\overline{L}_{isn}}{\overline{L}_{is}}}$$

$$\Omega_{isn} \equiv \left(1 - \frac{1}{1+\eta_N} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\right) \Psi_{isn}.$$

Then some tedious algebra implies

$$\hat{L}_{isn}(\omega) = \Psi_{isn}\hat{a}_{isn}(\omega) - \eta_S \Omega_{isn} \left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1 + \eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}} \hat{a}_{isn'}(\omega) \right).$$

Then summing up

$$\hat{L}_{is}(\omega) = \left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1 + \eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}} \hat{a}_{isn'}(\omega) \right).$$

B.3 Adapted Propositions

Next, we need to compute $\Omega(m)$. While the final covariance form no longer holds, the formula

$$\begin{aligned} \Omega(m) = \mathbb{E}[\overline{a}_{isn}(\omega)] &+ (1-\eta) \int_{\mathcal{S}} \overline{L}_{is} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \operatorname{Cov}\left(\hat{L}_{isn}(\omega), \hat{a}_{isn}(\omega)\right) ds \\ &- \frac{1-\eta}{2}(\eta + \eta_N) \int_{\mathcal{S}} \overline{L}_{is} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \operatorname{Var}(\hat{L}_{isn}(\omega)) ds \\ &- \frac{1-\eta}{2} \eta_S \int_{\mathcal{S}} \overline{L}_{is} \operatorname{Var}(\hat{L}_{is}(\omega)) ds \end{aligned} \tag{24}$$

still holds. This is our adjusted Proposition 1.

Proposition 1′. *For small ex-post shocks and oligopsonistic firms, $Y_i(\ell, m)$ features increasing retuns to scale if $\Omega(m)$ is increasing in m, where $\Omega(m)$ is defined as in equation (24). In particular,*

$$Y_i(\ell, m) \approx \ell^{1-\eta} m^{\eta} \Omega(m).$$

Next, we need to go through and calculate each of the statistics in order to recover a version of proposition 2. We have

$$\begin{aligned}
X_1 &\equiv \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \operatorname{Cov}(\hat{L}_{isn}(\omega), \hat{a}_{isn}(\omega)) \\
&= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn} \operatorname{Var}(\hat{a}_{isn}(\omega)) \\
&\quad - \eta_S \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Omega_{isn} \left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}} \operatorname{Cov}(\hat{a}_{isn}(\omega), \hat{a}_{isn'}(\omega)) \right)
\end{aligned}$$

Note that $Var(\hat{a}_{isn}(\omega)) = \sigma_S^2 + \sigma_N^2$. For $n \neq n'$

$$\begin{aligned}
Cov(\hat{a}_{isn}(\omega), \hat{a}_{isn'}(\omega)) &= Cov\left(\log \tilde{A}_{is}(\omega) + \log \tilde{a}_{isn}(\omega), \log \tilde{A}_{is}(\omega) + \log \tilde{a}_{isn'}(\omega)\right) \\
&= \sigma_S^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
X_1 = (\sigma_N^2 + \sigma_S^2) &\left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn} \right) \\
&- \eta_S \sigma_S^2 \left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Omega_{isn} \right) \left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}} \right) \\
&- \eta_S \sigma_N^2 \left(\sum_{n'} \left(\frac{\overline{L}_{isn'}}{\overline{L}_{is}} \right)^2 \frac{\Omega_{isn'} \Psi_{isn'}}{1+\eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}} \right).
\end{aligned}$$

Next we do the variances.

$$
\begin{aligned}
X_2 &= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \operatorname{Var}(\hat{L}_{isn}(\omega)) \\
&= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \operatorname{Cov}(\Psi_{isn}\hat{a}_{isn}(\omega) - \eta_S\Omega_{isn}\hat{L}_{is}(\omega), \Psi_{isn}\hat{a}_{isn}(\omega) - \eta_S\Omega_{isn}\hat{L}_{is}(\omega)0 \\
&= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn}^2 \operatorname{Cov}(\hat{a}_{isn}(\omega), \hat{a}_{isn}(\omega)) \\
&\quad - 2\eta_S \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn}\Omega_{isn} \operatorname{Cov}(\hat{a}_{isn}(\omega), \hat{L}_{is}(\omega)) \\
&\quad + \eta_S^2 \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Omega_{isn}^2 \operatorname{Cov}(\hat{L}_{is}(\omega), \hat{L}_{is}(\omega)) \\
&= \left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn}^2\right)(\sigma_S^2 + \sigma_N^2) \\
&\quad - 2\eta_S \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn}\Omega_{isn} \operatorname{Cov}\left(\hat{a}_{isn}(\omega), \sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}} \hat{a}_{isn'}(\omega)\right) \\
&\quad + \eta_S^2 \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Omega_{isn}^2 \operatorname{Var}(\hat{L}_{is}(\omega)) \\
&= \left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn}^2\right)(\sigma_S^2 + \sigma_N^2) \\
&\quad - 2\eta_S\sigma_S^2 \left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Psi_{isn}\Omega_{isn}\right)\left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right) \\
&\quad - 2\eta_S\sigma_N^2 \left(\sum_{n'} \left(\frac{\overline{L}_{isn'}}{\overline{L}_{is}}\right)^2 \frac{\Omega_{isn'}\Psi_{isn'}^2}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right) \\
&\quad + \eta_S^2 \sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}} \Omega_{isn}^2 \operatorname{Var}(\hat{L}_{is}(\omega))
\end{aligned}
$$

We next need to find $\mathrm{Var}(\hat{L}_{is}(\omega))$. We have

$$\begin{aligned}
X_3 &= \mathrm{Var}(\hat{L}_{is}(\omega)) \\
&= \mathrm{Cov}\left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}} \hat{a}_{isn'}(\omega), \sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}} \hat{a}_{isn'}(\omega)\right) \\
&= \sigma_S^2 \left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}}\right)^2 \\
&\quad + \sigma_N^2 \sum_{n' \in \mathcal{N}_{is}} \left(\frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S \sum_{n''} \frac{\overline{L}_{isn''}}{\overline{L}_{is}} \Omega_{isn''}}\right)^2 .
\end{aligned}$$

Putting everything together

$$
\begin{aligned}
\Omega(m) = \mathbb{E}[\bar{a}_{isn}(\omega)] &+ (1-\eta)(\sigma_N^2+\sigma_S^2)\int_{\mathcal{S}} \overline{L}_{is}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\Psi_{isn}\right) ds \\
&- (1-\eta)\eta_S\sigma_S^2 \int_{\mathcal{S}} \overline{L}_{is}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\Omega_{isn}\right)\left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right) ds \\
&- (1-\eta)\eta_S\sigma_N^2 \int_{\mathcal{S}} \overline{L}_{is}\left(\sum_{n'} \left(\frac{\overline{L}_{isn'}}{\overline{L}_{is}}\right)^2 \frac{\Omega_{isn'}\Psi_{isn'}}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right) ds \\
&- \frac{1-\eta}{2}(\eta+\eta_N)(\sigma_S^2+\sigma_N^2)\int_{\mathcal{S}} \overline{L}_{is}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\Psi_{isn}^2\right) ds \\
&+ (1-\eta)(\eta+\eta_N)\eta_S\sigma_S^2 \int_{\mathcal{S}} \overline{L}_{is}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{L}_{isn}}{\overline{L}_{is}}\Psi_{isn}\Omega_{isn}\right)\left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right) ds \\
&+ (1-\eta)(\eta+\eta_N)\eta_S\sigma_N^2 \int_{\mathcal{S}} \overline{L}_{is}\left(\sum_{n'} \left(\frac{\overline{L}_{isn'}}{\overline{L}_{is}}\right)^2 \frac{\Omega_{isn'}\Psi_{isn'}^2}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right) ds \\
&- \frac{1-\eta}{2}\eta_S\left((\eta+\eta_N)\eta_S-1\right)\sigma_S^2 \int_{\mathcal{S}} \overline{L}_{is}\left(\sum_{n'} \frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right)^2 ds \\
&- \frac{1-\eta}{2}\eta_S\left((\eta+\eta_N)\eta_S-1\right)\sigma_N^2 \int_{\mathcal{S}} \overline{L}_{is}\sum_{n'\in\mathcal{N}_{is}}\left(\frac{\overline{L}_{isn'}}{\overline{L}_{is}} \frac{\Psi_{isn'}}{1+\eta_S\sum_{n''}\frac{\overline{L}_{isn''}}{\overline{L}_{is}}\Omega_{isn''}}\right)^2 ds.
\end{aligned}
\tag{25}
$$

Then instead of the assumption that average HHI is decreasing in the number of firms, we need to assume that the market power adjusted HHI is decreasing in the number of firms.

Proposition 2′. *Suppose equation (25) is decreasing in m. Then average wages are increasing in population, i.e.* $\frac{d\log w_i}{d\log \ell_i} > 0$. *And as the population goes to infinity, these agglomeration effects disappear, i.e.* $\lim_{\ell_i\to\infty}\frac{d\log w_i}{d\log \ell_i} = 0$.

C Calibration

We calibrate the model using observed variance of log earnings. We calculate those values in the model here. Throughout this section we will use $\hat{x}$ for log deviations.

C.1 Competitive Equilibrium

Recall that $\hat{w}_{isn}(\omega) = \hat{a}_{isn}(\omega) - \eta \hat{\ell}_{isn}(\omega)$. Therefore

$$\begin{aligned}
\mathrm{Var}\left(\log\left(\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)\right)\right) &\approx \mathrm{Var}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega)\right)\right) \\
&= \mathrm{Var}\left(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\hat{a}_{isn}(\omega) + (1-\eta)\hat{\ell}_{isn}(\omega)\right)\right) \\
&= \mathrm{Var}\Bigg(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\Bigg(\hat{a}_{isn}(\omega) \\
&\qquad + \frac{1-\eta}{\eta+\eta_N}\left[\hat{a}_{isn}(\omega) - \frac{\eta_S}{\eta+\eta_N+\eta_S}\left(\sum_{n'\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\hat{a}_{isn'(\omega)}\right)\right]\Bigg)\Bigg) \\
&= \mathrm{Var}\Bigg(\sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\Bigg(\frac{1+\eta_N}{\eta+\eta_N}\hat{a}_{isn}(\omega) \\
&\qquad - \frac{1-\eta}{\eta+\eta_N}\frac{\eta_S}{\eta+\eta_N+\eta_S}\left(\sum_{n'\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\hat{a}_{isn'(\omega)}\right)\Bigg)\Bigg) \\
&= \mathrm{Var}\left(\left[\frac{1+\eta_N}{\eta+\eta_N} - \frac{1-\eta}{\eta+\eta_N}\frac{\eta_S}{\eta+\eta_N+\eta_S}\right]\left(\sum_{n'\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\hat{a}_{isn'(\omega)}\right)\right)
\end{aligned}$$

Then we break this up using covariances. Notice that

$$\mathrm{Cov}\left(\sum_{n'\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\hat{a}_{isn'}(\omega), \sum_{n'\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\hat{a}_{isn'}(\omega)\right) = \sigma_S^2 + \sum_{n\in\mathcal{N}_{is}}\left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 \sigma_N^2.$$

Thus,

$$\mathrm{Var}\left(\log\left(\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)\right)\right) = \left[\frac{1+\eta_N}{\eta+\eta_N} - \frac{1-\eta}{\eta+\eta_N}\frac{\eta_S}{\eta+\eta_N+\eta_S}\right]^2\left[\sigma_S^2 + \left(\int_{\mathcal{S}}\frac{\overline{\ell}_{is}}{\overline{\ell}_i}HHI_{is}ds\right)\sigma_N^2\right].$$

C.2 Oligopsony

Note that

$$\begin{aligned}
\log\left(\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)\right) &\approx \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_i}\left(\hat{w}_{isn}(\omega)+\hat{\ell}_{isn}(\omega)\right)\\
&= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_i}\left(\eta_S\hat{\ell}_{is}(\omega)+\eta_N\hat{\ell}_{isn}(\omega)+\hat{\ell}_{isn}(\omega)\right)\\
&= \eta_S\hat{\ell}_{is}(\omega)+(1+\eta_N)\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{\ell}_{isn}(\omega)\\
&= (1+\eta_N+\eta_S)\hat{\ell}_{is}(\omega)\\
&= (1+\eta_N+\eta_S)\left(\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\Psi_{isn}}{1+\eta_S\sum_{n'}\frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\Omega_{isn'}}\hat{a}_{isn}(\omega)\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{\mathcal{S}} \overline{L}_{is}\,\mathrm{Var}\left(\log\left(\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)\right)\right) \approx{}& (1+\eta_N+\eta_S)^2\left[\int_{\mathcal{S}}\overline{L}_{is}\left(\sum_n\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\Psi_{isn}}{1+\eta_S\sum_{n'}\frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\Omega_{isn'}}\right)^2 ds\right]\sigma_S^2\\
&+(1+\eta_N+\eta_S)^2\left[\int_{\mathcal{S}}\overline{L}_{is}\sum_n\left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\Psi_{isn}}{1+\eta_S\sum_{n'}\frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}}\Omega_{isn'}}\right)^2 ds\right]\sigma_N^2
\end{aligned}$$

D Solving the Quantitative Model

D.1 Summarizing the Equations

As shown above, production is given by

$$Y_i = z_i(\ell_i)^{1-\eta}(m_i)^{\eta}\Omega(m_i), \tag{26}$$

where Ω depends on if it is a competitive equilibrium or oligopsony. As payments to labor are given by

$$w_i\ell_i = (1-\eta)Y_i, \tag{27}$$

and firms get the remaining production. Average profits must be equal to the fixed cost of entering, therefore

$$\psi_i = \frac{\eta Y_i}{m_i}. \tag{28}$$

Utility of living in location i is,

$$U_i = u_i w_i. \tag{29}$$

People are free to move wherever they would like so that

$$\ell_i = \left(\frac{U_i}{U}\right)^{\theta} \ell, \tag{30}$$

where

$$U = \left[\sum_{i\in\mathcal{I}} (U_i)^{\theta}\right]^{\frac{1}{\theta}}. \tag{31}$$

Then spillovers in amenities and production are

$$u_i = \overline{u}_i (\ell_i)^{\gamma_u}, \tag{32}$$

and

$$z_i = \overline{z}_i (\ell_i)^{\gamma_z}. \tag{33}$$

D.2 Exact Hat Algebra

Rewriting the model into changes from an initial equilibrium, we get the following.

$$\hat{Y}_i = \hat{\overline{z}}_i (\hat{\ell}_i)^{1-\eta+\gamma_z} (\hat{m}_i)^{\eta} \frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)} \tag{34}$$

$$\hat{w}_i \hat{\ell}_i = \hat{Y}_i \tag{35}$$

$$\hat{\psi}_i = \frac{\hat{Y}_i}{\hat{m}_i} \tag{36}$$

$$\hat{U}_i = \hat{\overline{u}}_i (\hat{\ell}_i)^{\gamma_u} \hat{w}_i \tag{37}$$

$$\hat{\ell}_i = \left(\frac{\hat{U}_i}{\hat{U}}\right)^{\theta} \hat{\ell} \tag{38}$$

$$\hat{U} = \left[\sum_{i\in\mathcal{I}} \frac{\ell_i}{\ell} (\hat{U}_i)^{\theta}\right]^{\frac{1}{\theta}} \tag{39}$$

D.3 Counterfactual Computational Algorithm

We choose the equilibrium that has the most number of locations with a nonzero amount of labor. We take as given $\hat{U}$, and we find what that implies for labor and entry in location n. Notice that, combining the equations, we can write

$$
\begin{aligned}
\hat{\ell}_i &= \hat{U}_i^{\theta}\hat{U}^{-\theta}\hat{\ell} \\
&= \left(\hat{\hat{u}}_i(\hat{\ell}_i)^{\gamma_u}\hat{w}_i\right)^{\theta}\hat{U}^{-\theta}\hat{\ell} \\
&= \left(\hat{\hat{u}}_i(\hat{\ell}_i)^{\gamma_u-1}\hat{Y}_i\right)^{\theta}\hat{U}^{-\theta}\hat{\ell} \\
&= \left(\hat{\hat{z}}_i\hat{\hat{u}}_i(\hat{\ell}_i)^{\gamma_u-\eta+\gamma_z}(\hat{m}_i)^{\eta}\frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)}\right)^{\theta}\hat{U}^{-\theta}\hat{\ell}.
\end{aligned}
$$

Then solving for $\hat{\ell}_i$,

$$
\hat{\ell}_i = \left(\hat{\hat{z}}_i\hat{\hat{u}}_i(\hat{m}_i)^{\eta}\frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)}\hat{U}^{-1}\hat{\ell}^{\frac{1}{\theta}}\right)^{\frac{1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}.
$$

We then return to the free entry condition,

$$
\begin{aligned}
\hat{\psi}_i &= \frac{\hat{Y}_i}{\hat{m}_i} \\
&= \frac{\hat{\hat{z}}_i(\hat{\ell}_i)^{1-\eta+\gamma_z}(\hat{m}_i)^{\eta}\frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)}}{\hat{m}_i} \\
&= \hat{\hat{z}}_i\left(\hat{\hat{z}}_i\hat{\hat{u}}_i(\hat{m}_i)^{\eta}\frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)}\hat{U}^{-1}\hat{\ell}^{\frac{1}{\theta}}\right)^{\frac{1-\eta+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\hat{m}_i^{\eta-1}\frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)} \\
&= \hat{\hat{z}}_i\left(\hat{\hat{z}}_i\hat{\hat{u}}_i\hat{U}^{-1}\hat{\ell}^{\frac{1}{\theta}}\right)^{\frac{1-\eta+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\hat{m}_i^{\frac{\left(\gamma_u-\frac{1}{\theta}\right)(1-\eta)+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\left(\frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)}\right)^{\frac{\frac{1}{\theta}-\gamma_u+1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} \\
&= \hat{\hat{z}}_i\left(\hat{\hat{z}}_i\hat{\hat{u}}_i\hat{U}^{-1}\hat{\ell}^{\frac{1}{\theta}}\right)^{\frac{1-\eta+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} m_i^{-\frac{\left(\gamma_u-\frac{1}{\theta}\right)(1-\eta)+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\Omega(m_i)^{-\frac{\frac{1}{\theta}-\gamma_u+1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} \\
&\quad \cdot (\hat{m}_i m_i)^{\frac{\left(\gamma_u-\frac{1}{\theta}\right)(1-\eta)+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\Omega(\hat{m}_i m_i)^{\frac{\frac{1}{\theta}-\gamma_u+1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}
\end{aligned}
$$

Then we set up the algorithm.

(i) Guess $\hat{U}$.

(ii) For each location i, find $\hat{m}_i$ such that

$$\frac{\hat{\psi}_i}{\hat{\hat{z}}_i \left(\hat{\hat{z}}_i \hat{\hat{u}}_i \hat{U}^{-1} \hat{\ell}^{\frac{1}{\theta}}\right)^{\frac{1-\eta+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} m_i^{-\frac{\left(\gamma_u-\frac{1}{\theta}\right)(1-\eta)+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} \Omega(m_i)^{-\frac{\frac{1}{\theta}-\gamma_u+1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}}$$

$$= (\hat{m}_i m_i)^{\frac{\left(\gamma_u-\frac{1}{\theta}\right)(1-\eta)+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} \Omega(\hat{m}_i m_i)^{\frac{\frac{1}{\theta}-\gamma_u+1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}.$$

(iii) Then we can find the implied labor in location i,

$$\hat{\ell}_i = \left(\hat{\hat{z}}_i \hat{\hat{u}}_i (\hat{m}_i)^{\eta} \frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)} \hat{U}^{-1} \hat{\ell}^{\frac{1}{\theta}}\right)^{\frac{1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}},$$

and the implied utility

$$\hat{U}_i = \hat{\hat{u}}_i (\hat{\ell}_i)^{\gamma_u-\eta+\gamma_z} \hat{\hat{z}}_i (\hat{m}_i)^{\eta} \frac{\Omega(\hat{m}_i m_i)}{\Omega(m_i)}.$$

(iv) Then we can back out how $\hat{U}'$ must have changed based on those implied changes in utility

$$\hat{U}' = \left[\sum_{i \in \mathcal{I}} \frac{\ell_i}{\ell} (\hat{U}_i)^{\theta}\right]^{\frac{1}{\theta}}$$

(v) Update the guess of $\hat{U}$ and iterate until convergence

$$\hat{U}^{new} = \alpha \hat{U}' + (1-\alpha)\hat{U}.$$

D.4 Optimal Policy - Exact Hat Algebra

In the first best, profits should not equal the fixed cost of entering. Instead, the marginal product of another firm should equal the fixed cost. That is

$$\begin{aligned}
\psi_i &= \frac{d}{dm_i}\left[z_i(\ell_i)^{1-\eta}(m_i)^{\eta}\Omega(m_i)\right] \\
&= \eta z_i(\ell_i)^{1-\eta}(m_i)^{\eta-1}\Omega(m_i) + z_i(\ell_i)^{1-\eta}(m_i)^{\eta}\Omega'(m_i) \\
&= \frac{\eta Y_i}{m_i}\left(1 + \frac{1}{\eta}\frac{\partial \log \Omega(m_i)}{\partial \log m_i}\right).
\end{aligned}$$

We use $\tilde{x}$ to denote differences from the original equilibrium. We assume that this is paid for using a tax proportional to income. Therefore, the wages derived here are correct but the utility needs to be adjusted to take into account the taxes necessary to subsidize firm entry.

$$\tilde{Y}_i = \tilde{\tilde{z}}_i(\tilde{\ell}_i)^{1-\eta+\gamma_z}(\tilde{m}_i)^{\eta}\frac{\Omega(\tilde{m}_i m_i)}{\Omega(m_i)} \tag{40}$$

$$\tilde{w}_i\tilde{\ell}_i = \tilde{Y}_i \tag{41}$$

$$\tilde{\psi}_i = \frac{\tilde{Y}_i}{\tilde{m}_i}\left(1+\frac{1}{\eta}\frac{\partial\log\Omega(m_i)}{\partial\log m_i}\right) \tag{42}$$

$$\tilde{U}_i = \tilde{\tilde{u}}_i(\tilde{\ell}_i)^{\gamma_u}\tilde{w}_i \tag{43}$$

$$\tilde{\ell}_i = \left(\frac{\tilde{U}_i}{\tilde{U}}\right)^{\theta}\tilde{\ell} \tag{44}$$

$$\tilde{U} = \left[\sum_{i\in\mathcal{I}}\frac{\ell_i}{\ell}(\tilde{U}_i)^{\theta}\right]^{\frac{1}{\theta}} \tag{45}$$

D.5 Optimal Policy Computational Algorithm

Algebra similar to that for the counterfactual exact hat algebra gives

$$\tilde{\ell}_i = \left(\tilde{\tilde{z}}_i\tilde{\tilde{u}}_i(\tilde{m}_i)^{\eta}\frac{\Omega(\tilde{m}_i m_i)}{\Omega(m_i)}\hat{U}^{-1}\hat{\ell}^{\frac{1}{\theta}}\right)^{\frac{1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}.$$

Free entry implies

$$\begin{aligned}\tilde{\psi}_i = \tilde{\tilde{z}}_i\left(\tilde{\tilde{z}}_i\tilde{\tilde{u}}_i\tilde{U}^{-1}\tilde{\ell}^{\frac{1}{\theta}}\right)^{\frac{1-\eta+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} m_i^{-\frac{\left(\gamma_u-\frac{1}{\theta}\right)(1-\eta)+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\Omega(m_i)^{-\frac{\frac{1}{\theta}-\gamma_u+1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}} \\ \cdot(\hat{m}_i m_i)^{\frac{\left(\gamma_u-\frac{1}{\theta}\right)(1-\eta)+\gamma_z}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\Omega(\hat{m}_i m_i)^{\frac{\frac{1}{\theta}-\gamma_u+1}{\frac{1}{\theta}-\gamma_u+\eta-\gamma_z}}\left(1+\frac{1}{\eta}\frac{\partial\log\Omega(\tilde{m}_i m_i)}{\partial\log m_i}\right)\end{aligned}$$

The computational algorithm then follows the counterfactual one closely, but with the adjusted entry condition.

(i) Guess $\hat{U}$.

(ii) Find $\tilde{m}_i$ such that

$$\frac{\tilde{\psi}_i}{\tilde{\tilde{z}}_i\left(\tilde{\tilde{z}}_i\tilde{\tilde{u}}_i\tilde{U}^{-1}\tilde{\ell}^{\frac{1}{\theta}}\right)^{\frac{1-\eta+\gamma z}{\frac{1}{\theta}-\gamma u+\eta-\gamma z}} m_i^{-\frac{\left(\gamma u-\frac{1}{\theta}\right)(1-\eta)+\gamma z}{\frac{1}{\theta}-\gamma u+\eta-\gamma z}}\Omega(m_i)^{-\frac{\frac{1}{\theta}-\gamma u+1}{\frac{1}{\theta}-\gamma u+\eta-\gamma z}}}$$
$$= (\tilde{m}_i m_i)^{\frac{\left(\gamma u-\frac{1}{\theta}\right)(1-\eta)+\gamma z}{\frac{1}{\theta}-\gamma u+\eta-\gamma z}}\Omega(\tilde{m}_i m_i)^{\frac{\frac{1}{\theta}-\gamma u+1}{\frac{1}{\theta}-\gamma u+\eta-\gamma z}}\left(1+\frac{1}{\eta}\frac{\partial\log\Omega(\tilde{m}_i m_i)}{\partial\log m_i}\right).$$

(iii) Then we can find the implied labor in location i,

$$\tilde{\ell}_i = \left(\tilde{\tilde{z}}_i\tilde{\tilde{u}}_i(\tilde{m}_i)^{\eta}\frac{\Omega(\tilde{m}_i m_i)}{\Omega(m_i)}\hat{U}^{-1}\tilde{\ell}^{\frac{1}{\theta}}\right)^{\frac{1}{\frac{1}{\theta}-\gamma u+\eta-\gamma z}},$$

and the implied utility

$$\tilde{U}_i = \tilde{\tilde{u}}_i(\tilde{\ell}_i)^{\gamma_u-\eta+\gamma_z}\tilde{\tilde{z}}_i(\tilde{m}_i)^{\eta}\frac{\Omega(\tilde{m}_i m_i)}{\Omega(m_i)}.$$

(iv) Then we can back out how $\hat{U}'$ must have changed based on those implied changes in utility

$$\tilde{U}' = \left[\sum_{i\in\mathcal{I}}\frac{\ell_i}{\ell}(\tilde{U}_i)^{\theta}\right]^{\frac{1}{\theta}}$$

(v) Update the guess of $\tilde{U}$ and iterate until convergence

$$\tilde{U}^{new} = \alpha\tilde{U}' + (1-\alpha)\tilde{U}.$$

E Data Appendix

E.1 Comtrade Data

First, we take the annual value of traded goods from 1980 to 2016 across 4-digit product categories in SITC Rev. 2. Second, we convert them into the HS code, using a cleaner provided by Feenstra and Romalis (2014). Their cleaner resolves inconsistencies in favor of importer's reports, corrects values known to be inaccurate, accounts for re-exports of Chinese goods through Hong Kong, and includes Taiwan as a trading partner.[22] Third, we combine countries that unify or report jointly for subsets of years in the database.[23] Fourth, we convert these SITC Rev.2, 4-digit industrial categories into HS 2007, 6-digit using the crosswalk provided by the United Nations.[24] Finally, we convert these data in the 6-digit HS 2007 product code into 6-digit product categories used in the CoM data. We construct new crosswalk files based on the crosswalk provided in Baek et al. (2021).

Sample Construction of Establishment-level Panel Data Our main sample in the regression is a panel of establishments in Japan. First, We drop establishments with less than 4 employees since they are not sampled at an annual frequency.[25] Second, we use a converter provided by RIETI to link establishments across time starting in 1986. Finally, we keep establishments that appear at least 5 years consecutively. This is because we compute variance over time within establishments and need sufficient observations for each establishment.[26] Our final sample is an unbalanced panel of 724,417 unique establishments in manufacturing sectors from 1986 to 2016.

[22]Their cleaner is available here.

[23]The countries combined are East and West Germany; Belgium and Luxembourg; the islands that formed the Netherlands Antilles; North and South Yemen; and Sudan and South Sudan.

[24]The crosswalk is available in the UNSD web page here.

[25]The results are robust when we use all the establishments with at least 30 employees.

[26]Changing this threshold to a minimum of 10 years does not change our results.