Chapter 3
Agglomeration and Empirics

We have seen how agglomeration forces, modelled through a potential function, can give rise to concentration of economic activity in city centers, and shape the structure of cities. In this chapter, we delve deeper into the micro-foundations of these agglomeration forces.

We could start by outlining some microeconomic models of why locating together is advantageous, but you are probably tired of models from the previous two chapters. So instead, we will give an overview of the determinants of agglomeration in these models, and dig into some empirical papers that estimate the extent of agglomeration forces. For a review of theoretical models of agglomeration along with some models, see Duranton and Puga (2004).

Along the way, we will review some empirical strategies used in urban economics to obtain causal estimates. Baum-Snow and Ferreira (2015) provide an useful overview of empirical methods, with a focus in urban and regional economics.

3.1 Sharing, Matching and Learning

Duranton and Puga (2004) classify the microfoundations of agglomeration economies around three categories.

- Sharing: Firms may share indivisible goods, which have a large fixed cost, are shareable and excludable, subject to crowding. Think about a soccer stadium. No household would build it by themselves, but it is feasible to build it for a set of households. Once built, it satisfies all the criteria above. A city may organize around the stadium, and the equilibrium city size will be such that the stadium is not crowded.
  There are others kind of sharing. You may remember that in monopolistic competition the sector that aggregates the varieties produced by the intermediate supplier has increasing returns to scale. Aggregating these suppliers in a single space may be profitable for the firms. Duranton and Puga show in a simple multi-sector

model of monopolistic competition that a larger workforce leads to production of more varieties, such that there are increasing returns in the aggregate. They then show that with these increasing returns, each location must specialize in one sector. If that were not the case, moving one worker from the least productive sector to the most productive sector in each location would increase productivity. So all the firms in the same sector locate in the same place, and share the gains from variety.

Sharing increasing returns can also occur if there are increasing returns to scale at the firm level coming from increasing specialization of work. In that case, having a larger workforce in a single place may lead to larger than proportional production, because tasks are assigned more efficiently.

- Matching: Matching models of the labor market have workers and vacancies meeting each other. In directed search models and models of match-specific productivity, a larger pool of workers may increase the expected quality of each match, and also increase the probability of a match by firms.
- Learning: Clustering together may increase the diffusion of information between firms, and between workers. Human capital may spill over across workers when they are closer together. We will discuss many of these place-based spillovers when we talk about place-based policies.

3.2 Estimating the Returns to Density

We would like to estimate how strong are these increasing returns arising from agglomeration. A good starting point to tackle these empirical issues is to think of the ideal experiment that would give us the estimate we want (Angrist and Pischke, 2008).

We would like to have a couple of comparable cities with some productive structure in place. Then, we would alter the density in one of them, maybe increasing the number of workers per square kilometer, and look at whether the increase in output per square km is more than proportional to the increase in workforce. In reality, our thought experiment is impossible to implement. With a panel dataset of cities, the best we can do is compare output or productivity across cities that have different density over time.

There are at least three issues with this approach:

1. It is not immediately clear what the independent variable for comparison should be. Should we compare cities of different density, or cities of different size? In Fujita and Ogawa (1982), the specification of the potential function suggests that density (and not size) is what matters.
2. When comparing dense and sparse cities, we won't probably be holding everything else constant. It is likely that there are differences in worker productivity across cities, so the comparison will give us biased estimates of the effects of density. In terms of the monocentric model, you can think of this as comparing cities with different income per worker.

3. Even if we were to adequately control for observable worker productivity, we probably won't be able to control for unobservable differences in worker productivity across cities. In fact, economic theory suggests that these differences should be there and be important. Suppose that you have two types of workers: those who love programming and those who do not, and you can not observe which is which. The programmers should sort across space and move to Silicon Valley. You will see more of them in San Francisco and will also see larger productivity there, but this won't be the effect of density, just the effect of having better workers there.

All of these issues mean that we will have to be particularly creative when trying to estimate the returns to density. We will look at a couple classic papers on agglomeration effects, but you should be aware that this is a large branch of literature. Ahlfeldt and Pietrostefani (2019) have a recent synthesis about this.

3.3 Productivity and the density of economic activity

Ciccone and Hall (1996) estimate the effects on density on productivity using data for U.S. states.

Their starting point is a production function with increasing returns to scale at the county level:

$$q_c = a_c \left(\frac{n_c}{a_c}\right)^{\alpha} \left(\frac{q_c}{a_c}\right)^{\frac{\lambda-1}{\lambda}} \tag{3.1}$$

Here, n_c is labor per acre, q_c is output at the county level and a_c is the county's area. So production is a decreasing returns to scale function of labor, but there are some effects of density that increase this production. Solving for q_c yields:

$$\frac{q_c}{a_c} = \left(\frac{n_c}{a_c}\right)^{\gamma}, \tag{3.2}$$

with $\gamma = \alpha\lambda$. α measure the traditional marginal productivity of labor, and λ measures the agglomeration effects. If we now add this up over counties to the state level, we get:

$$\frac{Q_s}{N_s} = \frac{\sum c \in C_s n_c^{\gamma} a_c^{-(\gamma-1)}}{N_s} \tag{3.3}$$

The right hand side of this equation can be thought of as a density index. To see this, think that the state has two counties with equal area such that $a_c^{-(\gamma-1)} = 1$. You have one unit of labor and have to split into α for county 1 and $1-\alpha$ for county 2. The rigth hand side can now be written as:

$$D_s(\gamma) = \alpha^{\gamma} + (1-\alpha)^{\gamma} \tag{3.4}$$

For $\gamma < 1$, this is maximized for $\alpha = 1/2$. Agglomeration effects are low, so it is optimum to have employment spread out between the two counties. For $\gamma > 1$, this is minimized at $\alpha = 1/2$, and it is optimum to have employment in a single county.

Ciccone and Hall extend this basic model to account for physical and human capital, but the basic intuition remains the same. The estimating equation is the following:

$$log\frac{Q_s}{N_s} = log\phi + logD_s(\theta,\eta) + u_s \tag{3.5}$$

$$D_s(\theta,\eta) = \frac{\sum_{c\in C_s}\left(n_c h_c^{\eta}\right)^{\theta} a_c^{1-\theta}}{N_s}, \tag{3.6}$$

with θ playing the role of the density elasticity that γ played before. h_c are workers' average years of education per county.

3.3.1 Estimation issues

The first issue with estimating these equations should be the non-linearity. OLS won't work here. This can be circumvented by using a non-linear estimation method.

The larger issue is the endogeneity. Barring the non-linearity issues for a second, this can only be estimated if u_s is uncorrelated with $D_s(\theta,\eta$. But this is unlikely to be the case. u_s must be playing a role here because in the data, we observe that in equilibrium counties have different densities. But in the previous models we saw that if $\theta > 1$, then employment should be agglomerated in a single state. This means that workers must have some preferences for locating in states. To the extent that these preferences are related with worker and state productivity, it is unrealistic to assume that u_s is not correlated with density.

3.3.2 Aside: Instrumental Variables and Non-linear estimation

Here is a deliberately informal reminder on how to deal with these endogeneity issues. For a formal treatment, see Hayashi (2000). For a treatment within a potential outcomes framework, and for a coverage of causal estimation in urban economics, see Baum-Snow and Ferreira (2015). Consider the following setting. We want to estimate β in the following equation:

$$Y = X\beta + u. \tag{3.7}$$

We usually do this by pre-multiplying by X'. Under the assumption of no correlation between X and u, we solve for β:

$$\begin{aligned} X'Y &= X'X\beta + X'u \\ &= X'X\beta \\ \hat{\beta} &= (X'X)^{-1}(X'Y). \end{aligned} \tag{3.8}$$

This wouldn't work if $X'u \neq 0$. This is selection bias, that may arise because there are unobservable components in u that are correlated with X, or because there is an omitted variable in u.

If we have and instrumental variable Z that satisfies the following two conditions:

1. Relevance: $Z'X \neq 0$
2. Exclusion: $Z'u = 0$

Then we could estimate β by pre-multiplying by Z:

$$\begin{aligned} Z'Y &= Z'X\beta + Z'u \\ &= Z'X, \\ \hat{\beta}^{IV} &= (Z'X)^{-1}Z'Y. \end{aligned} \tag{3.9}$$

This is the instrumental variables estimator in a nutshell. An instrumental variable affects the outcome (relevance) only through its influence on the variable of interest X, and not through any other variables (exclusion).

A widely-used example of instrumental variables estimation in Urban Economics is Baum-Snow (2007). He is interested in the effect of highways on population decentralization in cities in the US. The endogeneity here arises from reverse causality: the existence of new highway towards the suburbs may cause population to move to the suburbs, but at the same time, the highway may have been built because population were in the suburbs already. Baum-Snow circumvents this by instrumenting the number of radial highways in each city with how many of these were planned in a 1947 federal highway plan. The key assumption for identification is that federal highway plan affected future highway construction, but was not related to suburbanization. As it turns out, the 1974 highway plan was built with military defense reasons in mind, and did not consider local features. So it seems that in this case, the instrument satisfies the requirements above.

In general, instrumental variables estimation is a bit of a risky business, because there are many reasons why the instrument may not meet the exclusion restriction.

Now, let us consider an extension to deal with non-linear functions, which is what Ciconne and Hall use in their estimation. Write the outcome as a non-linear function of the parameters:

$$y = f(\beta) + u. \tag{3.10}$$

With an instrumental variable in hand, the orthogonality assumption continues to be that Z and u are not correlated. Consider minimizing the following function in β:

$$[y - f(\beta)]' Z \Omega Z' [y - f(\beta)]. \tag{3.11}$$

This is a quadratic form in $Z'u = [y - f(\beta)]'Z$. By minimizing this function, we want to set $[y - f(\beta)]'Z$ as close to 0 as possible. If you have more than one instrumental variable available, you can still implement this, and the weighting matrix Ω will give different weights to each one of the instruments. Ciconne and Hall choose Ω as $(Z'Z)^-1$, which is a natural choice because it weights each instrument by the inverse of its variance.

We will learn about all these econometric issues more deeply when we start learning about quantitative urban models in the next section of the course, where we will frame linear and non-linear OLS and IV as different versions of the generalized method of moments estimator.

3.3.3 Estimation and Results

Turning back to Ciconne and Hall's equation of interest

$$log \frac{Q_s}{N_s} = log\phi + logD_s(\theta, \eta) + u_s \tag{3.12}$$

$$D_s(\theta, \eta) = \frac{\sum_{c \in C_s} \left(n_c h_c^{\eta}\right)^{\theta} a_c^{1-\theta}}{N_s}, \tag{3.13}$$

we see that to estimate this, we need data on state output and employment, which Ciconne and Hall get from the census and a Gross State Output dataset for 1988. Education comes from the CPS.

The endogenous variable here is D_s, for which Ciconne and Hall use 4 instruments:

1. Presence of absence of a railroad in the state in 1860
2. Population in 1950
3. Population density in 1880
4. Distance from the sea

All of these are past sources of agglomeration. The identifying assumption is that these measures of agglomeration do not affect 1988 productivity but still affect 1988 agglomeration.

Figure 3.1 shows the main results from their estimation. Their estimates hover around 1.05 for θ and 0.4 for η. These imply that doubling a county's density increases TFP by about 5%. Doubling the education of the workforce increases TFP by about 40%.

There are a couple of issues with Ciconne and Hall's seminal paper. The first is the lack of a role for heterogenous effects of density. This may be an average effect for those states on which past agglomeration affected future agglomeration, but it may not be as informative for policy purposes. The second issue, which is an IV issue in general, is how much we believe the instruments. It is a bit surprising that

Table 3.1 Ciconne and Hall's Table 1

TABLE 1—ESTIMATION RESULTS

Instrument	Density elasticity, θ (standard error)	Education elasticity, η (standard error)	R^2
None (NLLS)	1.052	0.410	0.551
	(0.008)	(0.396)	
Eastern seaboard	1.055	0.460	0.548
	(0.017)	(0.51)	
Railroad in 1860	1.061	0.330	0.537
	(0.011)	(0.450)	
Population in 1850	1.060	0.350	0.539
	(0.015)	(0.510)	
Population density in 1880	1.051	0.530	0.549
	(0.019)	(0.550)	
All	1.06	0.060	0.536
	(0.01)	(0.82)	

Notes: The equation estimated is (24). The data are value added for 46 states and Washington DC. For the 46 states we have used data on employment and average years of education at the county level.

Source: Ciccone and Hall (1996)

the OLS and IV estimates are very close to each other, which would suggest there was not much endogeneity to begin with.

A third issue which is pervasive in research designs that rely on spatial variation is whether the control group is also affected. All that Ciconne and Hall's paper is doing is comparing dense counties to not-so-dense ones. Suppose for a moment that the not-so-dense county in the control group actually has a dense area in the border with the dense county. Then some of the agglomeration effects are being passed into the control group, which would bias down the estimates of the agglomeration effect.

3.4 Million Dollar Plants

Let us go back to the thought experiment we outlined at the beginning of the chapter. Ideally what we would like to do is to increase density exogenously and see what effects that has on the existing firms. What if we could just throw some firms inside a county?

Greenstone et al. (2010) approximate this through a quasi-experiment. Their design is fun: they compile a series of location rankings from a corporate real estate journal, which lists how "million dollar plants" which are deciding where to open rank different locations, and which one they ultimately chose. Apparently, this was not uncommon, although Amazon popularized it recently with its choice among cities to open their latest plant. Then, they. compare plants that barely lost to those who won. You can think of this as a regression discontinuity design where the latent variable is some location attractiveness.

The paper carries out a TFP estimation at the county level, and compares the narrow winners to the narrow losers. They also frame their results around a spatial equilibrium Roback (1982) model, which we will see later in the course.

3.4.1 Theory

Greenstone et al. develop a simple model to frame their results. We will not cover their model here, and will defer to our study of Roback (1982) to study this in more depth. Instead, here we will focus on their discussion of possible sources of agglomeration and their implications for empirical work.

1. Firms and workers agglomerate because of the size of the labor market. Larger cities may lead to improved matching between workers and firms. If this is the only source of agglomeration, however, then TFP, which is conditional on employment, should not vary with density.
2. Firms may agglomerate because of increasing returns to scale from reduced transportation costs. The easiest way to think about this for us development country natives is to think about a rural area, and a truck that delivers goods weekly from a nearby more urban town[1]. Without agglomeration, the truck has to visit every house in the rural area, which is costly in terms of time and money. But if households agglomerate, the truck only has to visit a single place. Thus households pay a smaller transportation cost mark-up in the price of their goods. Note thas this type of agglomeration does not change the production function, but it is likely to reduce production costs.
3. Firms may share knowledge, by sharing workers or skills. This type of agglomeration will lead to larger TFP in denser areas.
4. Firms may agglomerate around local amenities. This does not change TFP, but it does change wages (from spatial equilibrium).
5. Firms may agglomerate around natural advantages. In this case, firms in agglomerated areas will have higher TFP, but this should come from the fact that there is a natural advantage nearby. Controlling for the natural advantage, agglomeration should bear no relation to TFP-

It is likely that a combination of these forces is at play for real world firms, but finding evidence of TFP increases makes the case for knowledge externalities stronger.

3.4.2 Research Design

The article compares the "winner" counties where the million dollar plant opened, to the "losers" which are runner-up counties. The identifying assumption is that in absence of the plant opening, TFP in winner and loser counties would have the same time trend. This is a straightforward application of "differences-in-differences" which we review below.

[1] I thank Margarita Gáfaro for this example.

3.4.2.1 Aside: Differences in Differences

This treatment of diff-in-diff follows Angrist and Pischke (2008). Suppose that in absence of the plant opening, in time 0, incumbent plants in winner and loser counties have different permanent TFP levels and a random component. This is noted in column 1 of table 3.2.

County/Time	$t=0$	$t=1$
Winner	$\alpha_w+\varepsilon_{w0}$	$\alpha_w+t+\theta+\varepsilon_{w1}$
Loser	$\alpha_l+\varepsilon_{l0}$	$\alpha_l+t+\varepsilon_{l1}$

Table 3.2 Differences in Differences

As time passes, TFP changes in both counties, but it changes in the same way, per our identifying assumption. So, in time $t = 1$ both counties' TFP changes by t. This is in the second column of table 3.2.

In $t = 1$, something else changes in the winner county. The plant opens and because of agglomeration spillovers, incumbent plants in the winner county get a TFP boost θ. Our goal is to estimate this spillover. How can we do it?

The answer is to compare differences in the expectation of TFP over winner and loser counties, over time. We take the difference over time for each county, and then subtract these time differences. Alternatively, we take the county differences in each time period, and then subtract these county differences. The resulting difference-in-difference is an unbiased estimator of the effect of interest θ. This is shown in table 3.3.

County/Time	$t=0$	$t=1$	E(Time Difference)
Winner	$\alpha_w+\varepsilon_{w0}$	$\alpha_w+t+\theta+\varepsilon_{w1}$	t + θ
Loser	$\alpha_l+\varepsilon_{l0}$	$\alpha_l+t+\varepsilon_{l1}$	t
E(County Difference)	$\alpha_w-\alpha_t$	$\alpha_w-\alpha_t+\theta$	$DD=\theta$

Table 3.3 Differences in Differences: Estimation

We can estimate θ in a regression framework as follows:

$$TFP_{it} = Winner_i + Post_t + \theta \times Winner_i \times Post_t + \varepsilon_{it}. \tag{3.14}$$

The regression has the advantage of allowing us to use our standard econometric toolkit to estimate standard errors, include control variables, etc. In general, with many units $i = 1,\dots,i$ and time periods $t = 1,\dots,T$, a panel differences-in-differences model is:

$$Y_{it} = \alpha_i + \beta_t + \theta X_{it} + \gamma Z_{it} + \varepsilon_{it}, \tag{3.15}$$

where α_i are unit dummies, β_t are time dummies, X_{it} is the policy variable of interest, Z_it are (time-varying) control variables and ε_{it} is an error term. We will revisit this differences-in-differences design many times as we go along.

3.4.2.2 Estimating Equation

Greenstone et al. have data from a census and a survey of manufacturers from 1973 to 1978, and identify the incumbent plants in winner and loser counties. Table 3.4 has some descriptive statistics about their sample.

Table 3.4 Greenstone et al. Table 1

TABLE 1
THE MILLION DOLLAR PLANT SAMPLE

	(1)
Sample MDP openings:[a]	
Across all industries	47
Within same two-digit SIC	16
Across all industries:	
Number of loser counties per winner county:	
1	31
2+	16
Reported year − matched year:[b]	
−2 to −1	20
0	15
1 to 3	12
Reported year of MDP location:	
1981–85	11
1986–89	18
1990–93	18
MDP characteristics, 5 years after opening:[c]	
Output ($1,000s)	452,801
	(901,690)
Output, relative to county output 1 year prior	.086
	(.109)
Hours of labor (1,000s)	2,986
	(6,789)

[a] Million Dollar Plant openings that were matched to the Census data and for which there were incumbent plants in both winning and losing counties that are observed in each of the 8 years prior to the opening date (the opening date is defined as the earliest of the magazine reported year and the year observed in the SSEL). This sample is then restricted to include matches for which there were incumbent plants in the MDP's two-digit SIC in both locations.

[b] Only a few of these differences are 3. Census confidentiality rules prevent our being more specific.

[c] Of the original 47 cases, these statistics represent 28 cases. A few very large outlier plants were dropped so that the mean would be more representative of the entire distribution (those dropped had output greater than half of their county's previous output and sometimes much more). Of the remaining cases, most SSEL matches were found in the ASM or CM but not exactly 5 years after the opening date; a couple of SSEL matches in the 2xxx–3xxx SICs were never found in the ASM or CM; and a couple of SSEL matches not found were in the 4xxx SICs. The MDP characteristics are similar for cases identifying the effect within same two-digit SIC. Standard deviations are reported in parentheses. All monetary amounts are in 2006 U.S. dollars.

Source: Greenstone et al. (2010)

There are a few noteworthy features here. First, there are not that many plants. Second, these plants are large, with an output of 450 million dollars on average.

They build on a Cobb-Douglas production function:

$$Y = AL^{\beta_1} K_B^{\beta_2} K_E^{\beta_3} M^{\beta_4}. \tag{3.16}$$

Here, Y is production, K_B is building capital, K_E is equipment capital and M are materials. A is TFP, and it is allowed to have an agglomeration forces effect:

$$ln(A_{pijt}) = \alpha_p + \mu_{it} + \lambda_j + \varepsilon_{pijt} + A(N_{ijpt}). \tag{3.17}$$

Here, α_p are plant fixed-effects, μ_{it} are industry-time effects, λ_j are pair (winner-loser) effects and ε_{pijt} is an error term. $A(N_{ijpt})$ is the agglomeration effect, over which they assume a linear structure. The pair effects λ_j are important because they guarantee that the variation in plant openings is coming from within each winner-loser pair.

They estimate an unit-specific-trends version of the panel differences-in-differences model in equation 3.15.[2]

$$\begin{aligned} ln(A_{pijt} = {} & \delta Winner_{pj} + \psi Trend_{jt} + \Omega Trend_{jt} \times Winner_{pj} + \kappa Post_{j}t \\ & + \gamma Trend_{jt} \times Post_{j}t + \theta_1 Winner_{pj} \times Post_{jt} \\ & + \theta_2 Trend_{jt} \times Winner_{pj} \times Post_{jt} + \alpha_p + \mu_{it} + \lambda_p + \varepsilon_{pijt}. \end{aligned} \tag{3.18}$$

This specification allows plant openings to affect not only the level of TFP but also its trend. This makes sense, although extreme differences in trends may cast doubt on the validity of our research design. In fact, having control and treatment groups with their own trend may be problematic in dynamic difference-in-difference designs (Borusyak and Jaravel, 2017). It is a bit of an unfair critique though, because we are only starting to understand these issues now.

3.4.3 Results

Figure 3.1 shows the main results of Greenstone et al.'s analysis. This is a dynamic specification of the differences-in-differences model in equation (3.18). In this specification, a separate regression is ran on winner an loser counties, with dummies for the time to plant opening.This is known as a panel event study design. The figure plots the event-time dummies and their differences. These differences can be thought of estimates of θ_1 from (3.18) but with a single coefficient for each event-time.

On average, TFP increased by around 4 % in winner counties compared to loser counties. There is also some evidence of a trend-break. Accounting for the trend, the estimate grows to a whooping 12 %. The problem here is that the control group is allowed to be on a separate trend, so there is no variation to separately identify the

[2] For the panel design, the estimating equation would be

$$Y_{it} = \alpha_i + \beta_t + \theta X_{it} + \gamma Z_{it} + \delta_i t + \varepsilon_{it},$$

and the δ_i coefficients capture the different trends.

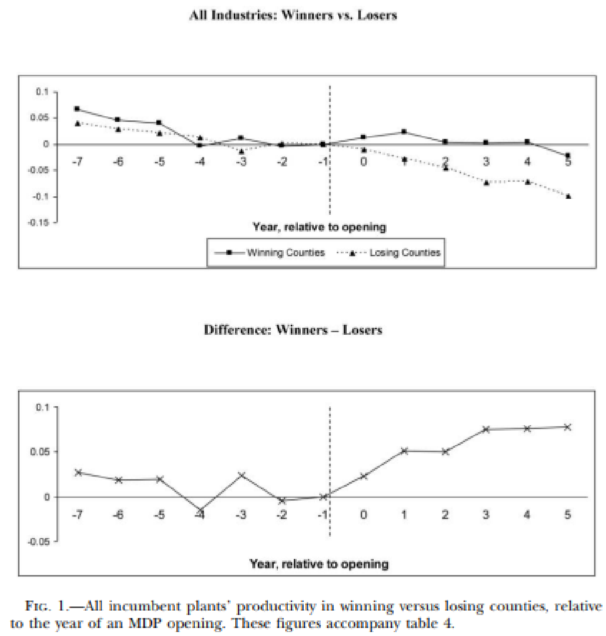

Fig. 3.1 Greenstone et al. (2010) Figure 1

trend and the event-time dummies (Borusyak and Jaravel, 2017). Still, the estimates are informative of the size of the agglomeration spillovers.

The extended analyses in Greenstone et al. show some additional facts about the spillovers:

- Spillovers seem to be larger within industry.
- The spillovers seem to increase the closer the incumbent firms are to the incoming plant, in terms of technology and worker transitions.
- New plants open after the incoming plant is built.
- Wages increase in the area.

The advantages of Greenstone et al.'s analysis over Ciconne and Hall are evident. The article's research design resembles the thought experiment more closely. Still, let us highlight a few shortcomings in this paper.

1. Inputs are also endogenous in equation (3.18). One way to fix this is by dynamic panel estimation, which they try in their robustness section.
2. There is a bit of selectivity in their sample. The million dollar plants seem to be opening in rust belt counties, so there are some external validity concerns.
3. We do not know if winners are really like losers. There must be a reason why firms choose to open the plants in the winner counties, and this is unobserved to the econometrician. To the extent that these unobserved features interact with the

performance of the incumbent firms, we may have bias. Moreover, these may be playing a big role here if the counties are really different. This is analogous to the issues with twin designs when studying the returns to education (Ashenfelter and Rouse, 1998).
4. There is not enough room to explore industry heterogeneity here, and these industry heterogeneity may be important to distinguish among theories of agglomeration.
5. What if plants relocated from the loser counties to the winner counties? This would definitely bias our estimates.
6. What if there is commuting? With commuting, the impacts of building the million dollar plants are not necessarily as local, as they may propagate to nearby counties, which could be in the control group. We will explore this issue further in the context of quantitative spatial models.

Greenstone et al.'s results suggest that information and knowledge transmission is a key driver of agglomeration effects. Recall that other explanations could be ruled out if TFP did not increase with density.

3.5 Information exchange and its spatial extent

Arzaghi and Henderson (2008) measure how strong these knowledge transmission "networking" benefits can be. Using a dataset of advertising agencies in Manhattan, they find that advertising firms have a higher willingness to pay and locate near other advertising firms. The existence of nearby firm within 250 meters increases the likelihood of a new firm opening by about 2 %. They also find that these effects are heavily localized, and die out at around 750 m of distance.

Arzaghi and Henderson highlight the advantages of their approach:

- They analyze advertising agencies instead of manufacturing, so they are not looking at rust belt cities.
- They look at finely-grained spatial data, that allows them to better measure the spatial extent of the agglomeration forces.
- Information exchange is the main source of agglomeration for advertising agencies.

3.6 Heterogeneity: Specialized and Diversified Cities

One question about agglomeration that we have not tackled is whether agglomeration occurs across or within sectors. It is reasonable to think that both the magnitude and spatial extent of the agglomeration forces changes, depending on whether the firm that receives the density benefits is located next to firms of the same sector.

However, entirely sector-specific agglomeration effects can not be the whole story. If this were the case, all the cities would be composed of firms of the same sector. In other words, all cities would be specialized. Duranton and Puga (2000) shows several facts about the distribution of cities that suggests that there are both intra-sector and inter-sector agglomeration effects.

1. There are specializaed and diversified cities, and they coexist.
2. Larger cities are diverse, smaller cities are specialized.
3. The distribution of city sizes is stable over time (Zipf's law).
4. Employment and population growth are related to specialization and diversity.
5. There is a lot of plant rotation. Plants relocate from big to small cities.

3.7 Summary

These papers give us a picture of the strength of agglomeration forces. You can interpret these estimates as measurements of β and α in Fujita and Ogawa (1982): the increase in output (or TFP) for an increase in density, and the decay of that extra productivity across space.

Greenstone et al.'s results are much larger that Ciconne and Hall's ones. This is counterintuitive, since we thought the selection bias was going in the other direction. In a recent meta-analysis, Ahlfeldt and Pietrostefani (2019) document a median elasticity of TFP to density of about 0.06, and a median labour productivity elasticity of 0.04. Their "recommended" estimate of the wage elasticity to density is 0.04.

We now emphasize that there seems to be a bit of a disconnect between the empirical evidence in this chapter and our previous analysis of the multicentric city model. If we are interested in estimating β and α, why not try to estimate the actual model? Turns out that these models are not very amenable to estimation. In the next chapter, we will examine a new breed of models that are more "data-ready". In doing so, we will introduce new sets of tools from econometrics and international trade.