# The Granular Origins of Agglomeration *

Shinnosuke Kikuchi
MIT

Daniel G. O'Connor
MIT

April 4, 2024

Preliminary

Click HERE for the most recent version

**Abstract**

A few large firms dominate many local labor markets. This leaves workers vulnerable to firm-specific shocks. If one firm has a bad productivity shock in a small market, workers will be stuck with that unproductive employer, while in a large labor market, workers can move to another firm. Building on that insight, we present a model of local labor markets with a finite number of firms subject to idiosyncratic shocks. We show that there are increasing returns to scale which disappear as the number of firms goes to infinity. We also show that there can be under-entry of firms, especially in small markets. We then test the main mechanism in Japanese administrative data. We first confirm that payroll is less volatile in larger, less concentrated local labor markets. We also show that establishments with larger payroll shares adjust their employment less in response to a demand shock. Finally, we propose a quantitative, granular model of economic geography with free entry of firms and costly mobility of workers across sectors and commuting zones that we use to quantify the mechanism and do counterfactuals.

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

# 1 Introduction

> ...[A] localized industry gains a great advantage from the fact that it offers a constant market for skill. Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skills as theirs and where therefore it is likely to find a good market.
>
> - Alfred Marshall, 1920

Individual firms play a key role in many labor markets. For example, the photography giant Kodak dominated the Rochester, New York labor market for much of the 20th century. At its peak, Kodak accounted for almost a quarter of the city payroll. Similarly, the international car manufacturing firm Toyota hires a large proportion of the workforce in its headquarter-city Toyota. These are extreme examples, but a firm does not need to be that large to have an outsized role in its labor market. Seattle is a big city, but even there, people with specialized skills like software engineers are at the mercy of a few, large firms like Microsoft.

This granularity implies that shocks to individual firms can have a large impact on their labor markets. If Kodak has a single bad year (or many bad years as it were), everyone in Rochester will be impacted. Many workers could end up unemployed or underemployed because there are not many other firms nearby to pick up the slack. By contrast, in a large city with a well-diversified labor market, people can move to one of the many other firms after getting laid off. This is bad for people if they are risk averse and so do not want to be exposed to idiosyncratic shocks, but it is also inefficient to have large amounts of unemployment because one firm has a bad year.

In this paper, we study the role that individual firms subject to idiosyncratic shocks have on the geography of economic activity. Our key theoretical result is to show that local labor markets feature aggregate increasing returns to scale for the reason that Alfred Marshall laid out in his discussion of labor market pooling. In small markets, there are very few firms so even if a firm has a bad productivity shock, people still work for it because there are not many other options. By contrast, large markets offer a "constant market for skill." If one firm has a bad shock, there are plenty of other firms nearby who would be happy to hire the workers. Therefore, larger labor markets are more productive and granularity gives a reason for people to agglomerate in cities. However, this source of increasing returns to scale disappears in the limit with an infinite number of firms so models with a continuum of firms miss it.

Importantly, micro-founding this source of agglomeration implies different optimal policies than those emphasized in the economic geography literature, i.e. Fajgelbaum and Gaubert (2020). Those papers take production in each location as a black box. Given the amount of labor in a location, that location produces some amount. Therefore, the only policies they can consider are policies that move labor around. We microfound the agglomeration which allows us to target the policy at the distortion. In our granular model, the distortion falls on the firm side, so the optimal policy would feature place-based firm entry subsidies like we see in the real world rather than location-specific wage subsidies.

Our key contribution is linking our conceptual framework to explicit empirical tests of the mechanism. This labor market pooling mechanism relies on two key features: (i) the variance of wages should be decreasing in the size of the labor market and (ii) firms should have an easier job finding workers in larger markets than in small ones. Prediction (i) is something of a necessary condition. Putting individual firms together in larger markets will only increase productivity if the firms are subject to idiosyncratic shocks and so workers could benefit from being able to move between them. If firms are subject to idiosyncratic shocks, we should see them "average out" to some extent in larger markets so that wages do not move around as much. Our second prediction speaks directly to the mechanism. The mechanism relies on workers moving to the more productive firms in larger markets. Therefore, we should see that firms in larger markets can expand more after a productivity shock. Furthermore, if firms are subject to similar-sized shocks, we should see that firms that are in larger markets should have a higher variance of log employment.

We test these empirical predictions using Japanese firm data. Following the theory, we use the number of firms in the labor market as our preferred measure of size. Defining a labor market as a 2-digit JSIC industry $\times$ commuting zone, we show that the variance of the log wage bill is decreasing in size. Not only that, but a log linearization of the model implies the log-linear relationship we see between the variance of the log wage bill and the log number of establishments. We also confirm that this is not being driven by workers coming in from other labor markets, and is instead mostly operating through the increase and decrease in the wage of the workers.

Having established that idiosyncratic shocks to firms are important, we see if firms take advantage of the diversified larger markets. Using data on each establishment's product mix along with data on what countries buy which products, we construct a measure of exposure of each establishment to real exchange rate movements in each country. With these shares, we construct shift share productivity shocks based on the real exchange rate movements of each country relative to the Japanese Yen. With this constructed demand shock, we see how much establishments respond in large markets relative to small markets. We show that firms that already hire a large portion of the market respond less to a demand shock, consistent with our theory.

That test considers very particular types of idiosyncratic demand shocks but leaves open the possibility that the idiosyncratic demand shocks necessary for our pooling mechanism do not matter much. To get some notion of quantitative importance, we construct a theory-consistent measure of the variance of log employment by market. We show in the data that this measure is increasing in the size of the market, consistent with firms finding the workers they need in response to many possible demand shocks.

We then present a quantitative model of granular economic geography. This idea of a finite number of firms subject to idiosyncratic shocks leading to a labor pooling reason for agglomeration has existed in the literature at least since Krugman (1992). However, it has stayed at the edge of the field, partially because models with a finite number of firms are difficult to work with. Our model remains tractable with an arbitrary number of locations each with a continuum of freely traded sectors. Firms freely enter but cannot direct their entry to any one sector. This allows us to keep a finite number of firms in each sector while a free entry condition holds with equality. After entering, firms then get ex-ante productivity shocks which determine their average size. Workers decide where they would like to live and invest in sector-specific skills. Then, ex-post productivity shocks for each firm are revealed and workers allocate their labor across firms and sectors subject to frictions. We show that our qualitative insights from the simple model hold in this quantitative framework. We also further characterize the optimal entry subsidies.

Calibrating the model using our regressions, we show that a large portion of the increasing returns to scale noted in the literature can be explained by this mechanism. To match an elasticity of wages to population of 0.03, our model only requires other externalities that have an elasticity of 0.005. For some small locations, our mechanism implies an elasticity of wages to population of 0.4. Furthermore, for those smallest locations, firm profits represent less than 80% of their contribution to production, suggesting a significant amount of under-entry. Finally, we consider a counterfactual where Japan's working age population drops by 10%. The remaining population becomes biased more towards large cities because the externality becomes stronger with a small population. Furthermore, 6 commuting zones see their manufacturing sector completely unravel.

### Related Literature

This paper builds on a large literature studying the agglomeration of economic activity going back at least to Marshall (1920). Marshall systematically explains how labor market pooling, sharing of inputs, and knowledge spillovers can all explain the patterns we see. Since then researchers have formalized these mechanisms, considered new mechanisms, and looked for evidence that they matter. Duranton and Puga (2004) provide a good review of the formal models for these (and other) mechanisms, and Rosenthal and Strange (2004) discuss the evidence.[1]

We focus on the labor market pooling mechanism. Our model builds fairly explicitly on the stylized model in the appendix of Krugman (1992). Krugman (1992) showed that when a finite number of firms are subject to idiosyncratic shocks, there are increasing returns to scale. Stahl and Walz (2001) extend the model to have multiple sectors to see the implications for coagglomeration. Compared to these papers, our model sacrifices some analytic tractability for realism. This ensures that we can get clear empirical predictions that can be brought directly to data and used to determine how important this mechanism is.

Several papers have tested some of the predictions from Krugman (1992). Overman and Puga (2010) found that sectors in the UK that experience greater idiosyncratic volatility are more spatially concentrated, as implied by the theory. de Almeida and de Moraes Rocha (2018) confirmed this in the Brazilian setting, and Nakajima and Okazaki (2012) showed it in Japan. More similar to our exercise, Andini et al. (2012) and Gan and Zhang (2006) look for direct evidence of the mechanism. Gan and Zhang (2006) shows that larger cities have shorter unemployment cycles, and Andini et al. (2012) shows that there is more worker turnover in larger labor markets. Compared to these papers, we allow for ex ante heterogeneous firms and do not impose structural assumptions on the productivity shocks. We then derive empirical predictions we test directly.

Our paper also relates to a growing literature on the importance of granularity in understanding the economy. Gabaix (2011) demonstrated that because the distribution of firms is thick-tailed, even though there are many firms, shocks to individual firms can have a sizable impact on the economy. We show that the thick-tailed firm size distribution implies that the labor market pooling externality is relevant for medium-sized cities, not just small towns.

The rest of the paper is laid out as follows. In section 2, we present a simple model of labor

[1] In the case of Japan, there are several papers, that study the source of agglomeration, including production externality (Nakamura, 1985), production and consumption variety gain (Tabuchi and Yoshida, 2000), firm-to-firm transaction (Nakajima et al., 2012; Miyauchi, 2018), labor market pooling (Nakajima and Okazaki, 2012), consumption access (Miyauchi et al., 2021).

market pooling taking as given the firms and workers. We test the main empirical predictions in section 3. Section 4 presents a quantitative version of the model which can be used to quantify the importance of the mechanism. Section 5 then calibrates the model and does the counterfactual, and Section 6 concludes.

# 2 Theoretical Framework

We present a model of a local labor market with a finite number of firms subject to idiosyncratic uncertainty. The model will demonstrate how granularity leads to increasing returns to scale and tease out some of the implications for optimal policy. We then discuss the testable implications of the model which speak directly to the mechanism.

## 2.1 The Model

We consider a small open region with a finite number $E$ of establishments, indexed by $e \in \mathcal{E}$, and a mass $\ell$ of workers.

**Establishments** Each establishment has decreasing returns to scale technology that only uses labor. An establishment $e$, produces $y_e(s)$ in state of the world $s$ according to

$$y_e(s) = z_e a_e(s) f(\ell_e(s))$$

where $z_e > 0$ is the ex-ante productivity of establishment $e$, $a_e(s) > 0$ is the idiosyncratic ex-post shock of establishment $e$, $f(x) = x^\eta$, and $\eta \in (0,1)$ is the elasticity of production to employment.[2] $a_e(s)$ are iid with cdf $G(\cdot)$. We assume that the first and second moments exist for both $a_e(s)$ and $\log a_e(s)$ and that $\mathbb{E}[\log a_e(s)] = 0$. We further assume that $\mathbb{E}[a_e(s)^{\frac{1}{1-\eta}}]$ exists and is finite.

Establishments hire labor in competitive markets after the state of the world $s$ is revealed. They maximize profits taking as given world prices and wages in its local labor market. Since sectoral prices are set by the rest of the world and do not move, we normalize productivity so that all prices are 1. Then

$$\ell_e(s) \in \underset{\ell'}{\operatorname{argmax}} \quad z_e a_e(s) f(\ell') - w(s)\ell' \tag{1}$$

for every establishment $e$ and state of the world $s$.

[2] The constant elasticity assumption is not necessary for the main theoretical results. It only simplifies our empirical predictions.

**Workers** Workers inelastically supply one unit of labor. Labor market clearing then requires that

$$\ell = \sum_{e \in \mathcal{E}} \ell_e(s) \tag{2}$$

in every state of the world $s$.

**Equilibrium** An equilibrium consists of wages in every state of the world $\{w(s)\}_{s\in\mathcal{S}}$ and labor choices for every establishment in every state of the world $\{\ell_e(s)\}_{e\in\mathcal{E},s\in\mathcal{S}}$ such that

- Establishments choose labor to maximize profits taking as given wages and prices (1); and
- The labor market clears in every state of the world (2).

## 2.2 The Granular Origins of Agglomeration

The equilibrium is simple and easily characterized. The establishment's first order condition for profit maximization (1) gives an expression for establishment labor demand: $z_e a_e(s) f'(\ell_e(s)) = w(s)$. We solve for $\ell_e(s)$ as a function of wages, and plug that into the labor market clearing condition (2) to get

$$w(s) = \eta \ell^{\eta-1} \left[ \sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}} \right]^{1-\eta} \tag{3}$$

This expression for wages along with establishment labor demand completely characterizes the equilibrium. But we are ultimately interested in production, so we plug the expression for $\ell_e(s)$ for all establishments back into the production function and sum across establishments. This gives production in each state of the world

$$Y(s) = \ell^{\eta} \left[ \sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}} \right]^{1-\eta}.$$

We denote by $Y(\ell, \mathcal{E})$ the expected production of a location with $\ell$ workers and the set $\mathcal{E}$ of establishments. It is given by

$$Y(\ell, \mathcal{E}) = \mathbb{E}\left[ \ell^{\eta} \left[ \sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}} \right]^{1-\eta} \right], \tag{4}$$

where expectations are taken over the state of the world $s$. Our model is competitive so not only is this the expected production in equilibrium, but it is also the maximum expected production a planner could achieve with $\ell$ workers and the set of establishments $\mathcal{E}$.

Before we confirm that there are increasing returns to scale, we need to define what it means for one labor market to have "more" firms when firms can differ in their ex-ante productivity.

**Definition 2.1.** Suppose that $\mathcal{E}$ and $\mathcal{E}'$ are sets of establishments and $\alpha \in \mathbb{R}_{>0}$. Then we will say that $\mathcal{E} = \alpha\mathcal{E}'$ if, for every possible ex-ante productivity $z \in \mathbb{R}_{>0}$, the number of establishments in $\mathcal{E}$ with ex-ante productivity $z$ is equal to $\alpha$ times the number of establishments in $\mathcal{E}'$ with ex-ante productivity $z$.

That is to say, $\alpha\mathcal{E}$ simply has $\alpha$ times more of every type of firm compared to $\mathcal{E}$. This gives a partial ordering over the size of labor markets, and our first proposition confirms that there are increasing returns to scale with respect to that ordering.

**Proposition 1.** *There are increasing returns to scale if idiosyncratic shocks have a positive variance. In math, for all $\ell > 0$, $\mathcal{E}$, and $\alpha > 1$, if $Var(a_e(s)) > 0$ then*

$$Y(\alpha\ell, \alpha\mathcal{E}) > \alpha Y(\ell, \mathcal{E}).$$

*Proof.* We leave a formal proof of proposition 1 for the appendix. Here, we sketch a proof when $\alpha = 2$.

We have twice the number of workers and twice the number of establishments. Suppose that we split them up into two separate markets that are exact copies of the original market. In that case, production would be $2Y(\ell, \mathcal{E})$. To prove that $Y(2\ell, 2\mathcal{E}) > 2Y(\ell, \mathcal{E})$, we simply need to show that we could increase production any amount over that separated labor market benchmark.

Notice that because $a_e(s)$ are non-degenerate and independent, there must be some set of times with positive measure where wages in the first labor market are higher than wages in the second labor market. Then we can simply move a small amount of labor from the second labor market to the first one during those times. Since wages were higher in the first market, the workers have a higher marginal product, and production must increase. Therefore, this deviation will produce more than the separated benchmark and $Y(2\ell, 2\mathcal{E}) > 2Y(\ell, \mathcal{E})$. □

In words, proposition (1) says that doubling the number of people and establishments more than doubles the average production. This happens because each establishment is different. They have different productivity shocks and are looking to hire people at different times. When you put more establishments in the same labor market, workers are able to move to the most productive one, while in a small labor market, workers are stuck working at a single establishment even if it is not productive.

We can also view this from the establishment's perspective. Consider again establishment $e$'s labor demand (1):

$$z_e a_e(s) f'(\ell_e(s)) = w(s). \tag{5}$$

That is, the marginal product of labor equals the marginal cost of a worker. Now, imagine that establishment $e$ has a positive productivity shock so that $a_e(s)$ is high. Holding the amount of labor fixed, this raises an establishment's marginal product of labor. Therefore, if wages remain constant, the establishment needs to hire more workers until the marginal product of labor falls to equal those wages. Conversely, if the establishment does not hire any more workers, then wages must rise. In most cases, the adjustment will not fall completely on one channel. Instead, the wages will rise to some extent, and the establishment will hire some more people. The incidence of adjustment will depend on the labor market.

If the establishment is the only one in the labor market, it always must hire everyone, even when it is unproductive. Therefore, wages will move around wildly and the firm will be relatively unproductive given the average number of people it hires. On the other hand, if the establishment is in a market so large that wages are basically constant, the establishment will hire more workers when it is productive, and those workers will find other work when the establishment is doing poorly. Therefore, the average productivity of workers at the establishment will be higher. Here, the productivity benefits of a large market are a natural consequence of establishments and workers finding each other when establishments want to expand.

This mechanism differs from many other agglomeration externalities in that it is not log-linear. Going from one establishment to two establishments significantly decreases the volatility of wages and thereby increases productivity. By contrast, a location with 100 establishments already has a low variance of log wages. Doubling the number of establishments to 200 does very little to make establishments more efficient since they had no issue finding workers to begin with. In the next proposition, we confirm that the importance of agglomeration is decreasing and actually disappears in the limit.

**Proposition 2.** *As the labor market becomes infinitely large, the aggregate production function approaches* *constant returns to scale. In math, suppose that $\ell > 0$, there is a set of establishments $\mathcal{E}$, and $\alpha > 1$. Then*

$$\frac{Y(\alpha\kappa\ell, \alpha\kappa\mathcal{E})}{\alpha Y(\kappa\ell, \kappa\mathcal{E})} \to 1$$

*as $\kappa \to \infty$.*

This has two main implications. First of all, this says that this force for agglomeration disappears in the limit with a large number of establishments. Therefore, traditional models with a continuum of establishments miss this force by necessity. Second, the fact that the strength of agglomeration is decreasing in the size of the market has important implications. If one were to move a few establishments and workers from a large market to a small market, the large market would be largely unaffected. Establishments will still have no issue finding workers when they need them. By contrast, the small market could see a large increase in observed productivity because of the decline in misallocation. We discuss more of the normative implications in the next section.

### 2.3 A New Reason for Spatial Policy

The model presented thus far takes as given the establishments and workers in the labor market and assumes they interact in competitive markets. Because of that, everything is efficient conditional on those factors of production, and there is no efficiency reason for direct intervention in the labor market.

However, we have shown that there are increasing returns to scale at the aggregate, local labor market level. It is therefore impossible for both firms and workers to be paid their marginal product. When there are constant returns to scale, the payments to the factors are exactly equal to total production by Euler's Homogenous Function Theorem. When there are decreasing returns to scale, the marginal product is lower than the average product so there will be profits leftover. When there are increasing returns to scale, the sector would need to pay out more than its total earnings, which it cannot do. The only remaining question is who is not capturing their full contribution: workers, firms, or a little bit of both.

It is very easy to calculate the marginal product of labor. It is simply the derivative of expected production with respect to labor. The expression is

$$\frac{\partial Y(\ell,\mathcal{E})}{\partial \ell} = \eta \ell^{\eta-1}\mathbb{E}\left[\left[\sum_{e\in\mathcal{E}}(z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right],$$

which is just expected wages. So workers are paid their full marginal product. It then follows that firm profits must be lower than their marginal product.

**Proposition 3.** *Adding new firms increases expected production more than the profits those firms would earn. In math, for $\alpha > 1$,*

$$\mathbb{E}[\sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}}\pi_e(s)] < Y(\ell,\alpha\mathcal{E}) - Y(\ell,\mathcal{E}),$$

*where $\pi_e(s) = z_e a_e(s)\ell_e(s)^\eta - w(s)\ell_e(s)$ are the profits earned when there are $\alpha\mathcal{E}$ set of firms operating.*

This fact alone is not enough to cause an inefficiency. We need that the supply of establishments to the local labor market is somewhat elastic so that the number of entrants is distorted downwards. Suppose that we were to follow Mankiw and Whinston (1986) and assume there are several potential entrants that can pay some fixed cost to enter. We order them by their ex-ante productivity so that the most productive firm enters first, then the next, and so on. They enter up until the point that all operating firms make higher profits than that fixed cost of entering, and if the marginal firm were to enter, it would make profits less than that cost. Proposition 3 says there will be times when that marginal firm would not enter when the planner would want it to! Conditional on entering, profits will be lower than the fixed cost even though expected production would increase by more than that cost.

Under-entry will be especially bad when the increasing returns are strongest: for example if the firms are subject to large amounts of idiosyncratic uncertainty or there are few firms. One can think of firm profits as the difference between total production and the marginal product of labor. As expected production gets closer to constant returns to scale, firm profits get closer to their marginal product. Therefore, when there are few idiosyncratic shocks, under-entry will not be too bad. Similarly, in really large markets where the agglomeration force weakens, there will be less under-entry.

Why does the First Welfare Theorem fail to hold in this setting? Because the free entry condition in Mankiw and Whinston (1986) is not Walrasian entry. In a competitive equilibrium, all firms need to take wages as given. That includes firms that have not currently entered. Under Walrasian entry, firms cannot ask, what would their profits be were they to enter, as they would in Mankiw and Whinston (1986)-style free entry. Instead, they need to ask, given the current distribution of wages, what would their profits be? In Mankiw and Whinston (1986)-style free entry, firms internalize that if they were to enter, wages would be higher. Not only would wages be higher on average, but wages would be higher precisely when the firm wants to hire more workers since their own shocks affect wages. That reduces expected profits and discourages entry relative to the competitive benchmark.[3]

Mankiw and Whinston (1986) show that if firms internalize their market power, there tends to be over-entry. In our model, if there are no ex-post shocks and firms were to internalize their monopsony power, that would also be the case. In that case, firms hire fewer workers, distorting wages down and profits up. Therefore, it is the workers who are not getting paid their full marginal product, and government policy would want to subsidize labor. When there are ex-post shocks, there could be either over-entry or under-entry of firms depending on the strength of our mechanism. This is a quantitative question that future work could look into.

### 2.4 Testable Implications of the Theory

Next, we turn to the testable implications of the theory. The productivity benefits rely on two main predictions: the variance of wages is decreasing in the size of the labor market and firms in larger labor markets have an easier time finding workers when they want them. We will discuss each of these predictions in turn.

#### 2.4.1 The Variance of Wages is Lower in Large Markets

One necessary condition for our mechanism to matter is that the variance of wages decreases in the size of the market. In the model, this happens because establishments are subject to idiosyncratic shocks which "average out" in large markets with a lot of establishments. If we did not see that the variance of wages is lower in larger markets, then firms must not be subject to idiosyncratic shocks in any meaningful way. Instead, we would say that productivity shocks happen at the local labor market level, and pooling more firms would not help with the misallocation

[3] One might ask if we could use competitive entry. In general, an equilibrium would not exist. That is because, with competitive entry, current operating firms need to earn non-negative profits while non-operating firms would need to earn non-positive profits given the current distribution of wages. However, the current distribution of wages is correlated with the productivity shocks of the current operating firms and uncorrelated with the shocks of the non-operating firms. All else being equal, that would imply the non-operating firm earns higher profits than the operating firm. The only way for the equilibrium to exist is if the current operating firms have sufficiently higher ex-ante productivities than the non-operating firms to swamp that correlation issue.

problem.

In our baseline model, firms are subject to idiosyncratic shocks. Therefore, the exposure of the market to firm-specific shocks is proportional to each firm's average payroll share.

**Proposition 4.** *To a first order log-linear approximation around $a_e(s) = 1$, we have that*

$$var(\log w(s)) = \left(\sum_{e \in \mathcal{E}} x_e^2\right) \sigma^2 \tag{6}$$

*where $x_e = \frac{w\ell_e}{\sum_{e'} w\ell_{e'}}$ is the firm's payroll share in the equilibrium with no shocks ($a_e(s) = 1$) and $\sigma^2 = var(\log a_e(s))$.*

This proposition relates the variance of log wages to the HHI of the labor market, but it leaves open the relationship to the overall size of the market. To relate this prediction to the size of markets, we follow Gabaix (2011) in assuming that the ex-ante distribution of productivity shocks is distributed Pareto.

**Proposition 5.** *Suppose that ex-ante productivity is drawn from a power law distribution*

$$\mathbb{P}[z_e > z] = az^{-\lambda},$$

*for $z_e > a^{1/\lambda}$. Then as $E \to \infty$, the variance of log wages follows*

$$var(\log w(s)) \sim \frac{v_\lambda}{(\log E)^2}\sigma^2 \qquad \text{if } \lambda(1-\eta) = 1,$$

$$var(\log w(s)) \sim \frac{v_\lambda}{\left(E^{1-\frac{1}{\lambda(1-\eta)}}\right)^2}\sigma^2 \quad \text{if } 1 < \lambda(1-\eta) < 2,$$

$$var(\log w(s)) \sim \frac{v_\lambda}{E}\sigma^2 \qquad \text{if } \lambda(1-\eta) \geq 2.$$

With a Pareto distribution, the log variance of log wages will actually be linear in the log number of firms. And the coefficient relating the two in regression will tell us something about the distribution of ex-ante firm shocks. If the distribution of ex-ante shocks is relatively weak-tailed, so that $\lambda(1-\eta) \geq 2$, then the variance of log wages decreases relatively quickly. Consistent with the Central Limit Theorem, the standard deviation declines at a rate $E^{-1/2}$, so the variance declines at the rate $E^{-1}$. But if the distribution of ex-ante shocks is so thick-tailed that it does not have a variance, the Central Limit Theorem fails. Then instead of declining at a rate $E^{-1}$, the variance shrinks at a rate $E^{-\phi}$ with $\phi \in (0,1)$. If the tail is very thick, i.e. $\lambda(1-\eta) = 1$, the variance decays even slower.

The coefficient also tells us how important our labor market pooling mechanism is for different-sized markets. The productivity gains are driven by firms being able to find workers when they need them. If the variance of log wages is really small, then firms will have no issue because the wages are not rising much when they have a good shock. So for our mechanism to matter, the variance of log wages must be sizable. When the distribution of firms is relatively thin-tailed, the idiosyncratic shocks to firms will average out for relatively small markets and our mechanism will not matter for medium or large markets.

When the distribution of ex-ante firm productivity is thick-tailed, then large firms continue to play an outsized role in moderate- and large-sized markets. Those firms continue to affect the variance of log wages and continue having a tough time finding workers even in those markets. Therefore, adding more firms can still increase productivity.

### 2.4.2 Firms Find Workers More Easily in Large Markets

The second prediction is that establishments in larger labor markets have an easier time finding workers when they need them. The fact that the variance of wages is lower in large markets is only a necessary condition. Suppose that workers could not move firms. In that case, there would be no productivity benefit of pooling the firms together because workers will not move to the more productive firm in their market. Yet, we would still see a decline in the variance of log wages if we are taking average wages at the market level.

Firms in larger markets need to be able to expand their employment in response to a productivity shock. We start by considering the comparative static of a firm getting an idiosyncratic shock.

**Proposition 6.** *In a first-order log approximation, employment of establishment e responds more to a productivity shock if it hires a small proportion of the workforce. In math,*

$$\Delta \log \ell_e(s) \approx \frac{1}{1-\eta} \left(1 - x_e(s)\right) \Delta \log a_e(s) \tag{7}$$

*where* $x_e(s) = \frac{w(s)\ell_e(s)}{\sum_{e'} w(s)\ell_{e'}(s)}$.

Intuitively, if a firm already hires a large proportion of the local labor market, it will need to raise the wage significantly to attract the few remaining workers away from the other firms.

However, as the wage rises, the firm becomes less interested in expanding. Therefore, the net effect is that firms that hire a large proportion of the labor force do not expand as much in response to a productivity shock.

In larger labor markets, there are more firms, so each individual firm will hire a smaller share of the workforce. That means those firms will have an easier time expanding in response to a positive productivity shock. We will test this comparative static directly below with a clean productivity shock. But we also want to have some idea of how important shocks are in general. To do that, we construct a measure of the average variance of log employment by market and show that that should increase in the size of the market.

**Proposition 7.** *To a first-order approximation around $a_e(s) = 1$, the weighted average variance of log employment is decreasing in HHI. In math,*

$$\sum_e x_e var(\log \ell_e(s)) \approx \frac{\sigma^2}{1-\eta}\left[1 - \left(\sum_e x_e^2\right)\right]$$

*where* $x_e = \frac{w\ell_e}{\sum_{e'} w\ell_{e'}}$.

Then we can transform this cross-section prediction on the payroll HHI into a prediction about the size of the markets by noting that $\sum_e x_e^2 \to E^{-\phi}$ as the number of establishments goes to infinity.

If these predictions hold true, we will have some direct evidence that firms have an easier time finding workers when they want to expand in larger markets. This mechanism is exactly what drives the agglomeration effects of granularity and so will be clear evidence that this mechanism matters.

### 2.5 Robustness to Alternate Labor Market Assumptions

So far, we have assumed that labor markets are perfectly competitive and workers can freely move between establishments. Here we consider how our results would change if we weaken those assumptions.

**Imperfect Mobility Across Establishments and Locations** The essential assumption for this mechanism is that people are more mobile between establishments within the same labor market than across labor markets. In the limit where location does not matter, this mechanism disappears, and there are no productivity benefits to being in larger labor markets. When we quantify the mechanism in Section 4, we allow for imperfect mobility across establishments but not across locations as it is not relevant on our time scale. The variance of observed log wages is still lower in larger labor markets. However, that does not give a good indication of the declining misallocation. Instead, one needs to look at the variance of employment to see how much easier it is for establishments to find workers in large labor markets.

**Labor Market Monopsony Power** Establishments are large in this model and thus can affect the equilibrium wages. So far, we have assumed that establishments simply ignore that market power, but they could internalize it. Then, as the labor markets become larger, the monopsony power weakens which improves efficiency. Therefore, monopsony power is another force making larger labor markets more efficient. As for the empirical predictions, when establishments have monopsony power, they will not respond as much to their productivity shocks. A really productive establishment would want to keep the equilibrium wage low by under-hiring while a less productive establishment has less market power and so does less under-hiring. Therefore, establishments that internalize their market power will have lower variance of wages and lower variance of employment in smaller labor markets. Comparing the variance of log wages in large labor markets to that in small labor markets will understate the productivity advantages of the large labor market. Comparing the variance of log employment of an establishment in a small market to that in a large market will show the reallocation benefits of the labor market pooling mechanism, which strengthens with declining market power.

**Labor Hoarding** Establishments with monopsony power engage in labor hoarding since they know that even though they are not productive today, they might be more productive in the future. At that point, they will want more employees. Therefore, they hold onto some employees now. The two features necessary for labor hoarding are monopsony power and friction in finding new workers. To the extent that enlarging a labor market decreases monopsony power and makes it easier to find workers when establishments need them, labor hoarding becomes less necessary, and the labor market becomes more efficient. Thus, the existence of labor hoarding increases the gains from this labor pooling mechanism. The empirical predictions mirror those when establishments have monopsony power. The declining wage variance will understate the productivity benefits while the increase in employment variance should be accurate.

**Wage Rigidity** Now suppose that wages are rigid. Large labor markets will not be affected much because there is very little variance in the demand for labor. By contrast, small labor markets have large swings in demand. When wages are rigid, this will lead to spikes in unemployment, which is even more unproductive than the case when wages adjust. Thus, the labor market pooling mechanism is stronger in a world with wage rigidity. For the empirical predictions, if wages are perfectly rigid then the variance of log wages will be zero. However, if wages imperfectly adjust, wage variance still decreases in the number of establishments but that understates the productivity benefits. Similarly for the variance of employment: when wages are perfectly rigid, there should be no difference between establishments in large markets and small. When wages adjust slowly, the variance of log wages will understate the benefits.

# 3 Empirical Evidence

We test our mechanism on manufacturing sectors in Japan where we have detailed establishment level data.

## 3.1 Data

**Japanese Census of Manufactures Data** Our primary data source is the Census of Manufacture (CoM) in Japan for the manufacturing sector. The Ministry of Economy, Trade, and Industry (METI) conducts the Census of Manufacture annually to gather information on the current status of establishments in the manufacturing sector. Specifically, this census covers all manufacturing establishments in years when the last digit of the survey year is 0, 3, 5, or 8. For other years, the census covers all establishments with at least 4 employees in Japan. The CoM survey was not conducted in 2012 and 2016, and instead, another survey, the Economic Census for Business Activity (ECBA) was conducted by METI and the Ministry of Internal Affairs and Communications for data in 2011 and 2015.[4] We used the ECBA survey to substitute the CoM survey in 2011 and 2015.

This data has two advantages. First, we observe panels of all the establishments with at least 4 employees. This feature allows us to compute volatility measures, for example, the variance of employment growth, within each establishment across periods and to compute local labor market

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

concentration measures, including several establishments and HHI across local labor markets.[5] Second, we observe yearly shipment values by detailed product categories for each establishment. This enables us to construct establishment-level exposure to foreign exchange rate changes using product shipment share and national, product-destination level export data.

**UN Comtrade Data** We supplement the CoM data with bilateral trade flows by product-year level from UN Comtrade data. First, we take annual values of traded goods from 1980 to 2016 across 4-digits product categories in SITC Rev. 2. We then convert them into HS Second, using a cleaner provided by Feenstra and Romalis (2014), we convert data at SITC Rev.2, 4-digit level across countries over time. This step gives primacy to importer's reports over exporter's reports where available, corrects values where UN values are known to be inaccurate, accounts for re-exports of Chinese goods through Hong Kong, and puts Taiwan back as an importer and an exporter.[6] Third, we combine some of the countries, which unify or report jointly for subsets of years in the database. we combined East and West Germany before reunification, Belgium and Luxembourg; the islands that formed the Netherlands Antilles; North and South Yemen; and Sudan and South Sudan. Fourth, we convert these SITC Rev.2, 4-digit industrial categories into HS 2007, 6-digit using the crosswalk provided by the United Nations.[7] Finally, we convert these data in the 6-digit HS 2007 product code into 6-digit product categories used in the CoM data. We newly construct crosswalk files based on the crosswalk provided in Baek et al. (2021).

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

**Sample Construction** First, we restrict samples to establishments with a minimum of 4 employees.[8] This is necessary to construct a panel of establishments at an annual frequency. Second, we construct a panel of establishments. While the CoM survey does not contain time-consistent establishment codes, RIETI provides a converter to enable researchers to link establishments across different years since 1986. Finally, we keep establishments, which appear in at least 5 years consecutively. This is because we compute variance over time within establishments and need sufficient

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

[6]Their cleaner is available here.

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

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

Table 1: Summary Statistics: Local Labor Market

| | Num. | Mean | Std. Dev. | p25 | p50 | p75 |
|---|---|---|---|---|---|---|
| Num. of Estab. | 4,782 | 55.32 | 184.23 | 3.00 | 13.00 | 45.00 |
| Employment HHI | 4,782 | 3169.53 | 3240.79 | 702.23 | 1808.67 | 4827.50 |
| Employment | 4,782 | 1702.27 | 4914.64 | 67.00 | 362.50 | 1434.00 |
| Log Avg. Wage | 4,782 | 5.78 | 0.37 | 5.58 | 5.82 | 6.02 |
| Log Variance of Log Wage Growth | 4,782 | -4.38 | 1.28 | -5.35 | -4.45 | -3.48 |
| Log Variance of Log Payroll Growth | 4,782 | -2.92 | 1.38 | -3.94 | -3.00 | -2.00 |
| Weighted Avg. of Log Emp. Growth | 4,782 | -3.40 | 0.68 | -3.67 | -3.35 | -3.06 |

*Note*: The tables show the summary statistics for the data used in the analysis across local labor markets. The variance of log wage growth, the variance of log payroll growth, and the weighted sum of log employment growth are in log units.

observations for each establishment.[9] Our final sample is an unbalanced panel of 724,417 unique establishments in manufacturing sectors from 1986 to 2016.

**Definition of Local Labor Markets** We define a local labor market as a pair of a JSIC 2-digit manufacturing industry and a commuting zone. To construct time-consistent industrial categories, we convert categories in each year into the one used in 2011, using a crosswalk file provided by RIETI. This leaves us with 23 unique 2-digit manufacturing industries and 256 commuting zones.[10] To construct time-consistent commuting zones from municipalities in Japan, we first follow Kondo (2023) to convert municipalities in each year into time-consistent municipality groups.[11] We then use the converter in Adachi et al. (2020) to convert these municipality groups into commuting zones.

### 3.2 Summary Statistics

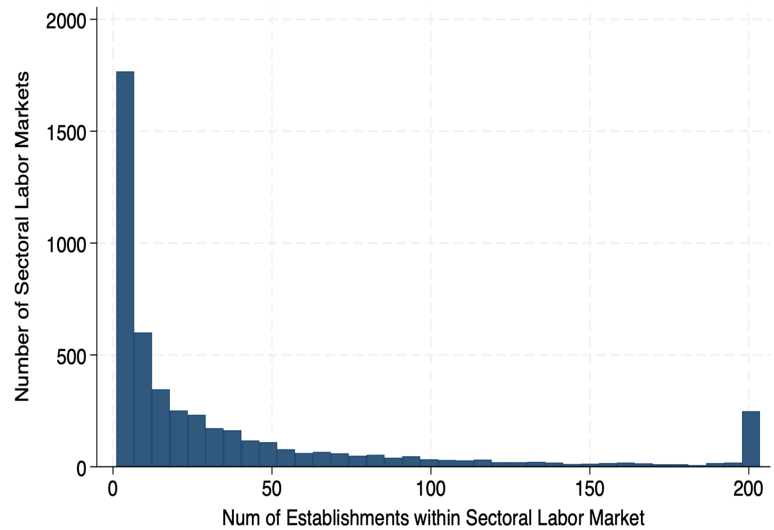
Table 1 shows the summary statistics for the data used in the analysis across local labor markets, which we show below. There are 4,782 local labor markets in our sample. The median local labor market has 13 firms and an employment HHI of 1809. With that few establishments, shocks to individual firms play a big role. Even the 75th percentile labor market only has 45 establishments. When one considers the skewness of the firm distribution, these markets are defined by a few important firms. Figure 1 shows the histogram of the number of establishments.

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

[10] For example, manufacturing of iron and steel industry is one of the 2-digit industrial categories.

[11] Japan has 1,724 municipalities as of June 2023, including 6 municipalities in the Northern Territories. We drop these 6 municipalities as the CoM data does not cover them.

Figure 1: Histogram of the Number of Establishments in each Local Labor Market




*Note*: The figure shows the histogram of the number of establishments in each local labor market. The unit of observation is the local labor market, a pair of a JSIC 2-digit industry, and a commuting zone. To be visible, I collapse all the local labor markets with the number of establishments larger than the top 5% percentile, 210, to be in the same bin.

### 3.3 Variance of Wages and Payrolls across Local Labor Markets

We start by checking if labor markets with more establishments have less volatile wages. This model prediction serves two key purposes. First, as we stated in the previous section, it confirms that establishments are subject to idiosyncratic shocks. If productivity shocks were only at the sectoral level, then adding in more establishments would not decrease the wage volatility in a labor market. And there would be no benefit to labor market pooling since establishments would demand labor at the same time.

The second purpose of this empirical prediction is quantitative. When labor markets are perfectly competitive, wage volatility is a measure of misallocation across states of the world. Therefore, the speed with which volatility drops with the size of the labor market is a sufficient statistic for the labor market pooling benefits of a larger labor market.

In our data, employers and employees are not linked, so we cannot get an estimate of the market wage by including worker fixed effects. Instead, we observe the total labor payroll. In the model, where the population of the labor market is fixed and labor is inelastically supplied, the variance of log labor payroll is the same as the variance of log wages

$$\text{Var}(\log(w(s)l)) = \text{Var}(\log w(s)).$$

In reality, people do move between commuting zones and in and out of employment. In the Japanese context, this is very rare, especially in response to unexpected shocks to the yearly growth rate that we focus on.[12]

The theory then guides our regression specification to test if the variance of log wages is decreasing in the size of the labor market. Equation (5) in Proposition 4, says that $Var(\log w(s)) \approx \sigma^2 / E^{\phi}$. Taking logs of both sides provides an estimable equation

$$\log\left(\text{Var}(\log(w_{nj}(s)l_{nj}))\right) \approx \log \sigma_j^2 - \phi \log E_{nj},$$

where $n$ is the commuting zone and $j$ is the industry.

Since $\sigma_j^2$ is not observable, we take it out by including industry fixed effects. In particular, we estimate the following log-linear model with industry-fixed effects and commuting-zone-fixed effects.

$$\log\left(\text{Var}(\log(w_{nj}(s)l_{nj}))\right) = \beta \ln E_{nj} + \mu_n + \mu_j + \varepsilon_{nj} \tag{8}$$

where $\mu_n$ and $\mu_j$ are commuting zone fixed effects and industry fixed effects, respectively.

For the empirical counterpart of the left-hand side variable, $\log\left(\text{Var}(\log(w_{nj}(s)l_{nj}))\right)$, we first compute total labor payroll in each local labor market $w_{nj}(t)l_{nj}(t)$. We then compute one-year log growth in each local labor market and take the variance of its growth over time. This is a conservative estimate of volatility since we take out persistent shocks to the growth rate.

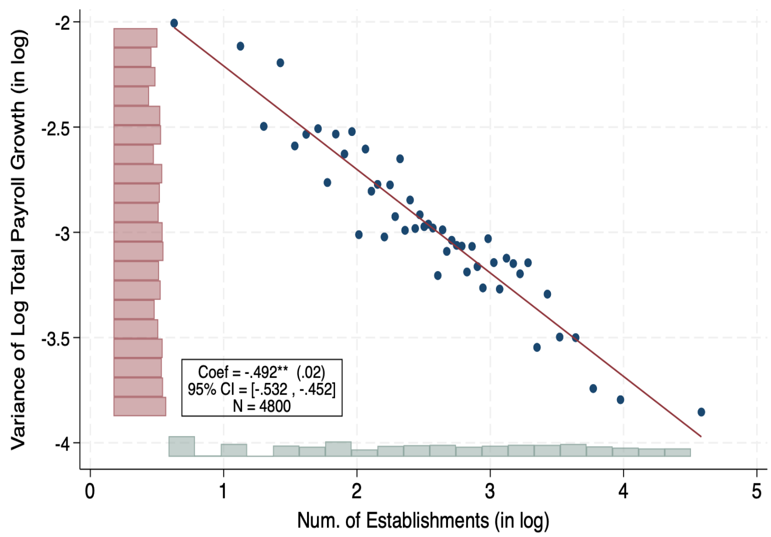
Figure 2a and 2b show the result in binned scatter plots for total labor payroll and average wages, respectively. The variances of log total payroll growth and log wage growth are both decreasing in the number of establishments, implying that the establishment-level idiosyncratic shocks are averaged out in local labor markets with a larger number of establishments. This supports the first prediction of Proposition 5. Furthermore, the relationship looks fairly linear further validating the model.[13]

[12]We show figures for both total labor payroll and average wages.

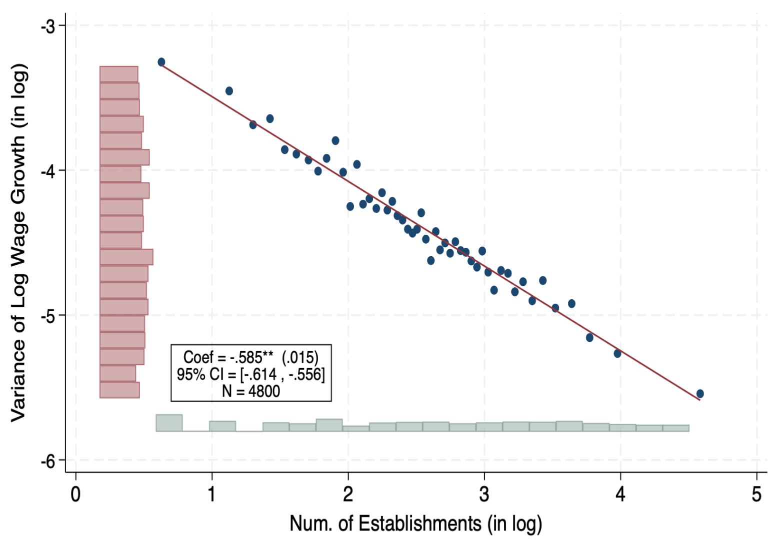
[13]See Figure B.1a where we use payroll HHI in the horizontal axis instead of the number of establishments.

Figure 2: Volatility of Payroll and Average Wage Growth and Number of Establishments

(a) Payroll




(b) Average Wage




*Note*: The panels show the binned scatter-plots and histograms of the relationship between volatility of log growth in total labor payment (top) and in average wage (bottom) and the number of establishments across local labor markets in Japan estimated by (8). We also show the histograms of both variables. The unit of observation is the local labor market, a pair of JSIC 2-digit industries, and a commuting zone. The vertical axis is the log variance of log growth in total labor payment (top) and average wage (bottom) over 1990-2016 in each local labor market. The horizontal axis is the number of establishments in log, averaged over 1986-2016 in each local labor market.

### 3.4 Variance of Establishment-level Employment

The second implication of the model is that establishments in larger labor markets have an easier time finding workers when they need them. First, we test the prediction of Proposition 6 that establishments with smaller payroll share within local labor markets respond more to establishment-level demand shock. Second, we examine the prediction of Proposition 7 that volatility of establishment-level employment is higher in local labor markets with more establishments (less concentration).

#### 3.4.1 Responses to Demand Shock across Establishments with Different Payroll Share

In this subsection, we examine how establishments with different payroll shares within local labor markets respond differently to establishment-level shock. In particular, we construct establishment-level demand shock from the establishment's product mix, country-level destinations, and destinations' real exchange rates. We then examine how that "adjusted real effective exchange rate shock (AREER)" affects an establishment's employment share within local labor markets.

**Specification** Our econometric specification is as follows:

$$\Delta \ln \ell_{e,t,t+1} = \beta_1 \Delta\mu_{e,t,t+1} + \beta_2 \left(\Delta\mu_{e,t,t+1} \cdot s_{e,t-1}\right) + \mathbf{X}'_{e,t}\Gamma + \zeta_e + \zeta_t + \varepsilon_{e,t}, \tag{9}$$

where $\Delta \ln \ell_{e,t,t+1}$ is the log change in the employment of establishment $e$ from year $t$ to $t+1$, $\Delta\mu_{e,t,t+1}$ is the negative demand shock for establishment $e$, $s_{e,t-1}$ is the payroll share of establishment $e$ in year $t-1$ within a local labor market. $\mathbf{X}_{e,t}$ is a vector of covariates at the establishment level, including an establishment age and its square, lagged payroll share, and lagged export ratio relative to total shipment.[14] $\zeta_e$ and $\zeta_t$ are fixed effects for establishments and year, respectively. $\varepsilon_{e,t}$ is the error term.

**Establishment-Specific Negative Demand Shock** We proxy the establishment-specific negative demand shock $\Delta\mu_{e,t,t+1}$ by an establishment-level exposure to a real effective exchange rate shock,

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

which we define as follows.[15]

$$\Delta \mu_{e,t,t+1} = \overline{\text{EXP}_e} \times \left( \sum_c \overline{\omega_{e,c}} \cdot \Delta REX^{JPN}_{c,t,t+1} \right) \tag{10}$$

where $\overline{\text{EXP}_e}$ is the median export share relative to the total shipment of establishment $e$ over the period.[16] $\overline{\omega_{e,c}}$ is the median exposure of establishment $e$ to country $c$ over time, where the time-specific exposure $\omega_{e,c,t} \equiv \sum_p \omega_{e,p,t} \cdot \omega_{p,c,t}$ is the share of product shipment of establishment $e$ to country $c$ in year $t$. Since CoM does not report establishment-specific export destinations, we use $\omega_{e,p,t}$, the share of shipment of establishment $e$ in 6-digit product category $p$ in year $t$ and $\omega_{p,c,t}$, the share of Japanese export to country $c$ out of total shipment in 6-digit product category $p$ from the UN Comtrade data. $\Delta REX^{JPN}_{c,t,t+1}$ is the change in the real exchange rate of Japanese Yen to the currency in the country $c$ from $t$ to $t+1$. Therefore, positive $\Delta REX^{JPN}_{c,t,t+1}$ means JPY appreciation against the currency in country $c$, so that $\Delta \mu_{p,t,t+1}$ and $\Delta \mu_{e,t,t+1}$ are *negative* demand shock for product and establishment, respectively.

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

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

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

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

**Summary Statistics: Establishment Panel** Table 2 shows the summary statistics for panels of the establishments in our analysis. The average and median establishments have 51 and 18 workers. The average payroll is 224 million in JPY (in 2015 value). On average, the share of non-regular workers is 34%. The changes in employment share are symmetric with the median of zero, but the volatility comes from non-regular employment with a standard deviation of 0.51, rather than regular employment with a standard deviation of 0.27. This is consistent with the findings in the previous literature in Japan.[17] The average share of employment as well as payroll within a local labor market is 2%.

Table 2: Summary Statistics: Establishment

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

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

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

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

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

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

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

**Results: Roles of Size in Response** After confirming that our measure of shock does actually affect the establishment's sales and non-regular employment, we examine the heterogeneous responses by the establishment's sizes. Table 4 shows the results. The dependent variable is the log changes in non-regular employment. The running variable is the adjusted real exchange rate shock at an establishment level. All columns include the following covariates: lagged share of non-regular workers, lagged payroll share within local labor markets, lagged log payroll, and the establishment's age square. All columns include establishment fixed effects and year-fixed effects. Regressions are weighted by the establishment's payroll. Standard errors are robust against heteroscedasticity.

Column (1) runs the regression without the interaction, which replicates Column (3) in Table 3 that a JPY appreciation decreases non-regular employment. Column (2) adds the interaction with payroll share within a local labor market. The positive estimate implies that the establishments with higher payroll share respond less to the shock, which is consistent with the prediction of Proposition 6. Quantitatively, if the plants differ in their payroll share by 10% pt, the elasticity decreases by 0.34, which is about 11% (=0.34/2.98) of the initial elasticity of 2.98.

However, this coefficient may not only contain our mechanism but also may contain the force that establishments with large sizes—regardless of the size relative to local labor markets—respond more or less. For example, larger establishments may adjust more because they have more capacity to pay fixed costs and replace workers by automation (Hubmer and Restrepo, 2021). Therefore, the estimate may pick up heterogeneity in responses by the absolute size, rather than relative size.

To address this concern, Column (3) adds the interaction term with the lagged log payroll of the establishment.[18] This means that we compare the elasticities of employment by payroll share within a local labor market, across establishments with the same absolute sizes. The estimate is now 8.26, which is larger.

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

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

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

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

### 3.4.2 Quantitative Importance across Local Labor Market

The prediction of Proposition 7 is that an establishment's variance of log employment would be larger if it were in a large labor market than if it were in a small one. While the labor pooling mechanism would drive large and risky establishments to locate in larger labor markets, we do

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

not want to capture those sorting effects. Instead, we try to compare similar establishments across labor markets. We do this by controlling for several characteristics found to be important for an establishment's employment volatility.

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

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

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

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

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

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

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

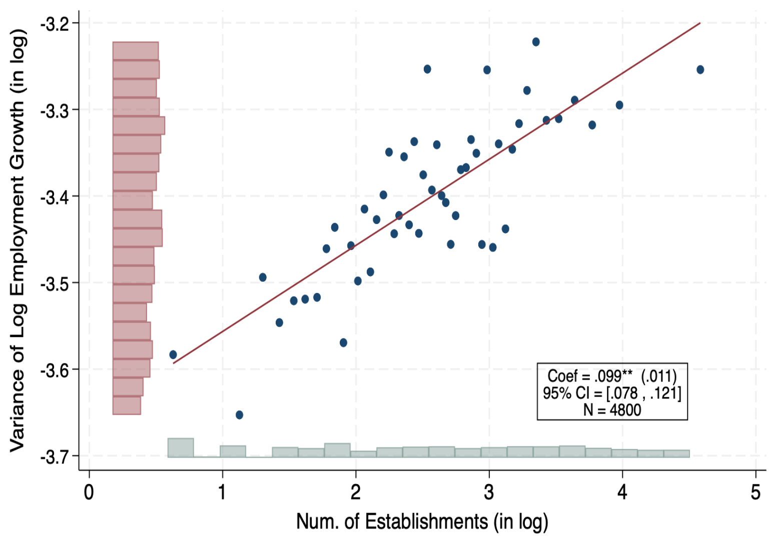
We estimate the following log-linear model

$$\sum_{e\in(n,j)} \omega_e \cdot \mathrm{Var}(\Delta\widehat{\varepsilon}^{\ell}_{e,t,t+1}) = \delta \ln E_{n,j} + \mu_n + \mu_j + \varepsilon_{n,j} \tag{11}$$

where $\ln E_{n,j}$ is the number of establishments in commuting zone $n$ and industry $j$. $\mu_n$ and $\mu_j$ are commuting zone fixed effects and industry fixed effects, respectively.

Figure 3 shows the result in binned scatter plots. One can see the clear positive relationship implying that comparing two similarly situated establishments, the establishment in the larger labor market will have a higher variance of log employment. And perhaps, if one squints, it looks like the relationship levels off as the number of establishments gets very large. This is also consistent with the theory which says that $Var(\log \ell_e(s)) \approx \frac{\sigma_j^2}{(1-\eta_j)^2}\left(1 - \frac{1}{E^{\phi}_{nj}}\right)$.

Figure 3: Volatility of Establishment-level Employment Growth and Number of Establishments




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

## 4 A Quantitative Model of Granular Economic Geography

In this section, we present a quantitative model of granular economic geography that features our mechanism but remains tractable with $N$ locations, many sectors, imperfect labor mobility, firm entry, and other externalities. We will then characterize the equilibrium and explore some of its properties in the next section.

There is a small open economy with $N$ regions we index by $n \in \mathcal{N}$ and a continuum of freely traded sectors we index by $j \in (0,1)$. In a pre-period, firms decide whether to enter and then get ex-ante productivity draws leading to a thick-tailed firm distribution as in Gabaix (2011). Seeing the firms, workers decide where to live and how much to invest in various sector-specific skills. Next, the state of the world $s \in \mathcal{S}$ is revealed, giving ex-post firm productivity shocks. Workers finally allocate their labor across firms and sectors subject to a friction. Firms produce with the labor they have, and pay their workers.

### 4.1 Environment

**Workers** There is a mass $\ell$ of workers who differ in how much they like the $N$ regions. We index these workers by how much they like each location $\boldsymbol{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_N\}$. The utility worker $\boldsymbol{\varepsilon}$ gets from living in location $n$ is

$$U_n(\boldsymbol{\varepsilon}) = U_n \varepsilon_n$$

where $U_n$ is the common utility of location $n$. We assume that the utility shocks are independent and distributed Fréchet with shape parameter $\theta$. Workers have linear preferences over a freely traded good $c_n$

$$U_n = u_n c_n,$$

where $u_n$ is the amenity of location $n$. Amenities take the form

$$u_n = \overline{u}_n (\ell_n)^{\gamma_u}, \tag{12}$$

where $\overline{u}_n$ is the fundamental amenity of location $n$, $\ell_n$ is the number of people living in location $n$, and $\gamma_u$ is the amenity spillovers. $\overline{u}_n$ captures the fact that certain locations might have nice weather or scenery, while $\gamma_u$ captures the fact that large numbers of people might lead to unpleasant congestion or pollution.

Workers are endowed with a single effective unit of labor that they supply inelastically. De-noting by $L_{nje}(s)$ the amount of labor a single worker supplies to establishment $e$ sector $j$,

$$L_{nj}(s) = \left[ \sum_{e \in \mathcal{E}_{nj}} b_{nje}^{-\frac{1}{\kappa}} L_{nje}(s)^{\frac{1+\kappa}{\kappa}} \right]^{\frac{\kappa}{1+\kappa}}$$

and

$$1 = \left[ \int_0^1 s_{nj}^{-\frac{1}{\nu}} L_{nj}(s)^{\frac{1+\nu}{\nu}} \right]^{\frac{\nu}{1+\nu}}$$

where $L_{nj}(s)$ is the effective labor supplied to sector $j$, $b_{nje}$ is the amenities of establishment $e$, and $s_{nj}$ is the amount of skills workers in location $n$ invested in sector $j$. Investing in sector-specific skills is costly since workers only have so much time. These skills must satisfy

$$1 = \int_0^1 s_{nj}^{\frac{1+\overline{\nu}}{\overline{\nu}-\nu}} dj.$$

We have parametrized this restriction so that the elasticity of labor to sectoral wages is $\nu$ if $s_{nj}$ is fixed and $\overline{\nu} > \nu$ if $s_{nj}$ can be chosen.

**Establishments** There is a continuum of potential establishment entrants in every location and sector. To attempt an entrance, an establishment must pay a fixed cost of $\psi_n > 0$ in terms of the freely traded final good.

If a mass of $m_n$ establishments attempt to enter in location $n$, the number of realized entrants in any sector $j$ is distributed Poisson with parameter $m_n$. That is

$$\mathbb{P}[E_{nj} = k] = \frac{(m_n)^k e^{-m_n}}{k!}$$

where $E_{nj}$ is the realized number of establishments in one particular sector $j$. Conditional on entering, an entrant gets an amenity draw $b_{nje}$ and an ex-ante productivity draw $z_{nje}$. We assume that $b_{nje} = \frac{1}{E_{nj}}$ to shut down the expanding variety gains that come from having more firms to work at and instead only leave our mechanism. $z_{nje}$ is drawn independently from Pareto distribution with shape parameter $\lambda$ and scale parameter $z_n$.

$z_n$ is the natural productivity of location $n$. It is subject to increasing returns to scale as our mechanism is not the only one driving agglomeration. It takes the form

$$z_n = \overline{z}_n(\ell_n)^{\gamma_z} \tag{13}$$

where $\gamma_z$ is the agglomeration and $\bar{z}_n$ is the fundamental productivity of location $n$.

Then, after ex-post productivity shocks $a_{nje}(s)$ are revealed, active establishments can produce a sector-specific good using a constant elasticity production function $\eta_j \in (0, 1)$,

$$y_{nje}(s) = z_{nje} a_{nje}(s) \ell_{nje}(s)^{\eta_j},$$

and sell it on world markets. The ex-post productivity shocks $a_{nje}(s)$ are distributed log-normal with mean parameter $\sigma^2/2$ and variance parameter $\sigma^2$ so that they have expectation 1.

**Market Clearing** The population in each region adds up to the population of the whole country:

$$\ell = \sum \ell_n.$$

The amount of labor supplied to establishment $e$ needs to equal demand

$$\ell_{nje}(s) = L_{nje}(s) \ell_n. \tag{14}$$

### 4.2 Market Structure and Equilibrium

**Workers** Workers take prices and wages as given, and decide where to live and what skills to invest in before the state of the world is revealed. Then after the state of the world is revealed, they decide where to work. Importantly, the continuum of sectors averages out any uncertainty under some technical assumptions left in the appendix. Therefore, any integrals over $j$ will get rid of any dependence on the state of the world $s$.

We will solve the worker problem backward. Once the state of the world is revealed, a worker in location $n$ allocates labor to maximize total earnings taking as given wages by each establishment. That is

$$\begin{aligned} L_{nje}(s), L_{nj}(s) \in \underset{L'_{je}, L'_j}{\operatorname{argmax}} \quad & \int_0^1 \left[ \sum_{e \in \mathcal{E}_{nj}} w_{nje}(s) L'_{je} \right] dj \\ s.t. \quad L'_j = & \left[ \sum_{e \in \mathcal{E}_{nj}} b_{nje}^{-\frac{1}{\kappa}} (L'_{je})^{\frac{1+\kappa}{\kappa}} \right]^{\frac{\kappa}{1+\kappa}} \\ 1 = & \left[ \int_0^1 s_{nj}^{-\frac{1}{\nu}} (L'_j)^{\frac{1+\nu}{\nu}} \right]^{\frac{\nu}{1+\nu}} \end{aligned} \tag{15}$$

We denote the value of the solution to this problem by $W_n(\{s_{nj}\})$, the wages a worker in location $n$ can earn given skill investments $s_{nj}$.

Then the workers choose skill investments to maximize expected wages in a pre-period

$$\begin{aligned} \{s_{nj}\}_{j\in(0,1)} \in \underset{s'_j}{\operatorname{argmax}} \quad & W_n(\{s'_j\}) \\ s.t. \quad 1 = & \int_0^1 (s'_j)^{\frac{1+\bar{\nu}}{\bar{\nu}-\nu}} dj. \end{aligned} \tag{16}$$

The solution to this gives $W_n$ as the maximum amount of wages a person could earn living in location $n$.

Workers buy freely traded goods with their earnings and a location-specific transfer from the government $T_n$. We normalize the price of the freely traded good to one so that utility is given by

$$U_n = u_n(W_n + T_n). \tag{17}$$

Workers are free to choose where to live. Therefore, the population in location $n$ is equal to the number of people for whom living in location $n$ is utility maximizing. Under our Fréchet assumption,

$$\ell_n = \left(\frac{U_n}{U}\right)^{\theta} \ell, \tag{18}$$

where

$$U = \left[\sum_n (U_n)^{\theta}\right]^{\frac{1}{\theta}}. \tag{19}$$

**Firms** Conditional on entry, firms hire labor to maximize profits taking as given wages and prices. In math, in every state of the world $s$, establishment $e$ solves

$$\ell_{nje}(s) \in \underset{\ell'}{\operatorname{argmax}} \quad z_{nje} a_{nje}(s)(\ell')^{\eta_j} - w_{nje}(s)\ell' \tag{20}$$

We denote the value of the solution to this by $\pi_{nje}(s)$ as the flow profits of the firm.

In each region, there is a continuum of potential entrants who are free to enter. Because of that, expected profits conditional on trying to enter must be equal to the fixed cost of entering. That is, total expected profits divided by the number of entrants must be equal to $\psi$. In math,

$$(1 - \tau_n)\, \psi_n = \frac{1}{m_n} \mathbb{E} \left[ \sum_{e \in \mathcal{E}_{nj}} \pi_{nje}(s) | m_n \right] \tag{21}$$

where $\tau_n$ is an entry subsidy on firms in location $n$, and expectations are taken over the probability of entering ($E_{nj}$), the ex-ante productivity shocks ($z_{nje}$), and the ex-post productivity shocks ($a_{nje}(s)$).

**Government** The government budget constraint needs to hold. In math,

$$0 = \sum_n T_n \ell_n + \sum_n \tau_n m_n \psi. \tag{22}$$

**Definition 4.1.** An *equilibrium* consists of wages $\{w_{nje}(s)\}_{n,j,e,s}$, labor supply decisions $\{L_{nje}(s)\}_{n,j,e,s}$, skill investment decisions $\{s_{nj}\}_{n,j}$, effective wages $\{W_n\}_n$, common utilities $\{U_n\}_n$, regional population $\{\ell_n\}_n$, and entry decisions $\{m_n\}_n$ such that

- Workers maximize expected utility taking as given wages as summarized by equations (15), (16), (17), (18), and (19);
- Conditional on entry, firms maximize profits taking wages and prices as given as summarized by equation (20);
- Firm entry is consistent with free entry (21);
- Utility and productivity are consistent with spillovers as summarized by equations (13) and (12);
- Markets clear (14); and
- The government budget constraint holds (22).

### 4.3 Social Welfare and the Planner's Problem

Social welfare is a weighted sum of household utility:

$$\mathcal{W} = \int \lambda(\varepsilon) U_{n^*(\varepsilon)}(\varepsilon) dF(\varepsilon), \tag{23}$$

where $n^*(\varepsilon)$ is the location choice of agent $\varepsilon$ and $\lambda(\varepsilon)$ is the planner's weight on $\varepsilon$. The planner chooses an equilibrium by choosing location-based taxes $T_n$ and entry subsidies $\tau_n$ to maximize social welfare.

### 4.4 Characterization of the Equilibrium

Here we turn to characterizing the solution. After the state is revealed, the worker solves a standard model of labor supply (15) from Berger et al. (2022) taking as given wages and the skill decisions she made in the pre-period $\{s_{nj}\}$. Defining a sector $j$ effective wage

$$W_{nj}(s) = \left( \sum_e b_{nje} w_{nje}(s)^{1+\kappa} \right)^{\frac{1}{1+\kappa}}, \tag{24}$$

labor supplied to establishment $e$ is given by

$$L_{nje}(s) = b_{nje} \left( \frac{w_{nje}(s)}{W_{nj}(s)} \right)^{\kappa} L_{nj}(s). \tag{25}$$

Similarly, we can define an effective aggregate wage

$$W_n(\{s_{nj}\}) = \left[ \int_0^1 s_{nj} W_{nj}(s)^{1+\nu} dj \right]^{\frac{1}{1+\nu}}, \tag{26}$$

and to find labor supply to a sector $j$

$$L_{nj}(s) = s_{nj} \left( \frac{W_{nj}(s)}{W_n} \right)^{\nu}. \tag{27}$$

In a pre-period, before the state of the world is revealed, the worker needs to choose skills $s_{nj}$. The expected utility is given by

$$W_n(\{s_{nj}\}) = \left[ \int_0^1 s_{nj} \mathbb{E}[W_{nj}(s)^{1+\nu}] dj \right]^{\frac{1}{1+\nu}}.$$

Then solving the optimal skill choice problem (16) gives the effective wage

$$W_n = \left( \int_0^1 \mathbb{E}[W_{nj}(s)^{1+\nu}]^{\frac{1+\bar{\nu}}{1+\nu}} dj \right)^{\frac{1}{1+\bar{\nu}}} \tag{28}$$

with

$$s_{nj} = \left( \frac{\mathbb{E}[W_{nj}(s)^{1+\nu}]^{\frac{1}{1+\nu}}}{W_n} \right)^{\bar{\nu}-\nu}. \tag{29}$$

We can then combine these labor supply equations with the labor demand equations

$$\eta z_{nje} a_{nje}(s) \ell_{nje}(s)^{\eta-1} = w_{nje}(s) \tag{30}$$

to characterize the equilibrium conditional on labor and firm entry. A lot of tedious algebra relegated to the appendix gives a familiar but more complicated expression for production in location $n$:

$$Y_n = (\ell_n)^{\eta} \Psi(m_n), \tag{31}$$

where

$$\Psi(m_n) = \left\{ \int_0^1 \mathbb{E} \left[ \left( \sum_{e \in \mathcal{E}_{nj}} b_{nje}^{\frac{\eta}{1+\kappa(1-\eta)}} z_{nje}^{\frac{1+\kappa}{1+\kappa(1-\eta)}} a_{nje}(s)^{\frac{1+\kappa}{1+\kappa(1-\eta)}} \right)^{\frac{1+\kappa(1-\eta)}{1+\kappa} \frac{1+\nu}{1+\nu(1-\eta)}} \right]^{\frac{1+\nu(1-\eta)}{1+\nu} \frac{1+\bar{\nu}}{1+\bar{\nu}(1-\eta)}} dj \right\}^{\frac{1+\bar{\nu}(1-\eta)}{1+\bar{\nu}}}. \tag{32}$$

Even more familiar are the expressions for wages and total profits. An $\eta$ share of earnings goes to workers and the rest goes to the firms

$$W_n \ell_n = \eta Y_n \tag{33}$$

and

$$\Pi_n = (1-\eta) Y_n. \tag{34}$$

Firm entry then implies that $(1-\tau_n)\psi_n m_n = \Pi_n$. The rest of the model is standard in the economic geography literature and can be solved using standard methods as long as the agglomeration forces are not too strong.

## 4.5 Properties of the Model

**Empirical Predictions** While the model presented in the previous section is more complicated than our initial theoretical framework. The basic qualitative insights from the empirical predictions remain the same. We start by confirming that the variance of the log wage should still be declining in the size of the local labor market. Furthermore, in this model where workers can move across sectors, we can also confirm that the variance of the log wage bill should be declining.

**Proposition 8.** *To a first-order log approximation around the equilibrium $a_e(s) = 1$, the variance of the log effective wage is given by:*

$$var(\log W_{nj}(s)) = \left(\frac{1}{1+\nu(1-\eta)}\right)^2 \left(\sum_e x_e^2\right) \sigma^2.$$

*The variance of the log wage bill is*

$$var\left(\log\left(\sum_e w_{nje}(s)\ell_{nje}(s)\right)\right) = \left(\frac{1+\nu}{1+\nu(1-\eta)}\right)^2 \left(\sum_e x_e^2\right) \sigma^2.$$

These are both consistent with the regressions in the empirical section. This also tells us that one can use the differences in those regressions to calibrate the short-run elasticity of substitution across sectors $\nu$.

Furthermore, the predictions about it being easier for firms to find labor in response to a productivity shock in a large labor market also survive in this quantitative framework.

**Proposition 9.** *To a first-order log approximation, labor responds according to*

$$\Delta \log \ell_{nje}(s) = \left[\frac{\kappa}{1+(1-\eta)\kappa} - \frac{\kappa-\nu}{(1+\nu(1-\eta))\,(1+(1-\eta)\kappa)} x_e\right] \Delta \log a_{nje}(s)$$

The proposition says that a firm that hires a larger proportion of the labor market will expand less in response to a productivity shock as long as $\kappa > \nu$. That is, as long as people can more easily move across firms within a sector than across. If $\kappa = \nu$, then workers move across sectors as easily as they move across firms. In that case, a firm's payroll share has no effect on its ability to find workers since even if it is large relative to the labor market, it is small relative to the entire commuting zone.

When $\kappa \approx \nu$, the firms in smaller markets will not expand as much as firms in larger markets but they will be similar. When $\nu \approx 0$, workers do not substitute across sectors easily and firms in smaller markets will have a very difficult time finding workers. The cross sectional prediction remains.

**Proposition 10.** *To a first-order log approximation around the equilibrium $a_e(s) = 1$, the weighted vari-*

*ance of the log variance of employment is given by:*

$$\sum_e x_e Var\left(\log \ell_{nje}(s)\right) = \sigma^2 \left(\frac{\kappa}{1+(1-\eta)\kappa}\right)^2 \left[1-(\kappa-\nu)\frac{\kappa+2\kappa\nu(1-\eta)+\nu}{\kappa^2\left(1+\nu(1-\eta)\right)^2}\left(\sum_e x_e^2\right)\right].$$

**The Planner's Solution** Here we turn to quantitatively characterizing the issue of under-entry in this model. Our first proposition characterizes the optimal policy necessary.

**Proposition 11.** *In any solution to the planner's problem,*

$$\psi_n = \frac{\Pi_n}{m_n}\frac{1}{1-\eta}\frac{\Psi'(m_n)m_n}{\Psi(m_n)}.$$

This simply says that in each location, the increase in production from another firm attempting to enter has to be equal to the fixed cost of that attempted entrance. We immediately get that in any Pareto optimal,

$$1-\tau_n = \frac{1-\eta}{\frac{\Psi'(m_n)m_n}{\Psi(m_n)}}.$$

However, just as before, there are increasing returns to scale so there will be too little entry. Therefore, there needs to be a subsidy on entry to achieve efficiency. Furthermore, it is varying in $m_n$.

**Proposition 12.** *In any solution to the planner's problem $\tau_n > 0$, that is, there is a subsidy on entry. Furthermore, $\tau_n$ is decreasing in $m_n$. That is to say, the optimal subsidy is decreasing in the size of the market.*

# 5 Quantification of Granular Agglomeration

We next turn to demonstrate the quantitative importance of granularity for the urban wage premium, optimal spatial policy, and thinking through counterfactuals.

## 5.1 Calibration

Many of the most important parameters in this model can be taken from the literature or read directly from our graphs and table regressions. The calibration is summarized in Table 5.

**Labor elasticities** We take the short-run labor elasticities from Berger et al. (2022). In that paper, they assume a labor market is a 3-digit industry within a particular commuting zone, and they

| Description | Parameter | Value | Source |
|---|---|---|---|
| Short run labor elasticity across sectors | $\nu$ | 0.42 | Berger et al. (2022) |
| Short run labor elasticity across firms | $\kappa$ | 10.85 | Berger et al. (2022) |
| Long run labor elasticity across sectors | $\overline{\nu}$ | 1 | Burstein et al. (2020) |
| Elasticity of production to labor | $\eta$ | 0.5 | Japanese man. labor share |
| Ex-ante firm prod. tail | $\lambda$ | 2.8 | Figure 2(a) |
| Ex-post shock log variance | $\sigma^2$ | 0.25 | Variance of log wages |
| Migration elasticity | $\theta$ | 3 | Redding (2016) |
| Congestion externality | $\gamma_u$ | -0.25 | Redding (2016) |
| Production externality | $\gamma_z$ | 0.0025 | Combes et al. (2011) |

Table 5: Calibration Summary

use yearly variation in taxes to get shocks to individual firms. If anything this will overstate the short-run elasticity of labor supply into a market, weakening our mechanism as we use a 2-digit industry within a particular commuting zone as our labor market.

We calibrate the long-run labor elasticity across sectors within a commuting zone to match the long-run elasticity across occupations in the US from Burstein et al. (2020).

**Firm production** We calibrate $\eta$ to match the average manufacturing labor share in Japan using equation (33). We then find the Pareto tail of ex-ante shocks. As suggesting by proposition 8,

$$\log\left(var\left[\log\left(\sum_e w_{nje}(s)\ell_{nje}(s)\right)\right]\right) = 2\log\left(\frac{1+\nu}{1+\nu(1-\eta)}\right) + \log\sigma^2 + \log\left(\sum_e x_e^2\right).$$

Furthermore, since ex-ante productivity draws are Pareto,

$$\left(\sum_e x_e^2\right) \to C\frac{1}{\left(E^{1-\frac{1}{\lambda(1-\eta)}}\right)^2}.$$

We can then back out $\lambda$ directly from figure 2(a) by setting

$$\hat{\beta} = 2\left(1 - \frac{1}{\lambda(1-\eta)}\right).$$

After we calibrate $\lambda$ we calibrate $\sigma^2$. Define

$$m(E) := \mathbb{E}\left[\left(\sum_e x_e^2\right)\middle| E\right].$$

We then choose $\sigma^2$ to minimize the sum of square errors

$$\sum_{n,j}\left[\log\left(var\left[\log\left(\sum_{e} w_{nje}(s)\ell_{nje}(s)\right)\right]\right) - 2\log\left(\frac{1+\nu}{1+\nu(1-\eta)}\right) - \log\sigma^2 - \log m(E_{nj})\right]^2$$

**Migration and Externalities** We take the migration and externality variables from the rest of the literature. We take the migration elasticity $\theta$ directly from Redding (2016) which is consistent with the evidence from Bryan and Morten (2019). We use $\gamma_u = -0.25$ to match average spending on housing.

We then try to match the estimated production externality of 0.03 reported in Combes et al. (2011). In our model, the externality is given by

$$\frac{d\log W_n}{d\log \ell_n} = \frac{\gamma_z + \frac{\Psi'(m_n)m_n}{\Psi(m_n)} - (1-\eta)}{1 - \frac{\Psi'(m_n)m_n}{\Psi(m_n)}},$$

which differs by market. We then choose $\gamma_z$ to match the average of these externalities weighted by population. That gives $\gamma_z = 0.0025$ suggesting a large portion of the externality can be explained by our mechanism.

**Location-specific parameters** We back out $\bar{u}_n$, $\bar{z}_n$, and $\psi_n$ to match the population, average wages, and the average number of firms in each location.

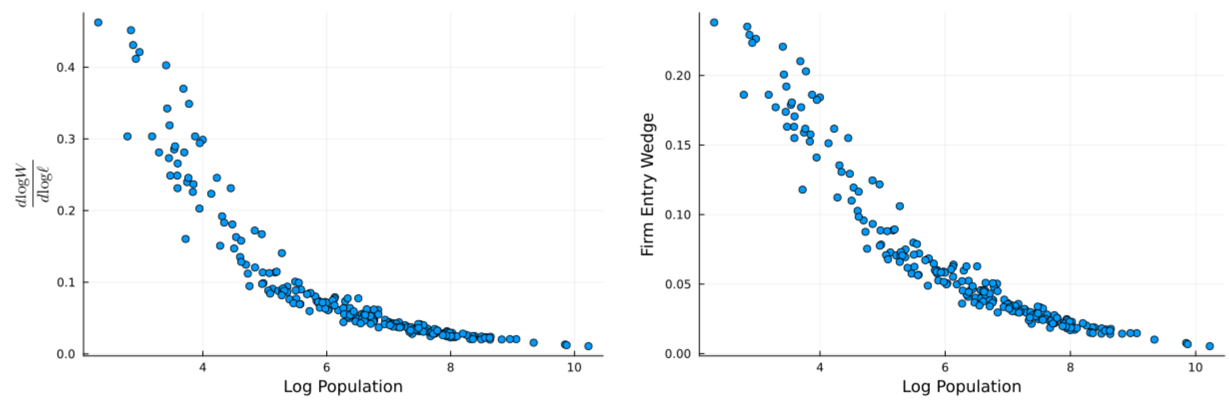
## 5.2 The Size of the Externality

In figure 4a, we plot the implied externality in this model for different commuting zones in Japan. In the limit, as log population goes to infinity, the granular reason for agglomeration disappears and the implied externality is simply $\frac{\gamma_z}{\eta} = 0.005$. The granular labor pooling mechanism plays a large role in commuting zones with a smaller population. The model suggests that wages could rise as much as 0.4% if the population increases by 1% in small locations, with most of that being driven by the labor market pooling externality.

The strength of the labor market pooling mechanism suggests that place-based industrial policy could have a large impact on welfare. We plot the firm entry wedge

$$\frac{1-\eta}{\frac{\Psi'(m_n)m_n}{\Psi(m_n)}} - 1,$$

Figure 4: The Importance of Labor Market Pooling




*Note*: Panel (a) plots the agglomeration externality in each commuting zone in Japan. It gives the elasticity of wages to the population in that location. Panel (b) plots the firm entry wedge for each commuting zone in Japan. This is the percent difference in the marginal benefit of another firm entering and the expected profits of that firm. We mathematically define this object in the text.

for each location $n$ in figure 4b. This gives the percentage difference between expected profits and the expected benefits on production. That is to say, firms in the smallest location capture less than 80% of their productivity benefits leading to significant under-entry.

### 5.3 Counterfactual

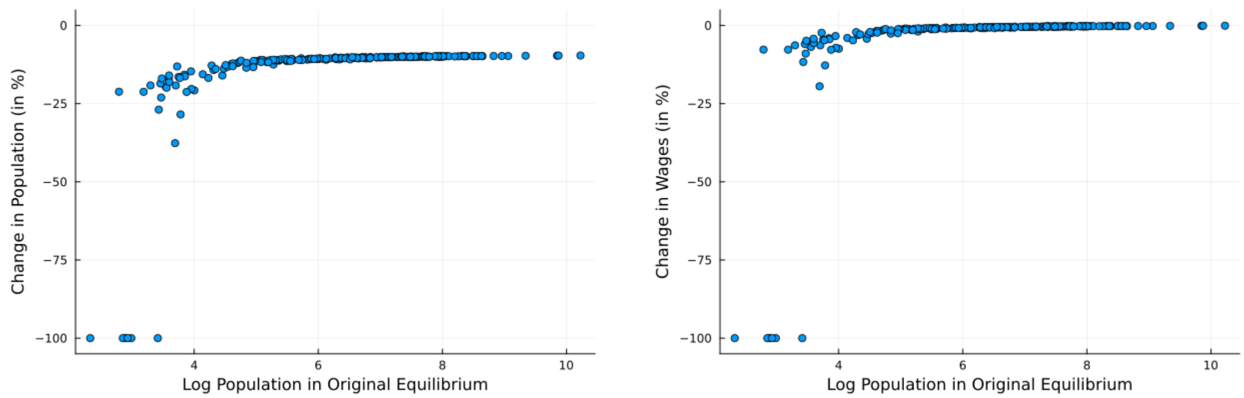
Japan's working-age population is decreasing. According to National Institute of Population and Social Security Research (2023), the working-age population (aged 15-64) peaked in 1995 at 87 million, and it decreased to 75 million in 2020 and is projected to continue decreasing to below 70 million in 2032. To shed light on how this might affect the geography of economic activity in Japan, we simulate a 10% drop in population.

This model features multiple stable equilibria. The strength of the externality becomes very strong as the number of firms trying to enter approaches zero. Therefore, having zero people and zero firms in a single location will always be stable. In our counterfactual, we choose the equilibrium where the most number of locations have a positive population.

The strength of the externality is decreasing in population so this drop in population will likely strengthen the externality, leading to a further urban bias. We plot the effects on population and wages in figure 5. The population drops in every location, but it drops an especially large amount in locations that were originally small. These locations had a few firms and so the pooling

Figure 5: Counterfactual with a 10% Decline in Population

(a) Percent change in population (b) Percent change in wages




*Note*: Panel (a) plots the percent change in population in each commuting zone in Japan. Panel (b) plots the percent change in average wages in each commuting zone in Japan.

externality was especially strong. Therefore, a drop in population has a large effect on wages driving more and more people to leave. 6 of the commuting zones we consider drop below the minimum necessary number of firms to support a manufacturing sector so their population drops to zero. In reality, because of the multiplicity of equilibria, more than 6 manufacturing sectors could unravel. It could also happen in commuting zones that an omniscient planner would rather keep active.

Every location also sees a drop in wages since there are agglomeration benefits. The largest locations, like the commuting zone that includes Tokyo, are in the parameter space where there is very little agglomeration externality. Therefore, while it loses 9.6% population, wages only drop 0.1%. Meanwhile, small locations, such as the commuting zone including Shari-cho, Kiyosato-cho, and Koshimizu-cho in Hokkaido, see a drop in population of 37.6% and wages drop of 19.5%.

While the 10% drop in population has a large impact on the geography of economic activity, it has less of an effect on aggregate variables. GDP per capita drops by only 0.2%. That is because, while small locations see huge drops in wages and population, the larger locations, where most economic activity takes place, do not see much drop in wages. Were we to focus on sectors with more specialized knowledge, those sectors would likely see wages drop even in the large locations. Furthermore, the larger locations have higher wages on average. Therefore, when people are driven out of smaller locations due to unraveling or very low wages, average wages do not change all that much. The workers still end up worse off because they are stuck working in a location they do not like as much.

## 6 Concluding Remarks

Determining the importance of a certain mechanism for explaining the productivity advantage of large cities has always been difficult. In this paper, we show that granularity is another reason for agglomeration. We give direct evidence of the mechanism, demonstrate its quantitative importance, explore some of the implications for optimal policy, and then show what it means for the future geography of economic activity in Japan.

This is far from the last word on labor market pooling: empirically or theoretically. We do not allow workers to work for multiple sectors. We suspect that commuting zones over-specialize and leave themselves too exposed to sector-specific shocks. We also suspect that the falling volatility of labor demand opens up other opportunities. Doctors can specialize more narrowly in large markets because they know that there will be a consistent demand for that narrow skill. This tying of Marshallian agglomeration with Smithian specialization could further explain the productivity benefits of living in a big city.

## References


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

Andini, M., G. d. Blasio, G. Duranton, and W. C. Strange (2012). Marshallian labor market pooling: evidence from italy. *IEB Working Paper 2012/27*.

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

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

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

Burstein, A., G. Hanson, L. Tian, and J. Vogel (2020). Tradability and the labor-market impact of immigration: Theory and evidence from the united states. *Econometrica 88*(3), 1071–1112.

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

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

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

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

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

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

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

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

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

Gan, L. and Q. Zhang (2006). The thick market effect on local unemployment rate fluctuations. *Journal of Econometrics 133*(1), 127–152.

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

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

Hubmer, J. and P. Restrepo (2021). Not a typical firm: The joint dynamics of firms, labor shares, and capital–labor substitution. Technical report, National Bureau of Economic Research.

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

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.

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

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

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

Miyauchi, Y., K. Nakajima, and S. J. Redding (2021). The economics of spatial mobility: Theory and evidence using smartphone data. Technical report, National Bureau of Economic Research.

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

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

Nakajima, K., Y. U. Saito, and I. Uesugi (2012). Measuring economic localization: Evidence from japanese firm-level data. *Journal of the Japanese and International Economies 26*(2), 201–220.

Nakamura, R. (1985). Agglomeration economies in urban manufacturing industries: a case of japanese cities. *Journal of Urban economics 17*(1), 108–124.

National Institute of Population and Social Security Research (2023). Population projections for japan: 2021–2070.

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

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

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

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

Stahl, K. O. and U. Walz (2001). Will there be a concentration of alikes? the impact of labor market structure on industry mix in the presence of product market shocks. Technical report, HWWA Discussion Paper.

Tabuchi, T. and A. Yoshida (2000). Separating urban agglomeration economies in consumption and production. *Journal of Urban Economics 48*(1), 70–84.

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

## A Theory Appendix

**Lemma 1.** *Suppose that $X_1$ and $X_2$ are iid random variables with positive support and $B$ is another random variable that is independent of $X_1$ and $X_2$ with positive support as well. Further suppose that $\alpha_1 > \alpha_2 > 0$ and $\beta \in (0,1)$. Then*

$$\mathbb{E}\left[\frac{X_2}{(\alpha_1 X_1 + \alpha_2 X_2 + B)^\beta}\right] > \mathbb{E}\left[\frac{X_1}{(\alpha_1 X_1 + \alpha_2 X_2 + B)^\beta}\right],$$

*if $Var(X_1) > 0$.*

*Proof.* To prove the result, we define a new random variable

$$Z(c_1, c_2) := \frac{X_2 - X_1}{(c_1 X_1 + c_2 X_2 + B)^\beta}.$$

Note that

$$\mathbb{E}[Z(\alpha_1, \alpha_2)] = \mathbb{E}\left[\frac{X_2}{(\alpha_1 X_1 + \alpha_2 X_2 + B)^\beta}\right] - \mathbb{E}\left[\frac{X_1}{(\alpha_1 X_1 + \alpha_2 X_2 + B)^\beta}\right].$$

Therefore, if we prove that $\mathbb{E}[Z(\alpha_1, \alpha_2)] > 0$, we prove the lemma. We start by noting that if $c_1 = c_2$, $\mathbb{E}[Z] = 0$:

$$\mathbb{E}[Z(c,c)] = \mathbb{E}\left[\frac{X_2}{(cX_1 + cX_2 + B)^\beta}\right] - \mathbb{E}\left[\frac{X_1}{(cX_1 + cX_2 + B)^\beta}\right] = 0.$$

Then

$$\frac{\partial Z}{\partial c_1} = -\beta \frac{X_2 - X_1}{(c_1 X_1 + c_2 X_2 + B)^{\beta+1}} X_1,$$

and

$$\frac{\partial Z}{\partial c_2} = -\beta \frac{X_2 - X_1}{(c_1 X_1 + c_2 X_2 + B)^{\beta+1}} X_2.$$

Then

$$\begin{aligned}
\mathbb{E}[Z(\alpha_1,\alpha_2)] &= \mathbb{E}[Z(\alpha_1,\alpha_2)] - \mathbb{E}\left[Z\left(\frac{\alpha_1+\alpha_2}{2}, \frac{\alpha_1+\alpha_2}{2}\right)\right] \\
&= \mathbb{E}\left[\int_0^{\frac{\alpha_1-\alpha_2}{2}} \frac{dZ\left(\frac{\alpha_1+\alpha_2}{2}+c, \frac{\alpha_1+\alpha_2}{2}-c\right)}{dc} dc\right] \\
&= \mathbb{E}\left[\int_0^{\frac{\alpha_1-\alpha_2}{2}} \left\{\frac{\partial Z\left(\frac{\alpha_1+\alpha_2}{2}+c, \frac{\alpha_1+\alpha_2}{2}-c\right)}{\partial c_1} - \frac{\partial Z\left(\frac{\alpha_1+\alpha_2}{2}+c, \frac{\alpha_1+\alpha_2}{2}-c\right)}{\partial c_2}\right\} dc\right] \\
&= \mathbb{E}\left[\int_0^{\frac{\alpha_1-\alpha_2}{2}} \beta \frac{(X_1-X_2)^2}{\left(\left(\frac{\alpha_1+\alpha_2}{2}+c\right)X_1 + \left(\frac{\alpha_1+\alpha_2}{2}-c\right)X_2 + B\right)^{\beta+1}} dc\right].
\end{aligned}$$

The denominator is always positive. The only remaining step is to make sure that $(X_1 - X_2)^2 > 0$ for a set of positive measure in order to ensure that $\mathbb{E}[Z(\alpha_1,\alpha_2)] > 0$. But

$$\begin{aligned}
\mathbb{E}[(X_1-X_2)^2] &= \mathbb{E}[X_1^2 - 2X_1X_2 + X_2^2] \\
&= Var(X_1) + \mathbb{E}[X_1]^2 - 2Cov(X_1,X_2) - 2\mathbb{E}[X_1]\mathbb{E}[X_2] + Var(X_2) + \mathbb{E}[X_2]^2.
\end{aligned}$$

Because $X_1$ and $X_2$ are independent, $Cov(X_1, X_2) = 0$. Since they are identically distributed, $Var(X_1) = Var(X_2)$ and $\mathbb{E}[X_1] = \mathbb{E}[X_2]$. Therefore,

$$\mathbb{E}[(X_1-X_2)^2] = 2Var(X_1),$$

so that if $Var(X_1) > 0$, $(X_1 - X_2)^2 > 0$ for a set of positive measure. □

**Proposition 13** (Proposition 1 in the main text). *For all $\ell > 0$, $\mathcal{E}$, and $\alpha > 1$, if* $\mathrm{Var}(a_e(s)) > 0$*, then*

$$Y(\alpha\ell, \alpha\mathcal{E}) > \alpha Y(\ell, \mathcal{E}).$$

*Proof.* We start by noting

$$\begin{aligned}
Y(\alpha\ell,\alpha\mathcal{E}) &= (\alpha\ell)^\eta\, \mathbb{E}\left[\left[\sum_{e\in\alpha\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right] \\
&= \alpha^\eta \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right]
\end{aligned}$$

Next, we will define new shocks $\tilde{a}_e(s)$. These will be a convex combination of firm $e$'s shocks and the average of the shocks of the ex-ante similar firms in $\mathcal{E}$. We denote $\mathcal{E}_z$ the firms in the initial grouping with initial productivity $z$. Then

$$\tilde{a}_e(\lambda, s)^{\frac{1}{1-\eta}} = (1-\lambda)\frac{\sum_{e' \in \mathcal{E}_{z_e}} a_{e'}(s)^{\frac{1}{1-\eta}}}{|\mathcal{E}_{z_e}|} + \lambda a_e(s)^{\frac{1}{1-\eta}}.$$

Then

$$\tilde{Y}(\lambda) \equiv \alpha^\eta \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} \tilde{a}_e(\lambda, s)^{\frac{1}{1-\eta}}\right]^{1-\eta}\right]$$

This gives us a good way to go between the two extremes

$$\begin{aligned}\tilde{Y}(1) &= \alpha^\eta \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right] \\ &= Y(\alpha\ell, \alpha\mathcal{E})\end{aligned}$$

and

$$\begin{aligned}\tilde{Y}(0) &= \alpha^\eta \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} \frac{\sum_{e' \in \mathcal{E}_{z_e}} a_{e'}(s)^{\frac{1}{1-\eta}}}{|\mathcal{E}_{z_e}|}\right]^{1-\eta}\right] \\ &= \alpha \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right] \\ &= \alpha Y(\ell, \mathcal{E}).\end{aligned}$$

Next, we will show that $\frac{\partial \tilde{Y}(\lambda)}{\partial \lambda} > 0$ for all $\lambda \in (0,1)$. It then follows that $\tilde{Y}(1) > \tilde{Y}(0)$ which implies the result. We have

$$\begin{aligned}\frac{\partial \tilde{Y}}{\partial \lambda} &= \frac{\partial}{\partial \lambda}\alpha^\eta \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} \tilde{a}_e(\lambda, s)^{\frac{1}{1-\eta}}\right]^{1-\eta}\right] \\ &= \frac{\partial}{\partial \lambda}\alpha^\eta \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} \left\{(1-\lambda)\frac{\sum_{e' \in \mathcal{E}_{z_e}} a_{e'}(s)^{\frac{1}{1-\eta}}}{|\mathcal{E}_{z_e}|} + \lambda a_e(s)^{\frac{1}{1-\eta}}\right\}\right]^{1-\eta}\right] \\ &= \frac{\partial}{\partial \lambda}\alpha^\eta \ell^\eta \mathbb{E}\left[\left[(1+(\alpha-1)(1-\lambda))\sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \lambda \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}}\right]^{1-\eta}\right]\end{aligned}$$

Since $\mathbb{E}\left[a_e(s)^{\frac{1}{1-\eta}}\right]$ exists, the Dominated Convergence Theorem allows us to bring the derivative inside the integral. We dominate the integral with

$$\left[\alpha \sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}}\right]^{1-\eta}.$$

Continuing the derivative

$$\begin{aligned}
\frac{\partial\tilde{Y}}{\partial\lambda} &= \frac{\partial}{\partial\lambda}\alpha^{\eta}\ell^{\eta}\mathbb{E}\left[\left[(1+(\alpha-1)(1-\lambda))\sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \lambda \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}}\right]^{1-\eta}\right] \\
&= \alpha^{\eta}\ell^{\eta}\mathbb{E}\Bigg[(1-\eta)\left[(1+(\alpha-1)(1-\lambda))\sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \lambda \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}}\right]^{-\eta} \\
&\qquad \left[-(\alpha-1)\sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}}\right]\Bigg] \\
&= \alpha^{\eta}\ell^{\eta}\mathbb{E}\left[(1-\eta)\frac{\sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}}\left[a_e(s)^{\frac{1}{1-\eta}} - \frac{\sum_{e'\in\mathcal{E}_{z_e}} a_{e'}(s)^{\frac{1}{1-\eta}}}{|\mathcal{E}_{z_e}|}\right]}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e\in\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}} + \lambda\sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_e)^{\frac{1}{1-\eta}} a_e(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right] \\
&= (1-\eta)\alpha^{\eta}\ell^{\eta}\sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} z_e^{\frac{1}{1-\eta}} \\
&\quad \times\Bigg\{\mathbb{E}\left[\frac{a_e(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}} + \lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right] \\
&\qquad - \sum_{e'\in\mathcal{E}_{z_e}} \frac{1}{|\mathcal{E}_{z_e}|}\mathbb{E}\left[\frac{a_{e'}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}} + \lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right]\Bigg\}
\end{aligned}$$

Notice that for $e'_0, e'_1 \in \mathcal{E}_{z_e}$,

$$\begin{aligned}
&\mathbb{E}\left[\frac{a_{e'_0}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}} + \lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right] \\
&\quad = \mathbb{E}\left[\frac{a_{e'_1}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}} + \lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}} (z_{e''})^{\frac{1}{1-\eta}} a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right],
\end{aligned}$$

since they enter symmetrically into the denominator and have the same distribution. Therefore,

$$\sum_{e' \in \mathcal{E}_{z_e}} \frac{1}{|\mathcal{E}_{z_e}|} \mathbb{E}\left[\frac{a_{e'}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}+\lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right]$$
$$= \mathbb{E}\left[\frac{a_{e'''}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}+\lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right],$$

for some $e''' \in \mathcal{E}_{z_e} \subseteq \mathcal{E}$.

Therefore,

$$\frac{\partial\tilde{Y}}{\partial\lambda} = (1-\eta)\alpha^{\eta}\ell^{\eta}\sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} z_e^{\frac{1}{1-\eta}}$$
$$\times\left\{\mathbb{E}\left[\frac{a_{e}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}+\lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right]\right.$$
$$\left.-\mathbb{E}\left[\frac{a_{e'''}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}+\lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right]\right\}.$$

To complete the proof, we simply need to show that for $e \in \alpha\mathcal{E} \setminus \mathcal{E}$ and $e''' \in \mathcal{E}_{z_e} \subseteq \mathcal{E}$,

$$\mathbb{E}\left[\frac{a_{e}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}+\lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right]$$
$$> \mathbb{E}\left[\frac{a_{e'''}(s)^{\frac{1}{1-\eta}}}{\left[(1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}+\lambda\sum_{e''\in\alpha\mathcal{E}\setminus\mathcal{E}}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}\right]^{\eta}}\right].$$

But this follows for $\lambda \in (0,1)$ from lemma 1 when $\text{Var}(a_e(s)) > 0$. We take $X_2 = a_e(s)^{\frac{1}{1-\eta}}$, $X_1 = a_{e'''}(s)^{\frac{1}{1-\eta}}$, $\alpha_1 = [1+(\alpha-1)(1-\lambda)](z_e)^{\frac{1}{1-\eta}}$ which is greater than $\alpha_2 = \lambda(z_e)^{\frac{1}{1-\eta}}$. And, we take $B$ to be equal to the denominator to the $1/\eta$, minus $\alpha_1 X_1 + \alpha_2 X_2$, or equivalently

$$B = (1+(\alpha-1)(1-\lambda))\sum_{e''\in\mathcal{E}\setminus e'''}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}} + \lambda\sum_{e''\in(\alpha\mathcal{E}\setminus\mathcal{E})\setminus e}(z_{e''})^{\frac{1}{1-\eta}}a_{e''}(s)^{\frac{1}{1-\eta}}.$$

Finally, we take $\beta$ in the lemma to be equal to $\eta$. It then follows that $Y(\alpha\ell, \alpha\mathcal{E}) = Y(1) > Y(0) = \alpha Y(\ell, \mathcal{E})$, completing the proof.

□

We next move to our result that the externalities are disappearing in the limit.

**Proposition 14** (Proposition 2 in the main text)**.** *Suppose that $\ell > 0$, there is a set of establishments $\mathcal{E}$, and $\alpha > 1$. Then*

$$\frac{Y(\alpha\kappa\ell, \alpha\kappa\mathcal{E})}{\alpha Y(\kappa\ell, \kappa\mathcal{E})} \to 1$$

*as $\kappa \to \infty$.*

*Proof.* The proof is relatively straightforward. We will consider $\kappa > 1$. Then $\kappa\mathcal{E}$ can be partitioned into sets $\Gamma_1(\kappa), \ldots, \Gamma_n(\kappa)$ where $|\Gamma_i(\kappa)| = \kappa$, $n = |\mathcal{E}|$, and for every $e \in \Gamma_i(\kappa)$, $z_e = \gamma_i$ where $\gamma_i > 0$ is some constant. Intuitively, we are partitioning the establishments into sets with the same ex-ante productivity shocks $z$. We can then look at the limit within $\Gamma_i(\kappa)$ as $\kappa \to \infty$. We have

$$\begin{aligned}
Y(\kappa\ell, \kappa\mathcal{E}) &= \kappa^\eta \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\kappa\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right] \\
&= \kappa^\eta \ell^\eta \mathbb{E}\left[\left(\sum_{i=1}^{n} \gamma_i^{\frac{1}{1-\eta}} \left[\sum_{e\in\Gamma_i(\kappa)} a_e(s)^{\frac{1}{1-\eta}}\right]\right)^{1-\eta}\right] \\
&= \kappa \ell^\eta \mathbb{E}\left[\left(\sum_{i=1}^{n} \gamma_i^{\frac{1}{1-\eta}} \frac{\left[\sum_{e\in\Gamma_i(\kappa)} a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right]
\end{aligned}$$

But then by the strong law of large numbers

$$\mathbb{P}\left[\lim_{\kappa\to\infty} \frac{\left[\sum_{e\in\Gamma_i(\kappa)} a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa} = \mu_\eta\right] = 1$$

where $\mu_\eta \equiv \mathbb{E}\left[a_e(s)^{\frac{1}{1-\eta}}\right]$. We can then take limits

$$\begin{aligned}
\lim_{\kappa\to\infty} \frac{Y(\kappa\ell, \kappa\mathcal{E})}{\kappa} &= \lim_{\kappa\to\infty} \ell^\eta \mathbb{E}\left[\left(\sum_{i=1}^{n} \gamma_i^{\frac{1}{1-\eta}} \frac{\left[\sum_{e\in\Gamma_i(\kappa)} a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right] \\
&= \ell^\eta \lim_{\kappa\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} \gamma_i^{\frac{1}{1-\eta}} \frac{\left[\sum_{e\in\Gamma_i(\kappa)} a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right].
\end{aligned}$$

We first need to confirm this limit exists. Jensen's inequality implies,

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right]\leq\left\{\mathbb{E}\left[\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right]\right\}^{1-\eta}$$
$$=\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\mu_\eta\right)^{1-\eta},$$

so it is bounded above. But then we had shown above that there are increasing returns to scale. Therefore,

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right]\leq\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\kappa+1)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa+1}\right)^{1-\eta}\right].$$

Then because it is increasing and bounded above it must converge. Then we have

$$Z\equiv\lim_{\kappa\to\infty}\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right].$$

Then

$$\frac{Y(\alpha\kappa\ell,\alpha\kappa\mathcal{E})}{\alpha Y(\kappa\ell,\kappa\mathcal{E})}=\frac{1}{\alpha}\frac{\alpha\kappa\ell^\eta\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\alpha\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\alpha\kappa}\right)^{1-\eta}\right]}{\kappa\ell^\eta\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right]}$$
$$=\frac{\ell^\eta\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\alpha\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\alpha\kappa}\right)^{1-\eta}\right]}{\ell^\eta\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_i^{\frac{1}{1-\eta}}\frac{\left[\sum_{e\in\Gamma_i(\kappa)}a_e(s)^{\frac{1}{1-\eta}}\right]}{\kappa}\right)^{1-\eta}\right]}$$
$$\to\frac{\ell^\eta Z}{\ell^\eta Z}=1.$$

Finishing the proof. □

Then the last result we have is that firms are under-rewarded.

**Proposition 15.** *The expected profits of adding a proportional number of establishments are lower than the contribution to expected production. In math, for $\alpha > 1$,*

$$\mathbb{E}\left[\sum_{e\in\alpha\mathcal{E}\setminus\mathcal{E}} \pi_e(s)\right] < Y(\ell,\alpha\mathcal{E}) - Y(\ell,\mathcal{E})$$

*where $\pi_e(s) = z_e a_e(s)\ell_e(s)^\eta - w(s)\ell_e(s)$.*

*Proof.* The proof is relatively straightforward. We will start by figuring out profits. Total production is

$$Y(\ell,\alpha\mathcal{E}) = \ell^\eta \mathbb{E}\left[\left[\sum_{e\in\alpha\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right]$$

Furthermore, the wages paid to workers is

$$w\ell = \eta\ell^\eta \mathbb{E}\left[\left[\sum_{e\in\alpha\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right]$$

So total profits are

$$\mathbb{E}\left[\sum_{e\in\alpha\mathcal{E}} \pi_e(s)\right] = (1-\eta)Y(\ell,\alpha\mathcal{E}).$$

Therefore, the profits of the firms that enter are $(1-\eta)\frac{\alpha-1}{\alpha}Y(\ell,\alpha\mathcal{E})$.We want to prove that

$$(1-\eta)\frac{\alpha-1}{\alpha}Y(\ell,\alpha\mathcal{E}) < Y(\ell,\alpha\mathcal{E}) - Y(\ell,\mathcal{E})$$

which is equivalent to

$$Y(\ell,\mathcal{E}) < \frac{1+(\alpha-1)\eta}{\alpha}Y(\ell,\alpha\mathcal{E}).$$

Notice that these two sides equal each other when $\alpha = 1$. But then

$$\begin{aligned}\frac{1+(\alpha-1)\eta}{\alpha}Y(\ell,\alpha\mathcal{E}) &> [1+(\alpha-1)\eta]\, Y\left(\frac{1}{\alpha}\ell,\mathcal{E}\right)\\ &= \frac{1+(\alpha-1)\eta}{\alpha^\eta}Y(\ell,\mathcal{E})\end{aligned}$$

using increasing returns to scale. Notice that when $\alpha = 1$, this is equal to $Y(\ell,\mathcal{E})$. Furthermore,

$$\frac{\partial}{\partial\alpha}\left[\frac{1+(\alpha-1)\eta}{\alpha^\eta}\right] = \frac{\eta}{\alpha^\eta} - \eta\frac{1+(\alpha-1)\eta}{\alpha^{\eta+1}}$$

$$
\begin{aligned}
&= \frac{1}{\alpha^{\eta+1}} \left[\eta\alpha - \eta - (\alpha - 1)\eta^2\right] \\
&= \frac{1}{\alpha^{\eta+1}} \eta(\alpha - 1) \left[1 - \eta\right] \\
&> 0
\end{aligned}
$$

Therefore,

$$
Y(\ell, \mathcal{E}) < \frac{1 + (\alpha - 1)\eta}{\alpha^\eta} Y(\ell, \mathcal{E})
$$

for $\alpha > 1$ and

$$
\frac{1 + (\alpha - 1)\eta}{\alpha^\eta} Y(\ell, \mathcal{E}) < \frac{1 + (\alpha - 1)\eta}{\alpha} Y(\ell, \alpha\mathcal{E})
$$

proving the result. □

We have a few different empirical predictions.

**Proposition 16.** *To a first-order log approximation*

$$
Var(\log w(s)) = \left(\sum_e x_e^2\right) \sigma^2
$$

*where* $x_e = \frac{z_e^{\frac{1}{1-\eta}}}{\sum_{e'} z_{e'}^{\frac{1}{1-\eta}}}$.

*Proof.* The proof is really straightforward. Recall that

$$
w(s)\ell = \eta\ell^\eta \mathbb{E}\left[\left[\sum_{e \in \mathcal{E}} \left(z_e a_e(s)\right)^{\frac{1}{1-\eta}}\right]^{1-\eta}\right]
$$

Therefore, to first order

$$
\log w(s) = \sum_e x_e \log a_e(s).
$$

And then taking the variance

$$
\begin{aligned}
Var(\log w(s)) &= \sum_e x_e^2 Var(\log a_e(s)) \\
&= \sum_e x_e^2 \sigma^2 \\
&= \left(\sum_e x_e^2\right) \sigma^2
\end{aligned}
$$

completing the proof. ☐

The next thing we want to do is relate this to the number of firms. This relies on Gabaix (2011).

**Proposition 17.** *Suppose that ex-ante productivity is drawn from a power law distribution*

$$\mathbb{P}\left[z_e > z\right] = a z^{-\zeta}$$

*for* $Z > a^{1/\zeta}$*. Then as* $N \to \infty$ *the variance of log wages follows*

$$\begin{aligned} Var \log w(s) &\sim \frac{v_\zeta}{(\log N)^2}\sigma^2 \text{ for } \zeta(1-\eta) = 1, \\ Var \log w(s) &\sim \frac{v_\zeta}{\left(N^{1-\frac{1}{\zeta(1-\eta)}}\right)^2}\sigma^2 \text{ for } 1 < \zeta(1-\eta) < 2, \\ Var \log w(s) &\sim \frac{v_\zeta}{N}\sigma^2 \text{ for } \zeta(1-\eta) \geq 2. \end{aligned}$$

*Proof.* This result follows immediately from Proposition 4 and Gabaix (2011) Proposition 2. The key thing to note is that if $z_e$ is distributed with power law $\zeta$, then $s_e = z_e^{\frac{1}{1-\eta}}$ is distributed according to a power law with parameter $\frac{\zeta}{1-\eta}$. To see this,

$$\begin{aligned} \mathbb{P}\left[s_e > s\right] &= \mathbb{P}\left[z_e^{\frac{1}{1-\eta}} > s\right] \\ &= \mathbb{P}\left[z_e > s^{1-\eta}\right] \\ &= a\left(s\right)^{-\zeta(1-\eta)}. \end{aligned}$$

Then we simply apply proposition 2 from Gabaix (2011). ☐

Here we go through and figure out how the variance of log employment varies across other possibilities.

**Proposition 18.** *In a first-order log approximation to a productivity shock,*

$$\Delta \log \ell_e(s) \approx \frac{1}{1-\eta}\left[1 - x_e\right] \Delta \log a_e(s)$$

*where* $x_e = \frac{w(s)\ell_e(s)}{\sum_{e'} w(s)\ell_{e'}(s)}$.

*Proof.* Again the proof is going to be very straightforward. The first-order condition for a single firm implies that

$$\eta z_e a_e(s)\ell_e(s)^{\eta-1} = w(s)$$

and

$$w(s)\ell = \eta \ell^{\eta} \mathbb{E}\left[\left[\sum_{e\in\mathcal{E}} (z_e a_e(s))^{\frac{1}{1-\eta}}\right]^{1-\eta}\right]$$

Then taking a first-order approximation for a single shock

$$d\log a_e(s) = d\log w(s) + (1-\eta) d\log \ell_e(s)$$

$$d\log w(s) = x_e d\log a_e(s).$$

Then solving

$$d\log \ell_e(s) = \frac{1}{1-\eta}\left(1-x_e\right) d\log a_e(s)$$

finishing the proof. □

Then we get the total variance.

**Proposition 19.** *To a first-order approximation, the variance of log employment is decreasing in HHI. In math,*

$$\sum_e x_e Var\left(\log \ell_e(s)\right) \approx \frac{\sigma^2}{1-\eta}\left[1-\left(\sum_e x_e^2\right)\right].$$

*Proof.* The proof is simple algebraic manipulations. Notice that

$$d\log \ell_e(s) = \frac{1}{1-\eta}\left[d\log a_e(s) - d\log w(s)\right]$$

Therefore,

$$\begin{aligned}
Var(\log \ell_e(s)) &= \frac{1}{(1-\eta)^2}\left[Var(\log a_e(s)) - 2Cov\left(\log a_e(s), \log w(s)\right) + Var(\log w(s))\right] \\
&= \frac{1}{(1-\eta)^2}\left[\sigma^2 - 2x_e\sigma^2 + \left(\sum_e x_e^2\right)\sigma^2\right] \\
&= \frac{\sigma^2}{(1-\eta)^2}\left[1 - 2x_e + \left(\sum_e x_e^2\right)\right]
\end{aligned}$$

We can then just take the weighted sum across the firms to get the result. □

# B Appendix for Empirical Results

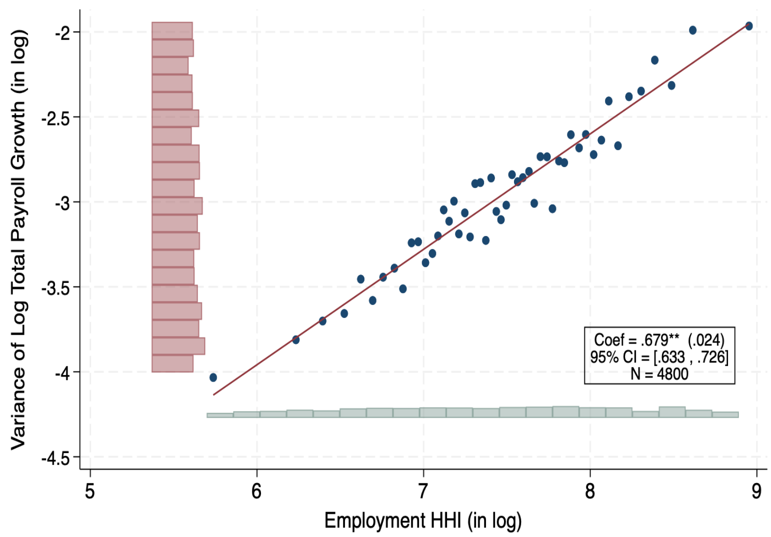
## B.1 Testing Model Predictions

In Figure 2a, Figure 2b, and Figure 3 in the main text, we use the number of establishments in the horizontal axis to proxy the environments of local labor markets. In the extended model with ex-ante heterogeneous establishments, however, the variance of log wage increases in payroll HHI in the local labor market. To accommodate this, Figure B.1a, Figure B.1b, and Figure B.2 use payroll HHI in the horizontal axis instead of the number of establishments.

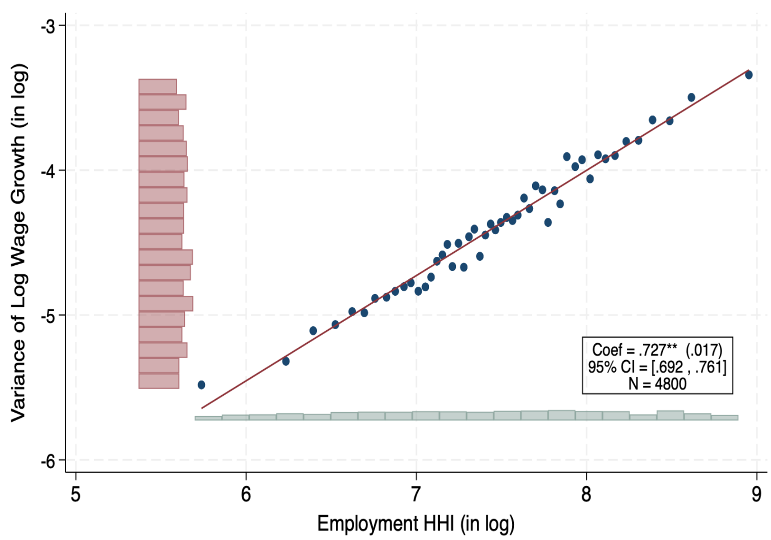
All the patterns in the main text hold—payroll and wages are more stable, and establishment-level employment is more volatile in less concentrated labor markets.

Figure B.1: Volatility of Payroll and Average Wage Growth and Number of Establishments

(a) Payroll



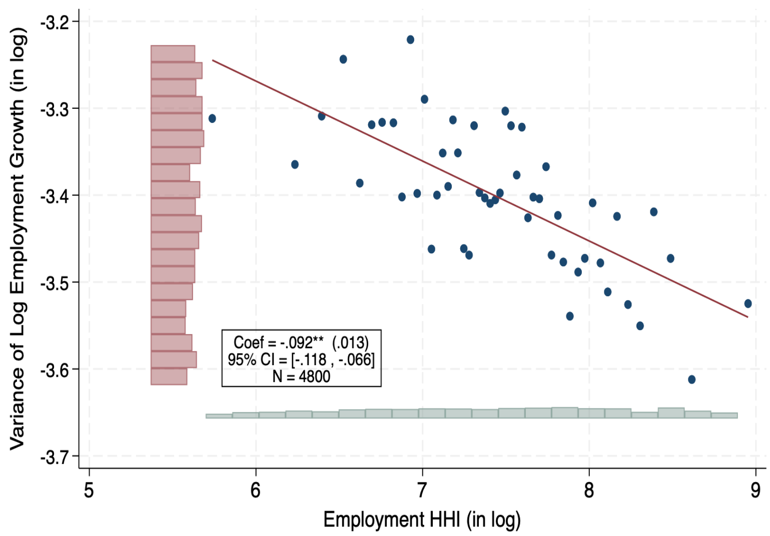

(b) Average Wage




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

## Prediction 2. Establishment Employment is More Volatile in Less Concentrated Labor Markets

Figure B.2: Volatility of Establishment-level Employment Growth and Local Labor Market Concentration




*Note*: The figures show the histograms and the binned scatter-plots of the relationship between volatility of establishment-level employment growth and employment HHI (in log) across local labor markets in Japan. The unit of observation is the local labor market, a pair of a JSIC 2-digit industry, and a commuting zone. The vertical axis is the average of establishment-level log variance of log growth of employment over 1986-2016 in each local labor market. The horizontal axis is the log of payroll HHI, averaged over 1986-2016 in each local labor market.

## B.2 Different Local Labor Market Definition: 3-digit Industry

In the main text of the paper, we define a local labor market as a pair of commuting zones and a JSIC 2-digit industry. In this subsection, we repeat our analysis when we instead define it as a pair of a commuting zone and a JSIC 3-digit industry. The average payroll share within each local labor market is now 8%, which is four times larger than the average when we use a 2-digit industry.

Table B.1 shows the result. Column (3) is our preferred specification where we replicate our main findings in Table 4 that plants with higher payroll share have lower elasticities.

Table B.1: Effects of JPY Appreciation on Changes in Non-Regular Employment: 3-digit Industry

| | Dep. Var.: Log Changes in Non-Regular Emp. | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| AREER Shock | -2.64 | -2.56 | -0.58 | -0.82 |
| | (0.23) | (0.30) | (0.44) | (0.48) |
| AREER Shock × Log Payroll | | | -0.85 | -0.82 |
| | | | (0.14) | (0.14) |
| AREER Shock × Payroll Share | | -0.31 | 1.83 | |
| | | (0.75) | (0.83) | |
| AREER Shock × (Payroll Share $>$ 3%) | | | | 0.80 |
| | | | | (0.53) |
| Observations | 1,164,359 | 1,164,359 | 1,164,359 | 1,164,359 |
| Covariates | ✓ | ✓ | ✓ | ✓ |
| Year FEs | ✓ | ✓ | ✓ | ✓ |
| Establishment FEs | ✓ | ✓ | ✓ | ✓ |

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