Plants and Productivity in Regional Agglomeration

Thomas J. Holmes* Wen-Tai Hsu† Sanghoon Lee‡

November 8, 2010

Abstract

This paper develops a model of agglommeration and regional trade that incorporates selection over plant productivity. It builds on the structure of Bernard, Eaton, Jensen, and Kortum (2003), by introducing entrepreneurship and endogenizing productivity distributions across locations. Two key ingredients of the model are: First that competition is head-to-head, with multiple firms competing to sell the same product. Second, that firms draw productivity from a "fat-tailed" distribution. We show the implications for productivity distributions with this structure are very different than those of Melitz-style models. In particular, there is no sense that productivity distributions become more compressed in large markets, as occurs in Melitz-style approaches. Moveover, in this environment, analysis of productivity distributions of surviving plants does not distinguish the relative importance of selection from knowledge spillovers in accounting for high average productivity in agglomerations.

*Department of Economics, the University of Minnesota, Federal Reserve Bank of Minneapolis, and the National Bureau of Economic Research.

†Department of Economics, the Chinese University of Hong Kong.

‡Sauder School of Business, the University of British Columbia.

1 Introduction

Models of firm heterogeneity that explicitly take into account exit by low productivity firms have played a prominent role in the international trade literature. The modeling approach of Melitz (2003) has been particularly influential. The alternative modeling approach of Eaton and Kortum (2002), and Bernard, Eaton, Jensen Kortum (2003) (hereafter BEJK) has also had significant impact. Recently, these ideas have been applied to models of regions rather than countries. Papers, including Baldwin and Okubo (2006), Combes, Duranton, Gobillon, Puga, and Roux (2009) (hereafter CDGPR), and Behrens, Duranton, and Robert-Nicoud (2010), follow the approach of Melitz or offshoots of Melitz like Melitz and Ottaviano (2008).

This paper develops a regional analysis that incorporates heterogeneity in productivity, but in contrast to other papers, it follows the BEJK approach. To understand what we do, it is first necessary to highlight two essential differences between the BEJK and the Melitz approaches. First, in BEJK, competition between firms is *head-to-head.* There is more than one potential producer of any given product and the different producers engage in Bertrand competition, market by market. In contrast, in the Melitz approach each firm has a monopoly over a particular differentiated product, as in Dixit and Stiglitz (1977). Second, in BEJK, firms draw their productivity distribution from a distribution with a *fat right tail.* In contrast, in the Melitz approach, it is not necessary that productivity be drawn from a fat tail.

This paper takes the two essential ingredients of the BEJK trade model: head-to-head competition and fat-tailed productivity draws. It adds to this: (1) labor mobility, and (2) a model of freely-mobile entrepreneurial activity. The end product is a model of selection. This arises because only one entrepreneur of a given variety at a given location survives the outcome of head-to-head competition. The end product is also a model of agglomeration and endogenous productivity distributions, in which productivity and agglomeration are explicitly linked. This arises from the equilibrium conditions of free mobility of entrepreneurs and workers, both across locations and across jobs. The main work of the paper is analysis of the endogenous productivity distributions.

We show that imposing the equilibrium mobility conditions in the BEJK structure has content. Average productivity is higher in agglomerated areas, a result that does not necessarily hold without free mobility in BEJK. We note that models with mobility conditions based on the Melitz approach also find that productivity is higher in agglomerated areas.

What sharply distinguishes our approach from Melitz-style approaches is what can be learned from how shapes of productivity distributions vary across locations. A recurring theme in the analysis of our model is that productivity varies in a "smooth way" across locations. Mean productivity may be higher in one location than the other, but the shape of the distribution doesn't change. The previous literature based on Melitz-style cutoff rules relies heavily on the fact that, in this class of models, productivity distributions do not vary in a smooth way. Analysis of productivity data generated in our model turns out to yield very different conclusions from what is found in the previous literature.

To explain the point, we start by comparing our results to Syverson (2004), which argues that because of increased selection, productivity distributions should be more compressed in large markets. In contrast, in our model, the variance of the level of productivity is actually larger in large markets. Looking instead at the variance of the log of productivity, it is constant across locations.

We can highlight the intuition for the divergence in results with a discussion of the case of a two-point distribution for productivity, low and high. This is a simple example that fits Syverson's case where there is no fat-tail. Small, rural locations will attract few entrants and so the threshold on productivity needed for survival will tend to be low; i.e. entrants that draw low productivity may survive. Large, urban locations will attract many entrants and this high degree of competition will increase the threshold for survival. Perhaps only high quality firms survive. The effect of more entry in large markets is to push firms up against the upper bound of productivity, compressing the distribution. It should now be clear why assuming a fat right tail on the productivity distribution changes things. Even though the increased selection in large locations does indeed weed out low productivity firms, there is no force of compression at the top of the distribution. This follows because there is always plenty room at the top; that is what a fat tail means. The distribution of survivors shifts up, but is not compressed.

Next we turn to CDGPR, which aims to disentangle the contributions to productivity of standard agglomeration economies, from those related to increased selection in large markets. Taking a Melitz-like approach, CDGPR's basic idea is that standard agglomeration economies will shift the entire distribution of productivity smoothly to the right, while increased selection in larger markets will improve the distribution by more extensively truncating it on the left, leaving the right portion of the distribution alone. The paper finds little evidence of increased truncation in larger markets, leading it to conclude that the higher average productivity found in larger markets must be mainly due to standard agglomeration economies.

In our model, changing parameters related to agglomeration and parameters related to selection both impact the productivity distribution in the same, smooth way. There is no sense that selection leads to a more truncated distribution. So the fact that empirically productivity distributions in bigger cities shift to the right in a smooth way, rather than through sharper truncation on the left, does not imply that selection is an unimportant contributor to cross-location differences in productivity. It is crucial for our result that competition is head-to-head; because of this there is no one uniform "productivity cutoff" at a given location. A particular firm might draw a high productivity level but get knocked out by a head-to-head rival at the same location that draws an even higher productivity. Another firm at the same location might draw low productivity and survive, if its head-to-head rivals, by chance, draw even lower. These random variations across different groups of head-to-head rivals leads the distribution of the survivor productivities aggregated over all plants at a location to be smooth, rather than truncated.

Last, we compare our results to Hsieh and Klenow (2009). This paper considers the distortions created by government policies that favor some firms over others. For example, suppose low productivity firms receive subsidies funded by taxes on high productivity firms. The strategy of Hsieh and Klenow (2009) is to use information on the variance of productivities to draw inferences about the extent of such distortions. Their approach works in the Melitz-style model they develop. Subsidies directed at low productivity firms reduce the survival cutoff on the left-side of the productivity distribution but leave the right side of the distribution alone. Analogous to what happens in Syverson as the cutoff increases, economies without such distortions have more compressed productivity distributions.

We consider a particular government policy with a productivity cutoff where firms above the cutoff pay a tax (a higher tax, the higher firm productivity) and firms below the cutoff receive a subsidy (a higher subsidy, the lower firm productivity). The resulting structure that we obtain is highly tractable, making it potentially useful for future empirical work on such distortions. We compare economies that differ in the extent of the distortion and show the distortion lowers average productivity in two ways. First, following the same logic as in Hsieh and Klenow (2009) and Rustucia and Rogerson (2008), the distortion misallocates production among a given set of firms. Second, we show the distortion reduces incentives for entrepreneurship. There is less entry in an economy with distortions and thus less selection. While the distortion lowers average productivity, the shape of the productivity distribution does not change. With the head-to-head competition taking place in our model, even high productivity firms can potentially be knocked out by inferior rivals receiving a distortionary subsidies. Distortions shift the distribution of productivity down throughout its support, and do not just move the truncation point of a fixed distribution. Put in another way, distortions do indeed lower mean productivity, but do not manifest themselves through increasing the dispersion of productivity.

As in BEJK, the Fréchet distribution is put to work extensively throughout this paper.

The Fréchet is the limiting distribution of the maximum of M independent draws from a fat-tailed distribution. In our model, there are many head-to-head competitors for each product, each with independent draws, and that is how we motivate going from a fat-tailed distribution to the Fréchet.[1] To motivate our use of a fat-tailed distribution for the underlying productivity draw, we note that these are quite common in the literature. In particular, Melitz and Ottaviano (2008), Chaney (2008), and Eaton, Kortum, and Kramarz (2008) all assume a Pareto distribution for underlying draws, the proto-typical example of a fat-tailed distribution.

The rest of the paper is organized as follows. Section 2 describes the model and the assumption on productivity draws. Section 3 studies the equilibrium given individuals' location and job choices. Section 4 studies job and location choices and determine when agglomeration arises. Section 5 studies equilibrium productivity distributions of survivors and the survival rates across locations. To study the impact of selection and agglomeration on productivity distribution separately, this section also develops a pure agglomeration model in which selection is completely shut down and compare this model with a case of our model in which the standard agglomeration economies are shut down. Section 6 studies the impact of policy distortion. Section 7 concludes.

2 Model

2.1 Description of the Model

There are two locations, $i = 1, 2$, that are ex ante identical. In the equilibrium of the model, it may happen that one location attracts more people than the other. We label things in such cases so that location 1 is the "big city" and location 2 is the "small city."

All agents have the same preferences for a composite good Q and land L; these preferences are represented by the utility function

$$U(Q, L) = Q^{\beta} L^{1-\beta}.$$

The composite is an aggregation of differentiated goods indexed by j on the unit interval. It follows the standard CES form,

$$Q = \left(\int_0^1 (q(j))^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}},$$

[1] Kortum (1997) motivates Fréchet by treating firms as getting many multple draws over time for the same firm. This is analogous to the maximum of many draws from different firms at a point in time.

where σ is the elasticity of substitution. We note that in setting up land as the force of dispersion, we are following the modeling approach in Helpman (1998). Redding and Sturm (2008) is a recent paper that calibrates a quantitative application of the Helpman (1998) approach and they interpret L broadly as a "nontradeable amenity."

There is a fixed supply of land (or nontradeable amenity) equal to $\bar{L}$ at each location. Analogous to what Redding and Sturm do, we assume the rents on the land at a particular location are distributed equally in a lump sum fashion among the population who lives at the location.

There is a measure $\bar{H}$ individuals in the economy. Individuals first choose whether to live and work in city 1 or live and work in city 2. Next, they choose whether to be an entrepreneur or to be employed as a worker. Let N_i be the number of individuals choosing to be a worker in city i and let M_i be the number of entrepreneurs. Let $H_i = N_i + M_i$. The resource constraint implies that

$$N_1 + M_1 + N_2 + M_2 = \bar{H}.$$

It will be convenient to work with fractions. Define these by

$$n_i = \frac{N_i}{\bar{H}}, \qquad m_i = \frac{M_i}{\bar{H}}, \qquad h_i = \frac{H_i}{\bar{H}} = n_i + m_i.$$

We now explain the process through which firms are created and productivities are determined, beginning with the arrival of M_i entrepreneurs at location i. Each entrepreneur picks a product $j \in [0,1]$ to attempt to enter. Let $S_i(j)$ be the density of entrepreneurs attempting to enter product j (the number of startups for this product). All entrepreneurs arriving at i pick some industry; i.e.,

$$M_i = \int_0^1 S_i(j)dj.$$

Each entrepreneur entering an industry obtains a plant. The productivity of a plant has two components that enter multiplicatively. First, there is a term A_i that is constant across all plants at location i and depends on the amount of knowledge spillovers. If there are H_i individuals located at i, then

$$A_i = H_i^{\zeta}.$$

The parameter ζ governs the significance of agglomeration spillovers. In particular, if $\zeta = 0$, then $A_i = 1$ and there are no spillovers. Second, there is a random term y to productivity that depends upon the entrepreneur's luck. An entrepreneur at location i with productivity

y for a particular good j can produce $A_i y$ units of good j, per unit of labor procured at location i.

We assume an iceberg transportation cost τ between the two locations. To deliver one unit of any differentiated good j to a different location, $\tau \geq 1$ units must be shipped.

There are three stages in the model. In stage 1, individuals choose where to live and what job to hold.

In stage 2, M_i entrepreneurs at location i allocate themselves across the product space $j \in [0,1]$ so that $S_i(j)$ is the density choosing good j. Each of the $S_i(j)$ entering product j at i obtains a single plant, with one draw of the random productivity term y. We impose as an equilibrium condition that the returns to entering each product j at a location are be equalized. We ignore integer constraints

In stage 3, the $S_1(j)$ plants at location 1 and the $S_2(j)$ plants at location 2 engage in Bertrand price competition for the product j market at each location. At the same time there is market clearing in the labor markets and land markets.

2.2 Distribution of Productivities

We turn now to our crucial assumption about the distribution of productivity draws. Let Y denote the random variable that is being drawn to determine productivity and let $G(y) : [0, \infty) \to [0, 1]$ be its distribution function. (We refer the distribution by either Y or $G(.)$.) As background for our assumption on $G(.)$, imagine taking M independent draws from $G(.)$ and selecting out the maximum, the *extreme value*. The Fisher-Tippet theorem states that when M goes to infinity, the distribution of a properly normalized maximum converges to one of the three type of limiting distributions.[2] We make

Assumption 1 *Y is in the class of distributions for which the limiting distribution of the normalized maximum is Fréchet (Type 2 extreme value distribution), that is, Y falls in the* domain of attraction *for Fréchet.*

The Fréchet distribution function on $[0, \infty)$ has the form of

$$e^{-\alpha y^{-\theta}}, \tag{1}$$

where α is the scale parameter and θ is the shape parameter.

Any distribution that falls in the domain of attraction for Fréchet is fat-tailed, i.e., its density declines at slower than an exponential rate. A distribution $G(.)$ is in the domain of

[2] See Embrechts et al. (1997) for a textbook treatment of Fisher-Tippet Theorem.

attraction for Fréchet if and only if it has a $\theta > 0$ such that for any constant $t > 0$,

$$\lim_{x \to \infty} \frac{1 - G(tx)}{1 - G(x)} = t^{-\theta}. \tag{2}$$

This implies that the smaller the θ, the slower the tail diminishes, and hence the fatter the tail. This tail parameter θ of $G(.)$ actually becomes the shape parameter of its limiting distribution. The Pareto is the best known example of a distribution in this domain of attraction. We add a restriction that $\theta > 2$, because this paper is concerned not only with the mean, but also the variance of the productivity distribution, and the Fréchet distribution has finite variance if and only if $\theta > 2$.

3 Equilibrium for Fixed Location and Job Choices

This section works out the equilibrium in stages 2 and 3, given the fractions (n_1, m_1, n_2, m_2) of individuals in the different jobs and locations determined in stage 1. The contribution of this section is to show that for fixed (n_1, m_1, n_2, m_2), our model maps directly into the international trade model of BEJK. This is an asymptotic result that applies as the size $\bar{H}$ of the economy gets large. We show this mapping and then go directly to BEJK to collect various formulas derived there that we later use.

We begin by analyzing expected profit of being a startup at location i, for a particular product j, as a function of the number of startup plants $S_1(j)$ and $S_2(j)$ at each location. Denote this as $v_i(S_1, S_2)$, leaving implicit the dependence of S_1 and S_2 on j. Conditional on the number of startup plants S_1 and S_2 at each location, the expected return $v_i(S_1, S_2)$ to being one of the startups does not depend upon j because of symmetry. It does depend upon the location i because demand may differ across locations (because of transportation costs) and because the wage and effective labor productivity may differ. Given a set of productivity draws for the S_1 and S_2 startups, the startups play a Bertrand price game at each location and the startup with the lowest delivered cost to serve each location wins, with a price equal to the minimum of the simple monopoly price and the second lowest delivered cost. The return $v_i(S_1, S_2)$ takes the expectation over the S_1 and S_2 productivity draws and combines the expected profit of serving both markets.

We impose the equilibrium condition for stage 2 that the return to entrepreneurial entry at a particular location i be equalized across products j, i.e.,

$$v_i(S_1(j), S_2(j)) = constant_i \text{ for all } j \in [0, 1]. \tag{3}$$

Our characterization of equilibrium builds on the following lemma.

Lemma 1 *(i) The expected return $v_i(S_1, S_2)$ to an entrepreneur for entering a particular product at i is strictly decreasing in the levels of entry S_1 and S_2 at the two locations for the product. (ii) Assume productivity draws are from the Fréchet given by (1). Suppose for (S_1, S_2) and (S_1', S_2') that $S_1 < S_1'$, $S_1 > S_1'$, and $v_2(S_1, S_2) = v_2(S_1', S_2')$. Then $v_1(S_1, S_2) > v_1(S_1', S_2')$.*

Proof. *Proof of (i).* Take as given a particular set of productivity draws for S_1 and S_2 startups at 1 and 2. Standard Bertrand competition arguments imply that if we add an additional startup to either location, the original S_1 and S_2 startups are weakly worse off. As the distribution of productivity is unbounded on the right, with positive probability the new startup displaces the original startups. Hence, in expectation the original startups are strictly worse off from additional entry. That is, $v_i(S_1, S_2)$ decreases in S_1 and S_2, as claimed. *Proof of (ii).* See the Appendix. ■

That result of part (i)—that the return to entry decreases in the total level of entry—is straightforward. The result of part (ii) is more complex. Suppose that the expected return to entry at location 2 were the same for two different entry patterns, (S_1, S_2) and (S_1', S_2'). To maintain indifference, there must be a tradeoff; i.e., if one pattern has less entry at location 1, this must be offset more entry at location 2. The result says that if an entrant at location 2 is indifferent between such a tradeoff, an entrant at location 1 would strictly prefer the alternative with less entry at location 1. This is intuitive, because we expect that entrants will value reductions in entry at their own locations more than entrants at the other locations would value them. While this is an intuitive result, we are only able to formally prove it for the case where the underlying draws are from the Fréchet.

We use Lemma 1 to show that in stage 2 when entrepreneurs pick products they spread themselves out evenly of the available unit measure of goods, i.e.,

$$S_i(j) = M_i. \tag{4}$$

To see this is an equilibrium, observe that an entrant at 1 gets $v_1(M_1, M_2)$ in the "spread out" equilibrium given by (4). If an entrant at 1 deviates from the "spread out" and "overloads" a particular product, the entrant gets $v_1(M_1 + 1, M_2)$ which is strictly worse, according to Part (i) of Lemma 1. Part (ii) of Lemma 1 implies that the "spread out" equilibrium is the unique equilibrium when the underlying draws are Fréchet. Suppose to the contrary we have an equilibrium where things are not evenly spread out, i.e., there is a j and j' such that $S_1(j) < S_1(j')$. Indifference across j and j' at location 2 implies $S_2(j) > S_2(j')$. But then

Part (ii) of Lemma 1 along with the Fréchet assumption implies at location 1 the return to choosing product j is greater than the return to product j', contradicting equilibrium condition (3). Without the Fréchet assumption, we don't have a formal proof that the spread out equilibrium is the unique equilibrium. That being said, we have no reason to suspect other equilibria exist in the general case. We focus on the spread out equilibrium given by (4) in what follows.

Thus for each product j, there are $S_i(j) = M_i$ plants. Only the one drawing the highest productivity in equilibrium will potentially produce in equilibrium. We will be interested then, in the distribution of the maximum productivity over the M_i draws, and, as we will see, the distribution of the second highest draw.

Formally, let

$$\widehat{Z}_{1i} \equiv \max\{y_1, y_2, ..., y_{M_i}\}$$

be the maximum over the productivity draws for the M_i startups, where these are *i.i.d.* draws from $G(.)$. Analogously, let $\widehat{Z}_{2i}$ be the second highest. The effective labor productivities of the highest and second highest productivity startups equal $A_i\widehat{Z}_{1i}$ and $A_i\widehat{Z}_{2i}$, where again the adjustment factor A_i depends upon knowledge spillovers, $A_i = H_i^{\zeta}$.

We determine the limiting distribution when $\bar{H}$ is large of the two highest productivity startups. If we hold fixed the underlying distribution $G(.)$ per draw and take $\bar{H}$ and hence M_i off to infinity, of course the highest productivity among the M_i draws goes off to infinity. Thus, for our asymptotic result, we rescale $\widehat{Z}_{1i}$ and $\widehat{Z}_{2i}$ to let the limiting distribution be independent of $\bar{H}$. This rescaling amounts to changing the unit of productivity as $\bar{H}$ gets large. In particular, for each $\bar{H}$, we select an $\alpha_{\bar{H}}$ such that

$$Z_{1i} = \alpha_{\bar{H}} A_i \widehat{Z}_{1i}, \qquad Z_{2i} = \alpha_{\bar{H}} A_i \widehat{Z}_{2i}. \tag{5}$$

In order to define the rescaling factor $\alpha_{\bar{H}}$, some simple notions of *regular variation* are needed.[3]

Definition 1 *A measurable, positive function $g(.)$ is regularly varying if there exists a $\theta \in \mathbf{R}$ such that for any $t > 0$,*

$$\lim_{x\to\infty} \frac{g(tx)}{g(x)} = t^{\theta}.$$

When $\theta = 0$, then $g(.)$ is said to be slowly varying.

The Fisher-Tippet theorem (Theorem. 3.2.3 in Embrechts et al. [1997]) and the further details about the domain of attraction for Fréchet (Thm 3.3.7 in Embrechts et al. [1997])

[3] See the Appendix of Embrechts et al. (1997) for a quick treatment of regular variation.

imply that if Z_1 is the maximum among M draws from $G(.)$, for $G(.)$ in the domain of attraction for Fréchet, then

$$\lim_{M\to\infty} \Pr\left[c_M^{-1} Z_1 < z\right] = e^{-z^{-\theta}}, \tag{6}$$

where the normalizing constant $c_M = \left(1/\overline{G}\right)^{-1}(M)$, and $\overline{G}(.) = 1 - G(.)$ is the tail probability.

In fact, $G(.)$ being in the domain of attraction for Fréchet is equivalent to $\overline{G}(.)$ being regularly varying; see (2). This implies that the normalizing constants c_M form a regularly-varying sequence, i.e.,

$$c_M = M^{1/\theta}\ell(M), \tag{7}$$

for some slowly varying function $\ell(.)$. Thus, for each $G(.)$, there exists some slowly varying $\ell(.)$, and we now define the rescaling factor as

$$\alpha_{\bar{H}} = \bar{H}^{-\zeta-1/\theta}\ell(\bar{H}). \tag{8}$$

Lemma 2 *Given $S_i(j) = M_i = m_i\bar{H}$, for any given j and for each $\bar{H}$, let Z_{1i} and Z_{2i} be rescaled highest and second highest productivity from M_i draws defined by (5) and (8). We have*

(i) As $\bar{H}$ goes to infinity, the distribution function of Z_{1i} converges pointwise to

$$F_i(z) = e^{-T_i z^{-\theta}}, \tag{9}$$

with scaling parameter

$$T_i = h_i^{\theta\zeta} m_i. \tag{10}$$

The joint distribution function of rescaled highest and second highest productivity Z_{1i} and Z_{2i} converges pointwise to

$$F_i(z_1, z_2) = [1 + T_i(z_2^{-\theta} - z_1^{-\theta})]e^{-T_i z_2^{-\theta}}, \tag{11}$$

for $0 \leq z_1 \leq z_2$.

(ii) If the distribution $G(.)$ is Fréchet given by (1), then (9) and (11) hold exactly for any value of $\bar{H}$.

Proof. See the Appendix. ■

Lemmas 1 and 2 deliver the key result of this section that once the location choices and job choices are determined, our model reduces to BEJK.

Proposition 1 *Our model beginning at stage 2 when the fractions* (n_1, m_1, n_2, m_2) *are determined maps into BEJK for any value of* $\bar{H}$ *if the underlying productivity draws are Fréchet. For more general underlying distributions satisfying Assumption 1, it maps into BEJK asymptotically as* $\bar{H}$ *is large.*

We now describe how the cost and price distributions are determined in BEJK and then collect several useful results from BEJK that we use.

For good j and for $k = 1, 2$, the unit cost of supplying to consumers in location n by the kth most efficient producers located in i is given by

$$C_{kni}(j) = \left(\frac{w_i}{Z_{ki}(j)}\right)\tau_{ni},$$

where $\tau_{ni} = 1$ if $n = i$, $\tau_{ni} = \tau$ if $n \neq i$, and the $Z_{1i}(j)$ and $Z_{2i}(j)$ are random variables whose joint distribution is given by from (11). The producer that actually serves market n has unit cost $C_{1n}(j) = \min_i\{C_{1ni}(j)\}$, and the second lowest cost supplying to market n is $C_{2n}(j) = \min\{C_{2ni^*}(j), \min_{i\neq i^*}\{C_{1ni}(j)\}\}$, where i^* is the region with the lowest cost supplier to n. Bertrand competition implies that the producer with $C_{1n}(j)$ charges $C_{2n}(j)$, and hence the markup is $C_{2n}(j)/C_{2n}(j)$. However, due to the CES utility, there is an upper bound of the markup, as no firm charges a higher markup than the monopoly markup $\mu = \sigma/(\sigma - 1)$, for $\sigma > 1$. (For $\sigma \leq 1$, $\mu = \infty$). Hence, $P_n(j) = \min\{C_{2n}(j), \mu C_{1n}(j)\}$. From (9) and (11), BEJK derive that the joint distribution function of the lowest cost C_{1n} and the second lowest cost C_{2n} is

$$K_n(c_1, c_2) = 1 - e^{-\Phi_n c_1^\theta} - \Phi_n c_1^\theta e^{-\Phi_n c_2^\theta}, \tag{12}$$

where

$$\Phi_n = \sum_{i=1}^{2} T_i(w_i\tau_{ni})^{-\theta}. \tag{13}$$

This parameter $\Phi_n = \sum_{i=1}^{2} T_i(w_i\tau_{ni})^{-\theta}$ distills the parameters of productivity distributions, wages, and the trade cost into one single term governing the cost and price distributions. The price distribution at location n is given by

$$K_n(p) = 1 - \left[1 + \Phi_n(1 - \mu^{-\theta})p^\theta\right] e^{-\Phi_n p^\theta}.$$

The BEJK's analytical results that are useful for our paper are listed as follows.

BEJK Result 1 The probability that location i provides a good at the lowest price in location n is

$$\pi_{ni} = \frac{T_i(w_i\tau_{ni})^{-\theta}}{\sum_{k=1}^{2} T_k(w_k\tau_{nk})^{-\theta}} = \frac{T_i(w_i\tau_{ni})^{-\theta}}{\Phi_n}.$$

BEJK Result 2 In any location n, the probability of buying a good with price lower than p is independent from where the good is purchased from. Letting X_{ni} be the total expenditure of location n on the goods from i, and X_n be the total expenditure, we have

$$X_{ni} = \pi_{ni} X_n.$$

BEJK Result 3 Assume that $\theta + 1 > \sigma$. Let Γ denote the gamma function; the price index is

$$\begin{aligned} P_n &= [\frac{1+\theta-\sigma+(\sigma-1)\mu^{-\theta}}{1+\theta-\sigma}\Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)]^{\frac{1}{1-\sigma}}\Phi_n^{-\frac{1}{\theta}} \\ &\equiv \gamma\Phi_n^{-\frac{1}{\theta}}. \end{aligned}$$

BEJK Result 4 A fraction $\theta/(1+\theta)$ of revenue goes to variable cost.

between workers and entrepreneurs in each location i, that is,

$$\frac{N_i}{M_i} = \frac{n_i}{m_i} = \theta. \tag{15}$$

We can use this result to rewrite the Fréchet scale parameter from equation (10) for the distribution of the maximum productivity plant at i as

$$T_i = h_i^{\theta\zeta} m_i = (1+\theta)^{\theta\zeta} m_i^{\theta\zeta+1}. \tag{16}$$

Since the utility is Cobb-Douglas in goods and land, the indirect utility for an individual choosing to locate at i given composite goods price P_i, land rental R_i, and wage i equals

$$U_i = \left(\frac{1-\beta}{\beta}\right)^{1-\beta} P_i^{-\beta} R_i^{-(1-\beta)} w_i. \tag{17}$$

Total expenditures on the $\bar{L}$ units of land at i must equal the land expenditure share times the total income at i,

$$R_i\bar{L} = (1-\beta)[R_i\bar{L} + (N_i + M_i)w_i].$$

solving for the rent R_i and substituting this into utility (17) yields

$$U_i = \gamma \left(\frac{w_i}{P_i}\right)^{\beta} \left(\frac{1}{N_i + M_i}\right)^{1-\beta}. \tag{18}$$

where γ is constant across the two locations.

For a fixed value of the population shares h_1 (and hence fixed $h_2 = 1 - h_1$), we can impose the equilibrium condition for job choice (15) and we can solve for the w_1, w_2, P_1 and P_2 that are consistent with equilibrium in the output market. Plug these into (18) and let $\tilde{U}_i(h_1)$ be the utility conditioned on h_1. Let $u(h_1) = \tilde{U}_1(h_1)/\tilde{U}_2(h_1)$ be the ratio of utilities. For an interior value $h_1 \in (0,1)$ to be an equilibrium of location choice, it must be that $u(h_1) = 1$.

As is standard in this literature, a *symmetric equilibrium* always exists at $h_1^* = 0.5$, where half of the individuals go to each location. Given symmetry, we can restrict attention to the range $h_1 \geq \frac{1}{2}$ where location 1 is weakly larger than location 2. An equilibrium with $h_1^* > 0.5$ is called an *agglomeration equilibrium.* If $u(1) \geq 1$, then $h_1^* = 1$ is an equilibrium where everyone goes to location 1. Call this a *black-hole equilibrium.* Define an interior equilibrium $h_1^* \in [\frac{1}{2}, 1)$ as *stable* if $du/dh_1 < 0$ at h_1^* and *unstable* if $du/dh_1 > 0$. Proposition 2 shows that depending on the parameters, there are three possibilities for how things can look. Figure 1 illustrates the three cases, showing how the utility ratio u depends upon the share h_1.[4]

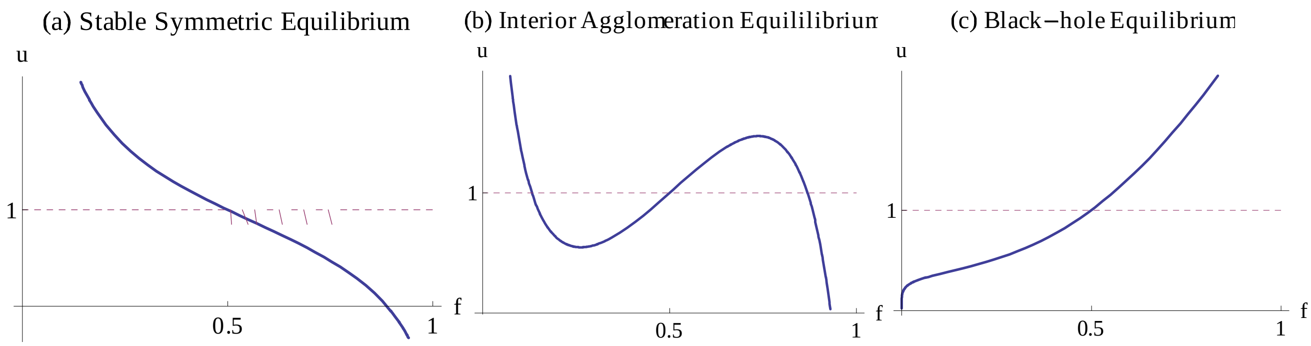

Figure 1: Characterization

Proposition 2 *Define two thresholds for β as a function of other model parameters,*

$$\hat{\beta}_L \equiv \frac{1}{\left(\zeta + 1 + \frac{1}{\theta}\right)} \frac{\tau^\theta(1+2\theta)+1}{\tau^\theta(1+2\theta)-1}, \qquad \hat{\beta}_H = \frac{1+\theta}{1+\theta+\theta\zeta} \quad .$$

(i) The parameter space $\{(\beta, \tau, \theta, \zeta) : \beta \in (0,1), \tau > 1, \theta > 1, \zeta \geq 0\}$ can be partitioned into the the following three subspaces each of which is associated with a distinct characterization of equilibria.

(a) When $\beta \leq \hat{\beta}_L$, the symmetric equilibrium is stable and the only equilibrium.

(b) When $\beta > \hat{\beta}_L$, the symmetric equilibrium is unstable, and there exists a unique agglomeration equilibrium and it is stable. If $\beta \in (\hat{\beta}_L, \hat{\beta}_H)$ the agglomeration equilibrium is interior, $h_1^* \in (0.5, 1)$, and if $\beta \geq \hat{\beta}_H$, the agglomeration is a black-hole, $h_1^* = 1$.

(ii) If there is an interior agglomeration equilibrium $h_1^* \in (0.5, 1)$, then h_1^* increases in β, τ, and ζ.

Proof. See the Appendix. ■

As shown in Helpman (1998), agglomeration increases with the weight β placed on goods consumption relative to land consumption and with the transportation cost τ. That agglomeration increases with the knowledge spillover parameter ζ is a standard finding.

Henceforth, assume the condition for an interior agglomeration holds so that there is agglomeration at location 1 but still there is some economic activity at location 2 that can be compared with location 1.

[4] The parameters for (a) (stable symmetric equilibrium) is $\beta = 0.7, \tau = 1.7, \theta = 1.5, \omega = 0.5, \xi = 0.7$. For (b) (interior agglomeration equilibrium) $\beta = 0.6, \tau = 1.7, \theta = 1.5, \omega = 1.5, \xi = 0.3$ For (c) (black-hole equilibrium), $\beta = 0.6, \tau = 1.7, \theta = 1.5, \omega = 3, \xi = 0.7$.

5 Productivity and Agglomeration

This section analyzes the distribution of productivities of surviving plants at the two locations. The first part derives the implications of the equilibrium conditions of worker and entrepreneurial choice. The conditions imply that the distribution of productivity is higher in the large city compared to the small city. The second part compares the observed productivity distributions to those generated in a pure agglomeration model where selection over productivity is shut down. The analysis shows it is impossible to distinguish these two cases with data on observed productivity distributions. The analysis also shows that it matters to distinguish the two cases—the two cases differ fundamentally in their underlying economics.

5.1 The Productivity Distribution is Higher in the Large City

We begin by deriving two equations from the equilibrium conditions. Recall that Φ_i defined in (13) is the summary statistic that pins down the distribution of prices at location i, gathering together all the various forces including the productivity distributions at each location. Define $\phi \equiv \Phi_1/\Phi_2$ as the ratio of this key statistic between the large and small cities and analogously define $h \equiv h_1/h_2$ to be the ratio of population shares. We derive two conditions linking ϕ and h. One equation uses indifference between wage work and entrepreneurship; the other uses indifference about where to live.

Using BEJK Results 1 and 2, expenditure by location 1 on goods from location 2 equals $X_{12} = X_1 T_2 (w_2 \tau)^{-\theta} / \Phi_1$. Using the analogous expression for X_{21} and the market clearing condition $X_{12} = X_{21}$ implies

$$\frac{X_1}{X_2} = \frac{T_1 w_1^{-\theta} \Phi_1}{T_2 w_2^{-\theta} \Phi_2} = \frac{m_1^{\theta\zeta+1}}{m_2^{\theta\zeta+1}} w^{-\theta} \phi,$$

where we substitute in equation (16) $T_i = (1+\theta)^{\theta\zeta} m_i^{\theta\zeta+1}$ for the productivity distribution scaling parameter and we let $w \equiv w_1/w_2$ be the wage ratio. The equilibrium job choice condition (15) implies $n_1/n_2 = m_1/m_2 = h$. Also, $wh = X_1/X_2$ (equation (14)). Plugging these in gives

$$w^{1+\theta} = h^{\theta\zeta} \phi. \tag{19}$$

Next we use the definition (13) of Φ_i to obtain

$$\phi = \frac{\Phi_1}{\Phi_2} = \frac{T_1 w_1^{-\theta} + T_2 w_2^{-\theta} \tau^{-\theta}}{T_1 w_1^{-\theta} \tau^{-\theta} + T_2 w_2^{-\theta}} = \frac{h^{\theta\zeta+1} w^{-\theta} + \tau^{-\theta}}{h^{\theta\zeta+1} w^{-\theta} \tau^{-\theta} + 1},$$

where again we use the job choice equilibrium condition to substitute in h for m_1/m_2. Solving this expression for w and substituting into (19) yields our first equation linking h and ϕ,

$$h^{\theta\zeta+\theta+1} = \left(\frac{\phi - \tau^{-\theta}}{1 - \tau^{-\theta}\phi}\right)^{1+\theta} \phi^{\theta}. \tag{20}$$

To derive the second equation in h and ϕ, we use the indifference condition from the individual's choice of where to live. This implies that the ratio of utilities $u(h_1) = 1$ for an interior equilibrium, $h_1^* < 1$, i.e.,

$$\begin{aligned} 1 &= \frac{U_1}{U_2} = \frac{\left(\frac{w_1}{P_1}\right)^{\beta} (N_1 + M_1)^{-(1-\beta)}}{\left(\frac{w_2}{P_2}\right)^{\beta} (N_2 + M_2)^{-(1-\beta)}} \\ &= w^{\beta} \phi^{\frac{\beta}{\theta}} h^{-(1-\beta)}. \end{aligned}$$

Using (19) to substitute in for w, we obtain our second equation

$$h^{(1-\beta)(1+\theta)-\beta\zeta\theta} = \phi^{\beta(1+2\theta)/\theta}. \tag{21}$$

With the derivation of conditions (20) and (21) complete, we can now analyze productivity and selection. If a plant for a particular product j at location i survives, it is necessarily the most efficient plant for product j at i. As derived in Section 3, the distribution of the most efficient plant at i for a given j is Fréchet, with density $f_i(z) = T_i \theta z^{-\theta-1} e^{-T_i z^{-\theta}}$, where T_i and θ are the scaling and shape parameters. In addition, for the plant to survive, its

θ, and scaling parameter equal to

$$\begin{aligned}\widehat{T}_1 &= w_1^\theta \Phi_1 \\ &= T_1 + T_2 \left(\frac{w_2 \tau}{w_1}\right)^{-\theta}. \end{aligned} \tag{22}$$

The selection induced by competition with the other location increases the scaling from T_1 to a higher level $\widehat{T}_1$; i.e., it shifts the distribution to the right. Similarly, the productivity distribution of survivors in location 2 has scale parameter $\widehat{T}_2 = w_2^\theta \Phi_2$, but otherwise has the same shape parameter θ. Therefore, the surviving plants at location 1 have a higher distribution than the survivors at 2 (in the sense of first-order stochastic dominance) if $\widehat{T}_1 > \widehat{T}_2$. Our result is

Proposition 3 *Suppose there is an agglomeration at location 1 (the large city), i.e., $h > 1$.*

(i) The productivity distributions of survivors are Fréchet at both locations with the same shape parameter θ, but the scaling parameter $\widehat{T}_1 = w_1^\theta \Phi_1$ at location 1 is strictly higher than the scaling parameter $\widehat{T}_2 = w_2^\theta \Phi_2$ at location 2.

(ii) The ratio of the mean productivities of the survivors equals

$$\frac{E\left[Z_1 | survive\right]}{E\left[Z_2 | survive\right]} = \frac{\frac{w_1}{P_1}}{\frac{w_2}{P_2}} = h^{\frac{1-\beta}{\beta}} \tag{23}$$

which is strictly greater than one. The variance of the productivity distribution of survivors is higher at the large city.

(iii) The mean of the logarithm of survivors' productivities is higher in the large city, while the variance of log productivity is constant across the two locations.

Proof. *Proof of (i).* We need to show that $\widehat{T}_1 / \widehat{T}_2 = w^\theta \phi > 1$. This equals

$$\begin{aligned} w^\theta \phi &= h^{\left(\frac{\theta\varsigma}{1+\theta}\right)\theta} \phi^{\frac{\theta}{1+\theta}} \phi \\ &= h^{\frac{(1-\beta)\theta}{\beta}} > 1 \end{aligned} \tag{24}$$

To obtain the first line, we substitute in for w using (19). To obtain the second line, we solve out (21) for ϕ in terms of h and substitute in. *Proof of (ii).* For the Fréchet with scaling parameter T, the mean and variance equal $T^{1/\theta}\Gamma\left((\theta-1)/\theta\right)$ and $T^{2/\theta}[\Gamma\left((\theta-2)/\theta\right) -$

$\Gamma^2\left((\theta-1)/\theta\right)]$. Hence, using (24),

$$\frac{E\left[Z_1|\text{survive}\right]}{E\left[Z_2|\text{survive}\right]} = \left(\frac{\widehat{T}_1}{\widehat{T}_2}\right)^{\frac{1}{\theta}} = w\phi^{1/\theta} = \frac{\frac{w_1}{P_1}}{\frac{w_2}{P_2}} = h^{\frac{(1-\beta)}{\beta}},$$

which is greater than 1. That variance is higher in the big city follows from the formula for variance. *Proof of (iii).* Note that the log of a Fréchet random variable becomes Gumbel, for which the distribution function is given by $e^{-\widehat{T}_i e^{-\theta y}}$, and the variance is given by $\pi^2/(6\theta^2)$, a constant independent of $\widehat{T}_i$. The mean of log productivity distribution in i, of course, increases when $\widehat{T}_i$ increases. ■

Proposition 3 uses the equilibrium conditions for where entrepreneurs and workers locate to show that the productivity distribution is strictly higher in the large city. The shape of the distribution is the same in the two cities and the variance of log productivity is identical across the two cities. In the introduction, we discussed Sverson's idea of looking for a compressed productivity distribution in large cities as evidence of the role of selection in productivity. Proposition 3 shows the idea does not work in this environment. In levels, the variance of productivity is actually higher in the large city. With a fat-tailed distribution, there is no sense that productivities get compressed against some upper bound on the right side of the distribution. With head-to-head competition, there is no common productivity threshold on the left-side of the distribution. Thus the productivity benefits of the large city manifest themselves in a smooth way.

There are two forces determining productivity in this model: knowledge spillovers and selection. Recall the productivity of a firm is $A_i y$. The knowledge spillover term $A_i = H_i^{\zeta}$ will be strictly larger in the large city if $\zeta > 0$. The variable y is the random productivity draw component and its average value at a location will depend upon how tough selection is at the location.

Recall that there is a measure M_i of entrepreneurs who start plants location i, each getting one draw of y from $G(\cdot)$. Of these, the measure that survive equals π_{ii}. As stated in BEJK Result 1, π_{ii} is the fraction of goods that location i sells to itself. (If a plant is not the lowest cost producer at its home location, it won't be the lowest cost producer at the other location either.) Hence, the survival rate for startups at location i equals

$$survival_i = \frac{\pi_{ii}}{M_i}.$$

Our result is

Proposition 4 *(Survival rates) At an agglomeration equilibrium, the survival rate is lower* *in the large city than in the small city if and only if*

$$\zeta < \frac{1-\beta}{\beta}.$$

Proof. The ratio of survival rates equals

$$\begin{aligned} \frac{survival_1}{survival_2} &= \frac{\pi_{11}}{\pi_{22}}\frac{M_2}{M_1} \\ &= \frac{T_1}{T_2}w^{-\theta}\phi^{-1}h^{-1} \\ &= \frac{h^{\theta\zeta+1}h^{-1}}{w^{\theta}\phi} = \frac{h^{\theta\zeta}}{h^{\frac{(1-\beta)\theta}{\beta}}}. \end{aligned}$$

The second line uses the formulas for π_{ii} and that $M_1/M_2 = h$. The third line first substitutes in $T_i = (1+\theta)^{\theta\zeta} m_i^{\theta\zeta+1}$ and then substitutes in for $w^{\theta}\phi$ using (24). The claim follows directly. ■

When the spillover parameter ζ is positive but not too big, there are two forces contributing to why plants tend to have high productivity in large cities. Not only is A_i higher, but also the random component y tends to be higher, since there is a lower chance of survival per unit draw of y. When ζ is large, the survival rate is actually higher in the large city. It is intuitive that when the labor market spillover is significant enough, a given low random draw of y in the large city will be able to beat a given high random draw of y in the small city. We note there are a variety of other models where survival rates are lower in large cities, including Asplund and Nocke (2006), Nocke (2006), Campbell (2010).

5.2 Shutting Down Selection: Does it Look Different and Does it Matter?

We highlight the role of selection in our model by contrasting a version of model in which spillovers are shut down with a *pure agglomeration model* in which selection is completely shut down. While the economics of the pure agglomeration model is different, the two models look the same in terms of observed productivity distributions.

The key assumption of the pure agglomeration model is that there is a monopoly entrepreneur for each product $j \in [0,1]$ rather than free entry and head-to-head competition. Each of the unit measure of entrepreneurs chooses where to locate. After the location decision is fixed, each entrepreneur draws a random productivity term y from the Fréchet with scale parameter T^b and shape parameter θ^b. (We use the superscript "b" to denote parameters of the pure agglomeration model.) There are a unit measure of workers that also choose whe647(w)15(h)11rhtwlch T647(w)1215(h)11(p)15(ro)10(s)87(d)121uohoh0(n(h)14(h)11n)

Note that $w_i = v_i$ no longer holds for the pure agglomeration model. However, observe that, using (17), we must have

$$U_1^n \equiv P_1^{\beta} R_1^{-(1-\beta)} w_1 = P_2^{\beta} R_2^{-(1-\beta)} w_2 \equiv U_2^n.$$

Similar equation $U_1^m = U_2^m$ for entrepreneurs holds with w_i replaced by entrepreneurs' expected payoff v_i. These imply that $w_1/w_2 = v_1/v_2$. As in other Dixit-Stiglitz type models, a fraction $1/\sigma^b$ of revenue goes to firms/entrepreneurs and $(\sigma^b - 1)/\sigma^b$ of the revenue goes to variable costs. This implies that

$$\begin{aligned} w_i &= \frac{\sigma^b - 1}{\sigma^b} \frac{X_i}{n_i}, \\ v_i &= \frac{1}{\sigma^b} \frac{X_i}{m_i}. \end{aligned}$$

The last two equations with the fact that $w_1/w_2 = v_1/v_2$ implies that $m_1/m_2 = n_1/n_2$, and hence $m_1/m_2 = h_1/h_2 = h$, a result that also holds true in the selection model. On the other hand, using (16) with $\zeta^s = 0$, we see that $T_i = m_i$ in the selection model. Hence, with a redefinition of $\theta = \sigma^b - 1$, the price indices of the two models are the same, given the same h. Hence, at the same h in both models, $P_1^{-\theta}/P_2^{-\theta} = \Phi_1/\Phi_2 \equiv \phi$ will be also the same. As we have shown earlier that the equilibrium conditions can be reduced to two equations linking h and ϕ, i.e., (20) and (21), it is easy to follow the same procedure to verify that these two conditions also determine equilibria for the pure agglomeration model. Therefore, an equilibrium h^* in the selection model must be also an equilibrium in the pure agglomeration model.

Proof of (ii). In the pure agglomeration model, the cumulative distribution function of productivity in location i is given by

$$\Pr[A_i y \leq z] = e^{-T^b A_i^{\theta} z^{-\theta}} = e^{-2^{\theta\zeta^b} T^b m_i^{\theta\zeta^b} z^{-\theta}}.$$

Hence, the productivity distribution at location i is Fréchet with shape parameter θ and scaling parameter

$$T_i^b = 2^{\theta\zeta^b} T^b m_i^{\theta\zeta^b}.$$

Using (24) and that $\zeta^b = \frac{1-\beta}{\beta}$, it is immediate to see that

$$\frac{\widehat{T}_1^s}{\widehat{T}_2^s} = \frac{T_1^b}{T_2^b} = h^{*\frac{(1-\beta)\theta}{\beta}}.$$

Thus, the ratio of means of two locations is the same in both models. With a proper choice of T^b, the productivity distribution at each location is the same in both models. ■

In the pure agglomeration model all productivity gains in the large city are due to knowledge spillovers. In the selection model (our model with $\zeta^s = 0$), all productivity gains in the large city are due to increased selection. Yet through a suitable choice of the ζ parameter for the pure agglomeration model, the data generated by the models regarding agglomeration and productivity distributions are exactly the same. In this respect, the models are observationally equivalent.

While these models look similar in the data they generate, the underlying economics are different and optimal policy is different. We illustrate this by considering the welfare impacts of a *zoning policy* that permits production only at location 1, i.e., suppose $h_1 = 1$ and $h_2 = 0$ are mandated by policy. (See Rossi-Hansberg (2004), for example.) The differing impacts of zoning in the two models are put in sharp contrast by an analysis of the limiting case where there is no transportation cost.

Proposition 6 *Assume $\tau = 1$. The equilibrium outcome is the same in both models, equal dispersion across the two locations, $h_1 = h_2 = \frac{1}{2}$. However, the welfare effect of the zoning policy differs across the two models. The zoning policy strictly decreases aggregate utility in the selection model for any value of the model parameters. In the pure agglomeration model, zoning increases aggregate utility if and only if*

$$\zeta^b > \frac{1-\beta}{\beta}.$$

Proof. That the zoning policy reduces aggregate utility in the selection model is immediate. In the pure agglomeration model, if h_1 and h_2 are the population fractions then aggregate utility equals

$$U^b = \left(h_1^{1+\zeta^b}\right)^\beta L^{1-\beta} + \left(h_2^{1+\zeta^b}\right)^\beta L^{1-\beta}$$

To see this, observe that if h_i locate at i, then goods production there equals $h_i^{1+\zeta^b}$, taking account of the knowledge spillover. Straightforward calculations show that if $\zeta^b > \frac{1-\beta}{\beta}$, this is maximized at $h_1 = 1$ and $h_2 = 0$, while if $\zeta^b < \frac{1-\beta}{\beta}$, this is maximized at $h_1 = h_2 = 0.5$. ■

The economics of this result is clear. In the selection model, when transportation costs are zero, it is not socially useful to concentrate production at location 1 because it creates congestion (the land at 2 is "wasted") with no benefit. In the pure agglomeration model, the cost of the congestion from the policy is offset by the gain of the knowledge spillovers and the policy can raise welfare. Agents in the pure agglomeration model do not internalize the externality of the spillover.

Let us now bring all the ideas of this subsection together. Imagine we have a data generating process that is either the pure agglomeration model or the selection model, we don't know which. Suppose we observe transportation cost and it is initially positive, $\tau > 1$, and we have an agglomeration at location 1. We observe the productivity distribution in both cities and see that it is higher at location 1, but, as we know from Proposition 5, access to the micro data doesn't help distinguish between the two models. Suppose transportation costs fall to zero, i.e., $\tau_{new} = 1$. Both models have the same prediction for equilibrium comparative statics: agglomeration disappears and we move to an equal split of population. However, the policy implications are very different in the two models. If selection is the true source of the productivity gains in the large city, then a zoning policy at the new transportation cost strictly decreases welfare. In contrast, if the source is knowledge spillovers, the policy is welfare neutral. (From Proposition 5, $\zeta^b = \frac{1-\beta}{\beta}$, and this value is the borderline case for welfare in Proposition 6.)

We can now follow up the discussion in the introduction of applying the idea in CDGPR to distinguish selection from spillovers. The idea is to look for truncation on the left as evidence of selection and any smooth overall rightward shift as evidence of spillovers. In their empirical analysis, CDGPR find little evidence of any kind of increased truncation in large cities. Rather they find productivity distributions shift to the right in a relatively smooth way. They conclude that spillovers must be what drives productivity gains. However, in our model, the pattern they document in the data is equally consistent with all productivity gains being due to selection and none being due to spillovers. As shown in Proposition 5, "Selection with No Spillover" and "Spillover with No Selection" generate the same kind of productivity data, shifting the distribution to the right in a smooth way in large cities.

Proposition 5 is related to a recent paper, Arkolakis, Costinot, and Rodriguez-Clare (2010), that argues that adding firm heterogeneity in either a Melitz or BEJK fashion does not add anything new in terms of aggregate impacts beyond what is already in the Dixit-Stiglitz symmetric firm model. They show these different models of international trade are equivalent at the aggregate level. The equivalence result between Dixit-Stiglitz and BEJK appears here as well. As noted above, the pure agglomeration model is essentially Helpman (1998), which is Dixit-Stigliz. The pure agglomeration model looks the same (with the same comparative statics from a change in transportation cost τ) as the selection model, which is BEJK in its underlying moving parts. While the connections noted by Arkolakis, Costinot, and Rodriguez-Clare (2010) show up here as well, the bottom line point that we get in our regional model with mobile labor and entrepreneurship is quite different from what they get in their trade model with fixed factors. The welfare effects of policies that impact the movement of factors of production do depend upon the underlying model. Whether it is the pure agglomeration or the selection model matters.

Finally, while the pure agglomeration and selection models are the same in their implications for the productivity distributions of producing plants, they may be different in terms of other variables that are potentially observable. Firm survival rates and how these vary across locations is an example.

6 The Impact of Distortion

Hsieh and Klenow (2009) and Rustucia and Rogerson (2008) are recent papers that connect policy distortions to aggregate productivity. In particular, suppose that successful firms are penalized through high taxes and unsuccessful firms are rewarded through subsidies. These policies will cause production to be misallocated across firms. As discussed in the introduction, Hsieh and Klenow (2009) propose using data on productivity distributions to infer the extent of such distortions. Adopting a Melitz-type modeling environment, they argue that these kinds of policies increase the dispersion of productivity, in addition to dragging down average productivity. The purpose of this section is to analyze the implications of such policies for productivity distributions in our model, where note again that the distinguishing features of our model are its head-to-head competition and fat-tails for underlying productivity draws.

The tax and subsidy scheme we consider takes a particular parametric form. A firm pays a tax or subsidy that depends upon its productivity. Specifically, a firm with productivity z makes a transfer to the government of

$$t(z) = \left(\frac{z}{\bar{z}}\right)^{\delta} - 1 \tag{25}$$

per unit of labor employed by the firm, for parameters $\delta \in (0,1)$ and $\bar{z} > 0$. The slope with respect to z is positive,

$$t'(z) = \delta z^{-(1-\delta)} \bar{z}^{-\delta} > 0.$$

The payment is denominated in units of labor. The parameter $\bar{z}$ is the cutoff separating who pays taxes and who receives subsidies. A firm with $z > \bar{z}$, pays a tax ($t(z)$ positive) that is higher the higher z. A firm with $z < \bar{z}$ receives a subsidy ($t(z)$ negative) that is higher (more negative) the lower is z. We assume the cutoff $\bar{z}$ is such that tax revenues are sufficient to offset the subsidies. Any excess of tax revenues over subsidy payments is destroyed, with free disposal. Finally, note that in the limiting case where $\delta = 0$, $t(z) = 0$ for all z and there is no distortion. So the parameter δ governs the extent of distortion.

While the functional form of the tax/subsidy scheme (25) is special, we think it is an interesting one to consider because (1) it is clear example of a policy that punishes success and rewards failure and (2) it yields a tractable formulation in which we are able to obtain a stark result.

To explain why the specification is tractable, we first note that a firm with productivity z incurs a cost per unit of output in labor units of

$$\frac{1}{z}[1 + t(z)] = \frac{1}{z^{(1-\delta)}\bar{z}^{\delta}},$$

including the transfer to the government. Hence, from the firm's perspective, it behaves as though it has a productivity of

$$\tilde{z} = \bar{z}^{\delta} z^{1-\delta}.$$

We will refer to $\tilde{z}$ as *perceived* productivity.

For simplicity, this section focuses on the symmetric equilibrium where the locations have equal population shares. Let $m = m_1 = m_2$ be the share becoming entrepreneurs at each location; let $n = n_1 = n_2$ be the share becoming workers at each location. These shares sum to one, $2m + 2n = 1$. It is convenient to start the analysis by taking m andqual ion-11.95J-1thz

Thus, if we define

$$\begin{aligned}\tilde{\theta} &\equiv \theta/(1-\delta), \\ \widetilde{T} &\equiv T\bar{z}^{\delta\tilde{\theta}},\end{aligned}$$

we see that the distribution of the highest perceived productivity at a given location is also Fréchet, with shape parameter $\tilde{\theta}$ and scale parameter $\tilde{T}$. It should now be clear why the choice of the tax/subsidy scheme (25) is tractable. It retains the Fréchet structure for perceived productivities. Hence, the analysis of equilibrium maps into BEJK like before, only with transformed parameters are $\tilde{\theta}$ and $\tilde{T}$. Note that since $1-\delta < 1$, $\tilde{\theta} > \theta$. This means that the perceived productivities are less dispersed than the actual productivities. The tax/subsidy scheme flattens things out. We can see that this is going to make selection play a smaller role in equilibrium.

Again, for a firm to survive, a firm not only has to be the best at its own location—it needs to beat the best from the other location. Using (22), the distribution of perceived productivity of survivors is Fréchet, with scale parameter given by

$$\widehat{\widetilde{T}} = \widetilde{T}\left(1+\tau^{-\tilde{\theta}}\right) = T\bar{z}^{\delta\tilde{\theta}}\left(1+\tau^{-\tilde{\theta}}\right) \tag{26}$$

and shape parameter $\tilde{\theta}$. Here we make use of the fact that with symmetry the ratio of wages between locations equals one, so this term drops out in (26).

With the distribution of perceived productivities of survivors in hand, we can now convert perceived to actual to determine the distribution of actual productivities. This is

$$\begin{aligned}\Pr[Z_{1i} &< z|\text{survival}] = \Pr[\left(\frac{\widetilde{Z}_{1i}}{\bar{z}^{\delta}}\right)^{\frac{1}{1-\delta}} < z|s] \\ &= e^{-\widehat{\widetilde{T}}\bar{z}^{-\delta\tilde{\theta}}z^{-(1-\delta)\tilde{\theta}}}\end{aligned}$$

Thus, using (26) the distribution is Fréchet with shape parameter $b\tilde{\theta} = \theta$ and scale parameter

$$\widehat{T} = \widehat{\widetilde{T}}\bar{z}^{-\delta\tilde{\theta}} = T\left(1+\tau^{-\theta/(1-\delta)}\right). \tag{27}$$

The crucial result here is that varying the tax/subsidy scheme through changes in the parameter δ has no impact on the shape parameter θ governing the distribution of surviving firms. It does impact the scale parameter. It is immediate in (27) that increasing distortions through increasing δ reduces the scale parameter T_s for surviving firms. Note that this effect is more pronounced the bigger the transportation cost τ. In fact, in the limiting case $\tau = 1$ with no transportation costs the effect of b drops out. With no transportation cost only the most efficient firm of a given good across the two locations survives. The tax/subsidy scheme flattens things out but it is still better to draw higher productivity as the advantage of higher productivity more than offsets the higher tax. The distortions kick in when transportation costs matter. In a free market, an inefficient firm at a particular location might be knocked out by the best firm at the other location. But with the help of a subsidy added to its transportation cost advantage, it might be able to survive in its home location.

The analysis so far has taken as fixed the population share m entering entrepreneurship. We now allow this to be endogenous. Using our earlier results (equation (15)), the share of the work force that enters entrepreneurship is $1/(1+\tilde{\theta})$, noting that the relevant shape parameter here is for the perceived distribution of productivity. As each location gets half of the entrepreneurs, we have

$$m = \frac{1}{2}\frac{1}{1+\tilde{\theta}} = \frac{1}{2}\frac{1}{1+\frac{\theta}{1-\delta}}.$$

We see that the greater the distortion through a higher level of δ, the lower the entrepreneurship rate. The distortion weakens the role of selection, diminishing the importance of entrepreneurship, making it a less attractive option in equilibrium. We conclude that the distortion decreases average productivity for two reasons. First, fixing the level of entrepreneurship, it induces inefficient firms that would otherwise exit to "hang on" and produce for their local market. Second, it reduces entrepreneurship.

Now consider economies that are identical except that they vary in the parameter δ of the tax/subsidy scheme, with $\delta = 0$ corresponding to the free market and higher δ implying more distortion. The analysis of this section has shown that varying δ makes no difference in the resulting shape parameter θ for surviving firms' productivities. Hence, the variance of log productivity is constant across δ, using the argument in Proposition 3. Increasing distortions by increasing δ does lower average productivity. In summary, we have shown.

Proposition 7 *In the symmetric equilibrium given tax/scheme parameter b, the distribution of productivity for surviving firms is Fréchet with scale parameter*

$$\widehat{T} = \frac{1}{2}\frac{1}{1+\frac{\theta}{1-\delta}}\left(1+\tau^{-\theta/(1-\delta)}\right),$$

which strictly decreases as δ increases (and distortions are increased). The shape parameter does not vary with δ, so the variance of log productivity does not depend upon δ.

We conclude that a strategy of using the dispersion of productivity to infer the extent of distortions does not work in this model.

7 Conclusion

This paper shows how the BEJK model of international trade can be embedded into a model of agglomeration and regional trade. The paper puts the model to work analyzing productivity distributions. In a nutshell, the paper shows how two economic factors shift around mean plant productivities, while leaving intact the shapes of the productivity distributions. In particular, the equilibrium conditions of entrepreneurial and labor mobility imply that mean productivity is higher in agglomerated areas. And, tax/subsidy policies that punish success and reward failure lower average productivity. Yet, neither factor changes the variance of log productivity. The results suggest strategies in the literature the rely on shapes of productivity distributions for identification do not work in this modeling structure.

For ease of exposition, we have focused on a simple, two location model. It is straightforward to generalize the model to allow for many heterogeneous locations with a rich transportation cost structure. Such a model can potentially be put to work in quantitative analysis, analogous to the way Redding and Sturm (2008) put the Helpman (1998) model to work on accounting for the population distribution in Germany. We expect that data on plant exit rates and how these vary across locations are likely to be particularly useful here. Directly measuring the selection taken place through plant exit can potentially shed much light on the importance of selection in determining productivity.

References

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2010), "New trade models, same old gains? ," NBER Working Paper No. w15628.

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1992), *A First Course in Order Statistics,* New York: John Wiley & Sons.

Asplund, M. and V. Nocke, "Firm turnover in imperfectly competitive markets," *Review of Economic Studies* (2006) 73, 295–327.

Baldwin, R. E. and T. Okubo (2006), "Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting," *Journal of Economic Geography*, 6(3): 323-346.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2010), 11(l)69E71 GsorRR-Nic:k:Srner92(k)11(i)6(n)1

[illegible]

Helpman, E. (1998), “The size of regions,” in D.Pines, E.Sadka and Y.Zilcha (eds.), *Topics in Public Economics*, pp. 33-54, New York: Cambridge University Press.

Hsieh, C.-T. and P. J. Klenow (2009), “Misallocation and aanufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124(4): 1403-1448.

Kortum, S. (1997), “Research, patenting, and technological change,” *Econometrica,* 65(6): 1389-1419.

Krugman, P. (1991), “Increasing returns and economic geography,” *Journal of Political Economy,* 99(3): 483-499.

Melitz, Marc J. (2003), “The impact of trade on intra-industry reallocations and aggregate industry productivity,” *Econometrica,* 71(6): 1695-1725.

Melitz, Marc J. and G. Ottaviano (2008), “Market size, trade, and productivity,” *Review of Economic Studies,* 75(1): 295-316.

Nocke, V. (2006), “A gap for me: Entrepreneurs and entry,” *Journal of the European Economic Association*, Sept 2006, 4(5):929-959.

Ottaviano, G., T. Tabuchi and J.-F. Thisse (2002), “Agglomeration and trade revisited,” *International Economic Review,* 43(2): 409-435.

Redding, S. and D. Sturm (2008), "The costs of remoteness: Evidence from German division and reunification" *American Economic Review*, 98(5): 1766-1797.

Restuccia, D. and R. Rogerson, “Policy distortions and aggregate productivity with heterogeneous plants,” *Review of Economic Dynamics*, 11 (2008), 707–720.

Rossi-Hansberg, E. (2004), “Optimal urban land use and zoning,” *Review of Economic Dynamics*, 7, 69-106, 707–720.

Syverson, C. (2004), “Market structure and productivity: A concrete example,” *Journal of Political Economy,* 112(6): 1181-1222.

Appendix

Proof of Part (ii) of Lemma 1.

For each good j, the number of draws from $G(.)$ is the number of entrepreneurs that enter j, $S_i(j)$. To simplify notations, we now leave out j and S_i (S_i') is understood as $S_i(j)$ ($S_i(j')$).We want to show that if (a) $S_1 < S_1'$, (b) $S_2 > S_2'$, and (c) $v_2(S_1, S_2) = v_2(S_1', S_2')$ all hold and transportation cost is strictly positive ($\tau > 1$), then $v_1(S_1, S_2) > v_1(S_1', S_2')$.

When $G(.)$ is Fréchet, i.e., (1), then the distribution function of the best productivity is exactly $e^{-\alpha S_i}$, and the joint distribution of the top two highest productivity is exactly that given in (11) with $T_i = \alpha S_i$. Without loss of generality, let $\alpha = 1$. Take the wage of the effective labor units $\widetilde{w}_i = w_i/A_i = w_i/H_i^{\zeta}$ as the input price in location i (see equation [**??**]), then the pricing behavior and the cost and price distributions for each good j will be exactly the same as in BEJK, now with $T_i(j) = S_i(j)$.

We first need to derive the expression of the expected profit $v_1(S_1, S_2)$ and $v_2(S_1, S_2)$. First note that $K_{1ni}(c) \equiv \Pr[C_{1ni} \le c] = \Pr[Z_{1i} \ge \frac{\widetilde{w}_i \tau_{ni}}{c}] = 1 - e^{-S_i(\widetilde{w}_i \tau_{ni})^{-\theta} c^{\theta}}$. Then, note that given $C_{1ni} = c$, the probability that the i's firm with the best efficiency for any particular good will open a business in n is $\prod_{k \ne i}[1 - K_{1nk}(c)]$. Let the event that $C_{1ni} = c$ and $C_{1nk} \ge c$, $\forall k \ne i$ be denoted as $B_i(c)$. It can be verified that $\Pr[C_{2ni} \ge p | C_{1ni} = c] = e^{-S_i(\widetilde{w}_i \tau_{ni})^{-\theta}(p^{\theta} - c^{\theta})}$ is a result of (12). Then,

$$\begin{aligned}
K_{2n|B_i(c)}(p) &\equiv \Pr[C_{2n} \le p | B_i(c)] \\
&= 1 - \Pr[C_{2n} > p | B_i(c)] \\
&= 1 - \frac{\Pr[C_{2ni} \ge p | C_{1ni} = c] \prod_{k \ne i}[1 - K_{1nk}(p)]}{\prod_{k \ne i}[1 - K_{1nk}(c)]} \\
&= 1 - \frac{e^{-S_i(\widetilde{w}_i \tau_{ni})^{-\theta}(p^{\theta} - c^{\theta})} \prod_{k \ne i} e^{-[\sum_{k \ne i} S_i(\tau_{nk} \widetilde{w}_k)^{-\theta}] p^{\theta}}}{\prod_{k \ne i} e^{-[\sum_{k \ne i} S_i(\tau_{nk} \widetilde{w}_k)^{-\theta}] c^{\theta}}} \\
&= 1 - e^{-\Phi_n(p^{\theta} - c^{\theta})},
\end{aligned}$$

where $\Phi_n = \sum_{i=1}^{2} S_i(\widetilde{w}_i \tau_{ni})^{-\theta}$. Now, observe that every entrepreneur has $1/S_i$ chance of being the best producer in location i, and hence

$$\begin{aligned}
v_i &= \frac{1}{S_i}\int_0^\infty \sum_{n=1}^{2}\prod_{k\neq i}[1-K_{1nk}(c)]\int_c^{\mu c}(p-c)\frac{X_n}{P_n^{-\sigma}}p^{-\sigma}dK_{2n|B_i(c)}(p)dK_{1ni}(c)+ \\
&\quad \frac{1}{S_i}\int_0^\infty \sum_{n=1}^{2}\prod_{k\neq i}[1-K_{1nk}(c)]\int_{\mu c}^\infty(\mu c-c)\frac{X_n}{P_n^{-\sigma}}(\mu c)^{-\sigma}dK_{2n|B_i(c)}(p)dK_{1ni}(c), \\
&= \theta^2\sum_{n=1}^{2}\frac{X_n}{P_n^{-\sigma}}(\tilde{w}_i\tau_{ni})^{-\theta}\Phi_n\int_0^\infty\int_c^{\mu c}c^{\theta-1}(p-c)p^{\theta-\sigma-1}e^{-\Phi_n p^\theta}dpdc+ \\
&\quad \theta(\mu^{1-\sigma}-\mu^{-\sigma})\sum_{n=1}^{2}\frac{X_n}{P_n^{-\sigma}}(\tilde{w}_i\tau_{ni})^{-\theta}\int_0^\infty c^{\theta-\sigma}e^{-\Phi_n(\mu c)^\theta}dc, \\
&\equiv \theta\sum_{n=1}^{2}\frac{X_n}{P_n^{-\sigma}}(\tilde{w}_i\tau_{ni})^{-\theta}\left[\theta\Phi_n V(\Phi_n)+(\mu^{1-\sigma}-\mu^{-\sigma})D(\Phi_n)\right], \\
&\equiv \theta\sum_{n=1}^{2}\frac{X_n}{P_n^{-\sigma}}(\tilde{w}_i\tau_{ni})^{-\theta}W(\Phi_n).
\end{aligned}$$

It is straightforward to verify that $dD(\Phi_n)/d\Phi_n<0$. Observe that

$$\begin{aligned}
\frac{d}{d\Phi_n}\Phi_n V(\Phi_n) &= \frac{d}{d\Phi_n}\Phi_n\int_0^\infty\left[\int_{p/\mu}^{p}c^{\theta-1}(p-c)dc\right]p^{\theta-\sigma-1}e^{-\Phi_n p^\theta}dp, \\
&= \left(\frac{\mu^{-\theta-1}}{1+\theta}-\frac{\mu^{-\theta}}{\theta}\right)\frac{d}{d\Phi_n}\left(\Phi_n\int_0^\infty p^{2\theta-\sigma}e^{-\Phi_n p^\theta}dp\right), \\
&= \left(\frac{\mu^{-\theta-1}}{1+\theta}-\frac{\mu^{-\theta}}{\theta}\right)\frac{\sigma-\theta-1}{\theta^2}\Phi_n^{\frac{\sigma-1-2\theta}{\theta}}\Gamma\left(\frac{1+2\theta-\sigma}{\theta}\right).
\end{aligned}$$

Under the parameter restriction in BEJK that $\sigma<1+\theta$, the above expression is negative. Hence, $dW(\Phi_n)/d\Phi_n<0$.

Suppose first that $\Phi_1\geq\Phi_1'$, or, equivalently,

$$S_1\widetilde{w}_1^{-\theta}+S_2\left(\widetilde{w}_2\tau\right)^{-\theta}\geq S_1'\widetilde{w}_1^{-\theta}+S_2'\left(\widetilde{w}_2\tau\right)^{-\theta}.$$

Then,

$$\left(S_2-S_2'\right)\left(\widetilde{w}_2\tau\right)^{-\theta}\geq\left(S_1'-S_1\right)\widetilde{w}_1^{-\theta}$$

By assumption $S_1<S_1'$ and $S_2>S_2'$ and $\tau>1$,

$$\left(S_2-S_2'\right)\widetilde{w}_2^{-\theta}>\left(S_1'-S_1\right)\left(\widetilde{w}_1\tau\right)^{-\theta},$$

or,

$$\Phi_2 > \Phi_2'.$$

But, $\Phi_1 \geq \Phi_1', \Phi_2 > \Phi_2'$ and the formulas above for profit contradict that a firm located at 2 is indifferent between (S_1, S_2) and (S_1', S_2'). Hence, $\Phi_1 < \Phi_1'$. To keep the firms in location 2 indifferent between j and j', it must be that $\Phi_2 > \Phi_2'$.

From $v_2(S_1, S_2) = v_2(S_1', S_2')$, we see that

$$\frac{X_1}{P_1^{-\sigma}} \tilde{w}_2^{-\theta} \tau^{-\theta} \left[W(\Phi_1) - W(\Phi_1')\right] = \frac{X_2}{P_2^{-\sigma}} \tilde{w}_2^{-\theta} \left[W(\Phi_2') - W(\Phi_2)\right],$$

or,

$$\frac{X_1}{P_1^{-\sigma}} \tilde{w}_1^{-\theta} \tau^{-\theta} \left[W(\Phi_1) - W(\Phi_1')\right] = \frac{X_2}{P_2^{-\sigma}} \tilde{w}_1^{-\theta} \left[W(\Phi_2') - W(\Phi_2)\right].$$

Since $\Phi_1 < \Phi_1'$, $\Phi_2 > \Phi_2'$, and $dW(\Phi_n)/d\Phi_n < 0$, both sides of the above equality is positive. Because $\tau > 1$, we have the following inequality.

$$\frac{X_1}{P_1^{-\sigma}} \tilde{w}_1^{-\theta} \left[W(\Phi_1) - W(\Phi_1')\right] > \frac{X_2}{P_2^{-\sigma}} \tilde{w}_1^{-\theta} \tau^{-\theta} \left[W(\Phi_2') - W(\Phi_2)\right],$$

or,

$$\frac{X_1}{P_1^{-\sigma}} \tilde{w}_1^{-\theta} W(\Phi_1) + \frac{X_2}{P_2^{-\sigma}} \tilde{w}_1^{-\theta} \tau^{-\theta} W(\Phi_2) > \frac{X_1}{P_1^{-\sigma}} \tilde{w}_1^{-\theta} W(\Phi_1') + \frac{X_2}{P_2^{-\sigma}} \tilde{w}_1^{-\theta} \tau^{-\theta} W(\Phi_2').$$

which is $v_1(S_1, S_2) > v_1(S_1', S_2')$.

Proof for Lemma 2

Proof. For part (i), observe that from (6) and (7), we have

$$\lim_{M \to \infty} \Pr\left[M^{-1/\theta} (\ell(M))^{-1} Z_1 < z\right] = e^{-z^{-\theta}}.$$

Also note that if $\ell(.)$ is slowly varying, then $1/\ell(.)$ is also slowly varying. Thus, we write, with a little abuse of notation of ℓ,

$$\lim_{\bar{H} \to \infty} \Pr\left[(m_i \bar{H})^{-1/\theta} \ell(m_i \bar{H}) Z_1 < z\right] = e^{-z^{-\theta}}. \tag{28}$$

Using (28), the distribution function of Z_{1i} is

$$\begin{aligned}\Pr[Z_{1i} &< z] = \Pr[\alpha_{\bar{H}} A_i \widehat{Z}_{1i} < z] \\ &= \Pr[(m_i \bar{H})^{-1/\theta} \ell(m_i \bar{H}) \widehat{Z}_{1i} < h_i^{-\zeta} m_i^{-1/\theta} \bar{H}^{-\zeta-1/\theta} \ell(m_i \bar{H}) \frac{z}{\alpha_{\bar{H}}}] \\ &\to \Pr[(m_i \bar{H})^{-1/\theta} \ell(m_i \bar{H}) \widehat{Z}_{1i} < h_i^{-\zeta} m_i^{-1/\theta} z], \qquad \bar{H} \to \infty \\ &\to e^{-h_i^{\theta\zeta} m_i z^{-\theta}}, \qquad \bar{H} \to \infty\end{aligned}$$

where the third line follows from the fact that $\lim_{\bar{H}\to\infty} \ell(m_i \bar{H})/\ell(\bar{H}) = 1$.Thus, $T_i = h_i^{\theta\zeta} m_i$. The joint distribution result (11) holds under the same condition for which (9) holds. See Theorem 8.4.2 in Arnold et al. (1992).

For part (ii), simply let $\alpha_{\bar{H}} = \alpha^{-1/\theta} \bar{H}^{-\zeta-1/\theta}$, and observe that

$$F_i(z) = \Pr[\alpha_{\bar{H}} A_i \widehat{Z}_{1i} < z] = \Pr[\widehat{Z}_{1i} < \frac{z}{\alpha_{\bar{H}} A_i}] = \left[e^{-\alpha\left(\frac{z}{\alpha_{\bar{H}} A_i}\right)^{-\theta}} \right]^{m_i \bar{H}} = e^{-h_i^{\theta\zeta} m_i z^{-\theta}}.$$

We actually chose a constant for $\ell(.)$ in this case.[5] This same $\alpha_{\bar{H}}$ can be used for the Pareto distribution, but the result only holds aymptotically, as stated in part (i). There are distributions in the domain of attraction for Fréchet that where $\ell(.)$ can not be a constant, e.g., Loggamma distribution; see Chapter 3.3 in Embrechts et al. (1997).

The Fréchet density is given by $f_i(z) = \theta\alpha z^{-\theta-1} e^{-\alpha z^{-\theta}}$.One can derive (11) from the joint density of first and second order statistics which is given by $f_i(z_1, z_2) = M_i(M_i - 1) f_i(z_1) f_i(z_2)$. ■

Proof for Proposition 2

To prove this proposition, it is most convenient to express the utility ratio $u(h_1)$ in ϕ, where $\phi \equiv \Phi_1/\Phi_2$. To this end, we first derive the equilibrium condition in terms of ϕ, given h_1, and study the properties of ϕ.

Using BEJK Results 1 and 2, expenditure by location 1 on goods from location 2 equals $X_{12} = X_1 T_2 (w_2 \tau)^{-\theta} / \Phi_1$. Using the analogous expession for X_{21} and the market clearing condition $X_{12} = X_{21}$ implies

$$\frac{X_1}{X_2} = \frac{T_1 w_1^{-\theta} \Phi_1}{T_2 w_2^{-\theta} \Phi_2} = \frac{m_1^{\theta\zeta+1}}{m_2^{\theta\zeta+1}} w^{-\theta} \phi,$$

[5] Alternatively, we can choose $\ell(.)$ according to the method described above, and it can be verified that such an $\ell(.)$ is *tail-equivalent* to a constant. A useful result is that if two distributions are tail-equivalent, then the same normalizing constant can be used. For tail-equivalence, see Chapter 3.3 in Embrechts et al. (1997).

where we substitute in equation (16) $T_i = (1+\theta)^{\theta\zeta} m_i^{\theta\zeta+1}$ for the productivity distribution scaling parameter and we let $w \equiv w_1/w_2$ be the wage ratio. The equilibrium job choice condition (15) implies $n_1/n_2 = m_1/m_2 = h_1/(1-h_1)$. Also, $wh_1/(1-h_1) = X_1/X_2$ (equation (14)). Plugging these in gives

$$w^{1+\theta} = \left(\frac{h_1}{1-h_1}\right)^{\theta\zeta} \phi. \tag{29}$$

Next, we use the definition (13) of Φ_i to obtain

$$\phi = \frac{\Phi_1}{\Phi_2} = \frac{T_1 w_1^{-\theta} + T_2 w_2^{-\theta} \tau^{-\theta}}{T_1 w_1^{-\theta} \tau^{-\theta} + \theta}$$

1210 1

Define $a(\phi)$ as

$$a(\phi) = -\tau^{\theta}\theta\left(\phi + \frac{1}{\phi}\right) + \tau^{2\theta}(1+2\theta) - 1.$$

The sign of $a(\phi)$ determine the sign of the derivative of the r.h.s. with respect to ϕ. Note that $a(\phi)$ achieves minimum at $\phi = \tau^{\theta}$ or $\tau^{-\theta}$. Then,

$$a(\phi) \geq a\left(\tau^{\theta}\right) = a\left(\tau^{-\theta}\right) = \left(\tau^{\theta} - 1\right)\left(1 + \tau^{\theta}\right)(1+\theta) > 0.$$

(iii) The values of ϕ at $h_1 = 0, 0.5, 1$ are straightforward by inspecting (30) and noting that $\phi = \Phi_1/\Phi_2$ must be finite and positive even when $h_1 = 0$ or $h_1 = 1$.
(iv) In equation (30), the l.h.s. explodes to ∞ as h_1 converges to 1. This implies that ϕ has to converge to τ^{θ}. The limit of ϕ when h_1 approaches 0 is similarly obtained. ■

Given Lemma 3, it is convenient to have

$$\frac{d\phi}{dh_1} = \frac{\frac{\theta\zeta+\theta+1}{1+\theta}\frac{\phi}{h_1(1-h_1)}}{\frac{1-\tau^{-2\theta}}{1-\tau^{-\theta}\phi}\frac{\phi}{\phi-\tau^{-\theta}} + \frac{\theta}{1+\theta}}. \tag{31}$$

To obtain the expression for the utility ratio u, observe that

$$u = \frac{U_1}{U_2} = \frac{\left(\frac{w_1}{P_1}\right)^{\beta}(N_1+M_1)^{-(1-\beta)}}{\left(\frac{w_2}{P_2}\right)^{\beta}(N_2+M_2)^{-(1-\beta)}} = w^{\beta}\phi^{\frac{\beta}{\theta}}\left(\frac{1-h_1}{h_1}\right)^{1-\beta}.$$

Using (29) to substitute in for w, we arrive at

$$u = \left(\frac{1-h_1}{h_1}\right)^{\frac{1+(1-\beta)\theta-\beta(\theta\zeta+1)}{1+\theta}}\phi^{\frac{\beta(1+2\theta)}{\theta(1+\theta)}}. \tag{32}$$

Denote the numerator of the exponent of the term $(1-h_1)/h_1$ in (32) as a composite parameter

$$\xi \equiv 1 + (1-\beta)\theta - \beta(\theta\zeta+1).$$

It can be verified that, from (31) and (32),

$$\frac{du}{dh_1} = \left(\frac{1-h_1}{h_1}\right)^{\frac{\xi}{1+\theta}}\frac{\phi^{\frac{\beta(1+2\theta)}{\theta(1+\theta)}}}{1+\theta}\frac{1}{h_1(1-h_1)}\left[\Psi(\phi) - \xi\right], \tag{33}$$

where

$$\Psi(\phi) = \frac{\beta(1+2\theta)[\theta\zeta+\theta+1]}{\theta(1+\theta)}\frac{1}{\frac{1-\tau^{-2\theta}}{1-\tau^{-\theta}\phi}\frac{\phi}{\phi-\tau^{-\theta}} + \frac{\theta}{1+\theta}}.$$

Since $\phi > 1$ for $h_1 > 0.5$,

$$\frac{d}{d\phi}\left(\frac{1-\tau^{-2\theta}}{1-\tau^{-\theta}\phi}\frac{\phi}{\phi-\tau^{-\theta}}\right) = \frac{\tau^{-\theta}(\phi^2-1)(1-\tau^{-2\theta})}{(1-\tau^{-\theta}\phi)^2(\phi-\tau^{-\theta})^2} > 0.$$

From Lemma 3, ϕ is strictly increasing over h_1 and ranges from 1 to τ^θ on $h_1 \in [0.5, 1]$. For $\phi \in [1, \tau^\theta]$, $\Psi'(\phi) < 0$, $\Psi(\phi) > 0$, $\Psi(\tau^\theta) = 0$, and

$$\eta \equiv \Psi(1) = \frac{\beta(1+2\theta)(1-\tau^{-\theta})[\theta\zeta+\theta+1]}{\theta(1+2\theta+\tau^{-\theta})} > 0.$$

From (33), the sign of du/dh_1 depends on the sign of $\Psi(\phi) - \xi$. Since $\Psi(\phi)$ decreases from η to 0 on $h_1 \in [0.5, 1]$, the two bounds of $\Psi(\phi)$, i.e., η and 0, provide two dividing lines which partition the parameter space into three: (a) $\xi \geq \eta$, (b) $\eta > \xi > 0$, and (c) $\xi \leq 0$. The two dividing lines have straightforward interpretations as follows.

That $\xi \geq \eta$ or not is equivalent to, by (33) and $\Psi(\phi) < \eta$, whether the utility ratio u on $h_1 \in [0.5, 1]$ is strictly decreasing or not. If yes, then symmetric equilibrium is stable and the only equilibrium. That $\xi \leq 0$ or not, by (33) and $\Psi(\phi) > 0$, is equivalent to whether u is strictly increasing or not. If yes, then a black-hole equilibrium exists and is the only agglomeration equilibrium. In fact, by inspecting (32), we know that if $\xi < 0$, $u \to \infty$ when $h_1 \to 1$, and if $\xi = 0$, u at $h_1 = 1$ is a finite number greater than 1.

In the in-between case, $\eta > \xi > 0$, there exists an $\hat{h}_1 \in (0.5, 1)$ such that $du/dh_1 > 0$ for $h_1 \in [0.5, \hat{h}_1)$, and $du/dh_1 < 0$ for $h_1 \in (\hat{h}_1, 1)$. In this case, u increases from 1 at 0.5 to a peak at $\hat{h}_1$ and then decrease after $\hat{h}_1$. In fact, $u \to 0$ as $h_1 \to 1$, since $\xi > 0$. Hence, there exists a unique agglomeration equilibrium $h_1^* \in (0.5, 1)$, and $h_1^* \in (\hat{h}_1, 1)$.

It is straightforward to verify that the three subspaces (1) $\xi \geq \eta$, (2) $\eta > \xi > 0$, and (3) $\xi \leq 0$ correspond to (1) $\beta \leq \hat{\beta}_L$, (2) $\beta \in (\hat{\beta}_L, \hat{\beta}_H)$, and (3) $\beta \geq \hat{\beta}_H$, respectively. At the interior agglomeration equilibrium h_1^*, $du/dh_1 < 0$ and is therefore stable. Any black-hole equilibrium is also stable. This proves part (i).

For part (ii): $du/dh_1 < 0$ at an interior agglomeration equilibrium h_1^*. Let κ denote a parameter generically. If $du/d\kappa > 0$ at h_1^*, then h_1^* increases. It suffices to show that $d\ln(u)/d\kappa > 0$ at h_1^*. From (32),

$$\ln(u) = \frac{\xi}{1+\theta}\ln\left(\frac{1-h_1}{h_1}\right) + \frac{\beta(1+2\theta)}{\theta(1+\theta)}\ln(\phi). \quad (34)$$

Differentiating (34) with respect to κ, we get

$$\frac{d\ln(u)}{d\kappa} = \frac{d}{d\kappa}\left(\frac{\xi}{1+\theta}\right)\ln\left(\frac{1-h_1}{h_1}\right) + \frac{d}{d\kappa}\left(\frac{\beta(1+2\theta)}{\theta(1+\theta)}\right)\ln(\phi) + \frac{\beta(1+2\theta)}{\theta(1+\theta)}\frac{d\ln(\phi)}{d\kappa}, \quad (35)$$

where

$$\begin{aligned}\frac{d\ln(\phi)}{d\kappa} =& \frac{1}{E}\{\theta\left[\frac{\theta\zeta+\theta+1}{1+\theta}\frac{d}{d\kappa}\left(\frac{1+\theta}{\theta}\right) - \frac{d}{d\kappa}\left(\frac{\theta\zeta+\theta+1}{\theta}\right)\right]\ln\left(\frac{1-h_1}{h_1}\right) \\ &+ \frac{\theta^2}{1+\theta}\frac{d}{d\kappa}\left(\frac{1+\theta}{\theta}\right)\ln(\phi) - (1+\theta)\left(\frac{1-\tau^{-\theta}\phi}{\phi-\tau^{-\theta}}\right)\frac{\partial}{\partial\kappa}\left(\frac{\phi-\tau^{-\theta}}{1-\tau^{-\theta}\phi}\right)\}, \end{aligned} \quad (36)$$

where

$$E = (1+\theta)\left(\frac{\phi}{\phi-\tau^{-\theta}}\frac{1-\tau^{-2\theta}}{1-\tau^{-\theta}\phi}\right) + \theta.$$

Obviously, $E > 0$. Evaluate both side of (34) at h_1^*, and we have

$$\ln\left(\frac{1-h_1^*}{h_1^*}\right) = -\frac{\beta(1+2\theta)}{\theta\xi}\ln(\phi^*). \quad (37)$$

Plugging (36) into (35) to simplify and using (37), we obtain

$$\frac{d\ln(u)}{d\kappa}|_{h_1=h_1^*} = \Sigma^*\ln(\phi^*) - \frac{\beta(1+2\theta)}{\theta E^*}\frac{1-\tau^{-\theta}\phi^*}{\phi^*-\tau^{-\theta}}\frac{\partial}{\partial\kappa}\left(\frac{\phi-\tau^{-\theta}}{1-\tau^{-\theta}\phi}\right)_{h_1=h_1^*}, \quad (38)$$

where

$$\begin{aligned}\Sigma^* =& \frac{d}{d\kappa}\left(\frac{\beta(1+2\theta)}{\theta(1+\theta)}\right) - \frac{\beta(1+2\theta)}{\theta\xi}\frac{d}{d\kappa}\left(\frac{\xi}{1+\theta}\right) \\ &+ \frac{\beta^2(1+2\theta)^2}{\theta^2\xi E^*}\frac{\theta}{1+\theta}\left[\frac{d}{d\kappa}\left(\frac{\theta\zeta+\theta+1}{\theta}\right) - \frac{\theta\zeta+\theta+1}{1+\theta}\frac{d}{d\kappa}\left(\frac{1+\theta}{\theta}\right)\right] \\ &+ \frac{\beta(1+2\theta)}{\theta E^*}\left(\frac{\theta}{1+\theta}\right)^2\frac{d}{d\kappa}\left(\frac{1+\theta}{\theta}\right).\end{aligned}$$

Note again that $\ln(\phi^*) > 0$ since $\phi^* > 1$ as $h_1^* > 0.5$. If $\kappa = \tau$, then $\Sigma^* = 0$, and

$$\frac{d\ln(u)}{d\kappa}|_{h_1=h_1^*} = -\frac{\beta(1+2\theta)}{\theta E^*}\frac{1-\tau^{-\theta}\phi^*}{\phi^*-\tau^{-\theta}}\frac{\theta\tau^{-\theta-1}(1-\phi^*)}{(1-\tau^{-\theta}\phi^*)^2} > 0.$$

If $\kappa = \zeta$ or β, the second term on the r.h.s. of (38) is zero, and it can be verified that $\Sigma^* > 0$. In sum, for $\kappa = \zeta, \tau$, or β, $\frac{d\ln(u)}{d\kappa}|_{h_1=h_1^*} > 0$, and hence h_1^* increases when either of ζ, τ, β increases.