A Spatial Knowledge Economy

Donald R. Davis
Columbia University and NBER
Preliminary. Not for circulation.

June 28, 2009

Abstract

Recent empirical research on cities has identified heterogeneity in size, in wage levels and inequality, and in the distribution of firm productivities as key targets for any model of a system of cities. Consistency with the system of cities approach should also allow for mobility of both firms and workers. To date, theoretical research has made progress on some of these problems, but often only with highly restrictive assumptions in other dimensions. The present paper presents a simple and unified approach to account for key facts considered.

1 Space, Knowledge, Hierarchies

The study of systems of cities is complex. The empirical literature has identified multiple dimensions of heterogeneity that theory should account for. First, of course, is to explain heterogeneity in the size of the cities themselves. Second is to understand why these cities differ in the type of workers they attract, the wages those workers earn, and the degree of wage inequality. Third is to understand why large and small cities differ in productivity, both in the aggregate and in the distribution across firms. Finally, to be consistent with the defining characteristic of the systems of cities literature, both firms and labor should be mobile across the cities.

The central fact about cities is agglomeration – size differences are dramatic. This is most frequently discussed in the context of Zipf's Law of Cities or the so-called rank-size rule, as discussed in Gabaix (1999). I don't address the specifics of Zipf per se. But in what are here termed "learning equilibria" symmetric equilibria will never be stable. That is, in spite of symmetric fundamentals, asymmetric city sizes will be generic.

Spatial patterns of wages also provide important insights to our understanding of other aspects of the spatial structure of the economy. Glaeser, Resseger, and Tobio (2008) recently documented a strong increasing relation between wage inequality, measured by the Gini coefficient, and measures of urbanization such as population density and size. They note that there remains a strong relation between the Ginis calculated directly and when they control for skill composition, leading them to point to variation in skill premia as an important source of inequality. They note, as well, that these premia are high in cities with a large share of people with a college degree.

A large literature documents aggregate productivity differences between large and small cities. [surveyed in Melo, Graham, and Noland (2009) and Rosenthal and Strange (2004)]. When Rosenthal and Strange (2004) asked what we could learn from these studies about microfoundations, their answer at the time was "Not much." In international trade, there is now considerable evidence that aggregate productivity gains from liberalization arise due to selection effects that redistribute demand from low to high productivity firms (e.g. Tybout 2003). Similarly, Syverson (2004) identifies selection effects strongly within the concrete industry. This still leaves open the question of the within-country spatial effect of selection, a question recently addressed by Combes, et al. (2008). Comparing firms across large and small cities in France, they find no evidence of left truncation of productivity, which they interpret as the absence of selection effects spatially. Instead what they find is that the distributions of firm productivity look remarkably alike in large and small cities at the lower tail and that the variation in average productivity emerges because of a "dilation" of productivity in which only the best firms are able to take advantage of agglomeration economies.

The theoretical literature has addressed heterogeneity in cities, workers and firms. Often, though, it addresses them one at a time. When efforts have been made to incorporate more than one dimension of heterogeneity, this has been done by specifying other aspects of the equilibrium exogenously. Thus, for example, we have papers that take the size of cities as exogenous and investigate the consequences for equilibrium firm types, allowing neither firms nor workers to move (e.g. Combes, et al. 2009). We have papers that allow for rural-urban migration of labor and endogenous heterogeneity of wages, but don't allow labor or firms to move between cities (e.g. Behrens and Robert-Nicoud 2008). While these papers provide important insights, these restrictions imply that they don't really provide a complete system of cities account.

The aim of the present paper, then, is to make sense of three types of heterogeneity – cities, workers, and firms – within a simple and unified framework. At the broadest level, we can divide the economy into landlords and labor. Landlords own a fixed supply of housing in equal supply across cities, and so function as a centrifugal force. Labor is heterogeneous, but to maintain simplicity the heterogeneity is uni-dimensional. A Roy Model segments labor into one sector (Services) of homogeneous productivity and a second (Manufacturing) in which labor heterogeneity has bite. The Roy character insures that the bottom of the labor productivity distribution self-selects into the homogeneous productivity Services sector. Since (at least) some of these services are non-traded, yet essential, the bottom of the income distribution is the same in real terms in all cities. The remaining labor finds employment in Manufacturing, a sector developed within the foundation of Antras, Garicano, Rossi-Hansberg (2006). Within manufacturing, labor divides endogenously with the top tier becoming managers and those below them becoming workers. As in AGRH, matching is one-to-many and assortative, with manager productivity the same as firm total factor productivity.

The force driving agglomeration is investment by managers in learning. As in AGRH, managers solve problems too hard for their workers. However, they now face a new choice. Each problem encountered can be solved directly, as in AGRH, or part of the manager's time can be allocated to an indirect method, in which the manager learns from other managers in the same city. Unlike the traditional modeling of Marshallian learning externalities, the productivity gain is not manna from heaven. It is an economic decision because learning comes at the sacrifice of time solving problems directly. As suggested by Lucas (1988), the average quality of others who choose to learn affects one's own productivity (as does the aggregate stock of knowledge). The specification makes use of the insight of Costinot (2009) on the key role of complementarities, applying it here to spatial sorting. In short, all managers find they have better learning opportunities when they are in the city with the high average level of learning. However the benefits are greatest for managers who are themselves of high ability. High land prices endogenously lead manager-worker teams to sort by type.

Spatial asymmetry is a generic feature of learning equilibria, so cities will have different sizes and the large city will have high housing prices. The large city will also be the one with the greater degree of inequality – the same real wages for the lowest paid workers in the service sector, but higher real (and nominal) wages that accrue to the top tier manufacturing teams locating there, as in Glaeser, et al. With manager productivity in manufacturing serving as firm TFP, we can also show that productivity will exhibit features akin to the dilation of Combes, et al. (2009).

The theory developed here does not rely on the central elements of the New Economic Geography (NEG). No increasing returns. No imperfect competition. No transport costs for tradable goods. It strips these out not because they are unimportant, but because they are inessential to the approach developed. It could prove interesting to introduce, at a later date, some of these elements of the NEG to understand the interactions with the model developed here.

2 An Isolated City

2.1 Consumers and the Aggregate Economy

We consider first the problem of a single isolated city. All consumers have identical Cobb-Douglas preferences over the homogeneous products Housing H, Manufactures M, and Services S. Manufactures serve as numeraire, with the relative price of housing given by P, the relative price of services given by W, and income given by I. These assumptions allow us to use a representative consumer for demand decisions. The representative consumer solves the following problem:

$$U = D_H^{\alpha} D_M^{\beta} D_S^{1-\alpha-\beta} \qquad s.t. \qquad PD_H + D_M + WD_S = I \tag{1}$$

This problem yields demand functions of the form:

$$D_H = \frac{\alpha I}{P} \text{ and } D_M = \beta I \text{ and } D_S = (1-\alpha-\beta)\frac{I}{W} \tag{2}$$

For any consumer j with income $I(j)$, the maximum attainable utility is given by an indirect utility function (with k a Cobb-Douglas constant):

$$V(j) = kP^{-\alpha}W^{-(1-\alpha-\beta)}I(j) \tag{3}$$

We can express the relative demands as

$$\frac{D_S}{D_M} = \frac{(1-\alpha-\beta)}{\beta W} \quad \text{and} \tag{4}$$
$$\frac{D_H}{D_M} = \frac{\alpha}{\beta P} \tag{5}$$

We can return to market clearing once we have looked more closely at the production side of the model.

2.2 Housing and the Allocation of Labor

The city has a fixed amount of housing given by $\overline{H}$. The housing is owned by resident landlords who supply it in a competitive market and collect rents but perform no other productive function.

The remaining goods draw on a common pool of labor. We start with a fundamental division between production processes of these goods. One is a good that can be produced at a uniform level of productivity by anyone employed in that sector and the other features individual level heterogeneity in productivity in that good. For concreteness, we will refer to these, respectively, as services S and manufactures M.

The aggregate labor force is of measure 1. We assume that labor quality is distributed uniformly, according to types $z \sim U[0,1]$. Since factor markets are competitive, the services sector, which features homogeneous labor productivity, will draw workers only at the bottom of the productivity distribution. Let the boundary worker in the service sector be z_S. Then all workers in the interval $z \sim [0, z_S]$ work in services and total service output is:

$$S = z_S \tag{6}$$

Production of manufactures is based on Garicano (2000) and Antras, Garicano, Rossi-Hansberg (AGRH 2006). Manufacturing is undertaken by teams of managers and workers, where the selection into each occupation is endogenous. Output depends on three elements. The first is the span of control of the manager, i.e. the number of workers a manager can supervise. The second is the productivity of the manager, in which native ability differences are amplified by endogenous learning choices. The third is the matching of managers and workers into teams. We consider these in turn.

The residual labor available to the manufacturing sector is distributed uniformly on the interval $z \in [z_S, 1]$. As we will confirm below, the lower tail of labor types, $z \in [z_S, z_m]$, become workers and the upper tail, $z \in [z_m, 1]$, become managers. Workers spend their time solving problems that arise in production. A worker of quality z_p encounters a unit of problems, solves a share z_p of these problems, and passes up the remaining $(1 - z_p)$ problems to a manager. The manager faces an aggregate time constraint normalized to unity. Each problem passed up to the manager requires h units of time to address, whether or not the problem is solved. Thus the manager's time constraint, with n workers, each of quality z_p, is such that $n(1 - z_p)h = 1$. Equivalently, the manager's span of control is given by:

$$n(z_p) = \frac{1}{h(1 - z_p)} \tag{7}$$

The span of control thus depends only on the specifics of worker quality z_p, not manager quality.

2.3 A Microfoundation for Knowledge Acquisition

Managers make two choices. One concerns the type of workers to hire, given the equilibrium wage schedule for worker types, considered in the next section. The other concerns knowledge acquisition, which may be considered separately, and to which we now turn.

Managers solve problems too hard for their workers. They do this by drawing on their native problem solving ability and also by spending time acquiring knowledge that is useful in production. A manager z is endowed with a unit of time. As discussed above, h units of time are absorbed in addressing any problem. However, here we allow the manager to divide these h units of time per question so that share β_z is devoted directly to producing and the remaining $(1 - \beta_z)$ is devoted to knowledge acquisition. Knowledge acquisition is costly because it comes at the sacrifice of direct effort in production, but it is beneficial because it raises manager z's productivity.

Managers acquire knowledge useful in production by interacting with other managers who also invest in knowledge acquisition. The efficiency of these efforts depend on a few key variables. The first is the aggregate stock of ideas, $A > 1$, which we take as exogenous. The second is the endogenously determined average quality z_a of those encountered while seeking knowledge. And finally, the efficiency of knowledge acquisition also depends on the quality z of the manager seeking knowledge.

We specify a simple form for the productivity $\widetilde{z}$ of manager z:

$$\widetilde{z} = \beta_z z \exp\left(\left(1 - \beta_z\right) A z_a z\right) \tag{8}$$

Choosing β_z to maximize productivity, the share of time devoted to knowledge acquisition is:

$$\left(1 - \beta_z\right) = \begin{cases} \left(1 - \frac{1}{A z_a z}\right) & \text{if} \quad A z_a z > 1 \\ 0 & \text{if} \quad A z_a z \leq 1 \end{cases} \tag{9}$$

The share of time devoted to knowledge acquisition for a manager z actively engaged in learning is higher when any of three variables is greater: the aggregate stock of ideas, A; the average quality z_a of those encountered while seeking knowledge; or the quality z of the knowledge seeker herself.

We now need to define the average quality of those seeking knowledge, z_a. We address a first case quickly. We will define $z_a = 0$ for the case in which no one invests in learning.[1] This will always be an equilibrium, since absent coordination activities, no individual will allocate time to learning from others when there are no others from whom to learn. While the no-learning equilibrium will not be the focus of our discussion, it does play an important role in the story. It underscores the fact that learning here is not manna from heaven but instead relies on a costly allocation of effort by all of those who seek to learn.[2] We address instead a learning equilibrium that focuses on the forces of agglomeration of interest.

There are a variety of cases that we can consider. We will focus on a case in which the aggregate stock of knowledge, A, is sufficiently high that all managers find it optimal to invest at least some time in learning. Managers who seek knowledge encounter a unit measure of other knowledge seekers drawn randomly from those seeking knowledge. Let z_m be the endogenously determined lowest quality manager. The average quality of these encounters, z_a, thus depends both on who is seeking knowledge and the intensity with which they are seeking it. Let $T_L = \int_{z_m}^{1} \left(1 - \beta_z\right) dz$ be the total effort devoted to knowledge acquisition. Then $\phi\left(z\right) \equiv \frac{\left(1-\beta_z\right)}{T_L}$ is a density with support on $[z_m, 1]$ that reflects the distribution of time devoted to knowledge acquistion. The average quality of the encounter is thus given by

[1] This is no longer an average, of course, but seems an appropriate definition that reflects the absence of opportunities to learn from others.

[2] If we allowed any interval of the most able to coordinate on moving to an equilibrium with learning, a move that would be in their collective interest, then we would also expect the no-learning equilibrium to be unstable.

$z_a = \int_{z_m}^{1} \phi(z) z dz$.

Making explicit the fact that the distribution of learning effort depends on z_a, the equilibrium average quality encountered in the market is the fixed point that solves:

$$z_a^* = \int_{z_m}^{1} \phi(z; z_a^*) z dz = \frac{1}{T_L} \int_{z_m}^{1} (1 - \beta_z^*) z dz \tag{10}$$

This value exists, is unique in the relevant interval, strictly increasing in z_m, and strictly interior to any positive interval of managers who learn. The existence of such a z_a^* is straightforward. The right hand side of the equation is a continuous function that is a weighted average of the values in the interval, hence lies strictly in the interior. Holding fixed the interval, the right hand side is also a decreasing function of z_a. To see this, note that relative weights are $\frac{\phi(z)}{\phi(z')} = \frac{(1-\beta_z)}{(1-\beta_{z'})} = \frac{z'}{z}\left(\frac{Az_a z-1}{Az_a z'-1}\right) \equiv g(z_a)$. We can also look at $\frac{dg(z_a)}{dz_a} = -\frac{z' A(z-z')}{z(Az_a z'-1)^2} < 0$ if $z > z'$. That is, although a rise in z_a leads all managers to increase learning, the increase raises the relative weight placed on the lower index z's. Hence as z_a ranges from z_m to 1 , it will equal the right hand side exactly once. This shows than any fixed interval has a unique z_a consistent with equilibrium.

It remains to demonstrate how the equilibrium value of z_a varies as the limits of the interval vary. This is also straightforward. As before, begin by considering an arbitrarily chosen initial interval $[z_m, z_M]$ and let $z_{a_{mM}}$ be the average value associated with optimal learning choices by managers in that interval. Now break this interval into two parts, given by $[z_m, z_{a_{mM}}]$ and $[z_{a_{mM}}, z_M]$. Let the respective average values be z_{a-} and z_{a+}, where $z_{a-} < z_{a_{mM}} < z_{a+}$ by the fact that each is in the interior of the interval. In order to understand how the equilibrium z_a changes as the interval changes, we can consider two experiments. The first is to consider a move from an interval $[z_m, z_{a_{mM}}]$ to $[z_m, z_M]$, i.e. a rise in the upper boundary of the interval. We have already seen that this leads to $z_{a-} < z_{a_{mM}}$, i.e. a rise in the upper boundary of the interval raises the equilibrium z_a. Similarly a move from $[z_{a_{mM}}, z_M]$ to $[z_m, z_M]$ shows the effect of reducing the lower bound of the interval, which lowers the equilibrium z_a from z_{a+} to $z_{a_{mM}}$. And since these results came from an arbitrarily chosen interval, they will hold generally.

Taking these facts together, there always exists a unique interior solution for the equilibrium average learning that is consistent with optimal firm choices in a learning equilibrium. This average value is interior to the range of learning managers and is strictly increasing in the upper bound of the range and strictly decreasing in the lower bound of the range. The share of time spent on learning is then $(1 - \beta_z) = \left(1 - \frac{1}{Az_a^* z}\right)$ and our initial assumption that all managers invest in knowledge acquisition will indeed be correct if $A > \frac{1}{z_a^* z_m}$.

We have a few key conclusions. Managers optimally choose to engage other managers in encounters from which they both learn. This learning takes time away from direct production, but maximizes total output of the managers by raising their productivity. Time devoted by any manager to learning is increasing in the aggregate stock of ideas, the average quality of other learners, and the quality of the manager herself.

With a microfoundation for knowledge acquisition in hand, we proceed in the following sections to introduce new choices for our managers, which includes a richer production model and later a spatial structure with cities. The aim is to explore how a microfounded model of learning married with worker heterogeneity affects spatial structure when location affects the bounds of learning interactions.[3]

2.4 Teams of Managers and Workers

We can now make explicit the full decision problem of managers. As in Antras, Garicano, and Rossi-Hansberg (2006), managers maximize their rents, here setting

$$\begin{array}{c} R\left(z_m\right) = \underset{z_p, \beta_z}{Max} \left(\widetilde{z}\left(\beta_z\right) - w\left(z_p\right)\right) n\left(z_p\right) \\ s.t. \qquad n\left(z_p\right) = \frac{1}{h(1-z_p)} \\ \widetilde{z} = \beta_z z \exp\left(\left(1-\beta_z\right) A z_a z\right) \end{array} \tag{11}$$

Because β_z affects only $\widetilde{z}$, the optimal choice $\widetilde{z}^*\left(\beta_z\right)$ can be determined separately, as per above. The first order condition with regard to z_p requires that:

$$w'\left(z_p\right) = \frac{\widetilde{z}\left(\beta_z\right) - w\left(z_p\right)}{\left(1 - z_p\right)} \tag{12}$$

This wage function insures that both workers and managers will be content with their match. We also require that $w\left(z_S\right) = W$, $w\left(z_m\right) = R\left(z_m\right)$, and $w'\left(z_m\right) < R'\left(z_m\right)$. The

[3] It is worth noting that the complementarities driving agglomeration would exist if manager productivity depended on a simple average of the quality of other managers in one's city, without need of a formal model of investment in learning. This simpler approach, however, would be vulnerable to the widespread critique of Marshallian learning as a "black box" rather than an economic choice (cf. e.g. Abdel-Rahman and Anas, 2004) that has kept learning at the margins of discussions in the New Economic Geography. The present approach avoids that pitfall, so hopefully will give impetus to further efforts focusing on the economics of idea exchange as a key force for agglomeration.

first insures that the worker who is marginal between the service and manufacturing sectors receive the same wage in each. The second provides a similar condition for the agent indifferent between being the most skilled worker and the least skilled manager. The last condition insures that all managers $z > z_m$ are content to be managers rather than workers, and as AGRH showed requires sufficiently high manager productivity (low h).

The productivity of workers and managers is complementary and so features positive assortative matching. Let z_m be the type of labor indifferent between being a manager and being a worker. Let $m(z)$ be the quality of a manager that matches with a worker of type z. Then $z_m = m(z_S)$, $1 = m(z_m)$ and $m'(z) > 0$. The exact matching uses these facts as well as the market clearing condition for each labor type, which requires that:

$$\int_{z_S}^{z_p} g(z)\, dz = \int_{m(z_S)}^{m(z_p)} n\left(m^{-1}(z)\right) g(z)\, dz \qquad \forall z_p \leq z_m \tag{13}$$

For a given z_S, the key to solving for the boundary z_m is that the arguments to the labor supply and demand conditions are pre-determined by the density of types and the span of control associated with worker types, which itself does not depend on the boundary.

2.5 Equilibrium

We can now turn to the determination of equilibrium. In this isolated city, the allocation decisions are straightforward. Services are determined directly by $S = z_S$. With z_S given, labor market clearing determines the boundary between workers and managers, z_m, and the matching between managers and workers. Managers choose learning optimally, which in combination with the foregoing determines manufacturing output as:

$$M = \int_{z_m}^{1} \widetilde{z} n\left(m^{-1}(z)\right) dz \tag{14}$$

The only element remaining is to make sure that the relative supplies of services and manufactures are consistent with goods market equilibrium and then so use this to solve for the housing market equilibrium. From our discussion of consumers above, $\frac{D_S}{D_M} = \frac{(1-\alpha-\beta)}{\beta W}$. Recall that W plays many roles – the price of services, the wage available in services, and the wage required by the least talented manufacturing worker. As W ranges over values from zero to infinity, relative demand for services declines monotonically from infinity to zero. On the supply side, and recalling that services always employs the lowest interval of workers, a rise in W raises the supply of workers in services (z_s rises), directly augmenting its supply. The consequences in manufacturing are more complicated. A rise in z_s removes some labor previously employed in manufacturing. Market clearing requires that z_m then rise as well. As we saw earlier, the rise in z_m will induce a rise in the equilibrium average quality of learners z_a in the pool, as each manager chooses to spend more time learning. Each surviving manager now matches with a better set of workers. Each continuing group of workers matches with a manager who has worse native abilities, but who now has better learning opportunities and so invests more in learning. It is sufficient for equilibrium that the net effect of moving the marginal labor from manufacturing to services raises the relative supply of services, which we take as the case of interest. In this case, W adjusts to equate relative demand and relative supply of services, thus also determining outputs of S and M. With M determined, so is income I. Since housing supply is exogenous, market clearing is given by $\overline{H} = D_H = \frac{\alpha I}{P}$, which thus determines the relative price of housing, the last element of the equilibrium.

3 A Two City Model

We now introduce a second city. Preferences of all agents in all locations are precisely as before. Housing is available in identical supply $\overline{H}$ in each location, but is immobile (as is the landlord-owner), hence the relative price of housing may differ across cities. Letting c be an index for a particular city, the representative consumer's problem yields:

$$D_{Hc} = \frac{\alpha I_c}{P_c} \text{ and } D_{Mc} = \beta I_c \text{ and } D_{Sc} = (1 - \alpha - \beta) \frac{I_c}{W_c} \tag{15}$$

In this section we will treat services as non-traded. Manufactured goods are in principle tradable but are not traded in this variant of the model because they are a homogeneous product and there is not yet any other tradable good (later we will introduce a tradable service). Thus while labor is wholly mobile across sectors and cities, goods market clearing will be city by city.

As before, efficiency given the competitive factor markets requires that services are undertaken by the least skilled labor in the economy. Let us denote the measure of labor employed in services as z_S, which is then also total services output, without yet designating where these services will be delivered. Since service workers will be present in both locations (owing to

Cobb-Douglas preferences) and are mobile across cities, they must have the same indirect utility in both cities. Recall that indirect utility is given by $V(j) = P_c^{-\alpha} W_c^{-(1-\alpha-\beta)} I_c(j)$ and that $I_c(j)$ for a service worker is W_c. Equating indirect utilities for cities 1 and 2, we find that:

$$\frac{W_2}{W_1} = \left(\frac{P_2}{P_1}\right)^{\frac{\alpha}{\alpha+\beta}} \tag{16}$$

With total employment in services z_s given for now, this leaves the labor interval $[z_S, 1]$ for employment in manufactures. With two cities, we now need to ask whether workers and managers need to co-locate. For now we will take the position that they do on the premise that proximate communication is essential in this knowledge-intensive, differential productivity sector.[4]

With z_S fixed, we showed in the single city framework how labor demand and supply in manufactures led to a division into managers and workers and the positive assortative matching of same. Leaving aside the question of the city in which teams locate, we need to know whether the same division and matching will occur. They do. Clearly there has to be assortative matching within each of the cities for the same reasons as in a single city. But with homothetic preferences, the incentives of managers and workers are aligned to move to a higher cost of living city only if the productivity advantages justify it. In that case, it gives the manager both the means and the incentive to make it worthwhile for the worker to move as well. Hence, with z_S given, we can take the division into managers and workers from the previous equation of demand and supply of workers and we again have assortative matching.

With the manufacturing teams fixed, we next investigate where they will locate. First, note that teams must choose a location to maximize joint income or an alternative location could make both workers and managers better off. Hence we can look at the indirect utility of a team z, which is given by $V_c(z) = P_c^{-\alpha} W_c^{-(1-\alpha-\beta)} I_c(z)$. Note that in equilibrium a team with manager z has $I_c(z) = \frac{1}{Az_{ac}} \exp(Az_{ac}z - 1)\, n(m^{-1}(z))$. Consider the relative attractiveness of cities 2 and 1 for team z:

[4] Later we will consider what happens if the manufactured good has attached to its production a tradable service (this last implying homogeneous productivity in this activity across workers). In a two city framework, we could also consider the consequences when workers and managers can separate at a cost, the magnitude of which is perhaps tied to the information intensity of the activity.

$$V_2(z)/V_1(z) = \frac{P_2^{-\alpha} W_2^{-(1-\alpha-\beta)} I_2(z)}{P_1^{-\alpha} W_1^{-(1-\alpha-\beta)} I_1(z)} = \frac{P_2^{-\alpha} W_2^{-(1-\alpha-\beta)}}{P_1^{-\alpha} W_1^{-(1-\alpha-\beta)}} \frac{z_{a1}}{z_{a2}} \exp(Az(z_{a2} - z_{a1})) \tag{17}$$

We can also ask how the relative attractiveness of cities 2 and 1 varies across teams with managers z and z'. This is given as:

$$\frac{V_2(z)/V_1(z)}{V_2(z')/V_1(z')} = \exp(A(z - z')(z_{a2} - z_{a1})) \tag{18}$$

Without loss of generality, let $z_{a2} > z_{a1}$ and $z > z'$. Then this says that there is a strict ordering in which high productivity teams have a stronger relative attraction to high average learning locations. Some manufacturing has to arise in both cities, otherwise the price of housing in the city sans manufacturing would go to zero and available utility there would go to infinity. Hence we need to find the cutoff determining which teams are in which location. This requires finding a team for which $V_2(z)/V_1(z) = 1$. Recalling that spatial equilibrium for service workers requires that $\frac{W_2}{W_1} = \left(\frac{P_2}{P_1}\right)^{\frac{\alpha}{\alpha+\beta}}$, we can substitute in to find the requirement for a boundary firm z_b that $\frac{z_{a1}}{z_{a2}} \exp(Az_b(z_{a2} - z_{a1})) = \left(\frac{P_2}{P_1}\right)^{\frac{\alpha}{\alpha+\beta}}$. We can also use the housing market equilibrium conditions, noting that there are equal supplies of housing in each city, which implies that housing prices are proportional to aggregate city income $\frac{P_2}{P_1} = \frac{I_2}{I_1}$ and with market clearing in manufactures also equal to relative output in manufactures, so $\frac{P_2}{P_1} = \frac{M_2}{M_1}$. Thus the condition for indifference of a team z_b is that:

$$\frac{z_{a1}}{z_{a2}} \exp(Az_b(z_{a2} - z_{a1})) = \left(\frac{M_2}{M_1}\right)^{\frac{\alpha}{\alpha+\beta}} \tag{19}$$

Note that the left hand side of this expression is a statement about the relative productivity of a team in two cities, while the right hand side is a statement about relative manufacturing output in the two cities. The right hand side goes from zero to infinity as we move teams across cities, while the left hand side is always positive and strictly bounded. This assures that we will find such a z_b.[5] With z_b determined, we can recover M_2 and M_1 separately, hence also I_2 and I_1, as well as all other prices, including W. We can then examine whether $W_1 = w(z_S)$, i.e. whether the wage paid in city 1 to a service worker equals the wage paid to the least skilled manufacturing worker z_S, who will also be located in city

[5] Further work is needed to identify sufficient conditions for uniqueness. For now, this will be assumed.

1. If the service wage is too low, we contract the posited z_S and vice versa if it is too high. Equality represents full closure of the model.

4 Discussion

4.1 Wage Patterns

We can now discuss the resulting patterns of nominal and real prices and wages across the different tasks and locations of labor. For illustration, we will term the large, high learning city as "New York," and the smaller, lower learning city as "Gary (Indiana)." New York will have a higher price index, reflecting higher prices both for housing and services. The high housing prices in New York relative to Gary are purely a function of the spatial equilibrium rather than any inherent difference between the two locations. Nominal wages for service workers are higher in New York, but these exactly compensate for the difference in the price indices; real wages are the same for service workers in both locations and are at the bottom of the over all wage scale. Nominal wages in New York are higher for both manufacturing workers and managers. For labor indifferent between the two locations, these represent pure compensation for the differences in price indices. For labor with a strict preference, this reflects both the higher quality of workers and managers in New York as well as the greater learning opportunities there. Figure 1 illustrates the differences in real wages across labor types. Implied nominal wages would adjust this figure by raising all segments associated with high price New York.

We note in passing that it would be straightforward to introduce a costlessly traded service, implying again an activity for which there is homogeneous productivity across labor types. As in the nontraded service, this would be undertaken by workers at the bottom end of the productivity distribution. The equilibrium would vary in just a few ways. First, since the service is traded, there is no reason to locate it in high price New York. Hence all workers in this sector would locate in low price Gary, receiving the Gary nominal wage for service workers. Equilibrium would need then also to feature a balance of trade, with New York exporting manufactures to Gary in exchange for the traded service. The resulting pattern of real wages is illustrated in Figure 2.

An important feature of the wage distribution is that it will be wider in New York. The reason is that the lowest productivity workers select into the activity with homogeneous activity. The real wage floor is the same in New York and Gary. But high wage managers and workers endogenously sort into New York, causing the greater range of observed real wages. This is in the same spirit as the evidence from Glaeser, Resseger, and Tobio (2008) of greater inequality in large and dense cities.

4.2 Dilation of Productivity

One of the empirical regularities emerging from the work of Combes, et al. (2009) is what they term the "dilation" of productivity. That is, the relative productivity advantage of large cities is negligible or non-existent at the bottom of the productivity distribution but becomes considerable as we move toward the most productive firms.

Addressing first the bottom of the distribution, it is indeed a puzzle why firms with a common low productivity would locate in large and expensive cities if the good is readily traded. However, if a good is difficult to deliver at a distance, then the firm can locate in the large and expensive city and recover its costs by marking up proportionally to the price index difference. This is precisely the role played here by the nontraded service, which has identical TFP, yet survives in both locations. These will also be measured as the least productive firms because they are hiring the bottom end of the labor productivity distribution (even though all labor has common productivity when engaged in this activity).

We need to be careful when interpreting total factor productivity in the manufacturing sector. Both workers and managers are heterogeneous in productivity. However the special structure of AGRH implies that it is only manager heterogeneity that matters for TFP. A first fact relevant for this is that due to the one-to-many matching structure of the firm, managers are of measure zero, so employment at the firm is equal to the measure of workers $n(z_p)$. With output given by $\widetilde{z}n(z_p)$ and workers $n(z_p)$ as the only positive measure input, firm TFP is directly given by the ratio $\frac{\widetilde{z}n(z_p)}{n(z_p)} = \widetilde{z}$.

We now turn to why the relative productivity gap between large and small cities grows as we move up the productivity distribution. There is not a unique way to interpret this observation, so we develop a couple of theoretical approaches to it. The difference in the two approaches is how to define the productivity gap within cities that is the basis for our comparison.

A first approach starts with the equilibrium productivity of a firm z compared to that of a firm $z + \triangle z$:

$\widetilde{z}(z) = \frac{1}{Az_a}\exp[Az_a z - 1]$

$\widetilde{z}\left(z+\triangle z\right)=\frac{1}{Az_a}\exp\left[Az_a\left(z+\triangle z\right)-1\right]$
Hence:
$\frac{\widetilde{z}(z+\triangle z)}{\widetilde{z}(z)}=\exp\left[Az_a\triangle z\right]$
We then consider this for any z' in another city:
$\frac{\widetilde{z}(z'+\triangle z)}{\widetilde{z}(z')}=\exp\left[Az'_a\triangle z\right]$,

Then the within city relative productivity gap, for any given $\triangle z$, is given by $\left(\frac{\widetilde{z}(z+\triangle z)}{\widetilde{z}(z)}\right)/\left(\frac{\widetilde{z}(z'+\triangle z)}{\widetilde{z}(z')}\right)$. We will say there is a "dilation" of productivity whenever this ratio exceeds unity for $\triangle z>0$ and $z_a>z'_a$. Calculating this directly, we find the ratio is given by:

$$\frac{\frac{\widetilde{z}(z+\triangle z)}{\widetilde{z}(z)}}{\frac{\widetilde{z}(z'+\triangle z)}{\widetilde{z}(z')}}=\exp\left[A\triangle z\left(z_a-z'_a\right)\right] \tag{20}$$

This exhibits dilation of productivity.

A second approach considers this for a factor λz, where $\lambda>1$. Then we would consider the equilibrium productivity of a firm z compared to that of a firm λz:
$\widetilde{z}\left(z\right)=\frac{1}{Az_a}\exp\left[Az_az-1\right]$
$\widetilde{z}\left(\lambda z\right)=\frac{1}{Az_a}\exp\left[Az_a\left(\lambda z\right)-1\right]$
Hence:
$\frac{\widetilde{z}(\lambda z)}{\widetilde{z}(z)}=\exp\left[Az_a\left(\lambda-1\right)\right]$
Consider this for any z' in another city:
$\frac{\widetilde{z}(\lambda z')}{\widetilde{z}(z')}=\exp\left[Az'_a\left(\lambda-1\right)\right]$,

Then the within city relative productivity gap, for any given $\triangle z$, is given by $\left(\frac{\widetilde{z}(\lambda z)}{\widetilde{z}(z)}\right)/\left(\frac{\widetilde{z}(\lambda z')}{\widetilde{z}(z')}\right)$. We will say there is a "dilation" of productivity whenever this ratio exceeds unity for $\triangle z>0$ and $z_a>z'_a$. Calculating this directly, we find the ratio is given by:

$$\left(\frac{\widetilde{z}\left(\lambda z\right)}{\widetilde{z}\left(z\right)}\right)/\left(\frac{\widetilde{z}\left(\lambda z'\right)}{\widetilde{z}\left(z'\right)}\right)=\exp\left[A\left(\lambda-1\right)\left(z_a-z'_a\right)\right] \tag{21}$$

This exhibits dilation of productivity.

In short, we have shown conditions under which large and small cities will have common TFP for firms at the bottom of the productivity distribution and a widening gap as we move up the distribution, as suggested by the observation of Combes, et al (2009) of a productivity "dilation."

5 Conclusions

This paper has developed a spatial knowledge economy. Complementarity between own type and learning opportunities makes high average learning cities attractive. All else equal, all types prefer to be in the high average learning city. But this puts pressure on local prices, rationing access. Since the higher quality types benefit relatively more from such access, they are the ones willing to pay for location, sustaining equilibrium. Low quality labor sorts into activities in which productivity is homogeneous, with a key role played when their products are non-traded.

This simple model allows us to replicate important features of data both for wage inequality and the distribution of firm productivities across large and small cities. Large cities have more inequality than small cities. Here this emerges because the bottom of the real wage distribution is common to all cities, pinned down by the wage in the homogeneous productivity non-traded services. But because of spatial sorting, the large city picks up the top of the wage distribution within both workers and managers. A parallel effect is at work with respect to firms, where we replicate a "dilation" of productivity. The bottom of the productivity distribution is the same in large and small cities, reflecting the presence of firms with common productivity specialized in non-traded services. The complementarity of own type and average learning opportunities that led to spatial concentration in the first place also leads to a relatively stronger stretching of productivities at the top end in larger cities.

6 References

To come.