EC1410 Topic #8

Agglomeration economies and why are there cities anyway?

Matthew A. Turner
Brown University
Spring 2025

(Updated December 30, 2024)

Outline

1. Why are there cities?
2. Returns to scale
3. Returns to scale and the planner's problem
4. Returns to scale and equilibrium
5. Measuring agglomerations
6. Measuring agglomeration economies, theory
7. Measuring agglomeration economies, empirical results
8. Mechanisms
9. (Some of) what we know about mechanisms
10. Conclusion

Why are there cities? I

The monocentric city model assumes that, for some reason, people earn a high enough wage in the CBD that many of them want to pay rent and commute in order to be near it. We have so far not considered why this might be.

Why do people want to crowd together badly enough to put up with the higher rents and commuting?

- An easy, if not very helpful answer: Because people like to be near each other.
- This raises three questions:
  - How much do they like to be near each other?
  - Why do people like to be near each other?
  - How does this desire for density affect how cities are organized?

Why are there cities? II

We have a pretty good answer for the first question. We'll review what is known about the second. We'll come back to the third point when we talk about systems of cities.

Returns to scale in production I

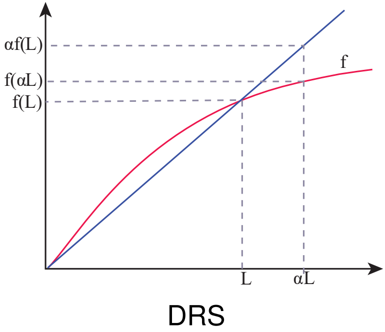
One of the main things people do in cities is work. Casual empiricism suggests that crowding makes people more productive. Why else would we have factories?

To consider this explanation more carefully, we need to talk about returns to scale in production.

Let production in a city be Y, labor/population be N and the production technology f, so that $Y = f(N)$. Say that f is

- Increasing Returns to Scale (IRS) if $\alpha f(N) < f(\alpha N)$ for any $\alpha > 1$.
- Constant Returns to Scale (CRS) if $\alpha f(N) = f(\alpha N)$ for any $\alpha > 1$.
- Decreasing Returns to Scale (DRS) if $\alpha f(N) > f(\alpha N)$ for any $\alpha > 1$.

In words, production is increasing returns if doubling inputs more than doubles outputs; decreasing returns if doubling inputs less than doubles output, and constant returns if doubling inputs exactly doubles output.



Returns to scale and location choice

Suppose we have two cities, A and B. In each city, labor is converted into output as above. We would like to divide 1 unit of population between the two cities to maximize aggregate output.

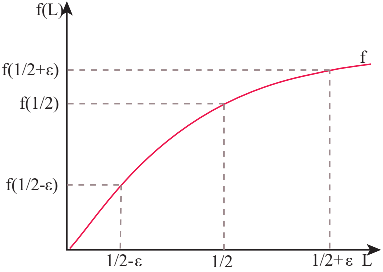
Note that this analysis is a 'planner's problem,' i.e., what could happen if a single agent were in charge of everyone. This need not be a spatial equilibrium. That is, a configuration where no one wants to move.

Because $N_A + N_B = 1$, without loss of generality, we can write

$$N_A = \frac{1}{2} - \varepsilon$$
$$N_B = \frac{1}{2} + \varepsilon$$

for $-\frac{1}{2} \leq \varepsilon \leq \frac{1}{2}$.

Suppose that both cities have a DRS technology,



If $\varepsilon > 0$ then by decreasing ε slightly, we shift from population from the larger to the smaller city.

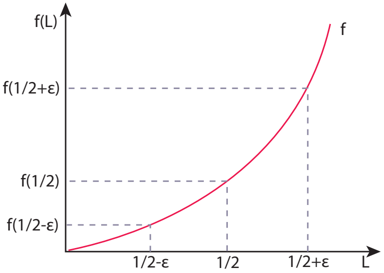
The marginal product of labor is lower, i.e. f' is flatter, in the larger than the smaller city.

It follows that decreasing ε increases total output.

A symmetric argument applies for $\varepsilon < 0$. Therefore, aggregate output is maximized when $N_A = N_B = \frac{1}{2}$.

With DRS, there is no incentive to gather together, to agglomerate. Rather production should be as dispersed as geography allows. This is just what we see in this example.

Suppose that both cities have an IRS technology,



If $\varepsilon > 0$ then by decreasing ε slightly, we shift from population from the larger to the smaller city.

The marginal product of labor is higher, i.e. f' is steeper, in the larger than the smaller city.

It follows that decreasing ε decreases total output. A symmetric argument applies for $\varepsilon < 0$.

- With IRS, there is an incentive for people to gather together, to 'agglomerate'. This problem has two solutions, $N_A = 1, N_B = 0$ or $N_A = 0, N_B = 1$. To maximize aggregate output, we want everyone in the same place.
- Multiple equilibria are common in models with IRS. They create three problems.
    - Fundamentally, they are an indication that the theory is incomplete. It does not provide a basis for selecting between multiple equilibria.
    - We do not understand how, or whether economies can switch between different equilibria.
    - We may not know which equilibria we are observing in measurements of cities. This can complicate making policy recommendations.

- We don't observe either complete dispersion or complete agglomeration. We observe some agglomeration, so there must be some IRS. But, IRS must attenuate, somehow, with city size. The moncentric city model suggests a mechanism for this. Even if labor is more productive in larger cities, more of it (or more labor income) are dissipated by commuting as cities get larger, too. This tension will help determine city sizes when we discuss systems of cities.

Returns to scale and spatial equilibrium

In spatial equilibrium, we need to think about the actions of a 'small' agent.

$$y_i \sim \text{output for small agent}, i = 1, ..., N$$
$$n_i \sim \text{labor for small agent}$$
$$N = \sum_j n_j$$
$$Y = \sum_j y_j$$

Suppose that our technology is,

$$y_i = AN^{\sigma} n_i, \; i = 1, ..., N$$

This implies an that the aggregate technology is,

$$
\begin{aligned}
Y_i = \sum_{j=1}^{N} y_i &= \sum_{j=1}^{N} AN^{\sigma} n_i \\
&= AN^{\sigma} \sum_{j=1}^{N} n_i \\
&= AN^{\sigma} N \\
&= AN^{1+\sigma}
\end{aligned}
$$

It is easy to check that this technology is IRS or DRS as σ is positive or negative. Urban economists often refer to σ as the agglomeration effect, or something similar.

Now consider the problem of a single agent choosing where to locate. We suppose that all firms and workers are ‘small enough’ that their affect on AN^{σ} is to small for them to perceive. This leads us to the following calculation of the marginal product of labor, from the perspective of a small firm or worker. We're going to assume that labor markets are competitive so that the wage is equal to this marginal product.

$$\begin{aligned}
\frac{\partial}{\partial n_i} y_i &= \frac{\partial}{\partial n_i} AN^{\sigma} n_i \\
&= \frac{\partial}{\partial n_i}(AN^{\sigma}) n_i + (AN^{\sigma}) \frac{\partial}{\partial n_i} n_i \\
&= \sigma(AN^{\sigma-1}) \frac{\partial}{\partial n_i} N_i n_i + (AN^{\sigma}) 1 \\
&= \sigma(AN^{\sigma-1}) 1 n_i + (AN^{\sigma}) 1 \\
&= \sigma(AN^{\sigma-1}) n_i + (AN^{\sigma})
\end{aligned}$$

As n gets small, the first term disappears, and we are left with

$$\frac{\partial}{\partial n_i} y_i = (AN^{\sigma})$$

That is, when agents are small, they ignore the effect that their location choice has on aggregate output.

Three comments:

1. When there are increasing aggregate returns, the individual incentive to crowd into a city is less that the social incentive. Another way of saying this, and one that is common in the literature, is that there is an 'agglomeration externality' or just an 'agglomeration effect'. An implication of this is that, at least sometimes, we should expect spatial equilibrium to lead to less agglomeration than is optimal, at least in the specific sense described in the planner's problem.

2. This example focuses on concentrations of people affect labor productivity. Similar logic allows us to ask how concentration of people affect firm productivity, or other measures of economic output.
3. This discussion focuses on agglomeration economies in production. It is also natural to think about agglomeration economies in consumption. There are clearly increasing returns to scale in the provision of, e.g., museums, sports teams and the latest, coolest stuff, while decreasing returns to scale probably operate for things like congestion, pollution, crime and park space.

4. Notice that the market wage we've just derived is actually the average product of labor if we account for the effect each person has on aggregate productivity. It is easy to see that the average product of labor is going to be less than the marginal product (accounting for aggregate effects).

How to measure agglomerations?

- One of the implications of agglomeration economies seems to be that production activity will concentrate in space.
- This is widely observed. Technology firms are very concentrated in Silicon Valley. Automobile firms are very concentrated in Detroit.
- On first look, this seems like strong evidence for increasing returns to scale. But, could it be random? If firms were distributed randomly, what would it look like, and would it look different enough from what we observe to convince us that firms wanted to be near each other?

Consider an economy with three firms, $i, j = 1, 2, 3$ choosing between two locations, $k = A, B$ at random. Half the time they choose A, half the time they choose B. What would we expect to observe?

outcome #	firm			# A	#B
	1	2	3		
1	A	B	B	1	2
2	A	A	B	2	1
3	A	B	A	2	1
4	A	A	A	3	0
5	B	B	B	0	3
6	B	A	B	1	2
7	B	B	A	1	2
8	B	A	A	2	1

This table describes all of the possible arrangements of firms. If firms choose randomly each outcome is equally likely.

We will be interested in the resulting distribution of pairwise distances between firms. To think about this, let d_{ij} be the pairwise distance between firms i and j. Six such distances are possible, one for each possible pair of firms, $(1,1),(1,2),(1,3),(2,2),(2,3),(3,3)$

Suppose that d_{ij} is one if firms are in different locations, and zero otherwise.

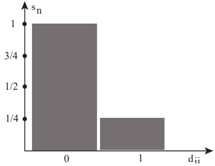
From the first table, we can easily calculate the share of outcomes, s_n, where one firm is in A and two are in B, etc.

s_n	#A	#B	$d_{ij}=0$	$d_{ij}=1$
$1/8$	3	0	6	0
$3/8$	2	1	4	2
$3/8$	1	2	4	2
$1/8$	0	3	6	0
Mean			$6\frac{1}{2}$	$1\frac{1}{2}$

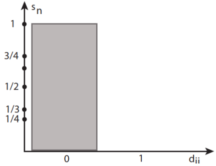
Firms are completely concentrated in one location one time in four and at least two thirds of firms are always in one location.

When should we conclude that the concentration of firms is not consistent with random choices across symmetric locations?

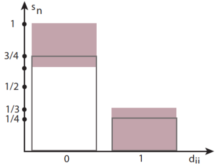
To think about this, plot the histograms of pairwise distance.



'dispersion'



'agglomeration'



'mean'

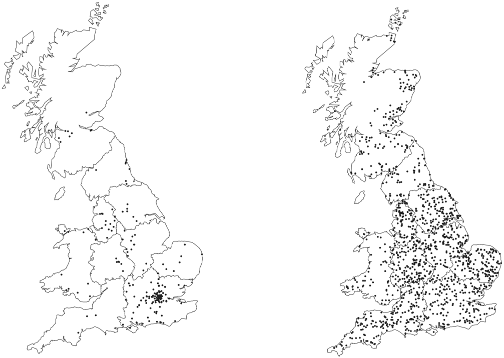
In the last panel, pink indicates the range of outcomes we can expect to see. If we see something out of this range, then firms are not choosing at random between identical locations. Either,

- The firms have a systematic preference for one location or the other (natural advantage).
- The firms coordinate to be near each other.

Note that it is not possible to tell these explanations apart with these data.

Duranton and Overman (2005) do an exercise very similar to the example above using data from the 1996 UK census of establishments.

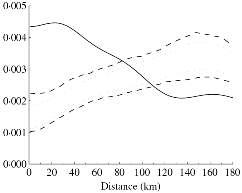
- ‘establishments’ are either ‘firms’ or, for firms with more than one facility, ‘plants’. Hence, these are ‘plant level data’.
- Data reports establishment post code, a little smaller than a US zipcode.
- Establishment Standard Industrial Classification (SIC). For example, SIC 2441 is ‘basic pharmaceuticals’ and 2932 is ‘other agricultural and forestry machinery’



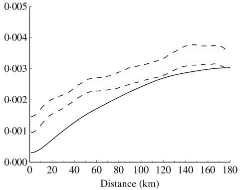
(a) Basic Pharmaceuticals (SIC2441)

(c) Other Agricultural and Forestry Machinery (SIC2932)

Each dot shows the location of an establishment in the particular SIC.



(a) Basic Pharmaceuticals (SIC2441)



(c) Other Agricultural and Forestry Machinery (SIC2932)

- These figures are constructed in much the same way as in our example.
- The solid line shows the realized histogram of pairwise distances.
- The dashed lines are like the ‘pink envelope’ in the figure based on our example. They show the range of outcomes we expect if firms chose randomly over the universe of ALL establishment location (over all SICs).

- Pharmaceuticals are ‘too close’ for randomness. Other Ag. etc., are a little too spread out.
- Of the 234 industries (with more that 10 establishments) in their data, 177 have distance profiles that don’t seem random. Industries that are ‘too close’ are much more common than ‘too dispersed’.
- Ellison and Glaeser (1997) conduct a qualitatively similar (but much more complicated) exercise using US data and reach a similar conclusion. Interestingly, they find that employment is more concentrated than firms. That is, bigger firms are more concentrated than smaller. (We don’t really have a theory for why, but this raises an important conceptual question: are bigger firms in cities because cities make them more productive (and therefore bigger), or do bigger firms locate in cities?)

How to measure agglomeration economies

We just argued that when we are considering small agents, and agents are paid the marginal product of their labor, in a spatial equilibrium we should have

$$w = \frac{\partial}{\partial n_i} y_i = (AN^{\sigma})$$

If we take logs, this gives us

$$\ln(w_i) = \ln(AN^{\sigma})$$
$$\ln(w_i) = \ln(A) + \sigma \ln(N)$$

This means that we can learn about the extent of aggregate returns to scale by looking at the relationship between log wage and log of total employment.

This leads to the following regression

$$\ln(w_{ij}) = B + \sigma \ln(N_j) + \varepsilon_{ij}$$

using data describing lots of workers, i, distributed across many cities, j.

We can do this same exercise with other measures of worker output, e.g., patents per tech worker. We can also do something similar with firms, e.g. output per unit of inputs. The idea is similar, but the details are different.

This seems straightforward, and variations of this regression have been estimated in hundreds, if not thousands, of academic research papers. And it remains an area of active research.

There are three reasons this has attracted so much attention.

- Anything to do with productivity and productivity growth, sheds light on the remarkable increase in human wealth and population since the industrial revolution. This is surely one of the most important things to happen in the world in the last several thousand years, so understanding it better, rightly, attracts a lot of attention.
- Understanding why people are drawn together in cities is central to understanding why we have cities, so understanding the foundations of agglomeration economies is central to the study of cities.

- A number of difficult econometric problems arise when we try to estimate σ, firms, people and cities all have unobserved traits that contribute to economic output, and there has been a lot of effort devoted to resolving the effects of city attributes on productivity from the effects of these unobservable city attributes.
- Why agglomeration effects arise is really important. If there is an externality, it creates a role for policy. If there is not, then we expect a decentralized (spatial) equilibrium to have good welfare properties. More on this later.

There are also a lot of details and practical problems that need to be addressed. For example,

- Does returns to scale vary with employment or population? In practice, there is not a lot of variation in labor force participation across cities, so this is not important.

- Does returns to scale vary with employment in your own industry, in the industries that make your inputs and buy your output, or all industry?
- Does returns to scale vary with the total size of the city, or with the density of employment near a particular firm?
- Does increasing returns depend equally on all workers, or just those with particular attributes, e.g., college educated, tech workers?
- Are different production activities subject to different degrees of returns to scale, e.g., new vs mature industries?
- Do agglomeration economies affect different types of workers, e.g. college educated and not, differently?

The answer to these questions, pretty broadly, seems to be ‘yes’. The simple framework we used to develop some intuition about agglomeration economies, and which gives rise to simple wage regression, is clearly far too simple.

There is not one agglomeration economy, there are many, and they probably don’t operate at the same strength or through the same mechanism in all cities or for all people.

We have a pretty good idea of the magnitude of σ in an average city, and we know a little bit about how σ differs from industry to industry, worker type to worker type, and mechanism to mechanism.

Estimates of σ

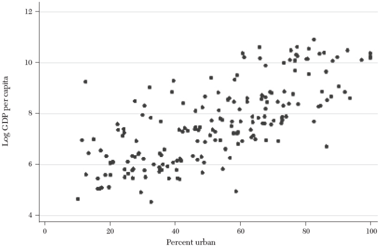
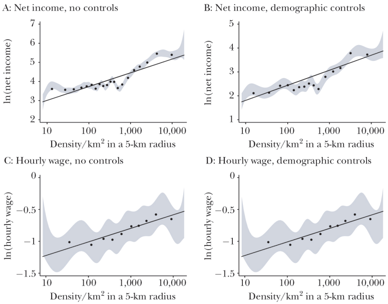


Figure 10. Income and Urbanization, 2006

Notes: Units of observation are countries. Data are from the World Development Indicators database. GDP per capita is measured in constant 2000 U.S. dollars.

This plot from Glaeser and Gottlieb (2009) shows *country* level data. The x-axis is share of urban population. The y-axis is log of per capita GDP. Countries with more people in cities are richer.



Note: Binscatter plots of LSMS net income of respondent household and of hourly wage, against the log of GHS population density in a 5km disk around the survey respondent. Log population density is censored below at about 8/km^2. Left panels have no controls. Right panels includes demographic controls and country fixed effects. Shading indicates 95 percent confidence band. Income includes wage income, net farm income, and net business income. For a small number of observations expenses exceed (monthly) incomes. We drop these observations to permit logarithmic scaling. LSMS survey countries are listed in table A2. Linear regression based on results in table A1a, which provides more details about the sample. Slope coefficients and standard errors of best linear fits are; (a) 0.313 (0.016) (b) 0.317 (0.014) (c) 0.118 (0.015) (d) 0.049 (0.009). Details in online Appendix.

In the developing world, Africa in particular, an increase in density increases wages and household income. Household income increases faster with density.

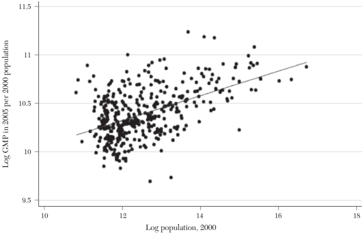


Figure 1. Productivity and City Size

Notes: Units of observation are Metropolitan Statistical Areas under the 2006 definitions. Population is from the Census, as described in the Data Appendix. Gross Metropolitan Product is from the Bureau of Economic Analysis.

The regression line is $\log \mathit{GMP\ per\ capita} = 0.13\ [0.01] \times \log \mathit{population} + 8.8\ [0.1]$. $R^2 = 0.25$ and $N = 363$.

This plot from Glaeser and Gottlieb (2009) shows that log per capita GDP increases with log of city population for US cities around 2005.

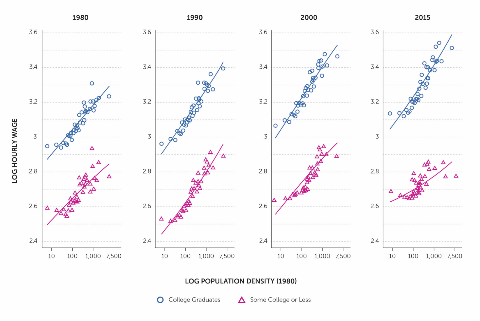


Figure plots real mean log hourly earnings among college graduates and workers with some college or lower education in 1980, 1990, 2000, and 2015. Wages are normalized to real 2015 levels using the Personal Consumption Expenditure deflator. Each plotted point represents

This remarkable plot from Autor (2020) shows the urban wage premium for high and low skilled workers over time. The density premium for the wages of low skilled workers decreased dramatically relative to high skilled workers between 2000 and 2015.

Rosenthal and Strange (2004) provide a comprehensive survey of the literature that estimates σ.

They find that for a when N is employment or population, most studies that estimate

$$\ln(w_{ij}) = B + \sigma \ln(N_j) + \varepsilon_{ij}$$

find that $\sigma \in [0.03, 0.08]$, with most estimates around 0.04

This means that doubling city employment increases labor productivity by about 4%. Moving from a town of 10,000 to a city of 1,300,000 means doubling N about 7 times. This implies that the wage of an identical worker increases by $(1.04)^7 = 1.32$.

Henderson et al. (1995) estimate agglomeration effects by industry. They find that wages in an industry respond to (1) own industry employment, and to a measure of how diverse is employment in the city across sectors, and (2) that different industries respond differently.

That is, agglomeration economies look like they are about as complicated as can be. Their effect varies by industry, and with the portfolio of industries in the city.

Spatial scale I

Rosenthal and Strange (2004) finds that when N is a measure employment or population density near the worker, typically in the same county, or within a disk of standard radius (e.g., 5 or 10 miles),

$$\ln(w_{ij}) = B + \sigma \ln(D_j) + \varepsilon_{ij},$$

find that $\sigma \in [0.04, 0.05]$, where D_j is the city measure of population or employment density.

Does this mean that there is a separate agglomeration effect for city size and density? Note that population is the product of area

Spatial scale II

and population density. Letting a_j denote the area of city j, this means that we can write

$$\begin{aligned}
\ln(w_{ij}) &= B + \sigma \ln(N_j) + \varepsilon_{ij} \\
\ln(w_{ij}) &= B + \sigma \ln(a_j D_j) + \varepsilon_{ij} \\
\ln(w_{ij}) &= B + \sigma \ln(D_j) + \sigma \ln(a_j) + \varepsilon_{ij}
\end{aligned}$$

So, if we regress log wages on city density and city population density, if the agglomeration effect is really from city size, we should find the coefficients on density and area are the same.

Using French data from 1976-1996, Combes et al. (2008) conduct this regression (more-or-less) and find that

$$\ln(w_{ij}) = B + 0.037 \ln(D_j) + 0.011 \ln(a_j),$$

Spatial scale III

and the estimates are precise enough to reject the hypothesis that the coefficient on density and area are the same.

These results suggest that the effect of population density is more important for worker productivity than is the total size of the city. If we double a city's population by doubling its area, density constant, wages increase by about 1%. If we double a city's population by doubling its density, area constant, wages increase by about 4%.

Note the slight cheat here. Density at the individual level and the city average density are not quite the same thing.

Arzaghi and Henderson (2008) and Rosenthal and Strange (2003) find strong evidence that at least some agglomeration effects are local. The effect of economic activity within a mile (Rosenthal and

Spatial scale IV

Strange) and within a few hundred yards (Arzaghi and Henderson) is important, but drops off rapidly.

Mechanisms I

Since productivity is so central to economic development, and so important a determinant of how cities are organized, we would like to understand why cities contribute more to productivity as they grow.

There are several candidate explanations.

- Labor market thickness. Labor markets are bigger in big cities and so the matching between employers and employees is cheaper and/or more productive in larger cities than small. If you want to hire an urban economist, your chances of finding a good one are better in a big city than a small one.
- Input market thickness. Input markets are more competitive and bigger in big cities, so firms can find exactly the right input and buy it more easily in larger cities than small.

Mechanisms II

- Knowledge spillovers or learning. If you are trying to learn how to make a new product, or refine your process for making an old one, it will be easier to figure out if there are lots of specialized people around for you to talk to.

These three have attracted a lot of attention because they are consistent with the sort of increasing returns to scale aggregate technology described above: As the city gets bigger, all three increase productivity (if they are present).

Note that, as we saw in our simple example, when there is an externality, people won't capture the full benefit of their decision to urbanize, and so we should worry that big cities will be underprovided in equilibrium.

Three other mechanisms are also important,

Mechanisms III

- Natural advantage. People agglomerate in a place because something makes it particularly productive, e.g., a natural harbor.
- Sorting. A city might be productive just because the people who live there (or the firms located there) are productive. Separating this effect from productivity effects caused by a place is difficult and important.
- Path dependence. Cities start at more or less random places, and stay there once they are established. This seems consistent with the presence of increasing returns and multiple equilibria. People need to agglomerate in a city *somewhere*, but where doesn't matter much.

Mechanisms IV

Notice that natural advantage and sorting don’t seem to involve external effects. If sorting or natural advantage is what is behind the observed relationships between agglomeration (however measured) and productivity (however measured) then there we have one less reason to worry that spatial equilibrium will fail to lead to an optimal arrangement of people and economic activity across locations.

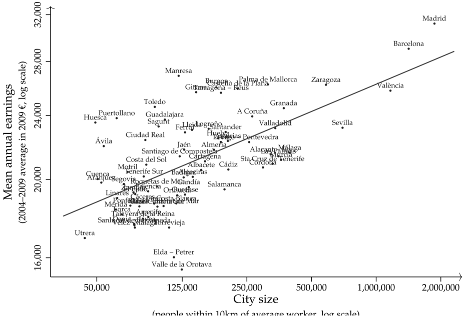
Input sharing I

Suppose a firm sells a widget for price p_{out}. To make this widget, the firm uses labor, capital, and p_{in} of intermediate inputs purchased from other firms.

If firms in denser cities are more productive because they can specialize into smaller parts of the production process, then we should see that $\frac{p_{in}}{p_{out}}$ increases with city size and density.

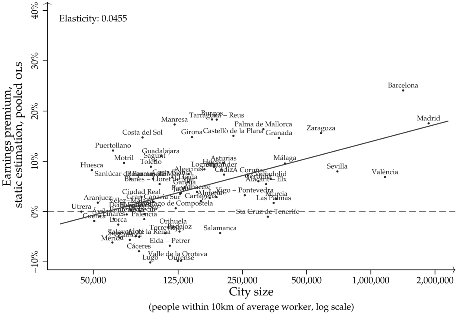
Holmes (1999) does exactly this calculation and finds that this happens for most industries. For a typical plant, increasing own industry employment in nearby counties from 10,000 to 25,000, increases $\frac{p_{in}}{p_{out}}$ by about 0.03. Since labor is about half of all firm expenditure, this is a 6% increase in the share of purchased inputs.

Sorting and learning I



People in bigger cities make more money. Going from 50k to 2m increases mean wage from 18,000 to 28,000 EU, an increase of 56%. How much of this is because more productive workers are in cities?

Sorting and learning II



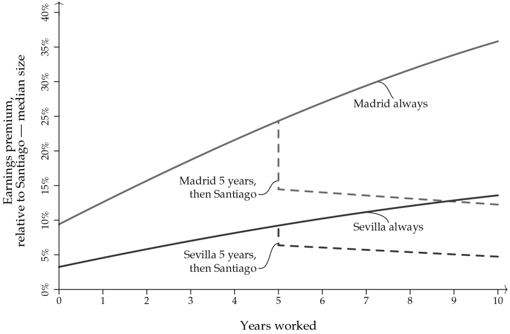
Statistically identical people (same sector, experience, tenure, skill, education) increase their wages by less than 20% when they move from a city of 50k to one of 2m.

Sorting and learning III

About half of agglomeration economies are due to sorting in Spanish data. Combes et al. (2008) finds about the same thing using French data.

Some of the effect is also due to learning. Someone who moves from Madrid to Santiago has higher wages, on average, than someone who moves from Sevilla to Santiago. More able people sort to big cities, and learn faster while they are there.

Sorting and learning IV



Is there an externality here? Yes. If I move to a big city, you learn faster.

Natural advantage and path dependence I

In the interior [South] the principal group of trade centers ... were those located at the head of navigation, or 'fall line,' on the larger rivers. To these points the planters and farmers brought their output for shipment, and there they procured their varied supplies... It was a great convenience to the producer to be able to sell his crop and buy his goods in the same market. Thus the towns at the heads of navigation grew into marked importance as collecting points for produce and distributing points for supplies of all sorts. (Philips (1905) quoted in Bleakley and Lin (2012))

Natural advantage and path dependence II

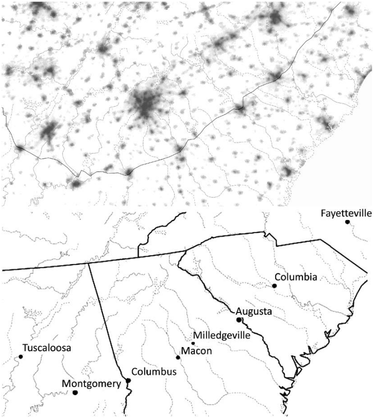
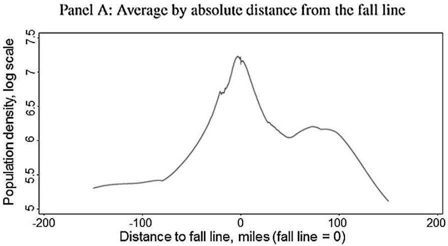


FIGURE II
Fall-Line Cities from Alabama to North Carolina

Natural advantage and path dependence III

Top images is lights at night in 2003. Solid line is ‘fall line’ Many cities started at the fall line, before railroads, are still important places today.



Plot of 2000 population density along fall line rivers as a function of distance to the fall line.

Natural advantage and path dependence IV

It looks like natural advantage and path dependence are both important for determining where people agglomerate.

Knowledge spillovers I

> *The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. (Marshall 1890)*

Something like this is surely going on, but this is a terrible basis for a theory. It sounds like magic and invites many questions about the mechanism. Does it matter which industries you are near? Do you learn by talking to people at lunch? Hiring knowledgeable co-workers?

There has been too much work on these questions to cover in detail. The data seem to make a strong case for some sort of knowledge spillovers.

Moretti (2021) makes about as strong a case for this sort of effect as we can hope for. Using panel data that reports the identity of

Knowledge spillovers II

inventors who file patents, there residential address he estimates the effect of city size on the number of patents.

To describe what he does, we need the following notation,

$$
\begin{aligned}
i, j, t &\sim \text{inventor, city, year} \\
y_{ijt} &\sim \text{patents filed} \\
\delta_i &\sim \text{Inventor fixed effects} \\
\gamma_j &\sim \text{city fixed effects} \\
x_{jt} &\sim \text{city size in year } t
\end{aligned}
$$

Knowledge spillovers III

He runs (more-or-less) the following regression for all workers within a sector,

$$y_{ijt} = \delta_i + \gamma_j + \sigma x_j t + \varepsilon_{ijt}$$

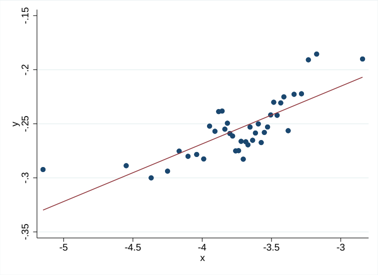
To understand this regression, take first differences. That is,

$$\begin{aligned} y_{ijt+1} &= \delta_i + \gamma_j + \sigma x_{jt+1} + \varepsilon_{ijt+1} \\ \underline{-y_{ijt}} &\underline{= -(\delta_i + \gamma_j + \sigma x_{jt} + \varepsilon_{ijt})} \\ \Delta y_{ijt+1} &= \sigma \Delta x_{jt+1} + \Delta\varepsilon_{ijt+1}) \end{aligned}$$

That is, σ reflects the effect of changes in city size on patenting, holding constant time invariant city and inventor traits. That is, controlling for sorting and natural advantage.

Knowledge spillovers IV

Figure 4. Average Log Number of Patents Per Inventor Per Year and Log Cluster Size — All Years and Fields



Notes: The slope is 0.053 (0.008). Controls include dummies for year, field and city.

For an average worker patenting increases with the number of other workers in that the same sector-city pair. Interestingly, σ is about 0.05, here, too.

This seems to suggest some sort of ‘knowledge spillover’ at work, even if we don’t know quite how it works.

Conclusion I

We now have the basis for answering the question we started with.

It is clear that agglomeration economies operate, at least at the industry level. People are willing to put up with commuting in order to be close to one another because it makes them more productive, and this productivity premium is increasing in the size of cities throughout the range observed city sizes.

Beyond this, things get complicated. The effects of agglomeration economies are complicated and it probably makes sense to think of there as being many agglomeration effects, not just one.

In particular, agglomeration effects are probably

- larger for density than city size.
- larger for the developing than the developed world.

Conclusion II

- larger for people who are more educated or have more experience.
- larger for knowledge intensive industries.
- about the same for levels and growth rates of productivity.
- contribute to human capital accumulation.

With that said, if you were going to guess that doubling city size improved your measure of output, whatever it is, by 5%, you would likely be within a factor of two of our best estimate, and probably much closer.

The mechanisms behind this have been the subject of speculation at least since Alfred Marshal in 1890. We have pretty good evidence for each of the following mechanisms for agglomeration economies.

Conclusion III

- Knowledge spillovers
- Labor market thickness
- Input market thickness
- Sorting of high ability people into larger cities
- Natural advantage
- Path dependence

Conclusion IV

We do not have a good sense for the extent to which these mechanisms interact. For example, it is not clear if the estimated effects from each of the mechanism sums to more or less than the total agglomeration effect, or whether the effect of city size operates through different mechanisms than the effect of density.

Given that empirical research is constrained by the fact that the samples of cities available for study number in the hundreds, it is not clear that it will be possible to make a lot of progress on these issues.

Arzaghi, M. and Henderson, J. V. (2008). Networking off madison avenue. *The Review of Economic Studies*, 75(4):1011–1038.

Autor, D. (2020). The faltering escalator of urban opportunity. *Securing Our Economic Future, Melissa S Kearney and Amy Ganz (eds.)*, pages 108–136.

Bleakley, H. and Lin, J. (2012). Portage and path dependence. *The quarterly journal of economics*, 127(2):587–644.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of urban economics*, 63(2):723–742.

Duranton, G. and Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106.

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

Glaeser, E. L. and Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of economic literature*, 47(4):983–1028.

Henderson, V., Kuncoro, A., and Turner, M. (1995). Industrial development in cities. *Journal of political economy*, 103(5):1067–1090.

Holmes, T. J. (1999). Localization of industry and vertical disintegration. *Review of Economics and Statistics*, 81(2):314–325.

Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review*, 111(10):3328–75.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85(2):377–393.

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