Granular Origins of Agglomeration [*]

Shinnosuke Kikuchi Daniel G. O'Connor
UCSD Princeton

April 10, 2026

Abstract

Large firms dominate many local labor markets. How does this granularity shape economic geography and optimal place-based policy? We develop an economic geography model with granular firms facing idiosyncratic shocks and show that average wages rise with labor market size. Using establishment-level data on Japanese manufacturing, we estimate the model and find evidence consistent with our mechanism. Granularity explains 10–20% of estimated agglomeration externalities in the smallest commuting zones, but only 2–4% in Tokyo. Optimal industrial and wage policy increases the population in the smallest cities; firm effects depend on labor market conduct.

---

[*] Kikuchi: shkikuchi@ucsd.edu. O'Connor: daniel.oconnor@princeton.edu. We thank Daron Acemoglu, Treb Allen, Kosuke Aoki, David Atkin, David Autor, John Becko, Arnaud Costinot, Dave Donaldson, Ryungha Oh, Yuhei Miyauchi, Enrico Moretti, Tomoya Mori, Kentaro Nakajima, Bradley Setzler, Bob Staiger, David Weinstein, Iván Werning, and Nathan Zorzi for their helpful comments. We also thank seminar participants at Boston, Chicago, Dartmouth, Hitotsubashi, NBER Japan Project Meeting, Minnesota, MIT, Princeton, RIETI, Rutgers, UCSD, 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) of Japan. Under that project, 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.

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 employs a large proportion of the workforce in its headquarters city, Toyota. Even in a large city like Seattle, software engineers are at the mercy of Microsoft and Amazon. With these giants, a shock to a single firm is a shock to 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 forced into 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 theoretically, empirically, and quantitatively. The key mechanism driving our results is a form of labor market pooling. If there is one firm in a labor market, and that firm has a negative shock, workers have nowhere to go. They are stuck at that unproductive firm. In contrast, if there are many firms in a market and one small firm becomes less productive, workers can move to another firm that is doing better and use their skills more productively. Thus, markets with many firms provide a "constant market for skill" as Marshall (1920) said and Krugman (1992) formalized. This labor market pooling mechanism implies that markets with many firms use labor more effectively than markets with a few dominant firms. Thus, larger markets are more productive when firms are granular, i.e., when individual firms matter for the aggregate local labor market.

Our primary contribution is a new, tractable model of local labor markets with endogenous entry of granular firms that is still rich enough to allow for normative, empirical, and quantitative analysis. The model remains tractable because we introduce a continuum of sectors in each location and allow each sector to have a finite number of (ex-ante heterogeneous) firms. We then assume that firm entry is not targeted towards a particular sector, but towards an entire location. This ensures that workers stuck in a sector are exposed to individual firm shocks in their sector. However, firm entry does not devolve into intractable inequality constraints or multiple equilibria.[1]

We start by presenting a model of a single location with an exogenous population. Firms freely enter, workers choose what sector they would like to work in, and then each firm is subject to a productivity shock and hires labor in its sectoral labor market. We show that, in this setting, wages are increasing in population because firms use labor

---

[1] See Bresnahan and Reiss (1991) for a discussion of the issues that arise in a model of firm entry with a finite number of firms.

more effectively in large labor 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, consider a local sectoral labor market with only one firm. Because workers are stuck in the short run, the firm does not change its employment in response to changes in productivity. Instead, it only drives up the market wage when it demands more labor. In contrast, if a firm hires a small share of the local sectoral labor market because the market is very large, it can expand by poaching workers from other firms in the same sector without affecting the wage much. We also show that the marginal benefits of increasing the size of the region disappear as the market becomes very large. That is because once a market is sufficiently large, firms no longer have any difficulty finding the workers they need. Therefore, increasing the size will not further improve productivity.

We go beyond this intuition and show that a statistic for how effectively a single firm uses labor is the covariance of that firm's log productivity and its log employment. That is to say, a firm uses labor more productively if it increases its employment when it has high productivity and decreases its employment when it has low productivity. 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. Our mechanism implies that too few firms enter in equilibrium. That is because, when firms are granular, they know that 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 its own attempt to increase its employment drives up the wages. That induces a correlation between wages and idiosyncratic firm shocks, which depresses firm profits since profit functions are convex in wages and productivity. Therefore, firms will "under-supply" their entry. A social planner can achieve a Pareto improvement using place-based subsidies on firm entry, especially in small locations where granularity matters.

We then enrich the model to test some of the predictions and to quantify the importance of our theoretical mechanism. We introduce imperfect mobility of labor across both firms and sectors, and we allow firms to internalize their market power, competing against each other à la Cournot. We embed this model into a standard model of economic geography in which workers are mobile across locations. Our main theoretical results hold in this richer setting when firms are competitive, but they need to be adjusted when firms are imperfectly competitive. As we show, imperfect competition strengthens the agglomeration benefits of granularity. However, the optimal policy is different. The social planner still provides larger subsidies to smaller locations, but that policy might be targeted towards workers in terms of wage subsidies rather than firms in terms of entry subsidies.

Our quantitative and empirical analysis focuses on Japan, where we have a rich panel of all manufacturing establishments with at least 4 employees every year.[2] These data include employment, payroll, and shipments by 6-digit product category. We define a sector in the model as a 3-digit JSIC industry and a location as a commuting zone. We then estimate the key parameters of the model using these data. In an important way, Japan is not an ideal setting for our mechanism *quantitatively* because the labor market is relatively illiquid. We use Japan because the data makes it perfect to test the key predictions and estimate our model, but we view our results as a lower bound on the importance of the mechanism in developed countries.

Before we use the estimated model to quantify our mechanism, we look to validate the model, both qualitatively and quantitatively, by considering some reduced-form evidence of our mechanism. We start by providing evidence that granularity matters. We show that the variance of log payroll in a local 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 then provide evidence consistent with our mechanism as suggested by the covariance statistic. We begin by showing that the variance of log employment at a single firm in a large local sectoral labor market is larger on average than that of a similarly situated firm in a small local sectoral labor market. That suggests that if those firms are subject to similar shocks, the firm in the larger market expands more in response to good shocks and shrinks more in response to bad ones. We also provide evidence that this is due to our mechanism rather than another mechanism, for instance, increasing returns to scale in the matching function. To provide more direct evidence, we construct revenue productivity shocks to each firm using that firm's exposure to national changes in demand for the products it produces. Consistent with our mechanism, both qualitatively and quantitatively, firms that already hire a large portion of the local sectoral labor market expand their employment less in response to these shocks.

Finally, we turn to demonstrating the quantitative importance of the mechanism. If firms are perfectly competitive, we find that the implied elasticity of wages to population

---

[2] For the analyses using establishment-level data, we restrict our samples to establishments with at least 10 employees.

gets as high as 0.0035 in the smallest locations, and is a much smaller 0.001 in Tokyo. In total, the mechanism implies that Tokyo is 1.3% more productive than the smallest city in Japan. If firms are Cournot competitive, the implied elasticity of wages to population reaches 0.004 in the smallest locations. Combes et al. (2011) find that most causal estimates of the urban wage premium are an elasticity between 0.02 and 0.05 when pooling across locations of all sizes. Therefore, granularity could explain as much as 20% (= 0.004/0.02) of the urban wage premium.

We then quantify the implications for optimal policy. We define the marginal product of labor and firms at the commuting zone level as the marginal contribution of another worker or firm to the output of the entire commuting zone. We then compare those values to the wages and profits in each location to see if workers and owners are properly compensated for their contributions. The results depend on assumptions about firm conduct. If firms are perfectly competitive, workers are paid their marginal product, and firm profits are only 98% of their marginal product in small locations. In large locations, firms capture 99% of their contribution to commuting zone production. Putting in place the optimal firm entry subsidies would increase the number of firms in the smallest locations by 1.7% and in Tokyo by 0.6%. This reallocation of firms leads a small number of workers to move, increasing population in small areas by about 0.3%.

If firms are Cournot competitive in labor markets, workers in small locations see an average wage markdown of 2%. That markdown implies that firms' profits are actually 8% higher than their marginal contribution to production in small locations. This leads to over-entry of firms. Therefore, putting in place the optimal policies will actually lead to almost a 10% reduction in the number of firms in the smallest locations. Even Tokyo sees a 5% reduction in the number of firms. However, the optimal policy also features a wage subsidy in the smallest locations to undo the wage markdown. On net, these policies end up increasing the population in those small areas by about 0.5%.

The rest of the paper is organized as follows. We give a short review of the literature below. In Section 2, we present the baseline model of a single location and show our theoretical results. We enrich the model in Section 3 and estimate it in Section 4. We validate the model and our mechanism in Section 5 before presenting the quantitative results in Section 6. Section 7 concludes and suggests ways the model could be enriched to capture important real-world features of granular markets.

Related Literature. The literature on the spatial agglomeration of economic activity is rich. In an early contribution, Marshall (1920) proposes 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, 2024; Davis and Dingel, 2019). Duranton and Puga (2004) provide a new way to classify these 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 falls 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), which considers a setting with a finite number of ex-ante identical firms where labor is perfectly mobile within a labor market but not across. Overman and Puga (2010) extend Krugman's model to include multiple sectors and 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 model with ex-ante heterogeneous firms and a continuum of sectors that yields new theoretical results clarifying how granular firms generate agglomeration. We also derive the normative implications. Beyond the theoretical contribution, our model can be applied to the data to provide direct evidence of the 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 theory we lay out, though our evidence focuses on the firm response rather than the worker side. Conte et al. (2024) show 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) give 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) study what that implies for optimal policy. Schoefer and Ziv (2024)

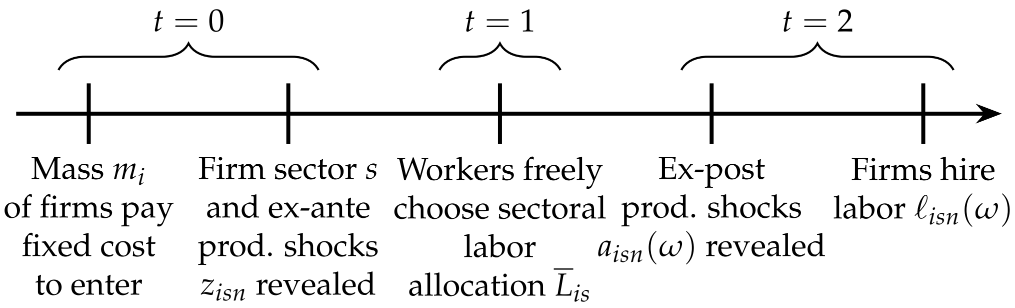


Figure 1: Timing of model

study how much of the observed differences across locations can be explained by granular shocks. We contribute to this literature by considering granularity's implications for the geography of economic activity and optimal place-based policy.

2 How Does Granularity Drive 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). In order to isolate the effects of granularity most transparently, we focus on the sectoral labor markets of a single location and consider a simplified environment that is neoclassical, conditional on firm entry.

We will enrich the model by endogenizing where people live, allowing for imperfect mobility across firms and sectors, and introducing imperfect competition in Section 3. Throughout the paper, we assume that workers and firms are homogeneous before making any training or entry decisions. Allowing for specialized skills used across different sectors and directed firm entry across regions and sectors is outside the scope of the current study. We discuss how including those features would likely change the analysis and how they could be incorporated into the model in Section 7.

2.1 Environment

There is a region i with a mass ℓ_i of workers and a continuum of sectors $s \in [0, 1]$. The sectors produce perfectly substitutable goods but hire in distinct sectoral labor markets.

Timing. There are three periods $t \in \{0, 1, 2\}$. In period 0, a mass m_i of firms 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 even though a non-integer mass of firms enter location i. This captures the long-run differences in firm size and will determine how exposed different sectoral markets are to short-run idiosyncratic firm shocks in period 2.

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

Then, in period 2, the state of the world $\omega \in \Omega$ is revealed. This determines the short-run productivity shocks of each firm. The worker cannot move labor across sectors at this time.[3] Instead, she supplies her sectoral labor inelastically, and firms hire labor in the sectoral labor market. Firms then produce and sell their goods. The fact that firms are better able to respond to these period 2 productivity shocks in larger markets will imply the agglomeration benefits. The model timing is summarized in Figure 1.

Workers. In location i, risk-neutral representative agents get utility from consuming a freely traded final good c_i and are endowed with one unit of labor that they supply to the market inelastically. In period 1, a worker freely allocates her units of labor across sectors $s \in [0, 1]$, taking as given the number of firms and each firm's ex-ante productivity z_{isn}. In particular, she chooses her vector of labor supply $\boldsymbol{L_i} \equiv \{L_{is}\}_s$ in the set of feasible labor allocations $\mathcal{L}$, i.e.

$$\boldsymbol{L_i} \in \mathcal{L} \equiv \left\{ \boldsymbol{L'_i} | \int_0^1 L'_{is} ds \leq 1 \right\}.$$

In period 2, the state of the world ω is revealed. The worker is unable to adjust labor at this point, and so inelastically supplies L_{is} labor to sector s.

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. We denote by $\mathcal{N}_{is}$ the set of firms operating in region i sector s and $N_{is} \equiv |\mathcal{N}_{is}|$ the (finite) number of firms. We assume that firms enter according to the "ball-and-urn model" so that N_{is} is distributed Poisson with mean m_i. That is, the probability mass function for the number of firms in a sector is $m_i^N e^{-m_i}/N!$. Firm n in sector s then gets an ex-ante productivity draw z_{isn} from a distribution $F_{iz}(\cdot)$.

In period 2, each firm n gets an ex-post idiosyncratic productivity shock, $\tilde{a}_{isn}(\omega)$, a

---

[3] This assumption is made for clarity; none of the key results depend on it. We allow for imperfect movement across both firms and sectors in period 2 in Section 3.

sector-wide productivity shock, $\tilde{A}_{is}(\omega)$, and produces a final good, $y_{isn}(\omega)$, according to,

$$y_{isn}(\omega) = z_{isn} a_{isn}(\omega) \ell_{isn}(\omega)^{1-\eta},$$

where $a_{isn}(\omega) \equiv \tilde{a}_{isn}(\omega)\tilde{A}_{is}(\omega)$ is the total productivity shock to firm n, $\ell_{isn}(\omega)$ is the total amount of labor firm n hires, and $\eta \in (0,1)$ is the degree of decreasing returns to scale. We formally state the assumptions on the productivity distributions below.

Assumption 1. The productivity distributions satisfy the following properties:

- Absent ex-post shocks, the expected firm size is finite, i.e. $\mathbb{E}[z_{isn}^{1/\eta}]$ exists and is finite;
- $\log \tilde{a}_{isn}(\omega)$ are iid with mean zero and a finite second moment σ_N^2; and
- $\log \tilde{A}_{is}(\omega)$ are iid with mean zero and a finite second moment σ_S^2.

Market Clearing. Total expected production in the location is

$$Y_i = \mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_{is}} z_{isn} a_{isn}(\omega) \ell_{isn}(\omega)^{1-\eta} ds\right],$$

where the expectation is taken with respect to ω, the number of firms in each sector, and the ex-ante 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 for investment must equal expected production,

$$c_i \ell_i + \psi_i m_i = Y_i. \tag{1}$$

In the labor market, labor demanded equals the individual labor supplied by each worker multiplied by the number of workers,

$$\sum_{n \in \mathcal{N}_{is}} \ell_{isn}(\omega) = L_{is} \ell_i, \quad \forall s, \omega. \tag{2}$$

2.2 Market Structure and Equilibrium

Labor Supply. Workers choose their labor allocation across sectors in period 1 to maximize their expected utility, taking wages as given. We normalize the price of the final good to 1, so workers solve the problem.

$$\boldsymbol{L_i} \in \underset{\boldsymbol{L_i'} \in \mathcal{L}}{\operatorname{argmax}} \quad \mathbb{E}\left[\int_0^1 w_{is}(\omega) L_{is}' ds\right], \tag{3}$$

where $w_{is}(\omega)$ is the equilibrium wage for sector s in state of the world ω. We will denote the maximum of (3) by w_i.[4]

Labor Demand. We normalize productivity so that the price of each good is 1. Then each active firm maximizes profits, taking wages and prices as given,

$$\ell_{isn}(\omega) \in \underset{\ell'}{\operatorname{argmax}} \quad z_{isn} a_{isn}(\omega)(\ell')^{1-\eta} - w_{is}(\omega)\ell'. \tag{4}$$

We will denote the maximum of (4) by $\pi_{isn}(\omega)$. We assume that firms are price takers, even though they are large relative to their market, for simplicity. In Appendix A.4, we show that nothing changes if firms commit to wage schedules in period 1 and compete à la Bertrand.

Entry. We assume that firms enter up to the point that expected profits are equal to the fixed cost of entering. After those entry decisions are made, all firms are randomly assigned to their sector and get their productivity draws. We can write this,

$$\psi_i = \frac{\mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_{is}} \pi_{isn}(\omega) ds\right]}{m_i}. \tag{5}$$

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.[5]

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

- Workers make labor supply decisions to maximize utility, taking wages as given, (3);

---

[4] In equilibrium, this will imply that average wages are equalized across sectors.

[5] An alternate entry decision would allow the marginal firm to observe the distribution of firms across sectors and their ex-ante productivity shocks before deciding whether or not to enter. We show that the free entry condition implied by that alternate entry game is also (5) if firms internalize that, were they to enter a sector with N firms, the sector would have $N+1$ firms in Appendix C.1.

- Conditional on entry, firms maximize profits taking prices and wages as given, (4);
- Firms enter up to the point that expected profits are equal to the fixed cost of entering, (5); and
- Goods and labor markets clear, (1) and (2).

2.3 Labor Market Pooling and Agglomeration

We now demonstrate how granularity implies that average wages increase with the number of workers. We will proceed in two steps. First, in this subsection, we will show that average wages increase with the number of workers if firms adjust their employment more in response to productivity shocks in larger markets. Throughout this paper, we will refer to this mechanism as labor market pooling, as it implies firms have an easier time finding the necessary workers to fill openings in larger markets.

Second, in Section 2.4, we will demonstrate how the existence of granular firms implies that firms can more easily expand their employment in response to shocks in larger markets. Combining this with Section 2.3, we conclude that granularity implies agglomeration benefits. Finally, in Section 2.5, we present the implications for optimal policy.

For tractability, we will derive all of our results using a log second-order approximation to expected production around the point with no ex-post shocks. We will use $\overline{x}$ to denote the value of a variable x in the absence of any ex-post shocks.

We start by introducing the *Regional Production Function* as it will be useful to organize the discussion. The regional production function gives the maximum possible production for location i, taking as given the mass of firms m and the number of workers ℓ:

$$Y_i(\ell, m) \equiv \max_{\ell'_{sn}(\omega), L_s} \left\{ \mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_s} z_{sn} a_{sn} \ell'_{sn}(\omega)^{1-\eta} ds \middle| m \right] \middle| \boldsymbol{L}' \in \mathcal{L}, \sum_{n \in \mathcal{N}_s} \ell'_{sn}(\omega) = L'_s \ell \right\}.$$

Since labor markets are perfectly competitive, this is not only the maximum level of production, but it also corresponds with the equilibrium production at the equilibrium levels of m and ℓ. The only difference between the regional production function and equilibrium production is that the number of firms is endogenously determined in equilibrium, while they are taken as given in the regional production function.

In the following Lemma, we characterize $Y_i(\ell, m)$ up to log second order.

Lemma 1. *The regional production function is* $Y_i(\ell, m) = z_i m^{\eta} \ell^{1-\eta} \Phi(m)$, *where* $z_i \equiv \mathbb{E}[z_{isn}^{1/\eta}]^{\eta}$

and $\Phi(m)$ is given by,

$$\Phi(m) \equiv \mathbb{E}[a_{sn}(\omega)] + \frac{1-\eta}{2} \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \operatorname{Cov}_\omega \left(\log a_{sn}(\omega), \log \ell_{sn}(\omega)\right) ds, \tag{6}$$

where the covariance is for a single firm across different states of the world.

The formal proof of Lemma 1 is in Appendix B, but we will provide a quick sketch here. We start by solving the maximization problem with no ex-post shocks. We then do a second-order approximation to the maximand, which implies $Y_i(\ell, m) = z_i \ell^{1-\eta} m^\eta \tilde{\Phi}(m)$ where

$$\begin{aligned} \tilde{\Phi}(m) \quad \equiv \quad & \mathbb{E}[a_{sn}(\omega)] + (1-\eta) \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \operatorname{Cov}\left(\log a_{sn}(\omega), \log \ell'_{sn}(\omega)\right) ds \\ & -\eta \frac{1-\eta}{2} \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \operatorname{Var}\left(\log \ell'_{sn}(\omega)\right) ds. \end{aligned} \tag{7}$$

That is, relative to keeping labor at its no-shock level in every state of the world, there are gains from increasing labor at firms that have good productivity shocks and decreasing labor at firms with bad productivity shocks. Those gains are tempered by the fact that there are decreasing returns to scale, so that productivity decreases if there is a very high variance of $\log \ell'_{sn}(\omega)$. When labor is efficiently allocated, either because the market is competitive or there is a planner,

$$\eta \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \operatorname{Var}\left(\log \ell'_{sn}(\omega)\right) ds = \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \operatorname{Cov}\left(\log a_{sn}(\omega), \log \ell'_{sn}(\omega)\right) ds, \tag{8}$$

proving the result. This leads to our first substantive result.

Corollary 1. *The regional production function features increasing returns to scale, i.e. $\frac{dY_i(\alpha\ell,\alpha m)}{d\alpha} > Y_i$, if and only if the employment weighted covariance between productivity and employment is increasing in the number of firms, i.e. $\frac{\partial\Phi(m)}{\partial m} > 0$.*

This result follows immediately from Lemma 1 since

$$\frac{dY_i(\alpha\ell, \alpha m)}{d\alpha} = \frac{d}{d\alpha}\left[\alpha z_i m^\eta \ell^{1-\eta} \Phi(\alpha m)\right] = Y_i(m, \ell)\left(1 + \frac{\partial \log \Phi(m)}{\partial \log m}\right).$$

Corollary 1 provides a clear interpretation of how labor market pooling implies that the regional production function features increasing returns to scale. Consider a firm n that never adjusts its workforce in response to idiosyncratic productivity shocks so that the covariance is 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 it is more productive. Thus, that firm could produce more goods on average while hiring the same average number of workers simply by increasing its covariance.

Corollary 1 then says that there are increasing returns to scale if increasing the size of the market improves how much 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.

Relation to Misallocation. This mechanism is closely related to the misallocation literature (Hsieh and Klenow, 2009). To see this in the math, note that for any firm on its labor demand curve, $w_s(\omega) = (1-\eta) z_{sn} a_{sn}(\omega) \ell_{sn}(\omega)^{-\eta}$. Therefore, looking at the variance of log wages,

$$\int_0^1 \frac{\overline{\ell}_s}{\overline{\ell}} \operatorname{Var}(\log w_s(\omega)) ds = \operatorname{Var}(\log a_{sn}(\omega)) \\ - \eta \int_0^1 \frac{\overline{\ell}_s}{\overline{\ell}} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \operatorname{Cov}(\log a_{sn}(\omega), \log \ell_{sn}(\omega)) ds,$$

using equation (8). That is, the employment weighted variance of log wages is decreasing in the employment weighted covariance of productivity and labor. This immediately implies the next corollary.

Corollary 2. *The regional production function features increasing returns to scale* $\frac{dY_i(\alpha\ell, \alpha m)}{d\alpha} > Y_i$ *if and only if the employment weighted variance of log wages is decreasing in the number of firms, i.e.* $\frac{\partial \Phi(m)}{\partial m} > 0$.

Corollary 2 says that larger markets are more productive than smaller markets if the average variance of log wages is lower in large markets. Loosely speaking, this is because there is less "misallocation" of labor. More precisely, the variance of log wages in sector s is a measure of how much more productive labor in sector s is in some states of the world than others. Therefore, it can be thought of as a measure of how unproductively labor is used on average.

Importantly, unlike misallocation, this variance is not a sign of inefficiency. Markets are perfectly competitive and are therefore efficient. Workers would love to reallocate some of their labor from states of the world in which the sectoral wage is low to states of the world where wages are high if they could. But that is not possible because the labor is stuck in each sector across all states of the world.

Agglomeration Benefits. We next turn to show how the increasing returns to scale of the regional production function $Y_i(\ell, m)$ imply that there are benefits to agglomeration. To do that, we need to relate the equilibrium number of firms to $Y_i(\ell, m)$. This is straightforward as wages are set competitively, so workers are paid their marginal product. That is,

$$w_i = \frac{\partial Y_i}{\partial \ell} = \frac{(1-\eta)Y_i}{\ell_i}. \tag{9}$$

Therefore, total profits are simply $Y_i - w_i \ell_i = \eta Y_i$, so that free entry (5) can be rewritten,

$$\psi_i = \frac{\eta Y_i}{m_i}. \tag{10}$$

Throughout the paper, we will focus on equilibria where the degrees of increasing returns to scale are small enough that profits are decreasing in the number of firms.

Combining equations (9) and (10) with Corollary 1 then implies the next proposition.

Proposition 1. *Average wages are increasing in the number of workers, i.e.* $\frac{d \log w_i}{d \log \ell_i} > 0$, *if and only if the employment weighted covariance between log firm productivity and log firm employment is increasing in m, i.e.* $\frac{\partial \log \Phi(m)}{\partial \log m}\Big|_{m=m_i} > 0$.

The proof proceeds in two steps. Equation (10) implies that the mass of firms increases more than proportionally with the increase in labor, $\frac{d \log m_i}{d \log \ell_i} = \frac{1-\eta}{1-\eta-\frac{\partial \log \Phi(m)}{\partial \log m}} > 1$, as average productivity increases with more entry. Then equation (9) implies average wages increase,

$$\frac{d \log w_i}{d \log \ell_i} = \frac{\frac{\partial \log \Phi(m)}{\partial \log m}}{1-\eta-\frac{\partial \log \Phi(m)}{\partial \log m}} > 0.$$

Proposition 1 summarizes the basic argument of labor market pooling. If firms have an easier time expanding in response to good productivity shocks in larger markets than in smaller markets, then average wages will increase with the size of the market.

2.4 Granularity and Labor Market Pooling

The previous subsection demonstrated how labor market pooling mechanisms can make larger markets more productive than smaller markets. In this subsection, we complete the argument that granularity implies agglomeration benefits by demonstrating how granularity implies those labor market pooling forces.

Proposition 2. *The average wage is increasing in the number of workers, i.e.,* $\frac{d \log w_i}{d \log \ell_i} > 0$*, if and only if idiosyncratic shocks have a positive variance,* $\sigma_N^2 > 0$*. Furthermore, the agglomeration benefits converge to zero as the size of the market goes to infinity, i.e.* $\frac{d \log w_i}{d \log \ell_i} \to 0$ *as* $\ell_i \to \infty$*.*

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, for a firm on its labor demand curve,

$$\Delta \log \ell_{isn}(\omega) = \frac{1}{\eta} \left(\Delta \log a_{isn}(\omega) - \Delta \log w_{is}(\omega)\right) = \frac{1}{\eta} \left(1 - \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right) \Delta \log a_{isn}(\omega),$$

where the last equality follows from the fact that the pass-through of a firm's productivity shock to sectoral wages is given by the share of the labor force that the firm hires $\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}$.

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 it cannot attract workers from other sectors. Instead, it drives up the wages in the sector. Therefore, the firm does not effectively scale up in response to a productivity shock and ends up using labor unproductively. In a market with a large mass of firms m_i, 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 poach workers from other firms in the sector, using that labor more productively.

The formal proof for Proposition 2 is in Appendix B, but we give a sketch of the proof here. Allowing for productivity shocks to every firm in sector s implies that, to first order, labor at firm n in sector s is given by

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

Then, some straightforward algebra reveals that the employment weighted average covariance is given by

$$\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \operatorname{Cov}(\log a_{isn}(\omega), \log \ell_{isn}(\omega)) ds = \frac{1}{\eta}\left(1 - \int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i} HHI_{is} ds\right)\sigma_N^2,$$

where $HHI_{is} \equiv \sum_{n\in\mathcal{N}_{is}} \left(\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\right)^2$. This covariance depends on the number of firms only through the HHI_{is}. Thus, the degree of increasing returns to scale is

$$\frac{\partial \log \Phi(m)}{\partial \log m} = -\frac{1}{2}\frac{1-\eta}{\eta}\frac{\frac{\partial}{\partial \log m}\left[\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i} HHI_{is} ds\right]}{\Phi(m)}\sigma_N^2. \tag{11}$$

If the average *HHI* across sectors decreases as the number of firms m increases, Proposition 1 is satisfied, and there are agglomeration benefits. We prove that the *HHI* of a 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 strange so that average HHI across sectors in location i is decreasing in the mass of firm entrants, m_i. For example, suppose that there were only two sectors with firms in them: one with two firms and another with one firm. If firm entry goes into unoccupied sectors, the average *HHI* would actually increase as the average number of firms in occupied sectors decreases. This, and other pathological cases, are ruled out with Poisson entry.

The speed with which the average *HHI* decreases depends on the distribution of entrants across sectors and the ex-ante productivity distribution $F_{is}(z)$. Entry 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 granular agglomeration forces are strong when there is 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, 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, HHI_{is} decreases at a rate of $N_{is}^{-\xi}$ where $\xi \in (0,1)$.[6] In that case, the weighted *HHI* does not fall as quickly, and large firms continue to be constrained in

---

[6] Formally, If $F_{is}(z)$ is a pareto distribution, i.e. $F_{is}(z) = 1 - az^{-\lambda}$, it is thick-tailed if $1 < \lambda\eta < 2$. Then $\xi = 2\left(1 - \frac{1}{\lambda\eta}\right)$.

their ability to respond to productivity shocks in larger markets. Thus, the externality is not as strong for small markets, but matters for medium-sized and even large cities. Nonetheless, the agglomeration force becomes less important as the number of firms increases, and the average *HHI* approaches zero.

2.5 Optimal Policy

In this subsection, we consider what this force for agglomeration means for policy. Conditional on firm entry, the model is competitive, so the only possible source of inefficiency is firm entry. Rewriting the goods market clearing condition (1), the first-best level of entry m_i^{FB} maximizes the total production net of firm entry costs,

$$m_i^{FB} \in \underset{m'}{\operatorname{argmax}} \quad Y_i(\ell_i, m') - \psi_i m'. \tag{12}$$

Taking the first order condition, we find that the first-best entry must satisfy

$$\psi_i = \left(1 + \frac{1}{\eta} \frac{\partial \log \Phi\left(m_i^{FB}\right)}{\partial \log m}\right) \frac{\eta Y_i}{m_i^{FB}}.$$

In contrast, equilibrium entry satisfies equation (10). Therefore, to implement the first best entry, the planner needs to enact a firm entry subsidy proportional to profits given by $\tau_i = \frac{1}{\eta} \frac{\partial \log \Phi(m)}{\partial \log m}$. And as $\frac{\partial \log \Phi(m)}{\partial \log m} \to 0$ as $m \to \infty$, the size of the optimal subsidy will also converge to zero in large markets. This implies our normative result.

Proposition 3. *If idiosyncratic shocks have a positive variance,* $\sigma_N^2 > 0$, *the optimal policy features a subsidy on firm entry proportional to firm profits given by* $\tau_i = \frac{1}{\eta} \frac{\partial \log \Phi(m)}{\partial \log m}\Big|_{m=m_i} > 0$. *Furthermore, the optimal subsidy converges to zero as the size of the market goes to infinity, i.e.* $\tau_i \to 0$ *as* $\ell_i \to \infty$.

One might wonder why there is any need for policy, given that wages are set competitively. The First Welfare Theorem fails in our model because firm entry is not Walrasian. In a competitive equilibrium, potential entrants take prices and wages as given when deciding whether or not they can make a profit by entering. In our model, by contrast, firms ask if current operating firms make a profit when deciding whether or not to enter. This condition is equivalent to one where a potential entrant internalizes that if it were to enter a sector with N firms, the sector would have $N + 1$ firms.[7]

---

[7] We show the equivalence in Appendix C.1.

Since firms are large relative to their sector, they know their entry will affect the wage distribution. A potential entrant understands that wages would be high precisely when it would like to hire more workers, since its own idiosyncratic shocks move market labor demand, and thereby, sectoral wages. Anticipating this endogenous wage response, a potential entrant is less likely to enter than a similarly situated Walrasian firm that takes the current, uncorrelated distribution of wages as given, since firm profit functions are convex. This under-entry force is strongest in small markets, where the marginal entrant has a large impact on the wage distribution.[8]

3 A Quantitative, Granular Model of Economic Geography

In Section 2, we presented a stylized model of sectoral labor markets within a single region where workers were freely mobile within a sector and unable to move across sectors. In this section, we model a country made up of I regions indexed by $i \in \mathcal{I} \equiv \{1, \ldots, I\}$. 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 location for the remaining periods.

We extend the model of the labor market in Section 2 so that in period 1, workers allocate their labor across sectors and firms in their location. Then, in period 2, workers can move their labor across both firms and sectors, subject to movement frictions. We further allow firms to internalize their labor market power so that they compete with other firms in their sector according to Cournot competition.[9] We show that the main results remain unchanged in this extended model with perfect competition, and the results only need to be adjusted slightly with imperfect competition. A more detailed description of the model, a characterization of the equilibrium, and proofs of the results are in Appendix A.

3.1 Extending the Baseline Model

Migration Across Regions. There is a mass ℓ of workers in the country. The fundamental utility of living in location i is $u_i = \overline{u}_i w_i$, where $\overline{u}_i$ are the local amenities. Each worker has an idiosyncratic preference for each location ε_i so that the utility the worker gets from living in location i is $u_i \varepsilon_i$.

---

[8] This result does not depend on our assumption that wages are set competitively. In Appendix A.4, we show that the under-entry result still holds if firms do not take wages as given and instead compete against each other à la Bertrand.

[9] In the appendix, we also present the model when firms compete in Bertrand.

We assume that ε_i are distributed Fréchet with shape parameter $\theta > 0$. Therefore, when people are free to live where they would like, the number of people who live in location i is given by $\ell_i = (u_i/u)^{\theta}\,\ell$, where $u = \left(\sum_{i\in\mathcal{I}}(u_i)^{\theta}\right)^{\frac{1}{\theta}}$.

Imperfect Mobility across Firms and Sectors. In period 1, the representative worker freely allocates her units of labor across sectors $s \in [0,1]$ 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 $\boldsymbol{L_i} \equiv \{L_{isn}\}_{s,n}$ in the set of feasible labor allocations $\mathcal{L}$,

$$\boldsymbol{L_i} \in \mathcal{L} \equiv \left\{\boldsymbol{L'_i} | \int_0^1 \sum_{n\in\mathcal{N}_{is}} L'_{isn} ds \leq 1\right\}.$$

In period 2, the state of the world ω is revealed. This determines the ex-post productivity shocks for all firms. The worker then reallocates labor across firms, choosing a vector of labor supply $\boldsymbol{L_i(\omega)} \equiv \{L_{isn}(\omega)\}_{s,n}$ in the set of feasible labor allocations $\mathcal{L}_{\Omega}(\boldsymbol{L_i})$ which depends on the worker's labor choices in period 1. The set is given by

$$\begin{aligned}\mathcal{L}_{\Omega}(\boldsymbol{L_i}) \equiv \Bigg\{\boldsymbol{L_i(\omega)} | 1 = &\left(\int_0^1 L_{is}^{-\frac{1}{\nu}} L_{is}(\omega)^{\frac{1+\nu}{\nu}} ds\right)^{\frac{\nu}{1+\nu}}, \\ &L_{is}(\omega) = \left(\sum_{n\in\mathcal{N}_{is}} \left(\frac{L_{isn}}{L_{is}}\right)^{-\frac{1}{\kappa}} L_{isn}(\omega)^{\frac{1+\kappa}{\kappa}}\right)^{\frac{\kappa}{1+\kappa}} \Bigg\},\end{aligned}$$

where $L_{is} = \sum_{n\in\mathcal{N}_{is}} L_{isn}$.

This set implies that $L_{is}(\omega) = L_{is}$ and $L_{isn}(\omega) = L_{isn}$ is feasible, but the more that the worker deviates from period 1 labor choices, the more labor is lost in transition. $\kappa > 0$ is the short-run elasticity of substitution across firms within a sector, and $\nu > 0$ is the short-run elasticity of substitution across sectors. The worker in location i chooses $\boldsymbol{L_i}$ and $\{\boldsymbol{L_i(\omega)}\}_{\omega}$ to maximize expected utility taking wages as given.

Ex-ante Productivity Distribution. The ex-ante productivity distribution is Pareto with shape parameter λ and location parameter ζ_i that can differ across locations. We assume that $\lambda\eta > 1$ so that the expected firm size is finite.

Labor Market Competition. For the rest of the paper, we will present results under alternate assumptions on firm behavior. Under perfect competition, firms continue to maximize profits, taking wages as given. However, now that workers are imperfectly substitutable across firms, each firm has its own wage $w_{isn}(\omega)$ that it takes as given.

Under imperfect competition, firms internalize that they can affect the wages they face. We assume that the firm can commit to a wage schedule in period 1, so that the firm internalizes how wage choices in state of the world ω will affect how much period 1 labor the worker supplies to the firm. This captures the decision-making of a firm that could lower wages significantly today but knows its workers would leave in the long run as a result. This implies that, in the absence of ex-post shocks, wages will correspond with competitive wages as workers are freely mobile in period 1. With ex-post shocks, firms will sometimes hire a larger portion of the market and have more market power. Therefore, the firm can vary its markdown across different states of the world to increase profits. We assume Cournot competition, so that each firm takes as given the labor hired by the other firms in its sector and the work opportunities in the other sectors.

3.2 Theoretical Results with Perfect Competition

When firms engage in perfect competition, the key results, Lemma 1, Proposition 1, Proposition 2, and Proposition 3 do not change at all as long as workers are more mobile across firms within a sector than across sectors, $\kappa > \nu$. The only qualitative change is that Corollary 2, relating the mechanism to misallocation, no longer holds. That is because misallocation only measures how efficiently existing labor is used. With movement frictions, the worker also needs to consider how much labor is lost in moving labor across firms and sectors.

Quantitatively, numbers change as the covariance, determining the strength of agglomeration, is given by

$$\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \operatorname{Cov}(\log a_{isn}(\omega), \log \ell_{isn}(\omega)) ds$$
$$= \frac{1}{\eta+\frac{1}{\nu}}\sigma_S^2 + \frac{\eta+\frac{1}{\nu} - \left(\frac{1}{\nu}-\frac{1}{\kappa}\right)\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} HHI_{is} ds}{\left(\eta+\frac{1}{\kappa}\right)\left(\eta+\frac{1}{\nu}\right)}\sigma_N^2.$$

The size of the externality then also depends on how much harder it is for a firm to attract workers from other sectors compared to workers from other firms in its own sector, $\frac{1}{\nu}-\frac{1}{\kappa}$.

3.3 Theoretical Results with Imperfect Competition

In this subsection, we discuss how the theoretical results change when firms compete à la Cournot.

Most importantly, Lemma 1 does not hold with Cournot competition as the covariance expression relied on the market efficiently trading off the gains from moving workers to more productive firms against the loss of labor and decreasing returns to scale. However, the production function still has a simplified form to log second-order, as we confirm in the next lemma. We will use $f \in \{p, c\}$ to denote firm conduct, with p denoting perfect competition and c denoting Cournot.

Lemma 2. *For firms competing $f \in \{p, c\}$, the regional production function is $Y_i^f(\ell, m) = z_i m^{\eta} \ell^{1-\eta} \tilde{\Phi}^f(m)$, where $z_i \equiv \mathbb{E}[z_{isn}^{1/\eta}]^{\eta}$. Furthermore, $\tilde{\Phi}^c(m) \leq \tilde{\Phi}^p(m)$.*

Generically, $\tilde{\Phi}^c(m) < \tilde{\Phi}^p(m)$ because when firms are imperfectly competitive, misallocation due to varying wage markdowns decreases the productivity of the market relative to the efficient benchmark.

In practice, including imperfect competition increases the agglomeration benefits. The reason is that with imperfect competition, larger locations have two advantages: it is easier for a theoretical planner to move workers to their most productive use, and the equilibrium gets closer to the planner's solution. In the limit with an infinite number of firms, firms do not vary their wage markdown so that workers move to their efficient use. By contrast, in small locations, firms do not raise their wage as much as a planner would like in response to a good productivity shock, so firms expand too little. Thus, Proposition 2 only becomes stronger.

In contrast, the optimal policy changes with imperfect competition. The first-best policy would exactly replicate the perfect competition benchmark with firm- and state-dependent wage subsidies that exactly undo the wage markdown. Throughout this paper, we will assume that this policy is infeasible. Instead, we assume that the planner can give place-based transfers to workers and place-based entry subsidies to firms. Therefore, conditional on the firms and workers in a location, the planner takes as given how competition continues. In math, this implies that the planner will take as given the functions $\tilde{\Phi}^f(m)$.

Then, to explore optimal policy, we need to know what workers and firms are paid in equilibrium, and how that differs from their marginal contribution to production. With perfect competition, workers are paid their marginal product, so the optimal policy features firm entry subsidies, as firms cannot get paid their marginal product. If firms exploit their market power, there are wage markdowns, and workers are not paid their marginal product. Instead, their earnings are distorted downwards. The following lemma summarizes how much.

Lemma 3. *For firms competing Cournot, there exists a function $\Psi^c(m) \leq 0$ such that total wage compensation in location i can be written,*

$$w_i \ell_i = (1-\eta) z_i (m_i)^{\eta} (\ell_i)^{1-\eta} \left(\tilde{\Phi}^c(m_i) + \Psi^c(m_i) \right).$$

We derive $\Psi^c(m)$ in the online supplement and use it when considering our mechanism quantitatively in Section 6. Comparing the marginal product of labor and firms to their payments then implies our next result.

Proposition 4. *The optimal policy features a subsidy on earnings in location i proportional to average wages of τ_i^w and a subsidy (or tax) on firm entry proportional to profits τ_i^{π} that satisfies,*

$$1 + \tau_i^w = \frac{1}{1 + \frac{\Psi^f(m_i)}{\tilde{\Phi}^f(m_i)}}; \qquad 1 + \tau_i^{\pi} = \frac{1 + \frac{1}{\eta} \frac{\partial \log \tilde{\Phi}^f(m_i)}{\partial \log m}}{1 - \frac{1-\eta}{\eta} \frac{\Psi^f(m_i)}{\tilde{\Phi}^f(m_i)}}. \tag{13}$$

Without any more information, we know that the optimal policy features a subsidy on wages, $\tau_i^w > 0$. Those subsidies will undo the average markdown workers face in location i. The sign of τ_i^{π}, on the other hand, is ambiguous. As we showed in Proposition 3, granularity implies that there is under-entry. This force is represented in the numerator of equation (13) as $\frac{\partial \log \tilde{\Phi}^f(m)}{\partial \log m} > 0$ and pushes for the optimal subsidy to be positive. However, with imperfect competition, there is another force in the denominator. As discussed by Mankiw and Whinston (1986), when firms are imperfectly competitive, there is a tendency for too many firms to enter. In our model, this arises because firms markdown wages, biasing profits above their competitive levels, increasing the incentive to enter. The quantitative importance of this force is summarized by the size of the wage markdown, $\frac{\Psi^f(m)}{\tilde{\Phi}^f(m)}$. We return to how these forces interact quantitatively in Section 6.

4 Estimation of Granular Agglomeration

In Section 3, we presented a quantitative model of economic geography with granular firms. In this section, we estimate the model in order to quantify how important the granular mechanism is for agglomeration in Section 6. We focus on manufacturing sectors in Japan, where high-quality firm-level data across years enables us to examine the extent to which firms experience idiosyncratic shocks.

4.1 Data

Data Sources. Our primary data source is the Census of Manufacture (CoM) in Japan, conducted annually by the Ministry of Economy, Trade, and Industry (METI). The census covers all manufacturing establishments in years ending with 0, 3, 5, or 8, and those with at least four employees in other years. The CoM was not conducted in 2012 and 2016; instead, the Economic Census for Business Activity (ECBA) by METI and the Ministry of Internal Affairs and Communications provides data for 2011 and 2015.[10] We use the ECBA to substitute for CoM in 2011 and 2015.

We use data from 2002–2019, as product classification changed discontinuously in 2002. For establishment-level analysis, we restrict the sample to establishments with at least 10 employees and those surveyed in at least 10 of the 18 years (2002–2019).

These Japanese data have two advantages. First, we observe panels of all 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.[11] Second, we observe yearly shipment values by detailed, 6-digit product categories for each establishment-year. This enables us to construct establishment-level exposure to product-level demand shocks at an annual frequency to test some of the predictions.

Mapping from Model to Data. We interpret region i as a commuting zone and denote by $\mathcal{I}$ the set of 225 commuting zones in Japan.[12] We map sectors $s \in \mathcal{S}$ onto 148 manufacturing industries at the 3-digit level.[13] Each establishment in the data is treated as an independent firm, as multi-establishment firms are not modeled. We refer to each establishment as a "firm" throughout. We interpret the short run as 1 year, which is enough time for workers to move between firms but not across commuting zones.

---

[10] The ECBA covers all establishments, including non-manufacturing ones, but we focus on manufacturing to maintain consistency with the CoM.

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

[12] To construct time-consistent commuting zones, we follow Kondo (2023) to convert municipalities into time-consistent groups. Japan has 1,724 municipalities as of June 2023, including 6 in the Northern Territories, which we drop because the CoM does not cover them. We then use the converter in Adachi et al. (2020) to map these groups into commuting zones and retain those with at least 10 manufacturing establishments in 2019.

[13] We use a RIETI crosswalk to convert all categories into 2011 codes. In theory, sectors form a continuum, but we use 148 finite sectors because workers move freely within these broad groups, which best represent labor markets. This abstraction omits sectoral granularity, underestimating individual firm influence.

4.2 Labor Supply

Short Run Labor Supply Elasticity across Firms. We first estimate a short-run labor supply elasticity across firms, κ. As we show in (28) in the Appendix, the labor supply equation in period 2 is given by

$$\log \ell_{isn}(\omega) - \log \ell_{isn} = \kappa \log w_{isn}(\omega) - (\kappa - \nu) \log w_{is}(\omega) - \nu \log w_i(\omega) + \tilde{\varepsilon}^{w}_{isn}(\omega). \quad (14)$$

We interpret the short run of period 2 as one year. Therefore, to estimate κ, we take a one-year change of these variables and get the following equation:

$$\Delta \log \ell_{isnt} = \kappa \Delta \log w_{isnt} + \gamma_{ist} + \tilde{\varepsilon}^{w}_{isnt}, \quad (15)$$

where γ_{ist} is a market-time fixed effect.

The key threat to the identification of κ is that changes in wages of firm n in location i and sector s might be correlated with the changes in workers' taste for firm n. To address this concern, we instrument $\Delta \log w_{isnt}$ with a shift-share IV, Δd_{isnt}, constructed as follows.

$$\Delta d_{isnt} \equiv \sum_{p} \overline{s^{p}_{isn}} \cdot \Delta d^{\text{national}}_{pt}, \quad (16)$$

where $\overline{s^{p}_{isn}}$ is a time-invariant share of product p in total shipment from firm n. We take the average over our sample periods between 2002 and 2019. $\Delta d^{\text{national}}_{pt}$ is a one-year log change in the shipment of product p at a national level. We interpret this shift in demand as a shock to the revenue productivity of the firm, as that shift increases the effective price of the firm's output. As we have normalized the price of each good to one, this shift in price is loaded on the effective productivity $a_{isn}(\omega)$ in the model.

The results are in Table 1. All columns include market-time fixed effects, the sum of the shares ('incomplete shares'), and the interaction of the sum and time fixed effects as covariates. Columns (2), (3), and (4) include firm-level lagged log employment growth as an additional covariate. Column (3) drops the top 10 products (units of the shifts) in terms of the absolute values of Rotemberg weights, which we present in Table A2. Column (4) uses lagged shipment share, instead of the time-invariant average shipment share, $\overline{s^{p}_{isn}}$, to construct the shift-share IV in (16). Exposure-robust standard errors (Borusyak et al., 2025) are in parentheses.

The estimates across columns are not statistically different, and we take the estimate in Column (1), $\hat{\kappa} = 1.64$, as our baseline estimate. This is in line with many of the estimates

Table 1: Estimation of Short-run, Across-Firm Labor Supply Elasticity

Dep. Var.: Log Employment Growth	(1)	(2)	(3)	(4)
Log Wage Growth	1.64	2.02	1.79	2.00
	(0.70)	(0.95)	(0.73)	(0.98)
LLM-Year FE	✓	✓	✓	✓
Control Lagged Emp Growth		✓	✓	✓
Drop Top 10 Rotemberg			✓	
Use Lagged Share				✓
1st-stage F Stat.	45.72	32.37	30.25	27.91
Num of Unique LLMs	9,978	9,830	9,796	9,830
Num of Unique Products (Shifts)	1,829	1,829	1,819	1,829
Num of Unique Firms	117,031	115,313	113,510	115,313
Num of Years	17	16	16	16
Shift-Year Obs.	30,693	28,794	28,634	28,790
Firm-Year Obs.	1,367,957	1,229,293	1,198,961	1,229,293

Note: This table shows the estimates of short-run labor supply elasticity across firms within markets, κ, following (15). They include market-time fixed effects, the sum of the shares, and the interaction of the sum and time fixed effects as covariates. Columns (2), (3), and (4) include firm-level lagged log employment growth as an additional covariate. Column (3) drops the top 10 products (units of the shifts) in terms of the absolute values of Rotemberg weights, which we present in Table A2. Column (4) uses lagged shipment share instead of the time-invariant average shipment share to construct the shift-share IV in (16). Exposure-robust standard errors (Borusyak et al., 2025) are in parentheses.

in the literature. Lamadon et al. (2022) and Berger et al. (2022) estimate this elasticity to be 6.52 and 10.52, respectively, in the more fluid labor market of the United States. In Brazil, Felix (2024) finds $\hat{\kappa} = 1.02$.

Short Run Labor Supply Elasticity across Markets. Next, we estimate a short-run labor supply elasticity across markets, ν. The labor supply to sector s in period 2 can be written,

$$\log \ell_{is}(\omega) - \log \ell_{is} = \nu \log w_{is}(\omega) - \nu \log w_i(\omega) + \varepsilon^{w}_{is}(\omega), \tag{17}$$

where

$$w_{is}(\omega) \equiv \left[\sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}}{\ell_{is}} w_{isn}(\omega)^{1+\kappa} \right]^{\frac{1}{1+\kappa}}.$$

We then estimate ν following a very similar procedure to how we estimate κ. Taking a one-year difference of these variables, we get the following equation:

$$\Delta \log \ell_{ist} = \nu \Delta \log w_{ist} + \gamma_{it} + \gamma_{st} + \varepsilon^{w}_{ist}, \tag{18}$$

where γ_{it} is a location-time fixed effect, and γ_{st} is a sector-time fixed effect. To get theory-consistent measures of the market-level labor and market-level wages, we use the estimate $\hat{\kappa} = 1.64$, to construct the variables

$$\ell_{ist} = \left[\sum_{n \in \mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \right)^{-\frac{1}{\hat{\kappa}}} \ell_{isnt}^{\frac{1+\hat{\kappa}}{\hat{\kappa}}} \right]^{\frac{\hat{\kappa}}{1+\hat{\kappa}}}, \quad w_{ist} = \left[\sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} w_{isnt}^{1+\hat{\kappa}} \right]^{\frac{1}{1+\hat{\kappa}}}.$$

We then use equation (18) to estimate ν. Just as in the case of κ, we instrument for wages to avoid endogeneity issues. We construct market-wide versions of the same instrument used to estimate κ,

$$\Delta d_{ist} \equiv \sum_{n} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \cdot \Delta d_{isnt}, \tag{19}$$

where Δd_{isnt} is the shift-share IV constructed in (16), $\overline{\ell}_{isn}$ is the median of employment of firm n in location i and sector s, and $\overline{\ell}_{is}$ is the median of total employment in location i and sector s. In essence, Δd_{ist} is a weighted average of firm-level demand shock at a market level.

The results are in Table 2. All columns include sector-time fixed effects, cz-time fixed effects, the sum of the shares ('incomplete share'), and the interaction of the sum and time fixed effects as covariates. Columns (2), (3), and (4) add lagged employment growth as a new covariate. Column (3) use lagged shipment share when constructing the firm-level shift share IV in (16). Column (4) weights each local labor market using the median total employment over time. Exposure-robust standard errors are in parentheses.

The estimates across the specifications are similar, and we take the estimate in Column (1), $\hat{\nu} = 0.82$, as our baseline estimate. In the US setting, Lamadon et al. (2022) and Berger et al. (2022) estimate this elasticity to be 4.57 and 0.42, respectively. Felix (2024) finds $\hat{\nu} = 0.80$ in Brazil.

4.3 Labor Demand

Returns to scale. We set $\eta = 0.18$ to be the average firm profit share in 2019 using Financial Statements Statistics of Corporations (FSSC).[14]

Ex-Ante Productivity Distribution. We could estimate the shape parameter of the ex-ante productivity distribution λ in a couple of different ways. However, Equation (11)

---

[14] Our baseline model does not have capital, so all money not paid out in profits goes to workers. In Appendix C.2, we present an extension of the model with freely traded capital and show that it is equivalent to our baseline model with an adjusted η.

Table 2: Estimation of Short-run, Across-Market Labor Supply Elasticity

Dep. Var.: Log LLM Employment Growth				
	(1)	(2)	(3)	(4)
Log LLM Wage Growth	0.82	0.86	0.91	0.98
	(0.14)	(0.14)	(0.15)	(0.14)
Sector-Year FE	✓	✓	✓	✓
CZ-Year FE	✓	✓	✓	✓
Control Lagged Emp Growth		✓	✓	✓
Use Lagged Share			✓	
Weighted				✓
1st-stage F Stat.	410.01	427.86	371.85	1868.43
Num of Unique LLM	13,754	13,246	13,246	13,246
Num of Years	17	16	16	16
LLM-Year Obs.	189,761	172,353	172,353	172,353

Note: This table shows the estimates of short-run labor supply elasticity across markets, ν, following (18). The unit of observation is the local labor market (LLM)-year level. LLMs are defined as pairs of 3-digit sectors and commuting zones. All columns include sector-time fixed effects, cz-time fixed effects, the sum of the shares ('incomplete share'), and the interaction of the sum and time fixed effects as covariates. Columns (2), (3), and (4) add lagged employment growth as a new covariate. Column (3) used lagged shipment share when constructing the firm-level shift share IV in (16). Column (4) weights each local labor market using the median total employment over time. Exposure-robust standard errors (Borusyak et al., 2025) are in parentheses.

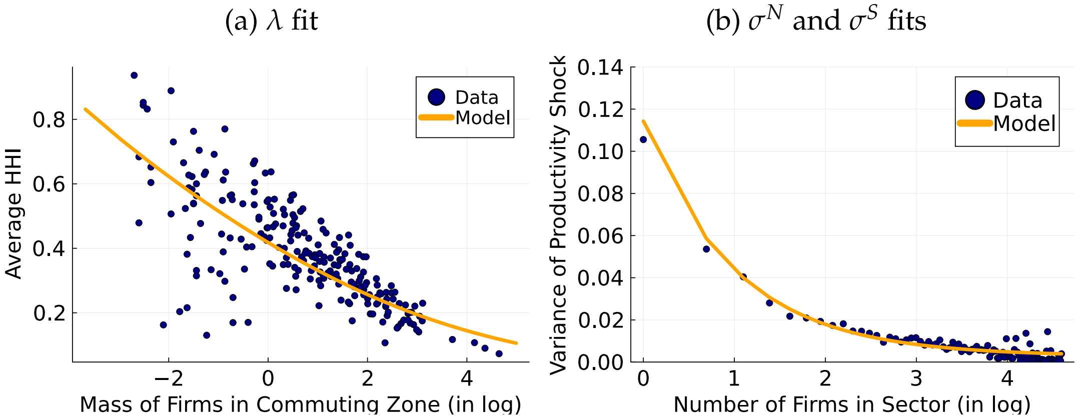
suggests that the key moment determining the strength of the externality is the average HHI across sectors within a location. Therefore, we estimate λ to match the average HHI across sectors for each location i, $\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i} HHI_{is} ds$, with a quadratic loss function. Intuitively, a higher λ implies that the firm distribution is less thick-tailed and average *HHI* will decrease quickly with the mass of firms in a location i. A lower λ implies that the *HHI* would decrease much more slowly as a few large firms will dominate most markets. The fit is depicted in Figure 2.

Ex-Post Productivity Distribution. To estimate the distribution of ex-post productivity shocks, we look at the distribution of the average shock to a sectoral labor market as a function of the number of firms in that market. Recall that production is $y_{isn}(\omega) = z_{isn} a_{isn}(\omega) \ell_{isn}(\omega)^{1-\eta}$. As we interpret the time scale of period 2 as one year, we take first differences to get

$$\Delta \log a_{isnt} = \Delta \log y_{isnt} - (1-\eta) \Delta \log \ell_{isnt}. \tag{20}$$

Using our estimate of η, we can estimate the productivity shock to firm n from equation (20). We then create a measure of the variance of the average productivity shock to

Figure 2: Supply Side - Model Fit


(a) λ fit

(b) σ^N and σ^S fits

Note: The figures show the fits of the model to the moments we target. The left panel shows the average HHI across sectors and the mass of firms in each commuting zone in log. The right panel shows the variance of the average productivity shock to sectors with N firms defined by (21) and the number of firms per sector in log. In both of the figures, the blue dots show the data for each commuting zone, and the orange line shows the model-implied relationship.

sectoral labor markets with N firms in sector s, $\sigma_s^2(N)$, defined as,

$$\sigma_s^2(N) = \text{Var}\left(\frac{\sum_{n \in \mathcal{N}_{is}} \Delta \log a_{isnt}}{N} \middle| N_{is} = N \right). \tag{21}$$

We then estimate σ_{Ss}^2 and σ_{Ns}^2 to match $\sigma_s^2(N)$ for $N = \{1, \ldots, 99\}$. We then average σ_{Ss}^2 and σ_{Ns}^2 across sectors s, weighted by national employment shares to get σ_S^2 and σ_N^2. We show the average fit in Figure 2. We find that σ_S^2 is very small. This comes from the fact that the average productivity shocks in large sectoral labor markets are very small, suggesting that sector-wide shocks are relatively unimportant on average.

4.4 Economic Geography Parameters

We take the migration elasticity $\theta = 3$ from the literature (Redding, 2016). Our key results on the size of agglomeration benefits and the optimal place-based policies do not depend on this parameter. It matters only for our results on how the population responds to the optimal policy change. However, similar migration elasticity estimates have been found in the United States (Hornbeck and Moretti, 2024) and Indonesia (Bryan and Morten, 2019). That leaves parameters summarizing the average firm productivity ζ_i, amenity $\overline{u}_i$, and fixed cost of opening a firm ψ_i in each location i. We calibrate these parameters to exactly match 2019 data on the population, average wages, and number of

Table 3: Summary of Estimated and Calibrated Parameters

Description	Parameter	Value	Source
A. Labor Supply			
Short run labor elasticity across firms	κ	1.64	Estimated (CoM)
Short run labor elasticity across markets	ν	0.82	Estimated (CoM)
B. Labor Demand			
Returns to scale	η	0.18	Profit Share (Data, FSSC)
Ex-ante firm prod. tail	λ	7.76	Estimated SMM (CoM)
Variance of sector shocks	σ_S^2	0.003	Estimated GMM (CoM)
Variance of idiosyncratic shocks	σ_N^2	0.11	Estimated GMM (CoM)
C. Economic Geography Parameters			
Migration elasticity	θ	3	Redding (2016)
Avg. Productivity, Amenity, Entry Cost	$\zeta_i, \overline{u}_i, \psi_i$		Exact hat algebra (2019)

Note: This table summarizes where the key parameters of the model come from. See the discussion in Section 4 for a more in-depth description.

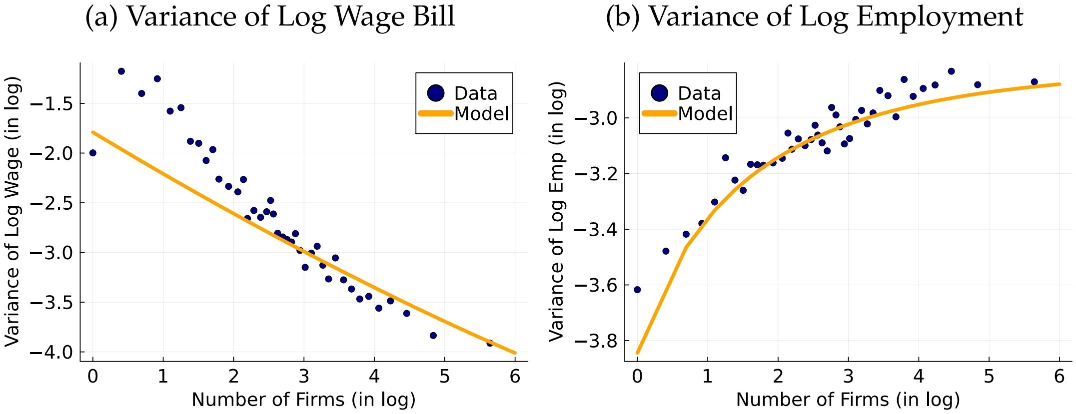
firms in each location.

5 Validation of the Mechanism

We presented a quantitative model of economic geography with granular firms in Section 3 and estimated it in Section 4. Before we show the quantitative importance of the mechanism, we present some reduced-form evidence of the mechanism and demonstrate that it is consistent with the model we have just quantified.

Our reduced-form evidence focuses on two key predictions intimately tied to the mechanism. First, we provide evidence that firms are subject to idiosyncratic shocks that "average out" in larger markets by looking at the variance of the wage bill across time in labor markets of different sizes. Second, we provide evidence that firms in larger markets expand their employment more than firms in small markets in response to positive productivity shocks and shrink more in response to negative shocks. We do this with cross-sectional and quasi-experimental evidence. For cross-sectional evidence, we show that firms in larger markets have a higher variance of log employment across time. We then do a quasi-experimental test of our mechanism. We construct productivity shocks and show that for similarly situated firms, firms that are small relative to their labor market expand more in response to the same shock.

Figure 3: Model Validation


(a) Variance of Log Wage Bill (b) Variance of Log Employment

Note: The figures show the fits of the model to the moments we do not target. The left panel shows the variance of the log wage bill for different sectoral local labor markets. The right panel shows the variance of firm-level employment aggregated to the sectoral local labor market level. Both statistics are plotted against the number of firms across local labor markets and in log units. In both of the figures, the blue dots show the binned data of each local labor market, and the orange line shows the model-implied relationship. The level of the y-axis (only the level, not the slope) is normalized so that the distance between the model and the data is minimized.

5.1 Shocks Average Out in Large Markets

Shocks to firm revenue are passed through to wages paid to workers. Thus, we can use the total payments to workers in a sectoral labor market as a measure of the shocks to the market.

If our theory is correct, firms are subject to different shocks, so the variance of the log wage bill at the sectoral market should be lower for markets with more firms. In math, workers are paid their marginal product of labor so that $w_{isn}(\omega) = (1-\eta) z_{isn} a_{isn}(\omega) \ell_{isn}(\omega)^{-\eta}$. And, the short-run labor supply is given by (14). Some straightforward algebra combining these equations then implies our first empirical prediction.

Proposition 5. *If $\sigma_N^2 > 0$, the variance of the log wage bill of a sectoral labor market decreases in the number of firms. In particular,*

$$\mathrm{Var}\left(\log\left(\sum_{n\in\mathcal{N}_{is}} w_{isn}(\omega)\ell_{isn}(\omega)\right)\right) = \left(\frac{1+\nu}{1+\eta\nu}\right)^2 \left(\sigma_S^2 + HHI_{is}\sigma_N^2\right).$$

We present a binned scatter plot of the log variance of the log wage bill by sectoral local labor market against the log number of firms in Figure 3(a). We include the variance of the log wage bill implied by the model as well.

Just as predicted by the model, the variance of the log wage bill is decreasing in the number of firms. The data suggests that the rate of that decrease is slightly faster than that suggested by the model. This is likely because our model assumes that every firm has the same variance of log productivity shocks, whereas in practice, larger firms have less volatile productivity. Therefore, in the real world, the market becomes less exposed to individual firm shocks at a faster rate than that suggested by the model.[15]

5.2 Firms Expand More in Larger Markets

We next turn to validating the model's key prediction for the mechanism: firms expand more in response to productivity shocks in larger markets. We provide two separate pieces of evidence for this prediction.

Cross-Sectional Evidence. We first give cross-sectional evidence for the prediction. In particular, we look at the variance of the log employment of individual firms in different-sized markets. If our theory is correct, and firms in larger markets are subject to similar shocks as those in smaller markets, then firms in larger markets should have a higher variance of log employment as they respond more to the same shocks.

In math, we can find how employment at an individual firm n responds to its own productivity shock and the shocks to every other firm to first order,

$$\Delta \log \ell_{isn}(\omega) = \frac{1}{\eta + \frac{1}{\kappa}} \left(\Delta \log a_{isn}(\omega) - \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \sum_{n' \in \mathcal{N}_{is}} \frac{\ell_{isn'}}{\ell_{is}} \Delta \log a_{isn'} \right). \tag{22}$$

Then, with some algebra, we can solve for the average variance of log employment, weighted by the average labor share. This implies our second empirical prediction.

Proposition 6. *If workers are more mobile across firms within a sector than across sectors, i.e.* $\kappa > \nu$, *and* $\sigma_N^2 > 0$, *then the weighted average variance of log employment increases with the number of firms in a sectoral labor market. In particular,*

$$\sum_{n \in \mathcal{N}_{is}} \frac{\ell_{isn}}{\ell_{is}} \operatorname{Var}\left(\log \ell_{isn}(\omega)\right) = \left(\frac{1}{\eta + \frac{1}{\nu}} \right)^2 \sigma_S^2 + \left(\frac{1}{\eta + \frac{1}{\kappa}} \right)^2 \sigma_N^2 - \left(\frac{1}{\nu} - \frac{1}{\kappa} \right) \frac{\eta + \frac{1}{\kappa} + \eta + \frac{1}{\nu}}{\left(\eta + \frac{1}{\kappa}\right)^2 \left(\eta + \frac{1}{\nu}\right)^2} HHI_{is} \sigma_N^2.$$

---

[15] Estimating λ, σ_N^2, and σ_S^2 to match this bin scatter rather than how we estimate them in Section 4 does not significantly change any results.

We plot the bin scatter for the log weighted variance of log employment for each firm against the log number of firms in Figure 3(b). We include the model-implied value as well.

Just as predicted by the model, the variance of log employment for individual firms is increasing in the number of firms. Furthermore, although we do not target it, our model matches the rate of that increase well. We even see evidence that the variance of log employment stops increasing as much for sectoral labor markets with a large number of firms as would be suggested by the model.

One thing we do not hit is the level of the variance of log employment. The figure normalizes the value to be 0 when the number of firms is one, but our model implies a higher variance in log employment than what we see in the data. This could be for a variety of reasons. One possible reason is that the employment data measures employment in June, while firms can adjust their employment throughout the year in response to shocks.

Employment Response to Demand Shocks. We conclude with a more direct test of the mechanism by examining whether firms that hire only a small share of the sectoral labor market expand employment more in response to a revenue productivity shock.

To test this prediction, we run the regression suggested by equation (22),

$$\Delta \log \ell_{isn,t} = \beta_1 \Delta d_{isn,t} + \beta_2 \left(\frac{\ell_{isn,t}}{\ell_{is,t}} \right) + \beta_3 \left(\Delta d_{isn,t} \times \frac{\ell_{isn,t}}{\ell_{is,t}} \right) + \zeta_t + \zeta_{isn} + \varepsilon_{isn,t} \tag{23}$$

where $\Delta d_{isn,t}$ is the firm-level shock constructed as in (16) for firm n in location i and sector s. ζ_t is the time fixed effect, and ζ_{isn} is the firm fixed effects. Importantly, we leave out any industry-time fixed effects because our mechanism follows from the fact that an idiosyncratic shock to a large firm is also an industry-wide shock.

The shock implied by the theory is in terms of revenue productivity, while the shock $\Delta d_{isn,t}$ is in terms of total sales of products produced by the firm. Therefore, without an elasticity relating the price of the products produced by firm n to a shift in demand for the products, there are no predictions on the size of β_1 or β_3. However, the theory predicts that the ratio is given by

$$\frac{\beta_3}{\beta_1} = -\frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \approx -0.435.$$

We report the results in Table 4. All columns include firm fixed effects. Columns (1), (2), and (3) include year fixed effects. Column (4) includes CZ-year fixed effects. Columns (2), (3), and (4) include lagged log employment growth as a covariate. Column (3) drops the top 10 products in Rotemberg weights. Exposure-robust standard errors (Borusyak

Table 4: Responses of Employment to Product-level Shocks

	Dep. Var.: Log Employment Growth			
	(1)	(2)	(3)	(4)
Shock	0.049	0.047	0.045	0.045
	(0.005)	(0.005)	(0.005)	(0.005)
Payroll Share	-0.026	-0.024	-0.023	-0.024
	(0.005)	(0.005)	(0.005)	(0.005)
Shock x Payroll Share	-0.025	-0.028	-0.026	-0.027
	(0.010)	(0.009)	(0.009)	(0.009)
Implied Ratio	-0.518	-0.585	-0.573	-0.597
95% CI	[-0.851, -0.185]	[-0.891, -0.278]	[-0.896, -0.250]	[-0.912, -0.282]
Firm FE	✓	✓	✓	✓
Year FE	✓	✓	✓	
CZ-Year FE				✓
Control Lag Emp Growth		✓	✓	✓
Drop Top 10 Rotemberg			✓	
Num of Unique LLMs	14,318	13,967	13,937	13,967
Num of Unique Firms	114,553	108,884	106,959	108,884
Num of Years	17	16	16	16
Firm-Year Obs	1,427,772	1,238,720	1,209,548	1,238,719

Note: This table shows the estimates of the impact of firm-level demand shock on employment, following (23). All columns include firm fixed effects. Columns (1), (2), and (3) include year fixed effects. Column (4) includes CZ-year fixed effects. Columns (2), (3), and (4) include lagged log employment growth as a covariate. Column (3) drops the top 10 products in Rotemberg weights. Exposure-robust standard errors (Borusyak et al., 2025) are in parentheses. The implied ratio is the ratio of the coefficient on the interaction term between the shock and the lagged payroll share within local labor markets, β_3, to the coefficient on the shock, β_1. The 95% confidence interval of that implied ratio is also reported.

et al., 2025) are in parentheses. The implied ratio is the ratio of the coefficient on the interaction term between the shock and the lagged payroll share within local labor markets, β_3, to the coefficient on the shock, β_1. The 95% confidence interval of that implied ratio is also reported.

First, the positive demand shock increases employment, and that effect is statistically significant across all specifications, consistent with the theory. Second, the elasticity of employment to the shock is smaller for firms with a larger payroll share within local labor markets. The estimated ratios range from -0.6 to -0.5 across specifications, all of which are close to, and not statistically different from, the theoretical prediction of -0.435. The model-implied ratio lies well within the 95 percent confidence intervals in every specification. These results indicate that the theoretical mechanism is both qualitatively and quantitatively consistent with the data.

One might be concerned that these results are driven by mechanisms other than our own. In Appendix D, we do a couple of tests to rule these out. First, we show that including industry-location-time fixed effects makes β_3 statistically insignificant, consistent with the theory. We also estimate the labor supply elasticity κ separately for large and small locations and find no evidence that they are statistically different from each other.

6 Quantification of Granular Agglomeration

In this section, we present the results demonstrating the quantitative contribution of granularity to agglomeration. For all of the results, we present the results for the cases of perfect competition and Cournot competition. This allows us to demonstrate which results are robust to assumptions on firm conduct and which results are not. The results for the Bertrand competition are in Appendix E.2 and generally fall between the two cases we present here.

6.1 Strength of Agglomeration Benefits

As we know from Proposition 1, the strength of the granular origins of agglomeration is determined by the shape of $\tilde{\Phi}^f(m)$. Thus, we plot $\log \tilde{\Phi}^f(m)$ in Figure 4 (a) and (b).

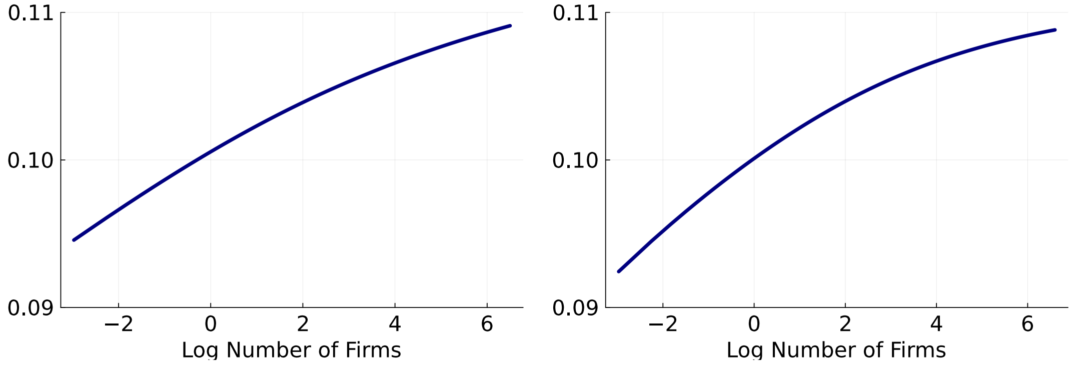
Consistent with Proposition 2, the curve is upward sloping, implying that larger markets are more productive than smaller markets. Also consistent with Proposition 2, the curve flattens out for larger m, suggesting that the marginal benefit of increasing the number of firms in a location that is already larger is small, while locations with few firms could see large productivity benefits from increasing the number of firms. The implied $\tilde{\Phi}^f(m)$ looks very similar whether firms are competing perfectly or according to Cournot.

In total, our estimated $\tilde{\Phi}^p(m)$ implies that the largest commuting zone in Japan (Tokyo, with an average of 105.8 firms per sector) is 1.3% more productive than the smallest commuting zone (with an average of 0.07 ($\approx 10/148$) firms per sector). Cournot competition implies the gap is 1.5%.

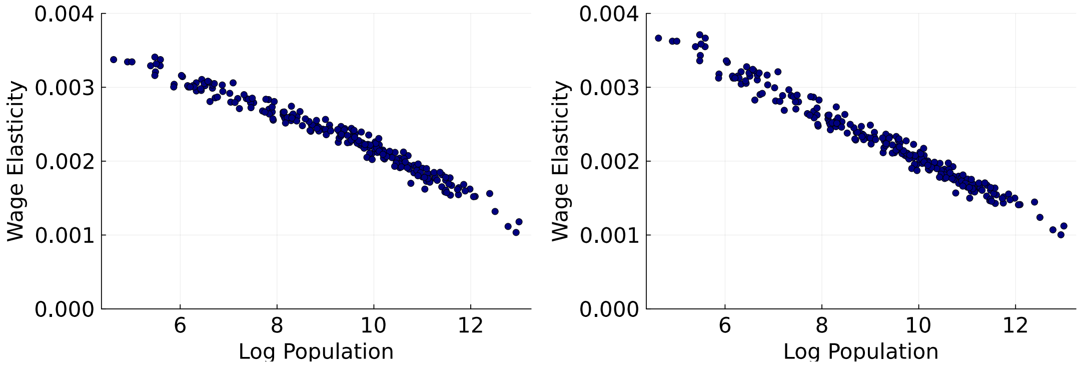
In Figure 4 (c) and (d), we plot the implied elasticity of wages to population due to our mechanism for each of the commuting zones in Japan against the population. In the perfect competition case, we find that the wage elasticity gets as high as 0.0035 for the smallest commuting zone. The implied elasticity is much smaller for the large locations: Tokyo has an implied elasticity of 0.001. If firms are competing à la Cournot, the elasticities range from 0.001 to 0.004.

For context, Combes et al. (2011) find that most causal estimates of the entire urban wage premium (due to all mechanisms) are the elasticities between 0.02 and 0.05 when

Figure 4: $\log \tilde{\Phi}^{f}(m)$ and Wage Elasticities


(a) Perfect Competition

(b) Cournot Competition


(c) Perfect Competition

(d) Cournot Competition

Note: The figures show $\log \Phi^{f}(m)$ against the log number of firms across different numbers of firms and wage elasticity across commuting zones in Japan. The left panel shows the case of perfect competition, and the right panel shows the case of Cournot competition.

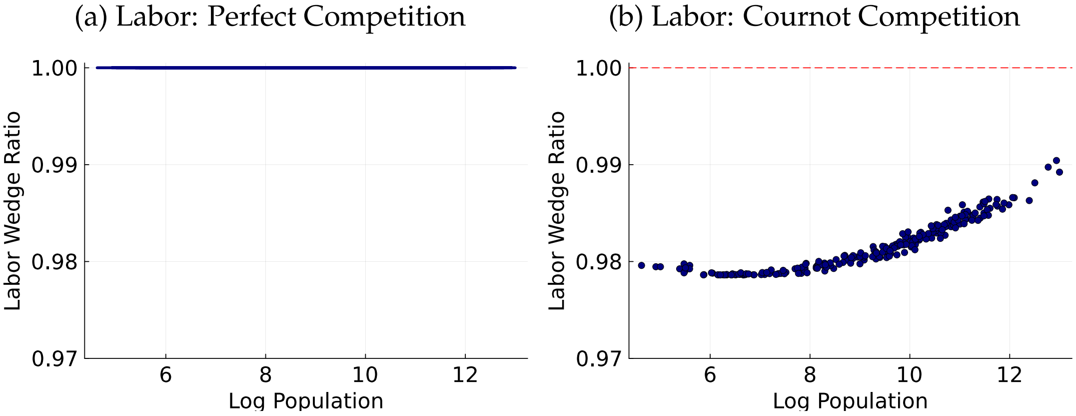
pooling across locations of all sizes. Thus, the granular mechanism could explain as much as 20% (=0.004/0.02) of the wage benefits of living in a large location. However, this mechanism cannot explain most of the urban wage premium.

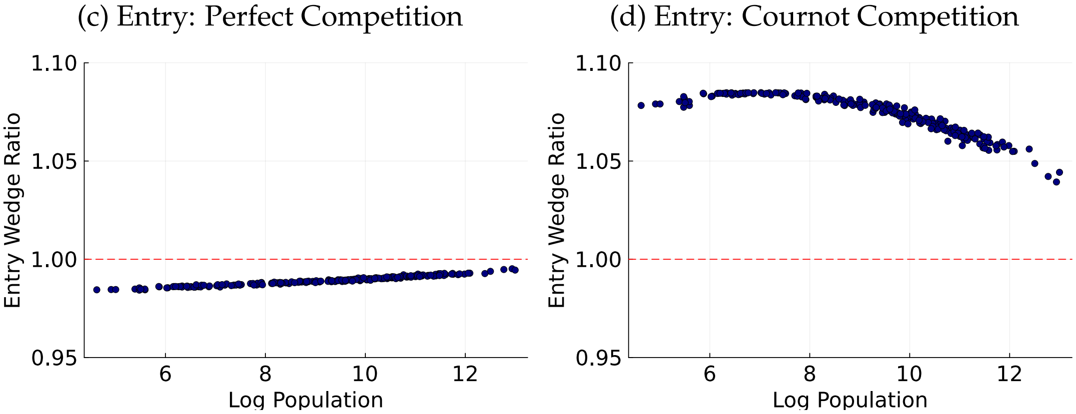
6.2 Factor Wedges

We next turn to quantify what our model implies for the degree of wage markdowns and under-entry.

Labor Wedges. We start by plotting the labor wedge ratio in Figure 5. This is the ratio of average wages to the marginal product of another worker. By assumption, the perfect competition case in Figure 5(a) is 1 for all locations. Under Cournot competi-

Figure 5: Factor Wedge: Labor and Firm Entry


(a) Labor: Perfect Competition
(b) Labor: Cournot Competition


(c) Entry: Perfect Competition
(d) Entry: Cournot Competition

Note: The figures show the labor wedge ratio and firm entry wedge across commuting zones in Japan. The left panel shows the case of perfect competition, and it is one by construction. The right panel shows the case of Cournot competition, and each dot represents a commuting zone in Japan.

tion, firms' market power increases following good shocks and decreases following bad shocks, which enables them to markdown wages even though workers are freely mobile in the long run. This short-run market power implies that workers face an average markdown of around 2% in small and medium-sized cities. Only in larger locations like Tokyo, do the wage markdowns become as small as 1%. Perhaps surprisingly, the smallest commuting zones do not have the largest markdowns. Instead, there is a local maximum around a log population of 6. That is because firms can only exploit their market power if their share of the sectoral local labor market varies. If there is only one firm in the sector, the firm only competes against other firms in the long run, when workers are freely mobile. If there are two firms in the sector, then the firm will see large swings in its own market power and can vary its markdown to take advantage of that.

Firm Entry Wedges. We next turn to the implications for firm entry. As stated in Proposition 3, granularity implies that too few firms enter in equilibrium. In Figure 5, we plot the ratio of firm profits to the marginal product of another firm for the location's output. Consistent with the proposition, Figure 5(c) shows that every firm is paid less than its marginal product if firms are perfectly competitive. In the smallest locations, firms capture 98.4% of their contribution to production, while in the largest location they capture a full 99.5%.

As we mentioned in Section 3, the efficiency of entry depends on firm conduct assumptions. By exploiting their market power, firms distort wages downward and profits upward. This can lead profits to increase above the marginal product of another firm, taking as given the conduct of the firms, conditional on firm entry. In Figure 5(d), we plot the ratio of firm profits to the marginal product of firms. We find that simply by varying their wage markdown, leading to an average wage markdown of 2%, firms can, in fact, increase profits so much that too many firms enter in equilibrium.

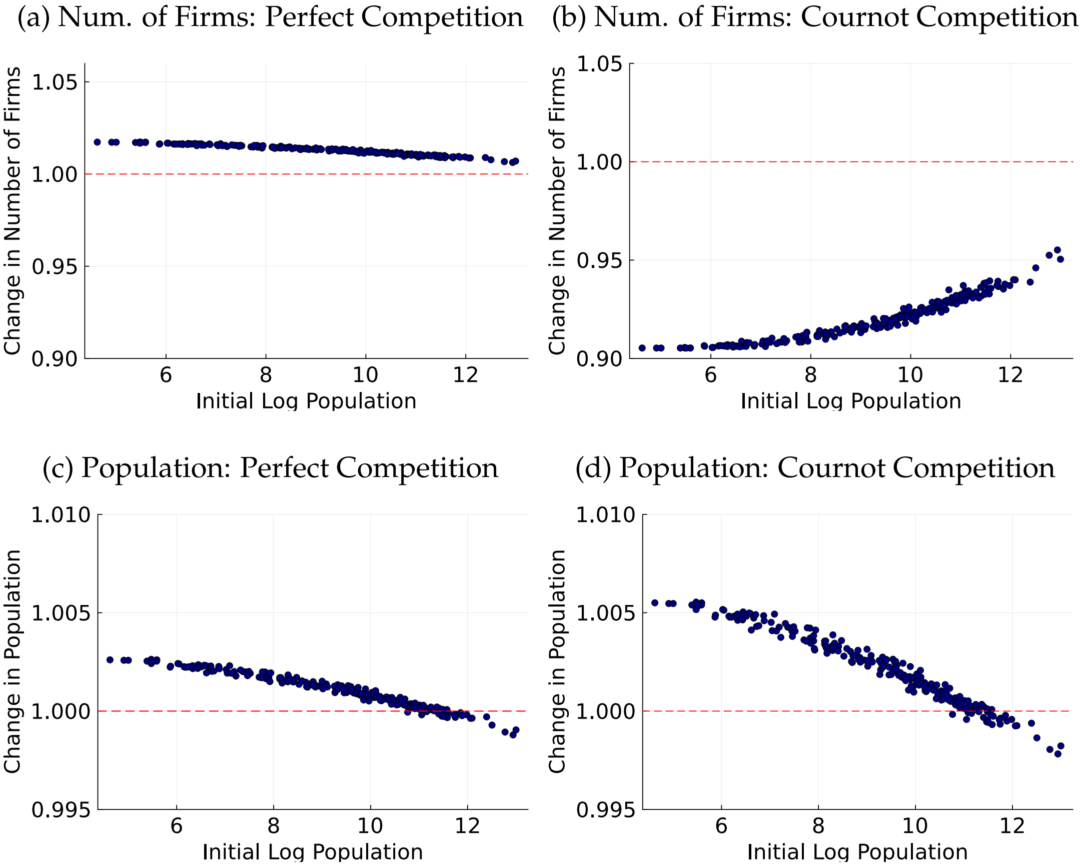
6.3 Implications of Optimal Policy

In the previous subsection, we showed the ratio of factor payments to the marginal product of that factor under different firm conduct assumptions. In this subsection, we consider how the equilibrium levels of population and mass of firms would change if a national government put in place the optimal place-based transfers and firm entry subsidies suggested by Proposition 4, paid for with a proportional increase in income taxes.

We plot how the optimal subsidies change the equilibrium number of firms and population in Figure 6. In Figure 6(a), we show that the equilibrium number of firms increases with the optimal subsidies if firms are competitive. The smallest locations see a 1.7% increase in the number of firms, while in Tokyo, the number of firms only increases by 0.6%. The effect of the optimal policy on the number of firms is very different when firms compete imperfectly. The over-entry implications of imperfect competition overrule the under-entry implications of granularity. Thus, Figure 6(b) shows that the optimal number of firms is 91% of the observed number of firms in the smallest locations, while it is 95% in Tokyo.

Figure 6 (c) and (d) plot the effect of the optimal policy on the population. If firms compete perfectly, the optimal policy only indirectly affects where people would like to live because the optimal policy will increase the number of firms, especially in the smallest locations. Therefore, the smallest locations see a 0.3% increase in population while

Figure 6: Changes in Number of Firms and Population


(a) Num. of Firms: Perfect Competition

(b) Num. of Firms: Cournot Competition

(c) Population: Perfect Competition

(d) Population: Cournot Competition

Note: The figures show the changes in the number of firms and population in response to the optimal subsidies. Each dot represents a commuting zone in Japan. The left panel shows the case of perfect competition, and the right panel shows the case of Cournot competition.

Tokyo sees a very small decline in population in Figure 6(c). The effects of optimal policy on population are more pronounced when firms are competing à la Cournot, as shown in Figure 6(d). Here, the optimal policy features a large tax on firm entry in the smallest locations, but also a subsidy on wages. And because the aggregate externality is still largest for the smallest locations, the net effect is that the population should increase in the small commuting zones. In fact, the population should increase even more than under perfect competition because the externality is larger with imperfect competition. Thus, the optimal population is more than 0.5% higher than the observed population in the smallest locations.

7 Concluding Remarks

The world is granular, and local sectoral labor markets are especially so. That granularity has important implications for the geography of economic activity and optimal policy. Larger locations are more productive because the firms in those areas can expand in response to productivity shocks. A simple model of this labor market pooling mechanism in the famously illiquid labor market of Japan suggests that it could explain as much as 20% of the wage premium of large cities. Thus, while most of the urban wage premium must come from other mechanisms, granularity has important explanatory power.

In formalizing the mechanism, we were forced to ignore certain important features of real firms and labor markets. For example, we abstract from inefficient labor market adjustment to shocks such as wage rigidity and search frictions. Including these features would likely increase the social cost of being over-exposed to individual firm-level shocks as they increase the cost of adjusting to shocks more generally. Thus, this would simply increase the importance of our mechanism.

We also assume that all workers are ex-ante homogeneous, whereas many workers have specialized skills that could be used in multiple sectors. The framework we laid out can be extended to include these features. We suspect that this extension will further increase the benefits of being in a large city due to granularity, as workers with very specialized skills will want to be in very large cities where there are enough firms that want them. The firms will then follow them as they cannot produce without those skillsets. This force will be even stronger if different skillsets are complementary in production.

On the firm side, we do not allow particular firms to direct their entry towards certain sectors or locations. Allowing for entry directed towards certain sectors could give new implications for which sectors will be overrepresented in large cities. Such a model that includes workers with specialized skills could also have important implications for co-agglomeration patterns of sectors that use workers with similar skillsets. Allowing firms to sort across locations could have still more interesting implications. Large, productive firms that greatly impact the wage distribution when they enter will have a greater incentive to locate in larger cities. This could further increase the productivity of large cities over and above what we find here.

Granularity has rich implications for the geography of economic activity, far beyond what we can explore in this paper. We think that our framework can serve as a starting point.

References

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

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

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

Benigno, P. and Woodford, M. (2003). Optimal monetary and fiscal policy: A linear-quadratic approach. *NBER Macroeconomics Annual*, 18:271–333.

Benigno, P. and Woodford, M. (2012). Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory*, 147(1):1–42.

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

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

Borusyak, K., Hull, P., and Jaravel, X. (2025). A practical guide to shift-share instruments. *Journal of Economic Perspectives*, 39(1):181–204.

Bresnahan, T. F. and Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, 99(5):977–1009.

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

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

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

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

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

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. Elsevier.

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

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

Felix, M. (2024). Trade, labor market concentration, and wages. Technical report, Working paper.

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

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

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

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

Hornbeck, R. and Moretti, E. (2024). Estimating who benefits from productivity growth: local and distant effects of city productivity growth on wages, rents, and inequality. *Review of Economics and Statistics*, 106(3):587–607.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Kline, P. and Moretti, E. (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.

Lamadon, T., Mogstad, M., and Setzler, B. (2022). Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review*, 112(1):169–212.

Mankiw, N. G. and Whinston, M. D. (1986). Free entry and social inefficiency. *The RAND Journal of Economics*, pages 48–58.

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

Miyauchi, Y. (2024). Matching and agglomeration: Theory and evidence from japanese firm-to-firm trade. *Econometrica*, 92(6):1869–1905.

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

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

Overman, H. G. and Puga, D. (2010). Labor pooling as a source of agglomeration: An empirical investigation. In *Agglomeration economics*, pages 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.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2119–2171. Elsevier.

Schoefer, B. and Ziv, O. (2024). Productivity, place, and plants. *Review of Economics and Statistics*, 106(5):1167–1186.

Appendix

A Details of the Quantitative, Granular Model of Economic Geography

This section lays out the details of the quantitative model introduced in Section 3.

A.1 Environment

The country is made up of I regions, indexed by $i \in \mathcal{I} \equiv \{1, \ldots, I\}$. There is a mass ℓ of workers and a continuum of sectors $s \in [0, 1]$. The sectors produce perfectly substitutable goods but hire in distinct sectoral labor markets.

Timing. There are four periods $t \in \{-1, 0, 1, 2\}$. In period -1, workers decide where to live and then stay there for the remaining periods, determining population ℓ_i. A mass m_i of firms pay a fixed cost of the traded final good in order to enter in period 0. Each firm is then randomly assigned a sector s and gets an ex-ante productivity draw z from some known distribution.

After observing those initial productivity draws, a representative worker freely allocates her labor L_{isn} across the sectors and firms in period 1. 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 worker can then reallocate her labor across the firms and sectors, subject to moving costs. Firms then produce and sell their goods.

Workers. The fundamental utility of living in location i, u_i is

$$u_i = \overline{u}_i c_i,$$

where $\overline{u}_i$ are the local amenities and c_i is consumption of the freely traded final good. Each worker has an idiosyncratic preference for each location ε_i, so that the utility the worker gets from living in location i is $u_i \varepsilon_i$. We assume that ε_i are distributed Fréchet with shape parameter $\theta > 0$.

Once the worker has decided on a location i, she needs to make her labor supply decisions. She is a risk-neutral representative agent, endowed with one unit of labor that she supplies to the market inelastically. In period 1, the worker freely allocates her units of labor across sectors and firms. In particular, she chooses her vector of labor supply

$\boldsymbol{L_i} \equiv \{L_{isn}\}_{s,n}$ in the set of feasible labor allocations $\mathcal{L}$,

$$\boldsymbol{L_i} \in \mathcal{L} \equiv \left\{ \boldsymbol{L'_i} | \int_0^1 \sum_{n \in \mathcal{N}_{is}} L'_{isn} ds \leq 1 \right\}.$$

In period 2, the state of the world ω is revealed. This determines the ex-post productivity shocks for all firms. The worker then reallocates her labor across firms, choosing a vector of labor supply $\boldsymbol{L_i(\omega)} \equiv \{L_{isn}(\omega)\}_{s,n}$ in the set of feasible labor allocations $\mathcal{L}_\Omega(\boldsymbol{L_i})$ which depends on the worker's labor choices in period 1. The set is given by

$$\mathcal{L}_\Omega(\boldsymbol{L_i}) \equiv \left\{ \boldsymbol{L_i}(\omega) | 1 = \left(\int_0^1 L_{is}^{-\frac{1}{\nu}} L_{is}(\omega)^{\frac{1+\nu}{\nu}} ds \right)^{\frac{\nu}{1+\nu}}, \right.$$
$$\left. L_{is}(\omega) = \left(\sum_{n \in \mathcal{N}_{is}} \left(\frac{L_{isn}}{L_{is}} \right)^{-\frac{1}{\kappa}} L_{isn}(\omega)^{\frac{1+\kappa}{\kappa}} \right)^{\frac{\kappa}{1+\kappa}} \right\},$$

where $L_{is} = \sum_{n \in \mathcal{N}_{is}} L_{isn}$.

Firms. Firms are the same as in the main text, so we do not restate the assumptions here.

Market Clearing. Total expected production in location i is still

$$Y_i = \mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_{is}} z_{isn} a_{isn}(\omega) \ell_{isn}(\omega)^{1-\eta} ds \right].$$

The final goods are freely traded, so the market-clearing condition holds at the national level,

$$\sum_{i \in \mathcal{I}} c_i \ell_i + \psi_i m_i = \sum_{i \in \mathcal{I}} Y_i. \tag{24}$$

In the labor market, labor demand needs to equal the individual labor supplied by each worker multiplied by the number of workers. However, this needs to hold for each individual firm as labor is imperfectly substitutable across firms,

$$\ell_{isn}(\omega) = L_{isn}(\omega) \ell_i, \quad \forall s, n, \omega. \tag{25}$$

A.2 Market Structure and Equilibrium

Labor Supply Decision. We will characterize the worker's decision using backward induction, starting with the labor supply decision in periods 1 and 2, and then characterizing the migration decision in period -1 in the next section.

Conditional on living in location i, workers choose their labor allocation across firms and sectors in periods 1 and 2 to maximize their expected utility, taking wages as given. We normalize the price of the final goods to 1, so workers solve the problem.

$$\boldsymbol{L_i}, \boldsymbol{L_i}(\boldsymbol{\omega}) \in \underset{\boldsymbol{L}_i' \in \mathcal{L}, \boldsymbol{L}_i'(\boldsymbol{\omega}) \in \mathcal{L}_\Omega(\boldsymbol{L}_i')}{\operatorname{argmax}} \mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_{is}} w_{isn}(\omega) L'_{isn}(\omega) ds\right], \tag{26}$$

where $w_{isn}(\omega)$ is the equilibrium wage for firm n in sector s in state of the world ω. We will denote the maximum of (26) by w_i.

Writing out the maximization problem, we get

$$\max_{\boldsymbol{L}_i', \boldsymbol{L}_i'(\boldsymbol{\omega})} \mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_{is}} w_{isn}(\omega) L'_{isn}(\omega) ds\right]$$

such that

$$L_{is} = \sum_{n \in \mathcal{N}_{is}} L_{isn} \tag{λ_{is}}$$

$$1 = \int_0^1 L_{is} ds \tag{λ_i}$$

$$L_{is}(\omega)^{\frac{1+\kappa}{\kappa}} = \sum_{n \in \mathcal{N}_{is}} \left(\frac{L_{isn}}{L_{is}}\right)^{-\frac{1}{\kappa}} L_{isn}(\omega)^{\frac{1+\kappa}{\kappa}} \tag{$\lambda_{is}(\omega)$}$$

$$1 = \int_0^1 \left(L_{is}\right)^{-\frac{1}{\nu}} L_{is}(\omega)^{\frac{1+\nu}{\nu}} ds. \tag{$\lambda_i(\omega)$}$$

Taking the first order conditions, we find that

$$\begin{aligned}
w_{isn}(\omega) &= \lambda_{is}(\omega) \left(\frac{L_{isn}}{L_{is}}\right)^{-\frac{1}{\kappa}} \frac{1+\kappa}{\kappa} L_{isn}(\omega)^{\frac{1}{\kappa}} \\
\lambda_{is}(\omega) &= \lambda_i(\omega) \left(L_{is}\right)^{-\frac{1}{\nu}} \frac{\kappa}{1+\kappa} \frac{1+\nu}{\nu} L_{is}(\omega)^{\frac{1}{\nu}-\frac{1}{\kappa}} \\
\lambda_{is} &= \mathbb{E}\left[\lambda_{is}(\omega) \frac{1}{\kappa} \left(\frac{L_{isn}}{L_{is}}\right)^{-\frac{1}{\kappa}} L_{isn}^{-1} L_{isn}(\omega)^{\frac{1+\kappa}{\kappa}}\right] \\
\lambda_i &= \lambda_{is} + \mathbb{E}\left[\lambda_i(\omega) \frac{1}{\nu} \left(L_{is}\right)^{-\frac{1}{\nu}} L_{is}^{-1} L_{is}(\omega)^{\frac{1+\nu}{\nu}}\right] - \frac{1}{\kappa} L_{is}^{-1} \mathbb{E}\left[\lambda_{is}(\omega) L_{is}(\omega)^{\frac{1+\kappa}{\kappa}}\right].
\end{aligned} \tag{27}$$

One can rewrite the short-run labor supply decision in the more familiar form

$$\frac{L_{isn}(\omega)}{L_{isn}} = \left(\frac{w_{isn}(\omega)}{w_{is}(\omega)}\right)^{\kappa} \frac{L_{is}(\omega)}{L_{is}}, \quad \frac{L_{is}(\omega)}{L_{is}} = \left(\frac{w_{is}(\omega)}{w_{i}(\omega)}\right)^{\nu},$$

where

$$w_{is}(\omega) \equiv \left(\sum_{n\in\mathcal{N}_{is}} \frac{L_{isn}}{L_{is}} w_{isn}(\omega)^{1+\kappa}\right)^{\frac{1}{1+\kappa}}, \quad w_i(\omega) \equiv \left(\int_0^1 L_{is} w_{is}(\omega)^{1+\nu} ds\right)^{\frac{1}{1+\nu}}.$$

Transforming the per capita labor supply, $L_{isn}(\omega)$, to the total labor supplied $\ell_{isn}(\omega)$ and taking logs implies,

$$\log \ell_{isn}(\omega) - \log \ell_{isn} = \kappa \log w_{isn}(\omega) - (\nu - \kappa) \log w_{is}(\omega) - \nu \log w_i(\omega). \tag{28}$$

Including firm-year specific amenities implies the equation (14).

Migration Decision. Each worker chooses the location that maximizes their utility. Therefore, the population in location i satisfies

$$\ell_i = \int_{\mathbb{R}^I} \mathbb{1}_{i\in \text{argmax}_{i'} \bar{u}_{i'} w_{i'}} dG(\varepsilon) \cdot \ell, \tag{29}$$

where G is the joint distribution of ε. This is a standard problem in the literature with a Fréchet distribution. It implies that $\ell_i = (u_i/u)^{\theta} \cdot \ell$, where $u = (\sum_i (u_i)^{\theta})^{\frac{1}{\theta}}$.

Labor Demand - Competitive. We will consider three different conduct assumptions on firms after they enter. The first assumption is that they are competitive. Then each active firm maximizes profits, taking wages and prices as given,

$$\ell_{isn}(\omega) \in \underset{\ell'}{\text{argmax}} \quad z_{isn} a_{isn}(\omega) (\ell')^{1-\eta} - w_{isn}(\omega) \ell'. \tag{30}$$

This implies that

$$z_{isn} a_{isn}(\omega) \ell_{isn}(\omega)^{-\eta} = w_{isn}(\omega). \tag{31}$$

Labor Demand - Cournot. Under Cournot competition, the firm takes as given the labor decisions of the other firms in its own sector. We assume that the firm then takes as given the workers' other options in other sectors. In the math, that will imply that the firm will take $\lambda_i(\omega)$ and λ_i in equation (27) as given. Combining some of the first-order necessary conditions of the worker's problem, we can write the firm problem, using $\boldsymbol{x}_{isn} \equiv \{w_{isn}(\omega), \ell_{isn}(\omega), \ell_{is}(\omega), \ell_{isn}, \ell_{is}\}$, as,

$$\begin{aligned}
\boldsymbol{x}_{isn} \in & \operatorname*{argmax}_{\boldsymbol{x}'_{isn}} \quad \mathbb{E}\left[z_{isn}a_{isn}(\omega)\ell'_{isn}(\omega)^{1-\eta} - w'_{isn}(\omega)\ell'_{isn}(\omega)\right] \\
& \text{s.t.} \quad (\ell'_{is})^{-\frac{1}{\kappa}}\ell'_{is}(\omega)^{\frac{1+\kappa}{\kappa}} = \left(\ell'_{isn}\right)^{-\frac{1}{\kappa}}\ell'_{isn}(\omega)^{\frac{1+\kappa}{\kappa}} + \sum_{n'\neq n}\left(\ell_{isn'}\right)^{-\frac{1}{\kappa}}\ell_{isn'}(\omega)^{\frac{1+\kappa}{\kappa}} \\
& \qquad \ell'_{is} = \ell'_{isn} + \sum_{n'\neq n}\ell_{isn'} \\
& \qquad w'_{isn}(\omega) = \lambda_i(\omega)\frac{1+\nu}{\nu}\left(\frac{\ell_{is}(\omega)}{\ell_{is}}\right)^{\frac{1}{\nu}-\frac{1}{\kappa}}\left(\frac{\ell_{isn}(\omega)}{\ell_{isn}}\right)^{\frac{1}{\kappa}} \\
& \qquad \lambda_i(1+\kappa)\nu = \mathbb{E}\left[\lambda_i(\omega)(1+\nu)\left(\frac{\ell_{is}(\omega)}{\ell_{is}}\right)^{\frac{1}{\nu}-\frac{1}{\kappa}}\left(\frac{\ell_{isn}(\omega)}{\ell_{isn}}\right)^{\frac{1+\kappa}{\kappa}}\right] \\
& \qquad\qquad + \mathbb{E}\left[\lambda_i(\omega)(\kappa-\nu)\left(\frac{\ell_{is}(\omega)}{\ell_{is}}\right)^{\frac{1+\nu}{\nu}}\right],
\end{aligned} \tag{32}$$

where we have transformed the per capita variables into the total amount of labor.

Taking the first order conditions and simplifying gives the following necessary conditions,

$$\begin{aligned}
0 =& \ \mathbb{E}[w_{isn}(\omega)\ell_{isn}(\omega)] - \ell_{isn}\mathbb{E}\left[\Lambda_{isn}(\omega)\left(\left(\frac{\ell_{isn}(\omega)}{\ell_{isn}}\right)^{\frac{1+\kappa}{\kappa}} - \left(\frac{\ell_{is}(\omega)}{\ell_{is}}\right)^{\frac{1+\kappa}{\kappa}}\right)\right] \\
& - (1-\eta)\mathbb{E}\left[z_{isn}a_{isn}(\omega)\ell_{isn}(\omega)^{1-\eta}\right] \\
0 =& \ \frac{1+\kappa}{\kappa}w_{isn}(\omega)\ell_{isn}(\omega) - \frac{1+\kappa}{\kappa}\Lambda_{isn}(\omega)\ell_{isn}(\omega)\left(\frac{\ell_{isn}(\omega)}{\ell_{isn}}\right)^{\frac{1}{\kappa}} \\
& - \frac{\Lambda^w_{isn}}{\lambda_i(\omega)}\frac{1+\kappa}{\kappa}\nu w_{isn}(\omega)\frac{\ell_{isn}(\omega)}{\ell_{isn}} - (1-\eta)z_{isn}a_{isn}(\omega)\ell_{isn}(\omega)^{1-\eta} \\
0 =& -\left(\frac{1}{\nu}-\frac{1}{\kappa}\right)\left[w_{isn}(\omega)\ell_{isn}(\omega) - \frac{\Lambda^w_{isn}}{\lambda_i(\omega)}\nu w_{isn}(\omega)\frac{\ell_{isn}(\omega)}{\ell_{isn}}\right. \\
& \left. - \kappa(1+\nu)\Lambda^w_{isn}\left(\frac{\ell_{is}(\omega)}{\ell_{is}}\right)^{\frac{1+\nu}{\nu}}\right] - \Lambda_{isn}(\omega)\ell_{is}\frac{1+\kappa}{\kappa}\left(\frac{\ell_{is}(\omega)}{\ell_{is}}\right)^{\frac{1+\kappa}{\kappa}}
\end{aligned} \tag{33}$$

Labor Demand - Bertrand. Under Bertrand competition, the firm takes as given the wage decisions of the other firms in its own sector. We assume that the firm then takes as given the workers' other options in other sectors. Just as for Cournot, that implies that the firm takes $\lambda_i(\omega)$ and λ_i in equation (27) as given. Combining some of the first order necessary conditions of the worker's problem, we can write the firm problem, using $x_{isn} \equiv \{w_{isn}(\omega), w_{is}(\omega), \ell_{isn}(\omega), \ell_{isn}, \ell_{is}\}$, as,

$$
\begin{aligned}
\boldsymbol{x}_{isn} \in & \underset{\boldsymbol{x}'_{isn}}{\operatorname{argmax}} \quad \mathbb{E}\left[z_{isn}a_{isn}(\omega)\ell'_{isn}(\omega)^{1-\eta} - w'_{isn}(\omega)\ell'_{isn}(\omega)\right] \\
& \text{s.t.} \quad w'_{is}(\omega)^{1+\kappa} = \frac{\ell'_{isn}}{\ell'_{is}}w'_{isn}(\omega)^{1+\kappa} + \sum_{n'\neq n}\frac{\ell'_{isn'}}{\ell'_{is}}w'_{isn'}(\omega)^{1+\kappa} \\
& \qquad \ell'_{is} = \ell'_{isn} + \sum_{n'\neq n}\ell_{isn'} \\
& \qquad \ell_{isn}(\omega)' = \lambda_i(\omega)^{-\nu}\left(\frac{\nu}{1+\nu}\right)\frac{\ell'_{isn}}{\ell'_i}w'_{is}(\omega)^{\nu-\kappa}w'_{isn}(\omega)^{\kappa} \\
& \qquad \lambda_i = \lambda_i(\omega)^{-\nu}\frac{1}{1+\kappa}\left(\frac{\nu}{1+\nu}\right)^{\nu}\ell_i^{-1}\mathbb{E}\left[w'_{is}(\omega)^{\nu\kappa\left(\frac{1}{\kappa}-\frac{1}{\nu}\right)}w'_{isn}(\omega)^{1+\kappa}\right] \\
& \qquad\qquad + \lambda_i(\omega)^{-\nu}\frac{\kappa}{1+\kappa}\left(\frac{1}{\nu}-\frac{1}{\kappa}\right)\left(\frac{\nu}{1+\nu}\right)^{1+\nu}\ell_i^{-1}\mathbb{E}\left[w'_{is}(\omega)^{1+\nu}\right].
\end{aligned}
\tag{34}
$$

where we have transformed the per capita variables into the total amount of labor.

Taking the first order conditions and simplifying gives the following necessary conditions,

$$
\begin{aligned}
0 = & \ \mathbb{E}\left[w_{isn}(\omega)\ell_{isn}(\omega)\right] - \ell_{isn}\mathbb{E}\left[\Lambda_{isn}(\omega)\left(w_{isn}(\omega)^{1+\kappa} - w_{is}(\omega)^{1+\kappa}\right)\right] \\
& - (1-\eta)\mathbb{E}\left[z_{isn}a_{isn}(\omega)\ell_{isn}(\omega)^{1-\eta}\right] \\
0 = & \ \frac{1+\kappa}{\kappa}w_{isn}(\omega)\ell_{isn}(\omega) - \frac{1+\kappa}{\kappa}\Lambda_{isn}(\omega)\ell_{isn}w_{isn}(\omega)^{1+\kappa} \\
& - \frac{\Lambda^{w}_{isn}}{\kappa}\lambda_i(\omega)^{-\nu}\left(\frac{\nu}{1+\nu}\right)^{\nu}\ell_i^{-1}w_{is}(\omega)^{\nu-\kappa}w_{isn}(\omega)^{1+\kappa} \\
& - (1-\eta)z_{isn}a_{isn}(\omega)\ell_{isn}(\omega)^{1-\eta} \\
0 = & -\frac{\kappa-\nu}{\kappa}\left[w_{isn}(\omega)\ell_{isn}(\omega) - \frac{1}{1+\kappa}\Lambda^{w}_{isn}\lambda_i(\omega)^{-\nu}\left(\frac{\nu}{1+\nu}\right)^{\nu}\ell_i^{-1}w_{is}(\omega)^{\nu-\kappa}w_{isn}(\omega)^{1+\kappa}\right. \\
& \left. - \Lambda^{w}_{isn}\lambda(\omega)^{-\nu}\frac{\kappa}{1+\kappa}\left(\frac{\nu}{1+\nu}\right)^{\nu}\ell_i^{-1}w_{is}(\omega)^{1+\nu}\right] \\
& - \frac{1+\kappa}{\kappa}\Lambda_{isn}(\omega)\left(\kappa\ell_{is}w_{is}(\omega)^{1+\kappa} + (\nu-\kappa)\ell_{isn}w_{isn}(\omega)^{1+\kappa}\right)
\end{aligned}
\tag{35}
$$

Entry. Entry is the same as in the baseline model. Thus, free entry implies that average profits are equal to the cost of entry, (5).

A.3 Characterizing Cournot Competition

In this subsection, we characterize the equilibrium under Cournot Competition. The expression for production given in Lemma 8 holds no matter how firms behave. Therefore, we go through and characterize what total wage payments are, which then implies profits.

Taking a second-order approximation to total wage payments and the firm FOCs associated with Cournot competition implies the next proposition.

Lemma 4. *If firms compete à la Cournot, total wage compensation in location i can be written,*

$$w_i \ell_i = (1-\eta) z_i (m_i)^{\eta} (\ell_i)^{1-\eta} \left(\tilde{\Phi}^c(m) + \Psi^c(m) \right),$$

where $\tilde{\Phi}^c(m)$ is defined as in Lemma 8 and

$$\Psi^c(m) \equiv \frac{1+\kappa}{\kappa} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \mathbb{E}\left[\frac{\hat{\Lambda}_{isn}(\omega)}{w_i} \left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn} - \hat{\ell}_{is}(\omega) + \hat{\ell}_{is} \right) \right] ds.$$

Proof. First, we note that because workers are freely mobile across all sectors in period 1, the equilibrium with no ex-post shocks and Cournot competition is the same as the equilibrium with no ex-post shocks and competitive firms. Doing a second-order approximation around the point with no ex-post shocks implies

$$\begin{aligned} &\mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_{is}} w_{isn}(\omega) \ell_{isn}(\omega) ds \right] \\ &\quad = \overline{w}_i \overline{\ell}_i \Bigg(\int_0^1 \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \left(1 + \mathbb{E}[\hat{w}_{isn}(\omega)] + \mathbb{E}[\hat{\ell}_{isn}(\omega)] + \frac{1}{2}\mathbb{E}[(\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega))^2] \right) ds \Bigg) \end{aligned}$$

We then take a second-order log approximation to the firm FOCs in equation (33). However, with no shocks, $\Lambda_{isn}(\omega) = 0$, so we leave that in levels. The first equation becomes

$$\begin{aligned} \mathbb{E}[\hat{a}_{isn}(\omega)] + (1-\eta)\mathbb{E}[\hat{\ell}_{isn}(\omega)] + \frac{1}{2}\mathbb{E}\left[(\hat{a}_{isn}(\omega) + (1-\eta)\hat{\ell}_{isn}(\omega))^2 \right] \\ = \mathbb{E}[\hat{w}_{isn}(\omega)] + \mathbb{E}[\hat{\ell}_{isn}(\omega)] + \frac{1}{2}\mathbb{E}\left[\left(\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega) \right)^2 \right] \\ - \frac{1+\kappa}{\kappa}\mathbb{E}\left[\frac{\hat{\Lambda}_{isn}(\omega)}{w_i} \left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn} - \hat{\ell}_{is}(\omega) + \hat{\ell}_{is} \right) \right]. \end{aligned}$$

We also need the second-order approximation to the labor constraints embedded in $\mathcal{L}$ and $\mathcal{L}_{\Omega}(\cdot)$,

$$\begin{aligned}
\hat{\ell}_{is}(\omega) &+ \frac{1}{2}\frac{\kappa}{1+\kappa}\left(-\frac{1}{\kappa}\hat{\ell}_{is}+\frac{1+\kappa}{\kappa}\hat{\ell}_{is}(\omega)\right)^2+\frac{1}{1+\kappa}\hat{\ell}_{is}^2 \\
&= \sum_{n\in\mathcal{N}_{is}}\frac{\ell_{isn}}{\overline{\ell}_{is}}\left(\hat{\ell}_{isn}(\omega)+\frac{1}{2}\frac{\kappa}{1+\kappa}\left(-\frac{1}{\kappa}\hat{\ell}_{isn}+\frac{1+\kappa}{\kappa}\hat{\ell}_{isn}(\omega)\right)^2+\frac{1}{1+\kappa}\hat{\ell}_{isn}^2\right) \\
0 &= \int_0^1\overline{\ell}_{is}\left(\hat{\ell}_{is}(\omega)+\frac{1}{2}\frac{\nu}{1+\nu}\left(-\frac{1}{\nu}\hat{\ell}_{is}+\frac{1+\nu}{\nu}\hat{\ell}_{is}(\omega)\right)^2+\frac{1}{1+\nu}\hat{\ell}_{is}^2\right)ds.
\end{aligned}$$

So then we can write

$$\begin{aligned}
&\int_0^1\frac{\overline{\ell}_{is}}{\overline{\ell}_i}\sum_{n\in\mathcal{N}_{is}}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\mathbb{E}\left[\hat{w}_{isn}(\omega)\right]+\mathbb{E}\left[\hat{\ell}_{isn}(\omega)\right]+\frac{1}{2}\mathbb{E}[(\hat{w}_{isn}(\omega)+\hat{\ell}_{isn}(\omega))^2]\right)ds \\
&= \frac{1}{2}\int_0^1\frac{\overline{\ell}_{is}}{\overline{\ell}_i}\sum_{n\in\mathcal{N}_{is}}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[(\hat{a}_{isn}(\omega)+(1-\eta)\hat{\ell}_{isn}(\omega))^2\right]ds \\
&\quad+\frac{1+\kappa}{\kappa}\int_0^1\frac{\overline{\ell}_{is}}{\overline{\ell}_i}\sum_{n\in\mathcal{N}_{is}}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[\frac{\hat{\Lambda}_{isn}(\omega)}{w_i}\left(\hat{\ell}_{isn}(\omega)-\hat{\ell}_{isn}-\hat{\ell}_{is}(\omega)+\hat{\ell}_{is}\right)\right]ds \\
&\quad+\int_0^1\frac{\overline{\ell}_{is}}{\overline{\ell}_i}\sum_{n\in\mathcal{N}_{is}}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\mathbb{E}[\hat{a}_{isn}(\omega)]+(1-\eta)\mathbb{E}[\hat{\ell}_{isn}(\omega)]\right)ds \\
&= \mathbb{E}[\hat{a}_{isn}(\omega)]+\frac{1}{2}\mathbb{E}[\hat{a}_{isn}(\omega)^2]+(1-\eta)\int_0^1\frac{\overline{\ell}_{is}}{\overline{\ell}_i}\sum_{n\in\mathcal{N}_{is}}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}[\hat{a}_{isn}(\omega)\hat{\ell}_{isn}(\omega)]ds \\
&\quad-\eta\frac{1-\eta}{2}\int_0^1\frac{\overline{\ell}_{is}}{\ell}\sum_{n\in\mathcal{N}_{is}}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}[\hat{\ell}_{isn}(\omega)^2]ds \\
&\quad-\frac{1}{\kappa}\frac{1-\eta}{2}\int_0^1\frac{\overline{\ell}_s}{\ell}\sum_{n\in\mathcal{N}_s}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[\left(\hat{\ell}_{sn}(\omega)-\hat{\ell}_{isn}\right)^2\right]ds \\
&\quad-\left(\frac{1}{\nu}-\frac{1}{\kappa}\right)\frac{1-\eta}{2}\int_0^1\frac{\overline{\ell}_{is}}{\ell}\mathbb{E}\left[\left(\hat{\ell}_{is}(\omega)-\hat{\ell}_{is}\right)^2\right]ds \\
&\quad+\frac{1+\kappa}{\kappa}\int_0^1\frac{\overline{\ell}_{is}}{\overline{\ell}_i}\sum_{n\in\mathcal{N}_{is}}\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[\frac{\hat{\Lambda}_{isn}(\omega)}{w_i}\left(\hat{\ell}_{isn}(\omega)-\hat{\ell}_{isn}-\hat{\ell}_{is}(\omega)+\hat{\ell}_{is}\right)\right]ds.
\end{aligned}$$

This completes the proof. □

Lemma 4 gives wages completely in second-order terms. Therefore, following Benigno and Woodford (2003) and Benigno and Woodford (2012), we can then approximate the labor constraints and firm first-order conditions with a log first-order approximation.

We then computationally evaluate $\tilde{\Phi}^c(m)$ and $\Psi^c(m)$ and use those functions in calculating the equilibrium.

A.4 Characterizing Bertrand Competition

In this subsection, we characterize the equilibrium under Bertrand Competition. The expression for production given in Lemma 8 holds no matter how firms behave. Therefore, we go through and characterize what total wage payments are, which then implies profits. We also show that in the limit as $\kappa \to \infty$, Bertrand competition converges to the perfect competition case. Thus, the original model in Section 2 is equivalent to a model with Bertrand competition.

Taking a second-order approximation to total wage payments and the firm FOCs associated with Cournot competition implies the next proposition.

Lemma 5. *If firms compete à la Bertrand, total wage compensation in location i can be written,*

$$w_i \ell_i = (1-\eta) z_i(m_i)^{\eta} (\ell_i)^{1-\eta} \left(\tilde{\Phi}^b(m) + \Psi^b(m) \right),$$

where $\tilde{\Phi}(m)$ is defined as in Lemma 8 and

$$\Psi^b(m) \equiv -(1+\kappa) w_i^{\kappa} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \mathbb{E}\left[\hat{\Lambda}_{isn}(\omega) \left(\hat{w}_{isn}(\omega) - \hat{w}_{is}(\omega) \right) \right] ds.$$

Proof. First, we note that because workers are freely mobile across all sectors in period 1, the equilibrium with no ex-post shocks and Bertrand competition is the same as the equilibrium with no ex-post shocks and competitive firms. Doing a second-order approximation around the point with no ex-post shocks implies,

$$\begin{aligned}
&\mathbb{E}\left[\int_0^1 \sum_{n \in \mathcal{N}_{is}} w_{isn}(\omega) \ell_{isn}(\omega) ds \right] \\
&= \overline{w}_i \overline{\ell}_i \left(\int_0^1 \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \left(1 + \mathbb{E}[\hat{w}_{isn}(\omega)] + \mathbb{E}[\hat{\ell}_{isn}(\omega)] + \frac{1}{2} \mathbb{E}[(\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega))^2] \right) ds \right)
\end{aligned}$$

We then take a second-order log approximation to the firm FOCs in equation (35). However, with no shocks, $\Lambda_{isn}(\omega) = 0$, so we leave that in levels. The first equation becomes

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

$$
\begin{aligned}
&= \mathbb{E}[\hat{w}_{isn}(\omega)] + \mathbb{E}[\hat{\ell}_{isn}(\omega)] + \frac{1}{2}\mathbb{E}\left[\left(\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega)\right)^2\right] \\
&\quad + (1+\kappa) w_i^{\kappa}\, \mathbb{E}\left[\hat{\Lambda}_{isn}(\omega)\left(\hat{w}_{isn}(\omega) - \hat{w}_{is}(\omega)\right)\right].
\end{aligned}
$$

We also need the second-order approximation to the labor constraints embedded in $\mathcal{L}$ and $\mathcal{L}_{\Omega}(\cdot)$,

$$
\begin{aligned}
&\hat{\ell}_{is}(\omega) + \frac{1}{2}\frac{\kappa}{1+\kappa}\left(-\frac{1}{\kappa}\hat{\ell}_{is} + \frac{1+\kappa}{\kappa}\hat{\ell}_{is}(\omega)\right)^2 + \frac{1}{1+\kappa}\hat{\ell}_{is}^2 \\
&\quad = \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\hat{\ell}_{isn}(\omega) + \frac{1}{2}\frac{\kappa}{1+\kappa}\left(-\frac{1}{\kappa}\hat{\ell}_{isn} + \frac{1+\kappa}{\kappa}\hat{\ell}_{isn}(\omega)\right)^2 + \frac{1}{1+\kappa}\hat{\ell}_{isn}^2\right) \\
&\quad 0 = \int_0^1 \overline{\ell}_{is}\left(\hat{\ell}_{is}(\omega) + \frac{1}{2}\frac{\nu}{1+\nu}\left(-\frac{1}{\nu}\hat{\ell}_{is} + \frac{1+\nu}{\nu}\hat{\ell}_{is}(\omega)\right)^2 + \frac{1}{1+\nu}\hat{\ell}_{is}^2\right) ds.
\end{aligned}
$$

So then we can write

$$
\begin{aligned}
&\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\mathbb{E}\left[\hat{w}_{isn}(\omega)\right] + \mathbb{E}\left[\hat{\ell}_{isn}(\omega)\right] + \frac{1}{2}\mathbb{E}[(\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega))^2]\right) ds \\
&\quad = \frac{1}{2}\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[(\hat{a}_{isn}(\omega) + (1-\eta)\hat{\ell}_{isn}(\omega))^2\right] ds \\
&\qquad - (1+\kappa) w_i^{\kappa} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[\hat{\Lambda}_{isn}(\omega)\left(\hat{w}_{isn}(\omega) - \hat{w}_{is}(\omega)\right)\right] ds \\
&\qquad + \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\mathbb{E}[\hat{a}_{isn}(\omega)] + (1-\eta)\mathbb{E}[\hat{\ell}_{isn}(\omega)]\right) ds \\
&\quad = \mathbb{E}[\hat{a}_{isn}(\omega)] + \frac{1}{2}\mathbb{E}[\hat{a}_{isn}(\omega)^2] + (1-\eta)\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}[\hat{a}_{isn}(\omega)\hat{\ell}_{isn}(\omega)] ds \\
&\qquad - \eta\frac{1-\eta}{2}\int_0^1 \frac{\overline{\ell}_{is}}{\ell} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}[\hat{\ell}_{isn}(\omega)^2] ds \\
&\qquad - \frac{1}{\kappa}\frac{1-\eta}{2}\int_0^1 \frac{\overline{\ell}_{s}}{\ell} \sum_{n\in\mathcal{N}_{s}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[\left(\hat{\ell}_{sn}(\omega) - \hat{\ell}_{isn}\right)^2\right] ds \\
&\qquad - \left(\frac{1}{\nu} - \frac{1}{\kappa}\right)\frac{1-\eta}{2}\int_0^1 \frac{\overline{\ell}_{is}}{\ell}\mathbb{E}\left[\left(\hat{\ell}_{is}(\omega) - \hat{\ell}_{is}\right)^2\right] ds \\
&\qquad - (1+\kappa)\overline{w}_i^{\kappa}\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\mathbb{E}\left[\hat{\Lambda}_{isn}(\omega)\left(\hat{w}_{isn}(\omega) - \hat{w}_{is}(\omega)\right)\right] ds.
\end{aligned}
$$

This completes the proof. □

Lemma 5 gives wages completely in second-order terms. Therefore, following Benigno and Woodford (2003) and Benigno and Woodford (2012), we can then approximate the labor constraints and firm first-order conditions with a log first-order approximation. We then computationally evaluate $\tilde{\Phi}^b(m)$ and $\Psi^b(m)$ and use those functions in calculating the equilibrium.

Next, we show the equivalence to the competitive equilibrium as $\kappa \to \infty$.

Lemma 6. *In the limit as $\kappa \to \infty$, labor and profits are the same in the Bertrand case and competitive equilibrium case.*

Proof. The system is characterized by the constraints in the firm maximization problem (34) and the FOCs (35). We start by solving the system in the absence of shocks and then doing a first order approximation around that point.

In the absence of shocks, wages are equalized across all firms. We will denote this wage by $\overline{w}_i$. Solving the system then implies that

$$\begin{aligned}
\overline{\lambda}_i(\omega)^{-\nu} &= \frac{1+\nu}{\nu}\ell_i \overline{w}_i^{-\nu} & \overline{w}_i &= (1-\eta) z_{isn} \overline{\ell}_{isn}^{-\eta} \\
\overline{\Lambda}_{isn}(\omega) &= 0 & \overline{\lambda}_i &= \frac{1}{\nu}\left(\frac{\nu}{1+\nu}\right)^{\nu} \overline{w}_i
\end{aligned}$$

$$\overline{\ell}_{isn} = \overline{\Lambda}^{w}_{isn}\left(\frac{\nu}{1+\nu}\right)^{\nu-1}.$$

Taking the log first order approximation implies the system of equations,

$$\begin{aligned}
\hat{\ell}_{is} + (1+\kappa)\hat{w}_{is}(\omega) &= \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\hat{\ell}_{isn} + (1+\kappa)\hat{w}_{isn}(\omega)\right) \\
\hat{\ell}_{isn}(\omega) &= -\nu\hat{\lambda}_i(\omega) + \hat{\ell}_{isn} + (\nu-\kappa)\hat{w}_{is}(\omega) + \kappa\hat{w}_{isn}(\omega) \\
\frac{1}{\nu}\left(\frac{\nu}{1+\nu}\right)^{\nu}\hat{\lambda}_i &= \frac{1}{1+\kappa}\mathbb{E}\left[-\nu\hat{\lambda}_i(\omega) + (\nu-\kappa)\hat{w}_{is}(\omega) + (1+\kappa)\hat{w}_{isn}(\omega)\right] \\
&\quad + \frac{\kappa}{1+\kappa}\left(\frac{1}{\nu} - \frac{1}{\kappa}\right)\left(\frac{\nu}{1+\nu}\right)^{\nu}\mathbb{E}[-\nu\hat{\lambda}_i(\omega) + (1+\nu)\hat{w}_{is}(\omega)] \\
\mathbb{E}[\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega)] &= \mathbb{E}[\hat{a}_{isn}(\omega) + (1-\eta)\hat{\ell}_{isn}(\omega)] \\
\frac{1+\kappa}{\kappa}\left(\hat{w}_{isn}(\omega) + \hat{\ell}_{isn}(\omega)\right) &= \frac{1+\kappa}{\kappa}\overline{w}_i^{\kappa}\hat{\Lambda}_{isn}(\omega) \\
&\quad + \frac{1}{\kappa}\left(\hat{\Lambda}^{w}_{isn} - \nu\hat{\lambda}_i(\omega) + (\nu-\kappa)\hat{w}_{is}(\omega) + (1+\kappa)\hat{w}_{isn}(\omega)\right) \\
&\quad + \hat{a}_{isn}(\omega) + (1-\eta)\hat{\ell}_{isn}(\omega) \\
0 &= -\frac{1+\kappa}{\kappa}\left(\kappa\overline{\ell}_{is} - (\nu-\kappa)\overline{\ell}_{isn}\right)\overline{w}_i^{\kappa}\hat{\Lambda}_{isn}(\omega)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{\kappa-\nu}{\kappa}\bar{\ell}_{isn}\left(\hat{w}_{isn}(\omega)+\hat{\ell}_{isn}(\omega)\right)\\
&+\frac{\kappa-\nu}{\kappa}\frac{1}{1+\kappa}\bar{\ell}_{isn}\left(\hat{\Lambda}^{w}_{isn}-\nu\hat{\lambda}_i(\omega)+(\nu-\kappa)\hat{w}_{is}(\omega)+(1+\kappa)\hat{w}_{isn}(\omega)\right)\\
&+\frac{\kappa-\nu}{\nu}\frac{\kappa}{1+\kappa}\bar{\ell}_{isn}\left(\hat{\Lambda}^{w}_{isn}-\nu\hat{\lambda}_i(\omega)+(1+\nu)\hat{w}_{is}(\omega)\right).
\end{aligned}
$$

We also have a first order approximation to the labor market clearing conditions,

$$
\begin{aligned}
\hat{\ell}_{is}(\omega) &= \sum_{n\in\mathcal{N}_{is}}\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}(\omega), & 0 &= \int_0^1\frac{\bar{\ell}_{is}}{\ell_i}\hat{\ell}_{is}(\omega)ds,\\
\hat{\ell}_{is} &= \sum_{n\in\mathcal{N}_{is}}\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}, & 0 &= \int_0^1\frac{\bar{\ell}_{is}}{\ell_i}\hat{\ell}_{is}ds.
\end{aligned}
$$

Note that $\mathbb{E}[\hat{a}_{isn}(\omega)] = 0$, so $\mathbb{E}[\hat{w}_{isn}(\omega)] = -\eta\mathbb{E}[\hat{\ell}_{isn}(\omega)]$. Taking the second equation, summing across firms within a sector gives

$$
\sum_n\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}(\omega) = -\nu\hat{\lambda}_i(\omega)+\hat{\ell}_{is}+\nu\hat{w}_{is}(\omega).
$$

Aggregating across sectors, and using the fact that $0=\int_0^1\bar{\ell}_{is}\hat{\ell}_{is}ds$,

$$
\int_0^1\frac{\bar{\ell}_{is}}{\ell_i}\sum_n\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}(\omega)ds = -\nu\hat{\lambda}_i(\omega)+\nu\int_0^1\frac{\bar{\ell}_{is}}{\ell_i}\hat{w}_{is}(\omega)ds
$$

Then, $\hat{\lambda}_i(\omega)$ is constant. Taking expectations, and using the fact that $\mathbb{E}[\hat{w}_{isn}(\omega)] = -\eta\mathbb{E}[\hat{\ell}_{isn}(\omega)]$,

$$
0 = (1+\nu\eta)\int_0^1\frac{\bar{\ell}_{is}}{\ell_i}\sum_n\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\mathbb{E}[\hat{\ell}_{isn}(\omega)]ds = -\nu\hat{\lambda}_i(\omega).
$$

Therefore, $\mathbb{E}[\hat{\ell}_{isn}(\omega)] = \mathbb{E}[\hat{w}_{isn}(\omega)] = 0$, and we also get $\hat{\ell}_{isn} = 0$. This further implies that $\hat{\lambda}_i = 0$. Rewriting the last two equations

$$
\begin{aligned}
\hat{w}_{isn}(\omega)+\eta\hat{\ell}_{isn}(\omega) &= \frac{1+\kappa}{\kappa}\overline{w}_i^{\kappa}\hat{\Lambda}_{isn}(\omega)+\frac{1}{\kappa}\hat{\Lambda}^{w}_{isn}+\hat{a}_{isn}(\omega)\\
\frac{1+\kappa}{\kappa}\left(\kappa\bar{\ell}_{is}+(\nu-\kappa)\bar{\ell}_{isn}\right)\frac{\overline{w}_i^{\kappa}}{\bar{\ell}_{isn}}\hat{\Lambda}_{isn}(\omega) &= \frac{\kappa-\nu}{\kappa}\frac{\kappa}{1+\kappa}\left(\hat{w}_{isn}(\omega)+\hat{\ell}_{isn}(\omega)-(1+\nu)\hat{w}_{is}(\omega)\right)\\
&\quad+\frac{\kappa-\nu}{1+\kappa}\frac{1}{\kappa}\hat{\Lambda}^{w}_{isn}.
\end{aligned}
$$

Taking expectations of the first equation implies that

$$\frac{1+\kappa}{\kappa}\overline{w}_i^{\kappa}\mathbb{E}[\hat{\Lambda}_{isn}(\omega)] = -\frac{1}{\kappa}\hat{\Lambda}^{w}_{isn}.$$

Then taking the expectations of the second equation implies that $\hat{\Lambda}^{w}_{isn} = 0$. Substituting in for $\hat{\ell}_{isn}(\omega) = (\nu-\kappa)\hat{w}_{is}(\omega) + \kappa\hat{w}_{isn}(\omega)$, in the second equation,

$$\left(\kappa + (\nu-\kappa)\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)\frac{1+\kappa}{\kappa}\overline{w}_i^{\kappa}\hat{\Lambda}_{isn}(\omega) = \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}(\kappa-\nu)(\hat{w}_{isn}(\omega)-\hat{w}_{is}(\omega)).$$

Substituting this into the first equation,

$$\hat{w}_{isn}(\omega)+\eta(\nu-\kappa)\hat{w}_{is}(\omega)+\eta\kappa\hat{w}_{isn}(\omega) = \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}}(\hat{w}_{isn}(\omega)-\hat{w}_{is}(\omega))+\hat{a}_{isn}(\omega).$$

Solving for $\hat{w}_{isn}(\omega)$,

$$\hat{w}_{isn}(\omega) = \frac{\hat{a}_{isn}(\omega)}{1+\eta\kappa-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}} + (\kappa-\nu)\frac{\eta-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{1}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}}{1+\eta\kappa-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}}\hat{w}_{is}(\omega).$$

Thus, in the limit as $\kappa\to\infty$, $\hat{w}_{isn}(\omega)=\hat{w}_{is}(\omega)$. Then the sectoral wages are

$$\begin{aligned}
\hat{w}_{is}(\omega) &= \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{w}_{isn}(\omega) \\
&= \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\hat{a}_{isn}(\omega)}{1+\eta\kappa-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}} \\
&\qquad + \hat{w}_{is}(\omega)\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\eta(\kappa-\nu)-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}}{1+\eta\kappa-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}} \\
\hat{w}_{is}(\omega) &= \frac{\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\hat{a}_{isn}(\omega)}{1+\eta\kappa-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}}}{1-\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\eta(\kappa-\nu)-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}}{1+\eta\kappa-\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\overline{\ell}_{isn}/\overline{\ell}_{is}}}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \frac{\hat{a}_{isn}(\omega)}{1+\eta\kappa-\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}}}}{\sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \frac{1+\eta\kappa-\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}}-\eta(\kappa-\nu)+\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}}}{1+\eta\kappa-\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}}}} \\
&= \frac{\sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \frac{\hat{a}_{isn}(\omega)}{1+\eta\kappa-\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}}}}{\sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \frac{1+\eta\nu}{1+\eta\kappa-\frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}}}}.
\end{aligned}
$$

Taking the limit as $\kappa \to \infty$, $\hat{w}_{is}(\omega) = \frac{1}{1+\eta\nu} \sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \hat{a}_{isn}(\omega)$. Finally, this also implies that $\hat{\ell}_{isn}(\omega) = \frac{1}{\eta}\left(\hat{a}_{isn}(\omega) - \hat{w}_{is}(\omega)\right)$, so that labor and wages converge to their competitive counterparts.

Next, we show that profits of firms also converge as $\kappa \to \infty$. Recall that

$$
\begin{aligned}
\Psi^b(m) &= -(1+\kappa) w_i^{\kappa} \int_0^1 \frac{\bar{\ell}_{is}}{\ell_i} \sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \mathbb{E}\left[\hat{\Lambda}_{isn}(\omega)(\hat{w}_{isn}(\omega) - \hat{w}_{is}(\omega))\right] ds \\
&= -\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i} \sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \frac{\kappa(\kappa-\nu)\bar{\ell}_{isn}/\bar{\ell}_{is}}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}} \mathbb{E}\left[(\hat{w}_{isn}(\omega) - \hat{w}_{is}(\omega))^2\right] ds \\
&= -\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i} \sum_n \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}} \frac{\kappa(\kappa-\nu)\bar{\ell}_{isn}/\bar{\ell}_{is}}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}} \frac{\mathbb{E}\left[(\hat{a}_{isn}(\omega) - (1+\eta\nu)\hat{w}_{is}(\omega))^2\right]}{\left(1+\eta\kappa - \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\frac{\kappa-\nu}{\kappa+(\nu-\kappa)\bar{\ell}_{isn}/\bar{\ell}_{is}}\right)^2} ds,
\end{aligned}
$$

which converges to 0 as $\kappa \to \infty$. Therefore, payments to labor converge to $(1-\eta)z_i m_i^{\eta} \ell_i^{1-\eta} \Phi^b(m)$, and profits converge to the competitive profits. □

B Proofs of Theoretical Results

This Appendix presents the proofs of the theoretical results in Section 2 and the technical lemmas used in those proofs. Throughout the appendix, we will use $\bar{x}$ to denote the value of x with no ex-post shocks, and we will use $\hat{x}$ to denote log deviations from that value. Furthermore, we prove the results for the general case with finite κ and ν. To get the results in Section 2, take the limit as $\kappa \to \infty$ and $\nu \to 0$.

B.1 Proofs of Technical Lemmas

In this section, we prove the Technical Lemmas used below to prove the results in the main text.

Lemma 7. *Average HHI in a location has the following properties:*

(i) $\frac{\partial}{\partial \log m}\left[\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 ds\right] < 0$*; and*

(ii) $\frac{\partial}{\partial \log m}\left[\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 ds\right] \to 0$ *as* $m \to \infty$*.*

Proof. We denote the average sectoral HHI by

$$H(m) \equiv \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n\in\mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 ds.$$

Recall that $\overline{\ell}_{isn} = \left(\frac{(1-\eta)z_{isn}}{\overline{w}_i}\right)^{\frac{1}{\eta}}$ where $\overline{w}_i = (1-\eta)\ell_i^{-\eta} m_i^{\eta} z_i$. Therefore,

$$\begin{aligned} H(m) &= \int_0^1 \frac{\ell_i m_i^{-1} z_i^{-\frac{1}{\eta}} \sum_{n\in\mathcal{N}_{is}} z_{isn}^{\frac{1}{\eta}}}{\ell_i} \sum_{n\in\mathcal{N}_{is}} \left(\frac{z_{isn}^{\frac{1}{\eta}}}{\sum_{n'\in\mathcal{N}_{is}} z_{isn'}^{\frac{1}{\eta}}}\right)^2 ds \\ &= \int_0^1 \frac{N_{is}}{m_i} \frac{\sum_{n\in\mathcal{N}_{is}} z_{isn}^{1/\eta}}{N_{is} z_i^{1/\eta}} \frac{\sum_{n\in\mathcal{N}_{is}} (z_{isn}^{1/\eta})^2}{\left(\sum_{n\in\mathcal{N}_{is}} z_{isn}^{1/\eta}\right)^2} ds \end{aligned}$$

We can then decompose this into an expectation over the number of firms in the sector and, conditional on the number of firms, the expected productivity shocks. Then

$$H(m) = \sum_{N=0}^{\infty} \frac{N}{m} \psi_N \frac{m^N e^{-m}}{N!},$$

where

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

for a given N. First note that $\frac{\sum_{n\in\mathcal{N}} (z_{isn}^{1/\eta})^2}{\left(\sum_{n\in\mathcal{N}} z_{isn}^{1/\eta}\right)^2} \in (0,1)$, so

$$\left|\frac{\sum_{n\in\mathcal{N}} z_{isn}^{1/\eta}}{N z_i^{1/\eta}} \frac{\sum_{n\in\mathcal{N}} (z_{isn}^{1/\eta})^2}{\left(\sum_{n\in\mathcal{N}} z_{isn}^{1/\eta}\right)^2}\right| \le \frac{\sum_{n\in\mathcal{N}} z_{isn}^{1/\eta}}{N z_i^{1/\eta}}.$$

Then as $\mathbb{E}\left[\frac{\sum_{n\in\mathcal{N}_{is}} z_{isn}^{1/\eta}}{N_{is}z_i^{1/\eta}}\right] = 1$, ψ_N exists for each N. We can then rewrite ψ_N,

$$\begin{aligned}
\psi_N &= \frac{1}{Nz_i^{1/\eta}}\mathbb{E}\left[\frac{\sum_{n\in\mathcal{N}}(z_{isn}^{1/\eta})^2}{\sum_{n\in\mathcal{N}} z_{isn}^{1/\eta}}\right] \\
&= z_i^{-1/\eta}\mathbb{E}\left[\frac{z_{is1}^{2/\eta}}{\sum_n z_{isn}^{1/\eta}}\right] \\
&= z_i^{-1/\eta}\mathbb{E}\left[\int_0^\infty z_{is1}^{2/\eta} e^{-t\left(\sum_n z_{isn}^{1/\eta}\right)} dt\right],
\end{aligned}$$

where we use the fact that $\frac{1}{x} = \int_0^\infty e^{-tx}$.

By Tonelli's theorem, we switch the orders of integration and use independence to write

$$\psi_N = z_i^{-1/\eta}\int_0^\infty \mathbb{E}[z_{is1}^{2/\eta}e^{-tz_{is1}^{1/\eta}}]\prod_{n\geq 2}\mathbb{E}[e^{-tz_{isn}^{1/\eta}}]dt.$$

We then define

$$g(t) \equiv \mathbb{E}[z_{is1}^{2/\eta}e^{-tz_{is1}^{1/\eta}}]; \quad \phi(t) \equiv \mathbb{E}[e^{-tz_{is2}^{1/\eta}}].$$

We have that $\phi(t)$ exists for all t as $e^{-tz_{is2}^{1/\eta}}$ is bounded above by 1. $g(t)$ exists for all $t > 0$ as $z_{is1}^{2/\eta}e^{-tz_{is1}^{1/\eta}}$ is maximized at $z_{isn}^{1/\eta} = \frac{2}{t}$, and so is bounded above as well. Therefore,

$$\psi_N = z_i^{-1/\eta}\int_0^\infty g(t)\phi(t)^{N-1}dt.$$

We then substitute this into the expression for $H(m)$,

$$\begin{aligned}
H(m) &= \sum_{N=0}^\infty \frac{N}{m}\psi_N\frac{m^N e^{-m}}{N!} \\
&= \lim_{K\to\infty}\sum_{N=1}^K\int_0^\infty g(t)\phi(t)^{N-1}\frac{m^{N-1}e^{-m}}{(N-1)!}dt \\
&= \lim_{K\to\infty}\int_0^\infty g(t)\sum_{k=0}^K\phi(t)^k\frac{m^k e^{-m}}{k!}dt \\
&= \int_0^\infty g(t)\sum_{k=0}^\infty\phi(t)^k\frac{m^k e^{-m}}{k!}dt \\
&= \int_0^\infty g(t)e^{-m}e^{m\phi(t)}dt \\
&= \int_0^\infty g(t)e^{-m(1-\phi(t))}dt,
\end{aligned}$$

where we use the monotone convergence theorem to swap the order of limits.

We complete the proof by taking the derivative with respect to m. The integrand is bounded above by $g(t)$ for all m so we can bring the derivative under the integral.

$$\frac{\partial H(m)}{\partial \log m} = mH'(m) = -\int_0^\infty (1-\phi(t))mg(t)e^{-m(1-\phi(t))}dt < 0,$$

using the fact that $\phi(t) \in (0,1)$. Next, we take the limit as $m \to \infty$. Note that $m(1-\phi(t))e^{-m(1-\phi(t))}g(t) \leq g(t)$, which is integrable. Thus, we can swap the limits to get,

$$\begin{aligned}\lim_{m\to\infty} \frac{\partial H(m)}{\partial \log m} &= -\lim_{m\to\infty}\int_0^\infty (1-\phi(t))mg(t)e^{-m(1-\phi(t))}dt \\ &= -\int_0^\infty g(t) \lim_{m\to\infty}(1-\phi(t))me^{-m(1-\phi(t))}dt \\ &= 0.\end{aligned}$$

□

Lemma 8. *The regional production function is* $Y_i(\ell, m) = z_i m^\eta \ell^{1-\eta}\tilde{\Phi}(m)$, *where* $z_i \equiv \mathbb{E}[z_{isn}^{1/\eta}]^\eta$ *and* $\tilde{\Phi}(m)$ *is given by,*

$$\begin{aligned}\tilde{\Phi}(m) \equiv{}& \mathbb{E}[a_{sn}(\omega)] + (1-\eta)\int_0^1 \frac{\overline{\ell}_s}{\ell}\sum_{n\in\mathcal{N}_s}\frac{\overline{\ell}_{sn}}{\overline{\ell}_s}\mathbb{E}[\hat{a}_{sn}(\omega)\hat{\ell}_{sn}(\omega)]ds \\ &-\eta\frac{1-\eta}{2}\int_0^1 \frac{\overline{\ell}_s}{\ell}\sum_{n\in\mathcal{N}_s}\frac{\overline{\ell}_{sn}}{\overline{\ell}_s}\mathbb{E}[\hat{\ell}_{sn}(\omega)^2]ds \\ &-\frac{1}{\kappa}\frac{1-\eta}{2}\int_0^1 \frac{\overline{\ell}_s}{\ell}\sum_{n\in\mathcal{N}_s}\frac{\overline{\ell}_{sn}}{\overline{\ell}_s}\mathbb{E}\left[\left(\hat{\ell}_{sn}(\omega)-\hat{\ell}_{sn}\right)^2\right]ds \\ &-\left(\frac{1}{\nu}-\frac{1}{\kappa}\right)\frac{1-\eta}{2}\int_0^1 \frac{\overline{\ell}_s}{\ell}\mathbb{E}\left[\left(\hat{\ell}_s(\omega)-\hat{\ell}_s\right)^2\right]ds. \end{aligned} \tag{36}$$

Proof. Expected production is $\int_0^1 \sum_{n\in\mathcal{N}_{is}} z_{isn}a_{isn}(\omega)\ell_{isn}(\omega)^{1-\eta}ds$. We will do a second order approximation around the point $\log \tilde{a}_{isn}(\omega) = \log \tilde{A}_{is}(\omega) = 0$.

We start by characterizing the solution at that point. We find that there is some $\overline{w}_i$ such that

$$\overline{w}_i = (1-\eta)z_{isn}(\overline{\ell}_{isn})^{-\eta},$$

so that $\overline{\ell}_{isn} = \left(\frac{(1-\eta)z_{isn}}{\overline{w}_i}\right)^{\frac{1}{\eta}}$. Then labor clearing requires

$$\ell_i = \int_0^1 \ell_i \overline{L}_{is} ds = \int_0^1 \sum_{n \in \mathcal{N}_{is}} \overline{\ell}_{isn} ds = \int_0^1 \sum_{n \in \mathcal{N}_{is}} \left(\frac{(1-\eta)z_{isn}}{\overline{w}_i}\right)^{\frac{1}{\eta}} ds.$$

This implies that $\overline{w}_i = (1-\eta)\ell_i^{-\eta} m_i^{\eta} \mathbb{E}\left[z_{isn}^{1/\eta}\right]^{\eta}$. Then production is

$$\overline{Y}_i = \int_0^1 \sum_{n \in \mathcal{N}_{is}} z_{isn} \left(\frac{(1-\eta)z_{isn}}{\overline{w}_i}\right)^{\frac{1-\eta}{\eta}} ds = z_i \ell_i^{1-\eta} m_i^{\eta},$$

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

Taking the log second-order approximation to production gives

$$Y_i \approx z_i (m_i)^{\eta} (\ell_i)^{1-\eta} \int_0^1 \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_i} \Big(1 + \left(\hat{a}_{isn}(\omega) + (1-\eta)\hat{\ell}_{isn}(\omega)\right) \\ + \frac{1}{2}\left(\hat{a}_{isn}(\omega) + (1-\eta)\hat{\ell}_{isn}(\omega)\right)^2 \Big) ds.$$

To transform this to be completely second order, we do a second-order approximation to the labor constraints,

$$-\frac{1}{\kappa}\hat{\ell}_{is} + \frac{1+\kappa}{\kappa}\hat{\ell}_{is}(\omega) + \frac{1}{2}\left(-\frac{1}{\kappa}\hat{\ell}_{is} + \frac{1+\kappa}{\kappa}\hat{\ell}_{is}(\omega)\right)^2 = \\ \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \Bigg[-\frac{1}{\kappa}\hat{\ell}_{isn} + \frac{1+\kappa}{\kappa}\hat{\ell}_{isn}(\omega) + \frac{1}{2}\left(-\frac{1}{\kappa}\hat{\ell}_{isn} + \frac{1+\kappa}{\kappa}\hat{\ell}_{isn}(\omega)\right)^2 \Bigg],$$

$$0 = \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i} \Bigg[-\frac{1}{\nu}\hat{\ell}_{is} + \frac{1+\nu}{\nu}\hat{\ell}_{is}(\omega) + \frac{1}{2}\left(-\frac{1}{\nu}\hat{\ell}_{is} + \frac{1+\nu}{\nu}\hat{\ell}_{is}(\omega)\right)^2 \Bigg] ds,$$

$$\hat{\ell}_{is} + \frac{1}{2}\hat{\ell}_{is}^2 = \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \left(\hat{\ell}_{isn} + \frac{1}{2}\hat{\ell}_{isn}^2\right),$$

and

$$0 = \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i} \left(\hat{\ell}_{is} + \frac{1}{2}\hat{\ell}_{is}^2\right) ds,$$

where we use the fact $\hat{\ell}_{isn}(\omega) = \hat{L}_{isn}(\omega)$.

Then we can transform $\int_0^1 \sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_i}\hat{\ell}_{isn}(\omega)ds$ to second order. That is,

$$\begin{aligned}
\int_0^1 \sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_i}\hat{\ell}_{isn}(\omega)ds &= -\frac{1}{2}\frac{\kappa}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\left(-\frac{1}{\kappa}\hat{\ell}_{isn}+\frac{1+\kappa}{\kappa}\hat{\ell}_{isn}(\omega)\right)^2 ds \\
&\quad + \frac{1}{2}\frac{\kappa}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\left(-\frac{1}{\kappa}\hat{\ell}_{is}+\frac{1+\kappa}{\kappa}\hat{\ell}_{is}(\omega)\right)^2 ds \\
&\quad + \frac{1}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}ds - \frac{1}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\hat{\ell}_{is}ds \\
&\quad + \int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\hat{\ell}_{is}(\omega)ds \\
&= -\frac{1}{2}\frac{\kappa}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\left(-\frac{1}{\kappa}\hat{\ell}_{isn}+\frac{1+\kappa}{\kappa}\hat{\ell}_{isn}(\omega)\right)^2 ds \\
&\quad + \frac{1}{2}\frac{\kappa}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\left(-\frac{1}{\kappa}\hat{\ell}_{is}+\frac{1+\kappa}{\kappa}\hat{\ell}_{is}(\omega)\right)^2 ds \\
&\quad - \frac{1}{2}\frac{\nu}{1+\nu}\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\left(-\frac{1}{\nu}\hat{\ell}_{is}+\frac{1+\nu}{\nu}\hat{\ell}_{is}(\omega)\right)^2 ds \\
&\quad + \frac{1}{1+\nu}\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\hat{\ell}_{is}ds \\
&\quad + \frac{1}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}ds - \frac{1}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\hat{\ell}_{is}ds \\
&= -\frac{1}{2}\frac{\kappa}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\left(-\frac{1}{\kappa}\hat{\ell}_{isn}+\frac{1+\kappa}{\kappa}\hat{\ell}_{isn}(\omega)\right)^2 ds \\
&\quad + \frac{1}{2}\frac{\kappa}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\left(-\frac{1}{\kappa}\hat{\ell}_{is}+\frac{1+\kappa}{\kappa}\hat{\ell}_{is}(\omega)\right)^2 ds \\
&\quad - \frac{1}{2}\frac{\nu}{1+\nu}\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\left(-\frac{1}{\nu}\hat{\ell}_{is}+\frac{1+\nu}{\nu}\hat{\ell}_{is}(\omega)\right)^2 ds \\
&\quad - \frac{1}{2}\frac{1}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}^2 ds \\
&\quad + \frac{1}{2}\frac{1}{1+\kappa}\int_0^1 \frac{\bar{\ell}_{is}}{\ell_i}\hat{\ell}_{is}^2 ds \\
&\quad - \frac{1}{2}\frac{1}{1+\nu}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\hat{\ell}_{is}^2 ds \\
&= -\frac{1}{2}\int_0^1 \frac{\bar{\ell}_{is}}{\bar{\ell}_i}\sum_{n\in\mathcal{N}_{is}} \frac{\bar{\ell}_{isn}}{\bar{\ell}_{is}}\hat{\ell}_{isn}(\omega)^2 ds
\end{aligned}$$

$$- \frac{1}{2}\frac{1}{\kappa} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn} \right)^2 ds$$
$$- \frac{1}{2} \left(\frac{1}{\nu} - \frac{1}{\kappa} \right) \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \left(\hat{\ell}_{is}(\omega) - \hat{\ell}_{is} \right)^2 ds.$$

Substituting this into the expression for the log second-order approximation to production gives

$$\begin{aligned}
\frac{Y_i}{z_i (m_i)^\eta (\ell_i)^{1-\eta}} &\approx 1 + \mathbb{E}[\hat{a}_{isn}(\omega)] + \frac{1}{2}\mathbb{E}[\hat{a}_{isn}(\omega)^2] \\
&\quad + (1-\eta) \int_0^1 \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \hat{a}_{isn}(\omega) \hat{\ell}_{isn}(\omega) ds \\
&\quad + \frac{(1-\eta)^2}{2} \int_0^1 \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \hat{\ell}_{isn}(\omega)^2 ds \\
&\quad - \frac{1-\eta}{2} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{\ell}_{isn}(\omega)^2 ds \\
&\quad - \frac{1-\eta}{2} \frac{1}{\kappa} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn} \right)^2 ds \\
&\quad - \frac{1-\eta}{2} \left(\frac{1}{\nu} - \frac{1}{\kappa} \right) \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \left(\hat{\ell}_{is}(\omega) - \hat{\ell}_{is} \right)^2 ds \\
&\approx \mathbb{E}[a_{isn}(\omega)] + (1-\eta) \int_0^1 \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\ell_i} \hat{a}_{isn}(\omega) \hat{\ell}_{isn}(\omega) ds \\
&\quad - \frac{1-\eta}{2} \eta \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{\ell}_{isn}(\omega)^2 ds \\
&\quad - \frac{1-\eta}{2} \frac{1}{\kappa} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn} \right)^2 ds \\
&\quad - \frac{1-\eta}{2} \left(\frac{1}{\nu} - \frac{1}{\kappa} \right) \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \left(\hat{\ell}_{is}(\omega) - \hat{\ell}_{is} \right)^2 ds,
\end{aligned}$$

where we use the fact that to log second order, $\mathbb{E}[a_{sn}(\omega)] \approx 1 + \mathbb{E}[\hat{a}_{sn}(\omega)] + \frac{1}{2}\mathbb{E}[\hat{a}_{sn}(\omega)^2]$.
□

B.2 Proofs of Propositions & Lemmas in Main Text

In this subsection, we prove the Propositions and Lemmas in the main text.

Proof of Lemma 1. By Lemma 8, the regional production can be written $Y_i(\ell, m) = z_i m^{\eta} \ell^{1-\eta} \tilde{\Phi}(m)$ where

$$\begin{aligned}
\tilde{\Phi}(m) \equiv \mathbb{E}[a_{sn}(\omega)] + (1-\eta) \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \mathbb{E}[\hat{a}_{sn}(\omega) \hat{\ell}_{sn}(\omega)] ds \\
- \eta \frac{1-\eta}{2} \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \mathbb{E}[\hat{\ell}_{sn}(\omega)^2] ds \\
- \frac{1}{\kappa} \frac{1-\eta}{2} \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \mathbb{E}\left[\left(\hat{\ell}_{sn}(\omega) - \hat{\ell}_{sn} \right)^2 \right] ds \\
- \left(\frac{1}{\nu} - \frac{1}{\kappa} \right) \frac{1-\eta}{2} \int_0^1 \frac{\overline{\ell}_s}{\ell} \mathbb{E}\left[\left(\hat{\ell}_s(\omega) - \hat{\ell}_s \right)^2 \right] ds
\end{aligned}$$

Therefore, the planner is looking to maximize this second-order production function subject to the labor constraints. Following Benigno and Woodford (2003) and Benigno and Woodford (2012), we can do a linear approximation to the labor constraints, embedded in the sets $\mathcal{L}$ and $\mathcal{L}_{\Omega}(\cdot)$,

$$\begin{aligned}
\hat{\ell}_{is}(\omega) &= \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{\ell}_{isn}(\omega), & 0 &= \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i} \hat{\ell}_{is}(\omega) ds, \\
\hat{\ell}_{is} &= \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{\ell}_{isn}, & 0 &= \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i} \hat{\ell}_{is} ds.
\end{aligned}$$

To simplify the explication, we rewrite the maximization problem with vector notation as

$$\begin{aligned}
\max_x \quad & ax - \frac{1}{2} x' b x \\
\text{s.t.} \quad & cx = 0
\end{aligned}$$

where x is the vector of labor supply decisions, a is the vector of productivity shocks so that

$$ax = (1-\eta) \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \mathbb{E}[\hat{a}_{sn}(\omega) \hat{\ell}_{sn}(\omega)] ds,$$

b is the self-adjoint operator (i.e., symmetric matrix) representing the loss function so that

$$x' b x = \eta(1-\eta) \int_0^1 \frac{\overline{\ell}_s}{\ell} \sum_{n \in \mathcal{N}_s} \frac{\overline{\ell}_{sn}}{\overline{\ell}_s} \mathbb{E}[\hat{\ell}_{sn}(\omega)^2] ds$$

$$
\begin{aligned}
&+\frac{1}{\kappa}(1-\eta)\int_0^1 \frac{\overline{\ell}_s}{\ell}\sum_{n\in\mathcal{N}_s}\frac{\overline{\ell}_{sn}}{\overline{\ell}_s}\mathbb{E}\left[\left(\hat{\ell}_{sn}(\omega)-\hat{\ell}_{sn}\right)^2\right]ds\\
&+\left(\frac{1}{\nu}-\frac{1}{\kappa}\right)(1-\eta)\int_0^1 \frac{\overline{\ell}_s}{\ell}\mathbb{E}\left[\left(\hat{\ell}_s(\omega)-\hat{\ell}_s\right)^2\right]ds,
\end{aligned}
$$

and c is the matrix representing the linear constraints. Forming the Lagrangian, we have

$$ax-\frac{1}{2}x'bx-\lambda cx.$$

Taking the FOCs, we get

$$a-(bx)'-\lambda c=0.$$

Therefore, $x=b^{-1}(a-\lambda c)'=b^{-1}(a'-c'\lambda')$. Using the constraint, we can solve for λ:

$$
\begin{aligned}
0&=cx\\
&=cb^{-1}(a'-c'\lambda')\\
\lambda'&=(cb^{-1}c')^{-1}cb^{-1}a'.
\end{aligned}
$$

Then, to complete the proof, we will show that $ax=x'bx$.

$$
\begin{aligned}
x'bx&=\left(b^{-1}(a'-c'(cb^{-1}c')^{-1}cb^{-1}a')\right)'bb^{-1}(a'-c'(cb^{-1}c')^{-1}cb^{-1}a')\\
&=(a-ab^{-1}c'(cb^{-1}c')^{-1}c)b^{-1}(a'-c'(cb^{-1}c')^{-1}cb^{-1}a')\\
&=ab^{-1}a'-ab^{-1}c'(cb^{-1}c')^{-1}cb^{-1}a'-ab^{-1}c'(cb^{-1}c')^{-1}cb^{-1}a'\\
&\quad+ab^{-1}c'(cb^{-1}c')^{-1}cb^{-1}c'(cb^{-1}c')^{-1}cb^{-1}a'\\
&=ab^{-1}a'-ab^{-1}c'(cb^{-1}c')^{-1}cb^{-1}a',
\end{aligned}
$$

where we use the fact that b and $cb^{-1}c'$ are self adjoint and therefore b^{-1} and $(cb^{-1}c')^{-1}$ are self adjoint. Similarly,

$$
\begin{aligned}
ax&=ab^{-1}(a'-c'(cb^{-1}c')^{-1}cb^{-1}a')\\
&=x'bx.
\end{aligned}
$$

Therefore, $ax-\frac{1}{2}x'bx=\frac{1}{2}ax$, completing the proof. □

Proof of Proposition 1. This result follows almost immediately from Lemma 1. Implicitly differentiating the free entry condition (10) implies that

$$\frac{d\log m_i}{d\log \ell_i} = \frac{1}{1 - \frac{1}{1-\eta}\frac{\partial \log \Phi(m_i)}{\partial \log m_i}}.$$

If profits are decreasing in the number of firms, the denominator must be positive. Therefore, differentiating the expression for wages gives

$$\begin{aligned}\frac{d\log w_i}{d\log \ell_i} &= \left(\eta + \frac{\partial \log \Phi(m_i)}{\partial \log m_i}\right)\frac{d\log m_i}{d\log \ell_i} - \eta \\ &= \frac{\frac{1}{1-\eta}\frac{\partial \log \Phi(m_i)}{\partial \log m_i}}{1 - \frac{1}{1-\eta}\frac{\partial \log \Phi(m_i)}{\partial \log m_i}},\end{aligned}$$

which is greater than zero if and only if $\frac{\partial \log \Phi(m_i)}{\partial \log m_i} > 0$. □

Proof of Proposition 2. We prove the result for finite κ and ν where labor is more substitutable across firms within a sector than across sectors, i.e., $\kappa > \nu$. Taking a log first order approximation to the labor supply FOCs (27) after substituting out $\lambda_{is}(\omega)$ and λ_{is} implies,

$$\begin{aligned}\hat{w}_{isn}(\omega) &= \hat{\lambda}(\omega) + \left(\frac{1}{\nu} - \frac{1}{\kappa}\right)\left(\hat{\ell}_{is}(\omega) - \hat{\ell}_{is}\right) + \frac{1}{\kappa}\left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn}\right) \\ \hat{\bar{\lambda}}_i &= \frac{1+\nu}{1+\kappa}\mathbb{E}\left[\hat{w}_{isn}(\omega)\right] + \frac{\kappa-\nu}{1+\kappa}\mathbb{E}\left[\hat{\lambda}(\omega)\right] + \frac{\kappa}{1+\kappa}\frac{1+\nu}{\nu}\mathbb{E}\left[\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn}\right],\end{aligned}$$

where we use the fact that $\hat{L}_{isn}(\omega) = \hat{\ell}_{isn}(\omega)$.

Taking a log first-order approximation to the labor constraints embedded in $\mathcal{L}$ and $\mathcal{L}_\Omega(\cdot)$ implies,

$$\begin{aligned}\hat{\ell}_{is}(\omega) &= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{\ell}_{isn}(\omega), & 0 &= \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i}\hat{\ell}_{is}(\omega)ds, \\ \hat{\ell}_{is} &= \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{\ell}_{isn}, & 0 &= \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i}\hat{\ell}_{is}ds.\end{aligned}$$

And the labor demand curve implies that $\hat{a}_{isn}(\omega) - \eta\hat{\ell}_{isn}(\omega) = \hat{w}_{isn}(\omega)$. First note that taking expectations $\mathbb{E}[\hat{a}_{isn}(\omega)] - \eta\mathbb{E}[\hat{\ell}_{isn}(\omega)] = \mathbb{E}[\hat{w}_{isn}(\omega)]$ which implies $\mathbb{E}[\hat{w}_{isn}(\omega)] = -\eta\mathbb{E}[\hat{\ell}_{isn}(\omega)]$. Therefore,

$$-\hat{\lambda}_i(\omega) = \left(\frac{1}{\nu} - \frac{1}{\kappa}\right)\left(\hat{\ell}_{is}(\omega) - \hat{\ell}_{is}\right) + \frac{1}{\kappa}\left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn}\right) + \eta\hat{\ell}_{isn}(\omega)$$

$$
\begin{aligned}
-\int_0^1 \frac{\overline{\ell}_{is}}{\ell_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{\lambda}(\omega) ds = \left(\frac{1}{\nu} - \frac{1}{\kappa}\right) \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i} \left(\hat{\ell}_{is}(\omega) - \hat{\ell}_{is}\right) ds \\
+ \int_0^1 \frac{\overline{\ell}_{is}}{\ell_i} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \left(\frac{1}{\kappa}\left(\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn}\right) + \eta \hat{\ell}_{isn}(\omega)\right) ds \\
\hat{\lambda}(\omega) = 0.
\end{aligned}
$$

Then

$$
\begin{aligned}
\hat{\lambda}_i &= \frac{1+\nu}{1+\kappa} \mathbb{E}\left[\hat{w}_{isn}(\omega)\right] + \frac{\kappa-\nu}{1+\kappa} \mathbb{E}\left[\hat{\lambda}(\omega)\right] + \frac{\kappa}{1+\kappa}\frac{1+\nu}{\nu} \mathbb{E}\left[\hat{\ell}_{isn}(\omega) - \hat{\ell}_{isn}\right] \\
\hat{\lambda}_i &= \frac{1+\nu}{1+\kappa}\left(\frac{\kappa}{\nu} - \eta\right) \mathbb{E}\left[\hat{\ell}_{isn}(\omega)\right] - \frac{\kappa}{1+\kappa}\frac{1+\nu}{\nu} \hat{\ell}_{isn}.
\end{aligned}
$$

Again, taking a weighted sum across all firms implies that $\hat{\lambda}_i = 0$. Since $\kappa/\nu > 1$ and $\eta \in (0,1)$, the right-hand side is a non-degenerate linear system of equations equal to 0. It then follows that $\hat{\ell}_{isn} = \mathbb{E}[\hat{\ell}_{isn}(\omega)] = \mathbb{E}[\hat{w}_{isn}(\omega)] = 0$ because otherwise it is not possible for $0 = \left(\frac{\kappa}{\nu} - \eta\right) \mathbb{E}\left[\hat{\ell}_{isn}(\omega)\right] - \frac{\kappa}{\nu}\hat{\ell}_{isn}$ and the labor constraints to hold.

Therefore, we can solve for sectoral labor

$$
\begin{aligned}
\hat{a}_{isn}(\omega) - \eta \hat{\ell}_{isn}(\omega) &= \left(\frac{1}{\nu} - \frac{1}{\kappa}\right) \hat{\ell}_{is}(\omega) + \frac{1}{\kappa} \hat{\ell}_{isn}(\omega) \\
\sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{a}_{isn}(\omega) &= \left(\frac{1}{\nu} - \frac{1}{\kappa}\right) \hat{\ell}_{is}(\omega) + \left(\frac{1}{\kappa} + \eta\right) \hat{\ell}_{is}(\omega) \\
\frac{1}{\eta + \frac{1}{\nu}} \sum_{n\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{a}_{isn}(\omega) &= \hat{\ell}_{is}(\omega).
\end{aligned}
$$

And individual firm labor is

$$
\hat{\ell}_{isn}(\omega) = \frac{1}{\eta + \frac{1}{\kappa}} \left(\hat{a}_{isn}(\omega) - \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \sum_{n'\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega)\right).
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}[\hat{a}_{isn}(\omega)\hat{\ell}_{isn}(\omega)] &= \mathbb{E}\left[\hat{a}_{isn}(\omega) \frac{1}{\eta + \frac{1}{\kappa}} \left(\hat{a}_{isn}(\omega) - \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \sum_{n'\in\mathcal{N}_{is}} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega)\right)\right] \\
&= \frac{1}{\eta + \frac{1}{\kappa}} \left(\sigma_S^2 + \sigma_N^2\right) - \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\left(\eta + \frac{1}{\kappa}\right)\left(\eta + \frac{1}{\nu}\right)} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \sigma_N^2 + \sigma_S^2\right)
\end{aligned}
$$

$$= \frac{1}{\eta + \frac{1}{\nu}} \sigma_S^2 + \frac{\eta + \frac{1}{\nu} - \left(\frac{1}{\nu} - \frac{1}{\kappa}\right) \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}}{\left(\eta + \frac{1}{\kappa}\right)\left(\eta + \frac{1}{\nu}\right)} \sigma_N^2.$$

Then we can calculate $\Phi(m)$,

$$\Phi(m) = \mathbb{E}[a_{isn}(\omega)] + \frac{1-\eta}{2} \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \mathbb{E}\left[\hat{a}_{isn}(\omega) \hat{\ell}_{isn}(\omega)\right] ds$$

$$= \mathbb{E}[a_{isn}(\omega)] + \frac{1-\eta}{2} \frac{1}{\eta + \frac{1}{\nu}} \sigma_S^2 + \frac{1-\eta}{2} \frac{\eta + \frac{1}{\nu} - \left(\frac{1}{\nu} - \frac{1}{\kappa}\right) \int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 ds}{\left(\eta + \frac{1}{\kappa}\right)\left(\eta + \frac{1}{\nu}\right)} \sigma_N^2.$$

Therefore,

$$\frac{\partial \log \Phi(m)}{\partial \log m} = -\frac{1-\eta}{2} \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\left(\eta + \frac{1}{\kappa}\right)\left(\eta + \frac{1}{\nu}\right)} \frac{\partial}{\partial \log m} \left[\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 ds\right] \frac{\sigma_N^2}{\Phi(m)}$$

By Lemma 7, $\frac{\partial}{\partial \log m} \left[\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 ds\right] < 0$. Then by Proposition 1, the first result will follow. Similarly, by Lemma 7, $\frac{\partial}{\partial \log m} \left[\int_0^1 \frac{\overline{\ell}_{is}}{\overline{\ell}_i} \sum_{n \in \mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2 ds\right] \to 0$ as $m \to \infty$ so

$$\frac{d \log w_i}{d \log \ell_i} = \frac{\frac{1}{1-\eta} \frac{\partial \log \Phi(m_i)}{\partial \log m_i}}{1 - \frac{1}{1-\eta} \frac{\partial \log \Phi(m_i)}{\partial \log m_i}} \to 0,$$

as $m \to \infty$. Noting that $m \to \infty$ as $\ell \to \infty$ completes the proof. □

Proof of Proposition 3. First-best entry must satisfy,

$$\psi_i = \left(1 + \frac{1}{\eta} \frac{\partial \log \Phi(m_i^{FB})}{\partial \log m}\right) \frac{\eta Y_i}{m_i^{FB}}.$$

In equilibrium, entry satisfies

$$\psi_i = (1 + \tau_i) \frac{\eta Y_i}{m_i},$$

where τ_i is the subsidy on entry. Thus, in any optimal equilibrium, $\tau_i = \frac{1}{\eta} \frac{\partial \log \Phi(m_i^{FB})}{\partial \log m}$. Furthermore, $\frac{\partial \log \Phi(m)}{\partial \log m} \to 0$ as $\ell \to \infty$ as we showed in Proposition 2. □

Proof of Lemma 2. This result follows immediately from Lemma 8 and noting that, as perfect competition is efficient, it must produce the most conditional on the number of firms and workers. ☐

Proof of Lemma 3. This result follows for Cournot and Bertrand competition from Lemmas 4 and 5 and noting that wages must have a markdown. ☐

Proof of Proposition 4. At the first best, workers need to earn their marginal product in each location. That is,

$$(1+\tau_i^w)w_i^{FB} = \frac{\partial Y_i(\ell_i^{FB}, m_i^{FB})}{\partial \ell} = \frac{(1-\eta)Y_i^{FB}}{\ell_i^{FB}}.$$

Furthermore, firms need to be paid their marginal product,

$$\psi_i = \frac{\partial Y_i(\ell_i^{FB}, m_i^{FB})}{\partial m} = \left(1 + \frac{1}{\eta}\frac{\partial \log \Phi^f(m_i^{FB})}{\partial \log m}\right)\frac{\eta Y_i^{FB}}{m_i^{FB}}.$$

In equilibrium, workers are paid

$$\begin{aligned} w_i &= \frac{(1-\eta)z_i(m_i)^\eta(\ell_i)^{1-\eta}(\Phi^f(m_i)+\Psi^f(m_i))}{\ell_i} \\ &= \left(1+\frac{\Psi^f(m_i)}{\Phi^f(m_i)}\right)\frac{(1-\eta)Y_i}{\ell_i}. \end{aligned}$$

Meanwhile, free entry implies

$$\begin{aligned} \psi_i &= (1+\tau_i^\pi)\frac{Y_i - w_i\ell_i}{m_i} \\ &= (1+\tau_i^\pi)\frac{z_i(m_i)^\eta(\ell_i)^{1-\eta}\Phi^f(m_i) - (1-\eta)z_i(m_i)^\eta(\ell_i)^{1-\eta}\left(\Phi^f(m_i)+\Psi^f(m_i)\right)}{m_i} \\ &= (1+\tau_i^\pi)\frac{\eta Y_i}{m_i}\left(1 - \frac{1-\eta}{\eta}\frac{\Psi^f(m_i)}{\Phi^f(m_i)}\right). \end{aligned}$$

Equating the equilibrium conditions to the first best optimality conditions implies the results. ☐

Proof of Proposition 5. Workers are paid their marginal product of labor. Thus, to log first order,

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

Then to a first order, the short labor supply elasticity across firms is

$$\hat{\ell}_{isn}(\omega) = \kappa \hat{w}_{isn}(\omega) - (\kappa - \nu)\hat{w}_{is}(\omega),$$

where

$$\hat{w}_{is}(\omega) = \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{w}_{isn}(\omega).$$

Summing across labor supply equations, weighted by $\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}$ implies $\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{\ell}_{isn}(\omega) = \nu \hat{w}_{is}(\omega)$. Therefore,

$$\begin{aligned}
\hat{w}_{is}(\omega) &= \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{w}_{isn}(\omega) \\
&= \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{a}_{isn}(\omega) - \eta \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{\ell}_{isn}(\omega) \\
&= \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{a}_{isn}(\omega) - \eta \nu \hat{w}_{is}(\omega) \\
&= \frac{1}{1+\eta\nu} \sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \hat{a}_{isn}(\omega).
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\ell}_{isn}(\omega) &= \kappa \hat{w}_{isn}(\omega) - (\kappa - \nu)\hat{w}_{is}(\omega) \\
&= \kappa \left(\hat{a}_{isn}(\omega) - \eta \hat{\ell}_{isn}(\omega) \right) - \frac{\kappa - \nu}{1 + \eta\nu} \sum_{n'} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega) \\
&= \frac{1}{\eta + \frac{1}{\kappa}} \left(\hat{a}_{isn}(\omega) - \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \sum_{n'} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega) \right).
\end{aligned}$$

With this, we turn to find the log wage bill.

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

$$
\begin{aligned}
&= \frac{\eta\left(1+\frac{1}{\kappa}\right)+\frac{1}{\nu}\left(1+\frac{1}{\kappa}\right)-\frac{1}{\nu}(1-\eta)+\frac{1}{\kappa}(1-\eta)}{\left(\eta+\frac{1}{\kappa}\right)\left(\eta+\frac{1}{\nu}\right)}\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{a}_{isn}(\omega) \\
&= \frac{\eta+\frac{1}{\nu}\frac{1}{\kappa}+\eta\frac{1}{\nu}+\frac{1}{\kappa}}{\left(\eta+\frac{1}{\kappa}\right)\left(\eta+\frac{1}{\nu}\right)}\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{a}_{isn}(\omega) \\
&= \frac{1+\frac{1}{\nu}}{\eta+\frac{1}{\nu}}\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{a}_{isn}(\omega) \\
&= \frac{1+\nu}{1+\eta\nu}\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{a}_{isn}(\omega).
\end{aligned}
$$

We then take the variance. We have,

$$
\begin{aligned}
\operatorname{Var}\left(\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\left(\hat{w}_{isn}(\omega)+\hat{\ell}_{isn}(\omega)\right)\right) &= \left(\frac{1+\nu}{1+\eta\nu}\right)^2 \operatorname{Var}\left(\textstyle\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{a}_{isn}(\omega)\right) \\
&= \left(\frac{1+\nu}{1+\eta\nu}\right)^2 \operatorname{Var}\left(\hat{\bar{A}}_{is}(\omega)+\textstyle\sum_n \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\hat{\tilde{a}}_{isn}(\omega)\right) \\
&= \left(\frac{1+\nu}{1+\eta\nu}\right)^2 \left(\sigma_S^2+\sum_{n\in\mathcal{N}_{is}}\left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}}\right)^2\sigma_N^2\right).
\end{aligned}
$$

To complete the proof, we need prove that expected HHI is decreasing in the number of firms.

Recall that $\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} = \frac{z_{isn}^{\frac{1}{\eta}}}{\sum_{n'} z_{isn'}^{\frac{1}{\eta}}}$. Denote $x_{isn} \equiv z_{isn}^{\frac{1}{\eta}}$. x_{isn} are independent pareto distributions. Then we will prove that

$$
\mathbb{E}\left[\frac{\sum_{n=1}^{N+1} x_{isn}^2}{\left(\sum_{n=1}^{N+1} x_{isn}\right)^2}\right] \leq \mathbb{E}\left[\frac{\sum_{n=1}^{N} x_{isn}^2}{\left(\sum_{n=1}^{N} x_{isn}\right)^2}\right].
$$

To prove this, consider some arbitrary m. Then we will find the expected HHI with $N+1$ firms when firm n' is the least productive firm and $x_{isn'} = m$. Note that

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{n=1}^{N+1} x_{isn}^2}{\left(\sum_{n=1}^{N+1} x_{isn}\right)^2}|m = x_{n'}, x_n \geq m, \forall n \neq n'\right] \\
&\qquad = \mathbb{E}\left[\frac{m^2+\sum_{n\neq n'} x_{isn}^2}{\left(m+\sum_{n\neq n'} x_{isn}\right)^2}|m = x_{n'}, x_n \geq m, \forall n \neq n'\right]
\end{aligned}
$$

$$= \mathbb{E}\left[\frac{1+\sum_{n\neq n'}\left(\frac{x_{isn}}{m}\right)^2}{\left(1+\sum_{n\neq n'}\frac{x_{isn}}{m}\right)^2}|m = x_{n'}, \frac{x_n}{m} \geq 1, \forall n \neq n'\right].$$

Then since x_{isn} is distributed Pareto, $\frac{x_{isn}}{m}$ is distributed Pareto. Furthermore, since $\frac{x_n}{m}$ is distributed Pareto, $\frac{x_n}{m}$ given it is greater than 1 is distributed Pareto with location parameter 1. Therefore,

$$\mathbb{E}\left[\frac{1+\sum_{n\neq n'}\left(\frac{x_{isn}}{m}\right)^2}{\left(1+\sum_{n\neq n'}\frac{x_{isn}}{m}\right)^2}|m = x_{n'}, \frac{x_n}{m} \geq 1, \forall n \neq n'\right] = \mathbb{E}\left[\frac{1+\sum_{n=1}^{N} y_n^2}{\left(1+\sum_{n=1}^{N} y_n\right)^2}\right]$$

for iid Pareto variables y_n with location parameter 1. Next, define $A \equiv \sum_{n=1}^{N} y_n^2$ amd $B \equiv \sum_{n=1}^{N} y_n$. Note that as $y_n \geq 1$, $A \geq B$. We will prove that

$$\frac{1+A}{(1+B)^2} \leq \frac{A}{B^2}.$$

Note,

$$\begin{aligned}\frac{1+A}{(1+B)^2} - \frac{A}{B^2} &= \frac{B^2 + AB^2 - A - 2AB - AB^2}{B^2(1+B)^2} \\ &= \frac{B^2 - A - 2AB}{B^2(1+B)^2} < 0.\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\frac{1+\sum_{n=1}^{N} y_n^2}{\left(1+\sum_{n=1}^{N} y_n\right)^2}\right] < \mathbb{E}\left[\frac{\sum_{n=1}^{N} y_n^2}{\left(\sum_{n=1}^{N} y_n\right)^2}\right],$$

and the right hand side is exactly the expected HHI of sectors with N firms as the location parameter does not affect HHI. Putting everything together,

$$\mathbb{E}\left[\frac{\sum_{n=1}^{N+1} x_{isn}^2}{\left(\sum_{n=1}^{N+1} x_{isn}\right)^2}|m = x_{n'}, x_n \geq m, \forall n \neq n'\right] < \mathbb{E}\left[\frac{\sum_{n=1}^{N} y_n^2}{\left(\sum_{n=1}^{N} y_n\right)^2}\right].$$

Taking expectations of both sides proves the result.

□

Proof of Proposition 6. From the previous proof,

$$\hat{\ell}_{isn}(\omega) = \frac{1}{\eta + \frac{1}{\kappa}} \left(\hat{a}_{isn}(\omega) - \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \sum_{n'} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega) \right).$$

Therefore, taking the variance,

$$\begin{aligned}
\text{Var}(\hat{\ell}_{isn}(\omega)) &= \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \mathbb{E}\left[\left(\hat{a}_{isn}(\omega) - \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \sum_{n'} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega) \right)^2 \right] \\
&= \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \mathbb{E}\left[\hat{a}_{isn}(\omega)^2 \right] \\
&\quad - 2 \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \mathbb{E}\left[\hat{a}_{isn}(\omega) \sum_{n'} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega) \right] \\
&\quad + \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \left(\frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \right)^2 \mathbb{E}\left[\left(\sum_{n'} \frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \hat{a}_{isn'}(\omega) \right)^2 \right] \\
&= \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \left(\sigma_N^2 + \sigma_S^2 \right) \\
&\quad - 2 \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \left(\sigma_S^2 + \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \sigma_N^2 \right) \\
&\quad + \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \left(\frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \right)^2 \left(\sigma_S^2 + \sum_{n' \in \mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \right)^2 \sigma_N^2 \right).
\end{aligned}$$

We then aggregate up to the sectoral level,

$$\begin{aligned}
\sum_{n \in \mathcal{N}_{is}} \frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \text{Var}(\hat{\ell}_{isn}(\omega)) &= \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \left(\sigma_N^2 + \sigma_S^2 \right) \\
&\quad - 2 \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \left(\sigma_S^2 + \sum_{n} \left(\frac{\overline{\ell}_{isn}}{\overline{\ell}_{is}} \right)^2 \sigma_N^2 \right) \\
&\quad + \frac{1}{\left(\eta + \frac{1}{\kappa}\right)^2} \left(\frac{\frac{1}{\nu} - \frac{1}{\kappa}}{\eta + \frac{1}{\nu}} \right)^2 \left(\sigma_S^2 + \sum_{n' \in \mathcal{N}_{is}} \left(\frac{\overline{\ell}_{isn'}}{\overline{\ell}_{is}} \right)^2 \sigma_N^2 \right)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{\eta+\frac{1}{\nu}}\right)^2 \sigma_S^2 + \left(\frac{1}{\eta+\frac{1}{\kappa}}\right)^2 \sigma_N^2 \\
&\qquad - \left(\frac{1}{\nu}-\frac{1}{\kappa}\right) \frac{\eta+\frac{1}{\kappa}+\eta+\frac{1}{\nu}}{\left(\eta+\frac{1}{\kappa}\right)^2\left(\eta+\frac{1}{\nu}\right)^2} HHI_{is}\sigma_N^2.
\end{aligned}
$$

The previous proposition then proves that the HHI is decreasing in N. □

C Robustness of Theory

In this section, we consider some extensions to the baseline model to show that the key theoretical takeaways are robust to alternate modeling assumptions.

C.1 Alternate Entry Condition

The free entry condition in the main text,

$$
\psi_i = \frac{\mathbb{E}\left[\int_0^1 \sum_{n\in\mathcal{N}_{is}} \pi_{isn}(\omega) ds\right]}{m_i}, \tag{37}
$$

imagines a staged entry game. All firms choose whether or not to enter the market, and then all of those firms are randomly assigned sectors and ex-ante productivity shocks. In this appendix, we consider an alternate entry game where a potential entrant sees the number of firms in each sector and the ex-ante shocks they have received and decides whether or not to enter. We will characterize this alternate free entry condition and prove that it is the same condition as the main text free entry condition (37).

Under the alternate entry game, the potential entrant can see that there is a mass $p(N, m_i)$ of sectors with N firms, where $p(N, m_i) = m_i^N e^{-m_i}/N!$ is the probability mass function (pmf). Therefore, the probability that the firm enters a sector s with N firms already in it is $p(N, m_i)$. The firm internalizes that if it were to enter such a sector s, that sector would not have N firms, but instead $N+1$ firms. Nothing else about the equilibrium would change as that sector is small relative to the entire location so the expected profits of the firm would be the average profits of a firm in a sector with $N+1$ firms, $\pi^e_{i,N+1} \equiv \mathbb{E}[\pi_{isn}(\omega)|N_{is} = N+1]$, where expectations are taken over ex-ante and ex-post shocks.

Therefore, the free entry condition can be written,

$$\psi_i = \sum_{N=0}^{\infty} \pi^e_{i,N+1} p(N, m_i). \tag{38}$$

Now we will show that this free entry condition (38) is equivalent to (37). Taking the original free entry condition, notice that,

$$\begin{aligned}\frac{\mathbb{E}\left[\int_0^1 \sum_{n\in\mathcal{N}_{is}} \pi_{isn}(\omega) ds\right]}{m_i} &= \frac{\sum_{N=0}^{\infty} \mathbb{E}\left[\sum_{n\in\mathcal{N}_{is}} \pi_{isn}(\omega) | N_{is} = N\right] p\left(N, m_i\right)}{m_i} \\ &= \sum_{N=0}^{\infty} \pi^e_{i,N} \frac{N}{m_i} p(N, m_i).\end{aligned}$$

Taking the alternate free entry condition, we have

$$\begin{aligned}\sum_{N=0}^{\infty} \pi^e_{i,N+1} p(N, m_i) &= \sum_{N=0}^{\infty} \pi^e_{i,N+1} \frac{N+1}{m_i} \frac{m_i^{N+1} e^{-m_i}}{(N+1)!} \\ &= \sum_{N=1}^{\infty} \pi^e_{i,N} \frac{N}{m_i} p(N, m_i) = \sum_{N=0}^{\infty} \pi^e_{i,N} \frac{N}{m_i} p(N, m_i),\end{aligned}$$

proving the equivalence.

C.2 Including Capital

In this section, we extend the baseline model to include capital. There are a few ways to include capital. One could assume that capital is completely stuck in each firm across all states of the world. This is captured in our current model, where the entry costs include capital. Alternatively, we could introduce capital that is subject to movement costs just like labor. If the movement costs are the same as labor, then the model does not change at all.

In this subsection, we introduce freely traded capital and show that it is equivalent to our baseline model with an adjusted η. We show the quantitative implications in Section E.1.

C.2.1 Environment

We assume that Japan is a small open economy and can rent capital from international capital markets for an exogenous price r on spot markets in every state of the world. Firms have a Cobb-Douglas technology over labor and capital. They maximize profits by taking wages, the rental rate of capital, and prices as given,

$$y_{isn}(\omega), k_{isn}(\omega) \in \underset{\ell', k'}{\operatorname{argmax}} \quad z_{isn}a_{isn}(\omega)\left((\ell')^{1-\alpha}(k')^{\alpha}\right)^{1-\eta} - w_{isn}(\omega)\ell' - rk'.$$

Denote by $k_i \equiv \int_0^1 \sum_{n \in \mathcal{N}_{is}} k_{isn}(\omega) ds$ the total capital used in location i. National market-clearing (24) is then adjusted to account for the goods that must be exported to pay for the capital used in production,

$$\sum_{i \in \mathcal{I}} c_i \ell_i + rk_i + \psi_i m_i = \sum_{i \in \mathcal{I}} Y_i. \tag{39}$$

C.2.2 Equivalence

We show the equivalence between this economy and an economy with no capital in two steps. We start by solving for optimal capital taking as given the labor choice. We then plug this into the production function to get an adjusted production function only as a function of labor.

Capital solves

$$\max_{k'} z_{isn}a_{isn}(\omega)\ell_{isn}(\omega)^{(1-\alpha)(1-\eta)}(k')^{\alpha(1-\eta)} - rk'.$$

Taking the first order condition, we find that,

$$rk_{isn}(\omega) = \alpha(1-\eta)y_{isn}(\omega).$$

Plugging this in, production is

$$\begin{aligned} y_{isn}(\omega) &= z_{isn}a_{isn}(\omega)\ell_{isn}(\omega)^{(1-\alpha)(1-\eta)}\left(\frac{\alpha(1-\eta)y_{isn}(\omega)}{r}\right)^{\alpha(1-\eta)} \\ &= (z_{isn}a_{isn}(\omega))^{\frac{1}{1-\alpha(1-\eta)}}\left(\frac{\alpha(1-\eta)}{r}\right)^{\frac{\alpha(1-\eta)}{1-\alpha(1-\eta)}}\ell_{isn}(\omega)^{\frac{(1-\alpha)(1-\eta)}{1-\alpha(1-\eta)}}. \end{aligned}$$

Furthermore, payments to labor are then

$$w_{isn}(\omega)\ell_{isn}(\omega) = (1-\alpha)(1-\eta)y_{isn}(\omega).$$

Then the economy is equivalent to one with no capital where the degree of decreasing returns to scale is $\tilde{\eta}$, defined by $1 - \tilde{\eta} \equiv \frac{(1-\alpha)(1-\eta)}{1-\alpha(1-\eta)}$, and production is $\tilde{y}_{isn}(\omega) = (1 -$

$\alpha(1-\eta))y_{isn}(\omega)$. Payments to labor are then

$$w_{isn}(\omega)\ell_{isn}(\omega) = (1-\tilde{\eta})\tilde{y}_{isn}(\omega).$$

Furthermore, $rk_i = \alpha(1-\eta)Y_i$. Therefore, as $\tilde{Y}_i \equiv (1-\alpha(1-\eta))Y_i$, the national market-clearing condition can be rewritten as in our baseline model,

$$\sum_{i\in\mathcal{I}} c_i\ell_i + \psi_i m_i = \sum_{i\in\mathcal{I}} \tilde{Y}_i.$$

D Robustness of Estimation

D.1 Short-run across-firm labor supply by market size.

Here, we examine whether the estimated short-run labor supply elasticity across firms, κ, varies systematically with the size of the local labor market (LLM). In our main analysis, we showed that firms' employment responses to idiosyncratic shocks depend on their relative size within the market—larger firms expand more when productive because they can attract workers from other firms in the same market. This mechanism implies that what matters is a firm's size relative to its local labor pool, not the absolute size of the labor market itself.

Nonetheless, larger LLMs may exhibit stronger aggregate employment responses to wage changes for reasons unrelated to our mechanism, such as greater worker mobility, more diverse industry composition, or different adjustment frictions. To ensure that our results are not driven by such differences across locations, we re-estimate κ separately for large and small LLMs.

Table A1 reports the results. We classify LLMs based on their annual number of firms between 2002 and 2019, defining large LLMs as those that maintained at least 11 firms (i.e., more than 10) in every year of the sample period, and small LLMs as the remainder. This classification yields 1,339 large and 13,228 small markets out of 14,567 in total, and group assignments are fixed over time according to this criterion. Column (1) of Table A1 replicates the full-sample estimate from the main text, while Columns (2)–(3) show the results for the two subsamples.

The estimated elasticities are 1.37 for large markets and 2.44 for small markets, although the latter is not statistically significant due to the weak first stage. The difference between the estimates for all markets and large markets is not statistically significant, and the latter point estimate, which excludes the small markets, is smaller. These results would work against our results which suggest firms in larger markets expand more for the same shocks. Changing the cutoff around this threshold does not materially affect the results.

Table A1: Estimation of Short-run Labor Supply Elasticity across Firms: Subsample by Market Size

Dep. Var.: Log Employment Growth			
	All LLM	Large LLM	Small LLM
Log Wage Growth	1.64	1.36	2.47
	(0.70)	(0.66)	(1.91)
Observations	1,367,957	825,483	542,474
Unique Num. of LLMs	14,567	1,339	13,228
1st Stage F-Stat.	45.72	38.91	8.71

Note: This table shows the estimates of short-run labor supply elasticities across firms, κ, by subsamples. The specification is the same as Column (1) of the Table in the main text. Column (1) reports results for the full sample, which replicates the result in the main text. Columns (2)–(3) split local labor markets by annual firm counts over 2002–2019. Column (2) includes markets that maintained *at least 11 firms* (i.e., more than 10) in *every* year from 2002 to 2019 (i.e., $\min_{t\in[2002,2019]} \#\text{firms}_{ist} \geq 11$); Column (3) contains the remainder. Group assignments are fixed over time for each market based on this criterion. Exposure-robust shift-share standard errors are in parentheses. Market-year fixed effects are included in all the columns. Covariates include the sum of the share ('incomplete share') and its interaction with time fixed effects.

D.2 Shift-Share Instrument Diagnostic

In the main text, we drop the top 10 products with the largest absolute Rotemberg weights and find that the estimates are very similar, suggesting that the results are robust. Table A2 lists these 10 products, with rotemberg weights, log growth in shipments, and the share of shipments. All values are the average between 2002 and 2019.

Table A2: Top 10 Rotemberg Weight Products

Product Name	Rotemberg wt	Shift	Shipment (%)
Steelworks	0.233	0.023	0.31%
Iron and steel prepared products	-0.087	0.046	0.12%
Medical material preparations	0.079	-0.010	0.21%
Sushi and box lunches	-0.078	0.029	0.3%
Fabrication of plastic film, leathers etc	-0.072	0.025	0.26%
Incandescent lamp fixtures	-0.063	0.047	0.07%
Misc. aliphatic intermediates	0.062	0.028	0.21%
Semi-finished green tea	0.059	-0.000	0.02%
Part of semiconductor mfg equipment	0.057	0.054	0.11%
Parts of internal combustion engines for motor vehicles	-0.054	0.029	1.04%

Note: This table reports the top 10 (in absolute values) product categories according to the Rotemberg weights used when constructing shift-share instruments in (16). All values are the average between 2002 and 2019. Shift is the log growth of shipments at the product level. Shipment (%) is the product shipment share in all products.

D.3 Labor Supply Elasticity over a Two-year Horizon

In the main text, we estimate labor supply elasticities (κ, ν) using one-year changes in employment and labor demand. Here, we estimate them using two-year changes. Tables A3 and A4 show the results. The estimates are larger over the longer horizon, as expected, since workers have more time to adjust. In particular, the baseline estimate of κ increases from 1.64 to 2.19, and the baseline estimate of ν increases from 0.82 to 1.37.

Table A3: Estimation of Short-run, Across-Firm Labor Supply Elasticity (2-year)

Dep. Var.: Log Employment Growth	(1)	(2)	(3)	(4)
Log Wage Growth	2.19	2.60	2.95	2.75
	(0.62)	(0.99)	(1.06)	(1.09)
LLM-Year FE	✓	✓	✓	✓
Control Lagged Emp Growth		✓	✓	✓
Drop Top 10 Rotemberg			✓	
Use Lagged Share				✓
1st-stage F Stat.	89.31	51.03	34.81	40.34
Num of Unique LLMs	9,935	9,684	9,645	9,684
Num of Unique Products (Shifts)	1,829	1,825	1,815	1,825
Num of Unique Firms	116,423	113,269	111,239	113,269
Num of Years	16	14	14	14
Shift-Year Obs.	28,866	25,141	25,001	25,137
Firm-Year Obs.	1,273,983	1,044,638	1,018,945	1,044,638

Note: This table shows the estimates of short-run labor supply elasticity across firms within markets, κ, following (15), using two-year changes. All columns include market-time fixed effects, the sum of the shares, and the interaction of the sum and time fixed effects as covariates. Columns (2), (3), and (4) include firm-level lagged log employment growth as an additional covariate. Column (3) drops the top 10 products (units of the shifts) in terms of the absolute values of Rotemberg weights, which we present in Table A2. Column (4) uses lagged shipment share instead of the time-invariant average shipment share to construct the shift-share IV in (16). Exposure-robust standard errors (Borusyak et al., 2025) are in parentheses.

Table A4: Estimation of Short-run, Across-Market Labor Supply Elasticity (2-year)

Dep. Var.: Log LLM Employment Growth				
	(1)	(2)	(3)	(4)
Log LLM Wage Growth	1.37	1.35	1.47	1.32
	(0.23)	(0.22)	(0.26)	(0.22)
Sector-Year FE	✓	✓	✓	✓
CZ-Year FE	✓	✓	✓	✓
Control Lagged Emp Growth		✓	✓	✓
Use Lagged Share			✓	
Weighted				✓
1st-stage F Stat.	195.03	205.86	165.19	1036.30
Num of Unique LLM	13,430	12,707	12,707	12,707
Num of Years	16	14	14	14
LLM-Year Obs.	175,982	146,976	146,976	146,976

Note: This table shows the estimates of short-run labor supply elasticity across markets, ν, following (18), using two-year changes. The unit of observation is the local labor market (LLM)-year level. LLMs are defined as pairs of 3-digit sectors and commuting zones. All columns include sector-time fixed effects, cz-time fixed effects, the sum of the shares ('incomplete share'), and the interaction of the sum and time fixed effects as covariates. Columns (2), (3), and (4) add lagged LLM employment growth as a covariate. Column (3) uses lagged shipment share when constructing the firm-level shift-share IV in (16). Column (4) weights each local labor market using the median total employment over time. Exposure-robust standard errors (Borusyak et al., 2025) are in parentheses.

D.4 Summary Statistics for the Baseline Reduced-form Regression Sample

Table A5 reports summary statistics for the exact estimation sample used in Column (1) of Table 4. We report the distributions separately for the beginning and end of the sample period using 2002 and 2018 as the base years.

Table A5: Summary Statistics: Baseline Reduced-form Regression Sample

	2002						2018					
	Mean	p10	p25	p50	p75	p90	Mean	p10	p25	p50	p75	p90
Establishment-level data												
Employment	61.7	11.0	15.0	24.0	49.0	115.0	71.1	11.0	15.0	26.0	57.0	135.0
Employment Growth (in log)	-0.007	-0.154	-0.069	0.000	0.054	0.143	-0.020	-0.154	-0.069	0.000	0.038	0.102
SSIV Shock (in log)	0.009	-0.072	-0.022	0.005	0.039	0.087	-0.012	-0.068	-0.030	-0.003	0.013	0.032
Observations	84,031						80,809					
LLM-level data												
Total Employment	448.9	17.0	37.0	116.0	366.0	1,017.0	502.1	17.0	39.0	123.0	391.0	1,111.0
Num. of Establishments	7.3	1.0	1.0	3.0	7.0	16.0	7.1	1.0	1.0	3.0	7.0	16.0
Payroll HHI	5,828	1,381	2,713	5,225	10,000	10,000	5,907	1,439	2,800	5,297	10,000	10,000
Observations	11,557						11,448					

Note: This table reports summary statistics for the exact estimation sample used in Column (1) of Table 4. Establishment-level changes use base-year-to-next-year changes, so 2002 corresponds to 2002–2003 and 2018 corresponds to 2018–2019. Local labor market (LLM) statistics aggregate sampled establishments within each LLM-year, where LLMs are defined as 3-digit industry by commuting zone cells.

D.5 Reduced-form Results with Various Fixed Effects

In Table 4 in the main text, we show the baseline reduced-form results. Here, we show how those results change when we vary the fixed effects in Table A6. The negative interaction term remains when we include year fixed effects or CZ-year fixed effects, but it becomes small and statistically insignificant once we include sector-year or local labor market-year fixed effects. This is exactly what we would expect, as those fixed effects absorb the variation in shocks across sectors and local labor markets that identify the mechanism.

Table A6: Responses of Employment to Product-level Shocks (Various Fixed Effects)

	Dep. Var.: Log Employment Growth			
	(1)	(2)	(3)	(4)
Shock	0.049	0.046	0.025	0.021
	(0.005)	(0.005)	(0.005)	(0.005)
Payroll Share	-0.026	-0.026	-0.045	-0.402
	(0.005)	(0.005)	(0.005)	(0.014)
Shock x Payroll Share	-0.025	-0.025	-0.001	0.002
	(0.010)	(0.010)	(0.009)	(0.018)
Implied Ratio	-0.518	-0.530	-0.037	0.097
95% CI	[-0.851, -0.185]	[-0.874, -0.185]	[-0.753, 0.678]	[-1.610, 1.804]
Firm FE	✓	✓	✓	✓
Year FE	✓			
CZ-Year FE		✓		
Sector-Year FE			✓	
LLM-Year FE				✓
Num of Unique LLMs	14,318	14,318	14,317	9,657
Num of Unique Firms	114,553	114,553	114,553	111,585
Num of Years	17	17	17	17
Firm-Year Obs	1,427,772	1,427,772	1,427,760	1,363,080

Note: This table shows the estimates of the impact of firm-level demand shock on employment, following (23). All columns include firm fixed effects. Column (1) includes year fixed effects. Column (2) includes CZ-year fixed effects. Column (3) includes sector-year fixed effects. Column (4) includes local labor market (LLM)-year fixed effects. The implied ratio is the ratio of the coefficient on the interaction term between the shock and payroll share within local labor markets, β_3, to the coefficient on the shock, β_1. The 95% confidence interval of that implied ratio is also reported. Exposure-robust standard errors (Borusyak et al., 2025) are in parentheses for the shock and interaction terms. Standard errors for payroll share are clustered at the firm level.

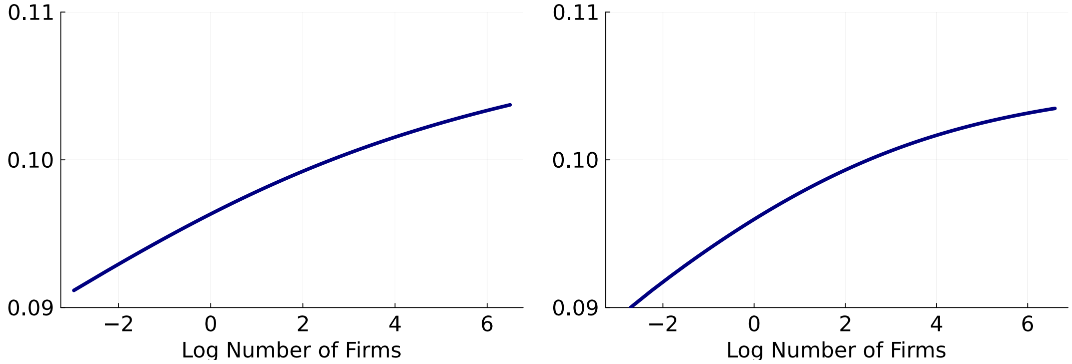
E Robustness of Quantification

E.1 Freely Traded Capital

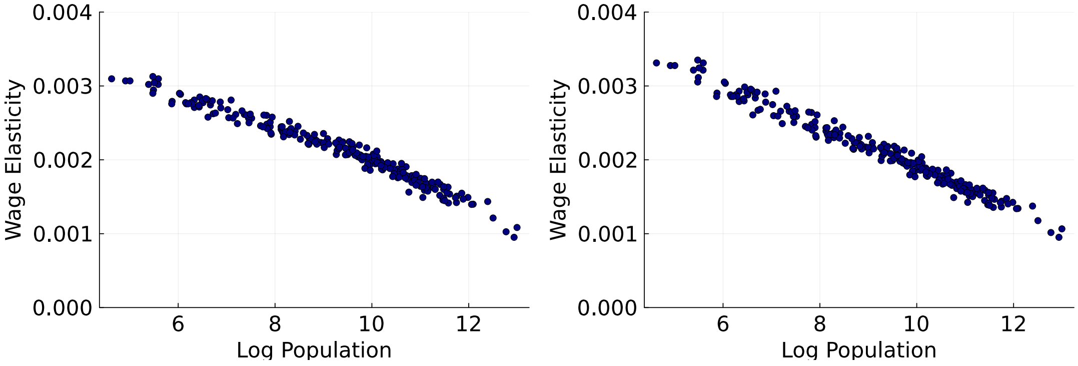
In this subsection, we use the alternate calibration with $1-\tilde{\eta}=\frac{(1-\alpha)(1-\eta)}{1-\alpha(1-\eta)}$, where we take α to match the capital share in Japan.

The results are qualitatively very similar to our baseline results. The implied agglomeration benefits, wage elasticities, and wedges are all slightly smaller, but the patterns across commuting zones and the differences across conduct assumptions remain essentially unchanged.

Figure A7: $\log \Phi^{f}(m)$ and Wage Elasticities (with capital)


(a) Perfect Competition

(b) Cournot Competition

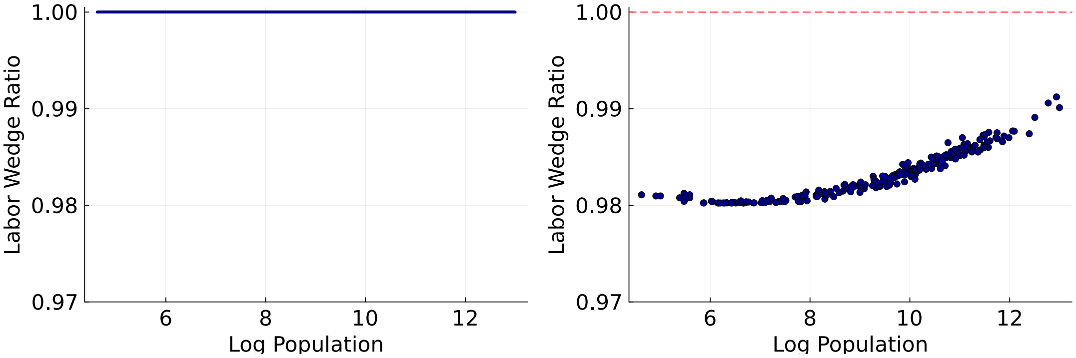

(c) Perfect Competition

(d) Cournot Competition

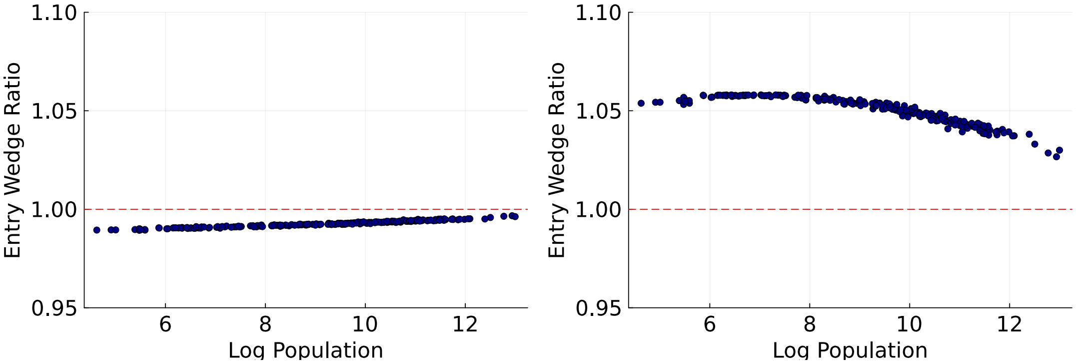
Note: The figures show $\log \Phi^{f}(m)$ against the log number of firms across different numbers of firms and wage elasticity across commuting zones in Japan. The model includes capital. The left panel shows the case of perfect competition, and the right panel shows the case of Cournot competition.

Figure A8: Factor Wedge: Labor and Firm Entry (with capital)

(a) Labor: Perfect Competition (b) Labor: Cournot Competition



(c) Entry: Perfect Competition (d) Entry: Cournot Competition

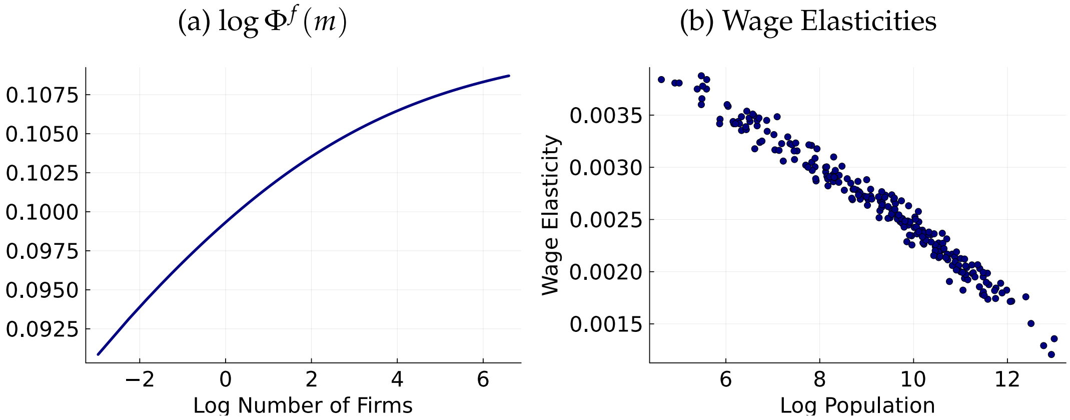


Note: The figures show the labor wedge ratio and firm entry wedge across commuting zones in Japan. The model includes capital. The left panel shows the case of perfect competition, and it is one by construction. The right panel shows the case of Cournot competition, and each dot represents a commuting zone in Japan.

E.2 Bertrand Competition

In this section, we present the results under Bertrand Competition. The results generally fall between the perfect competition and Cournot cases, so the qualitative conclusions are unchanged. This is true both for the strength of the agglomeration benefits in Figure A9 and for the wedges in Figure A10.

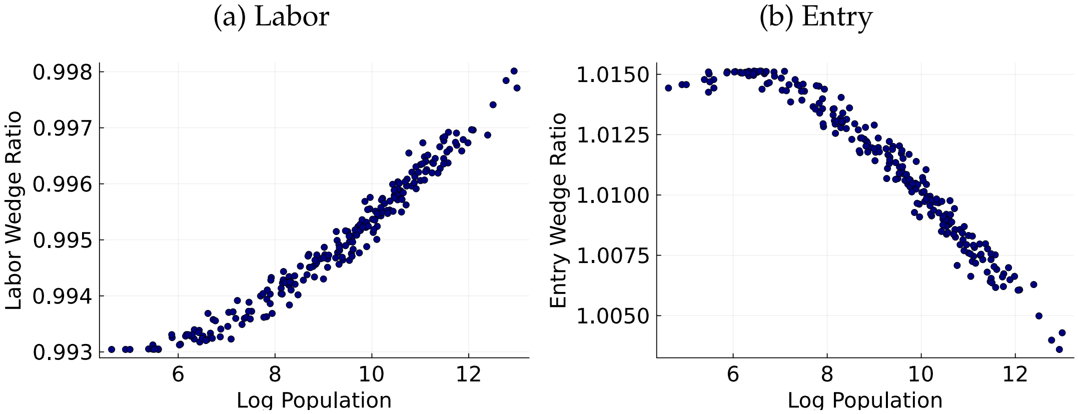
Figure A9: $\log \Phi^f(m)$ and Wage Elasticities (Bertrand)


(a) $\log \Phi^f(m)$

(b) Wage Elasticities

Note: The figures show $\log \Phi^f(m)$ against the log number of firms across different numbers of firms and wage elasticity across commuting zones in Japan, under the Bertrand competition case.

Figure A10: Factor Wedge: Labor and Firm Entry (Bertrand)


(a) Labor

(b) Entry

Note: The figures show the labor wedge ratio and firm entry wedge across commuting zones in Japan, under the Bertrand competition case. Each dot represents a commuting zone in Japan.