Introduction to spatial and urban economics: A mini course

Jordan Rosenthal-Kay[1]

Federal Reserve Bank of San Francisco
February 2026

[1]The views expressed herein do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System. All errors are my own.

All of this cribs heavily from Duranton and Puga (2004), Combes et al. (2011), Duranton and Puga (2015), and Bryan et al. (2020), among others.

This mini course consists of three lectures. The first two are on **Cities** *and* **Regions***. These lectures combine theory and highlights of the empirical literature, and focus on using models to interpret data. The last lecture is on* **Quantitative Spatial Economics***, and is aimed at teaching the building blocks of a quantitative spatial model, the type currently used in applied work, as well as highlight frontiers in using spatial (e.g., GIS) data.*

Contents

Lecture 1: Cities

1.1 Cities and economic development

The 'problem of economic development' Bob Lucas wrote, is 'the problem of accounting for the observed pattern [...] of rates of growth and per capita income' and that 'once one starts to think about [it], it is hard to think about anything else' (Lucas, 1988). Differences in per capita income across space are *staggering*. Take a look at Figure 1.1, which lines up countries by their GDP per capita and the share of their population living in cities. The horizontal axis is on a logarithmic scale: the average person in the United States roughly 60 times more income than the average person in, say, Niger. Macroeconomists have tried to account for these differences due to, e.g., capital deepening or total factor productivity. The lesson of this type of growth accounting is that much of the difference in living standards across places is due to 'technology:' residual, unexplained variance. The focus on this mini course is the single greatest technology ever invented: cities.

Why are cities a technology? To think about this, one must recognize what is meant by 'technology' in macroeconomics: not so much tools and equipment (we can account for these in capital) but rather (1) recipes, (2) process knowledge, (3) organizational efficiency.[1] By 'recipes' I mean instructions or blueprints, the production methods that convert capital and labor into output. These are ideas. 'Process knowledge' refers to the know-how embedded in people on how to follow a recipe; for example, this can include both management practices and tacit skills involved in machine operation. These 'skills' are nontransferable (not human capital) and are not part of the recipe. 'Organizational efficiency' refers to the environment in which production occurs, from supply chains and the spatial organization of production to the social and political

[1] What follows is my own conception of technology.

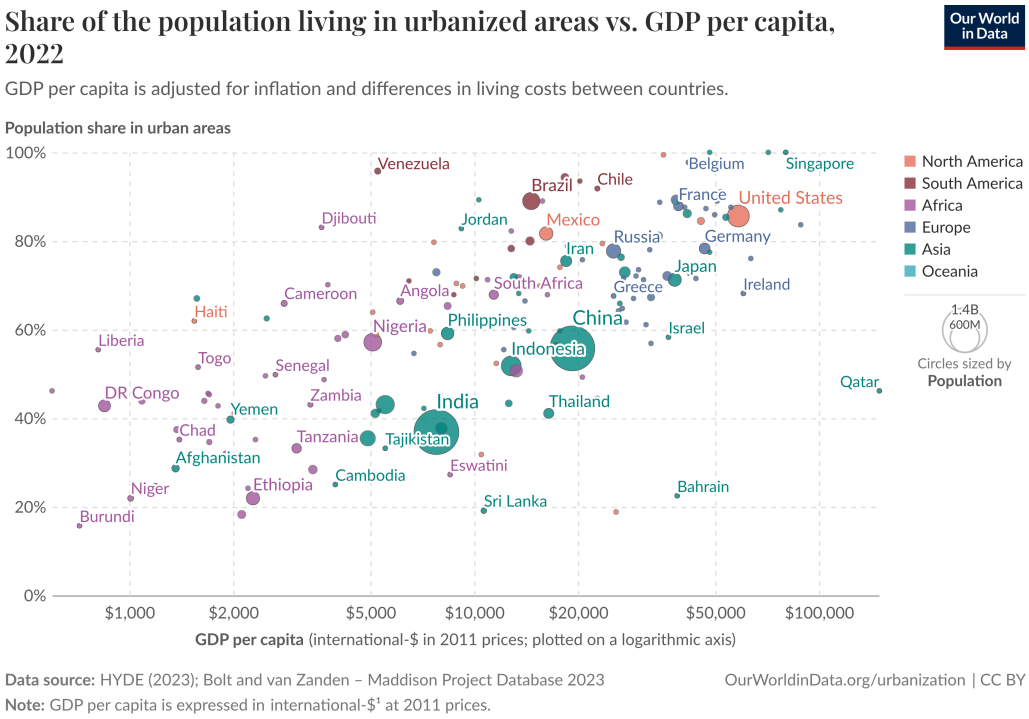

Figure 1.1: GDP/capita vs. urbanization

institutions that shape both the efficiency of firms as well as the (mis)allocation of resources across them.[2] Cities are places that generate recipes, diffuse process knowledge, and efficiently organize production.

Figure 1.1 suggests that cities may be important in explaining these income differences. The figure shows that as income per capita rises, so does urbanization. There are several competing explanations of the correlation between income per capita and urbanization. First, the causality may run from development to urbanization: As income rises, and as agricultural productivity increases, economies escape subsistence agriculture and Malthusian traps, and the excess food surplus can support nonagricultural labor and urban life.

These forces are most certainly at play: Figure 1.2 replicates a figure from Bryan, Glaeser, et al. (2020). It shows the share of labor *not* participating in agriculture against output per worker in agriculture and nonagricultural sectors. Poor countries with very low agricultural output per worker have massive amounts of their labor force participating in the production of food, compelling evidence of a subsistence trap. When the income elasticity of demand for agricultural goods is less than 1 (a necessity), agricultural productivity improvements increase aggregate income more than they increase demand for agricultural goods, redirecting

[2]For a more canonical reference, Robert E Hall and Jones (1999) refer to the institutional environment as 'social infrastructure.'

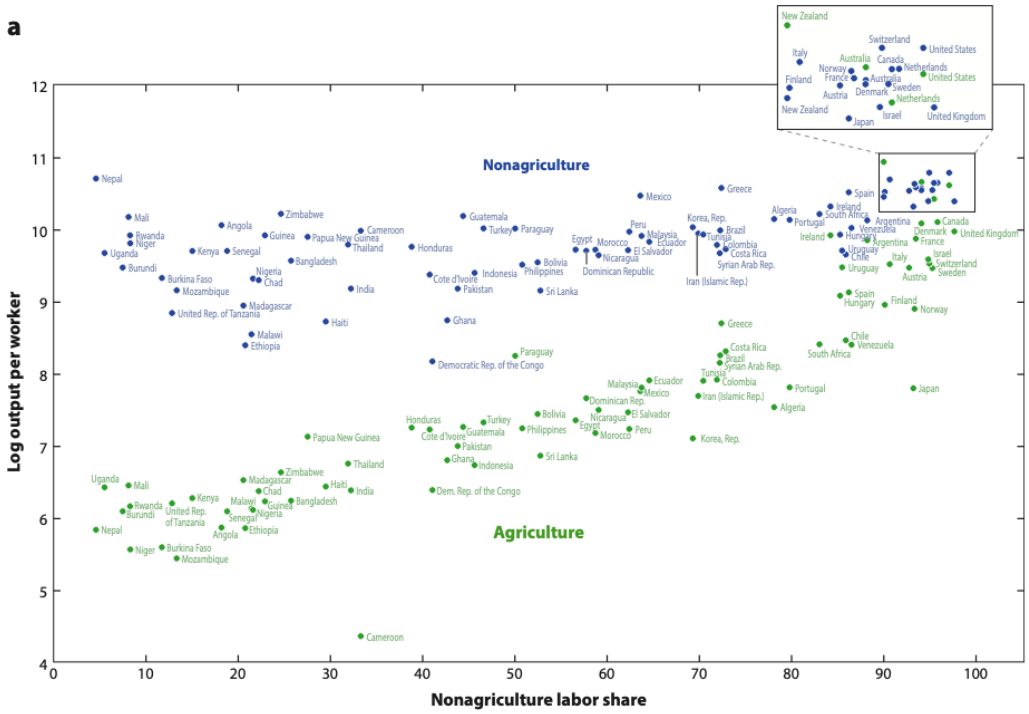

Figure 1.2: Reproduction of Figure 1a from Bryan, Glaeser, et al. (2020).

labor to other sectors; in effect, improving labor productivity in agriculture 'frees up' labor to work elsewhere, in cities. This process of structural transformation has been ongoing throughout history and naive conclusion would be that cities are not engines of economic growth, but rather urbanization is a consequence of growth.

This story probably sounds familiar from any ultra longue durée view of human history: the Neolithic Revolution was the first major economic transition; the invention of agriculture, animal husbandry, and so on, were the key technological advances that incentivized nomadic peoples to settle down and form small villages that preceded what we would call cities. Let's humor an alternative hypothesis: cities predated and helped spur the Neolithic Revolution. This is the view put forward in Jane Jacobs' *Economy of Cities* (Jacobs, 1969). Consider the Göbekli Tepe or Çatalhöyük: early Neolithic agglomerations that predate or codeveloped with early agriculture. What were cities doing at the time? Based on early anthropological evidence from Çatalhöyük, a city that existed in modern day Turkey 7500 BC to 5600 BC, Jacobs developed an alternative hypothesis. She hypothesized that these early 'cities' were like trading posts. Seed mixing occurred at these sites, potentially leading to the invention of domesticated, resilient and higher-yield varieties of cereals. Likewise, animal husbandry may have occurred, as well as small-scale specialization in the development of tools. There, there

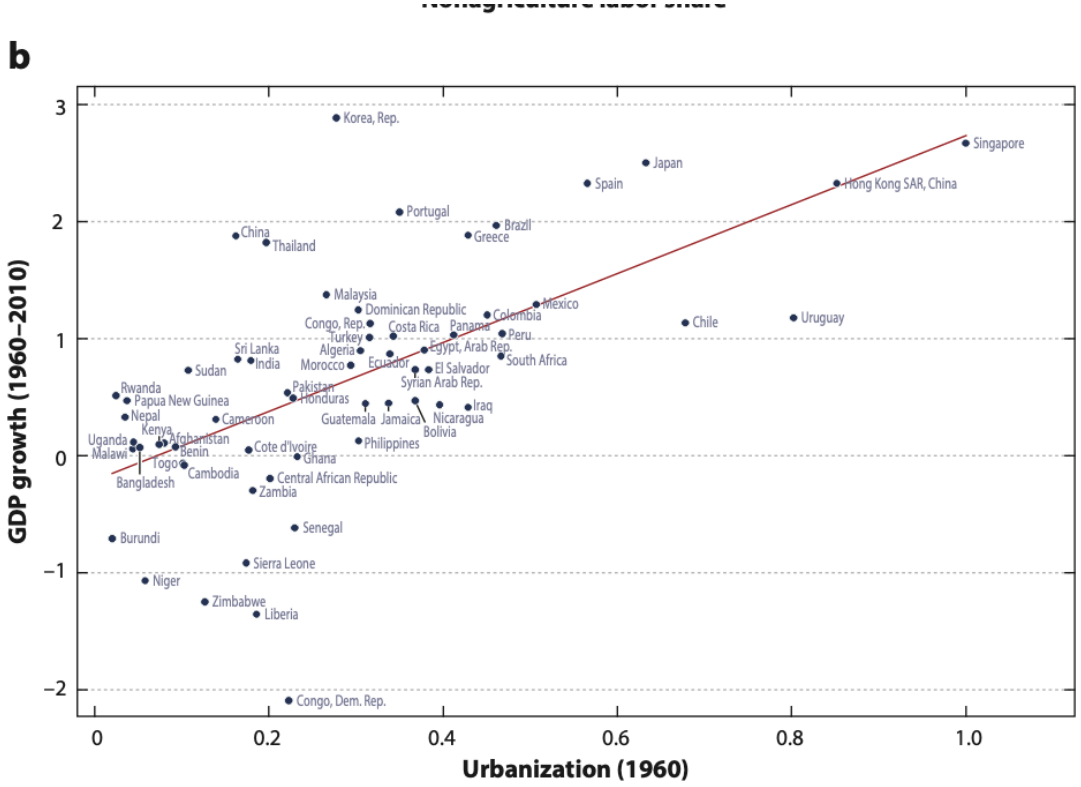

Figure 1.3: Reproduction of Figure 1b from Bryan, Glaeser, et al. (2020).

is evidence of the evolution of agriculture within the 'city:' there is evidence of both domesticated and wild cereals having been stored, consumed, grown, or traded there. Pieces of obsidian, shaped into knives or arrowheads, may have acted as a currency. While the anthropological evidence is mixed on these hypotheses, we cannot reject simultaneous or reverse causality in the agricultural-urbanization debate.

So why think that urbanization matters for growth today? Consider Figure 1.3, which shows that among poor countries in 1960, those that grew more 1960-2019 were more urbanized at baseline. To paraphrase Lucas (1988) and Bryan, Glaeser, et al. (2020), this suggests possibilities: the possibility that urbanization is an important causal driver of economic development, that by perhaps reducing rural-urban migration barriers or by improving the quality of cities themselves, we can improve human wellbeing.

1.2 Are cities a growth technology?

Now, I will try and convince you that cities are a growth technology. To do so, I will lay out some theories about what makes cities special. These are called 'agglomeration forces' or 'economies of density.' Simply put, they are a source of *increasing returns*. When you see something clustered in space, there must be economies of scale counteracting the incentive to diffuse production across space and avoid competition for land. After all, land is in nearly infinite supply, so why treat it as scarce? We will be theorizing why there must be returns to co-location to explain the existence and economic importance of cities. A great review is available in Duranton and Puga (2004).

I will then turn towards empirical analysis. One of the most important facts in urban economics is the fact that wages grow with city size. This suggests the potential for a causal impact of city size or density on productivity, consistent with agglomeration economies, and supporting these theories. However, a causal interpretation of this correlation is stymied by a host of confounding factors. To think about this fact, I'm going to riff heavily on Combes et al. (2011).

1.2.1 Sources of agglomeration economies

We call these gains of co-locating 'agglomeration economies' or 'economies of density,' and they are often thought to reflect a pure *technological externality*: i.e., an economic force that is *external* to agents' decision-making, and does not operate through the price system (a 'pecuniary externality'). Whatever it is about cities, it is, as Alfred Marshall remarked, as if 'something was in the air.' These externalities are thought to operate both in terms of proximity and size: hyper-local interactions ('spatial spillovers') and scale effects ('increasing returns').

We do not actually believe there is something aethereal about cities that makes them so productive. Alfred Marshall himself, in his 1890 work *Principles of Economics* highlighted three features that generate agglomeration economies: (1) labor market pooling, (2) knowledge spillovers, and (3) input-output linkages. These are sometimes referred to as the 'Marshallian Trinity.'

Labor market pooling refers to the fact that in larger labor markets, with more employers and job-searchers, more productive matches can be formed. Both employers and employees are somewhat specialized. From a worker's perspective, it is easier to find an employer that is a good fit when there's more employers to choose from. From an employer's perspective, they can interview more applicants and find the best match in bigger markets.

Knowledge spillovers refers to two things. One, the spatially localized diffusion of production knowledge, and two, the innovative returns to interacting with many people.

Production knowledge is often tacit and embodied in people and capital goods that are slow to move across space. That is, innovations in how to do things often spread out very slowly. Imagine the first coffee shop to start doing counter instead of waited service. This is a 'process innovation' that turns out to save the coffee shop more money on staff and serve more customers. This idea would spread out slowly, and another coffee shop owner might only have the sense to try it out themselves if they were located nearby and could see the benefit and figure out how to do it themselves. Two, there are returns to having many people together. There's more social interaction in larger places: e.g., meeting people in cafes, bars, and at parties. A bad but useful model of social interaction in cities is one of a nuclear reactor, where social fission causes interpersonal collision, and the exchange and production of new ideas. Density heightens the reaction, and the intensity of social forces present in a particular place. We learn from our peers, and so more social interaction means the greater the chance of a new idea is born. These are the dynamics brilliantly highlighted at a macro scale in the avant garde film *Koyaanisqatsi* (Dir. Reggio, 1982), which shows urban interaction, production, and commuting at a scale such that the city-as-biological system (or machine) is legible. At a more micro scale, *The Social Network* (Dir. Fincher, 2010) shows how ideas born through social interaction – ideas from the Winklevoss twins, a reminder of how chess Elo scoring works, and so on – compound to build (the) Facebook.

Input-output linkages refer to the fact that industries that co-locate and share inputs can economize on inputs by minimizing transportation costs between them. This is certainly a scale economy, but I'm not convinced it's a pure technological externality.

There are other reasons cities exhibit increasing returns, as posited by Jane Jacobs in her books *The Death and Life of Great American Cities* and *The Economy of Cities*. Jacobs posits that the scale achieved in large

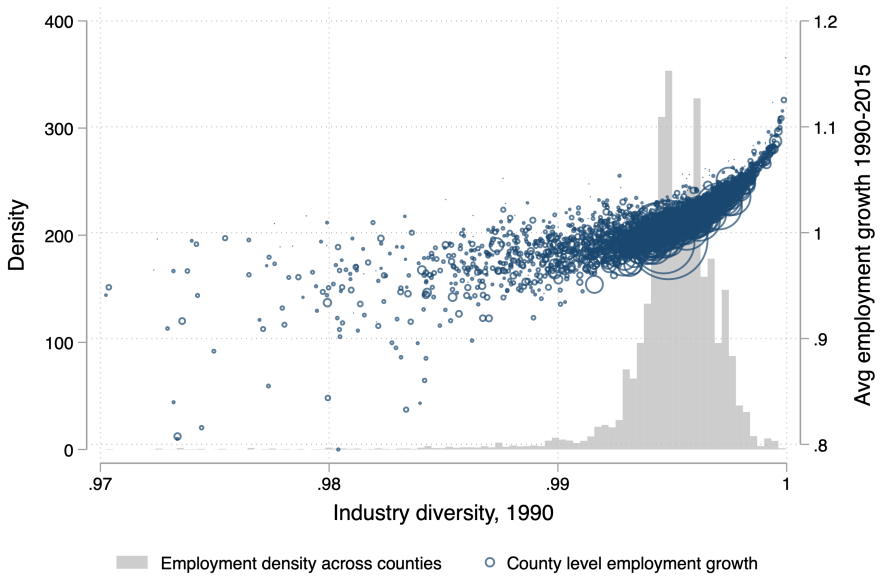

Figure 1.4: Quasi-replication of Glaeser, Kallal, et al. (1992) by me.

markets leads to new combinatorial possibilities, like the ability to mix seeds together and form new crop varieties, or the ability to take an innovation in one industry and apply it in an other. In Jacobs' model, new ideas are not simply productivity enhancing, but also generative of 'new work:' new product lines that can be exported. While Jacobs did not formalize her model, her work has inspired empirical tests of non-Marshallian agglomeration forces.

For example, in Glaeser, Kallal, et al. (1992), the authors test whether industrial diversity predicts growth. Replicating their analysis in more contemporary data, Figure 1.4 shows that cities with more diverse industrial ecosystems in 1990, as measured by an industry-employment HHI, grew faster. Testing the 'new work' hypothesis, Lin (2011) documents that larger cities are proportionality responsible for new occupational titles.

1.2.2 Measuring urban wage premia

That urban wages are higher than rural wages and that wages grow with city size are some of the most well-documented facts in urban economics. For example, in the United States, workers living in MSAs earn around 33% more than non MSA workers (Glaeser and Maré, 2001), and workers in bigger cities earn more

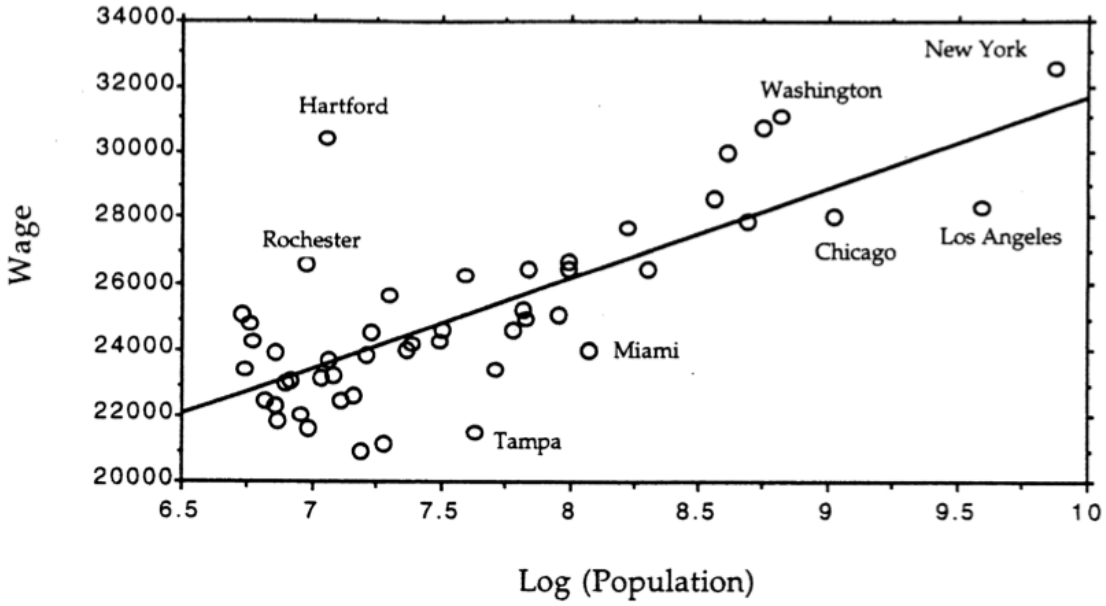

Figure 1.5: Reproduction of Figure 1 from Glaeser and Maré (2001).

than those in smaller cities. For example, Figure 1.5 reproduces the wage-population gradient across MSAs in 1992. Here, and in many other studies, the elasticity of wages with respect to city size or density tends to be around 0.05 (Ahlfeldt and Pietrostefani, 2019), and this holds internationally, e.g., in China, India, and Brazil (Chauvin et al., 2017). Does this reflect a causal effect of urbanization on productivity?

Consider what we learn from the cross section by examining a regression of individual wages on city size,

$$\ln w_{i,c(i)} = \alpha \ln L_{c(i)} + v_{i,c(i)}, \quad v_{i,c(i)} = \mu_i + \eta_c + e_{i,c(i)} \tag{1.1}$$

where i is an individual and $c(i)$ is the city of individual i. I've partitioned the error term into three components: a systematic city-level shock η_c, an individual shock μ_i and an idiosyncratic match effect $e_{i,c}$. Can we identify α from this regression? In other words, does the ordinary least squares estimate $\hat{\alpha}$ reveal the causal effect of increasing city size on workers' wages? No! What problems do we have? Well, consider what we

learn when we estimate this regression:

$$\begin{aligned}\operatorname{plim}\hat{\alpha} &= \alpha + \frac{\operatorname{Cov}(\ln L_{c(i)}, v_{i,c(i)})}{\operatorname{Var}(\ln L_{c(i)})} \\ &= \alpha + \frac{\operatorname{Cov}(\ln L_{c(i)}, \mu_i)}{\operatorname{Var}(\ln L_{c(i)})} + \frac{\operatorname{Cov}(\ln L_{c(i)}, \eta_c)}{\operatorname{Var}(\ln L_{c(i)})} + \frac{\operatorname{Cov}(\ln L_{c(i)}, e_{i,c(i)})}{\operatorname{Var}(\ln L_{c(i)})}.\end{aligned}$$

Each component of the error reflects a different kind of bias:

1. If there's sorting on unobserved ability, $\operatorname{Cov}(\ln L_{c(i)}, \mu_i) \neq 0$!

   For example, if inherently high wage earners (big μ_i) prefer to live in big cities, then $\hat{\alpha}$ is biased.

2. Simultaneity bias: if L_c depends on $\mathbb{E}[w_c]$ then $\operatorname{Cov}(\ln L_{c(i)}, \eta_c) \neq 0$.

   For example, suppose labor supply sloped up in the average wage:

   $$\ln L_{c(i)} = \varepsilon \ln \bar{w}_c + \xi_c$$

   Where $\bar{w}_c = \mathbb{E}[w_{i,c(i)}]$. Taking expectations and plugging in Equation (1.1), assuming that $\mathbb{E}[\mu_i + e_{ic}] = 0$,

   $$\ln L_c = \varepsilon\left(\alpha \ln L_c + \eta_c\right) + \xi_c \implies \ln L_c = \frac{\varepsilon\eta_c}{1-\varepsilon\alpha} + \frac{\xi_c}{1-\varepsilon\alpha}$$

   Then, $\operatorname{Var}(\ln L_c) = (1-\varepsilon\alpha)^{-2}\left(\varepsilon^2 \operatorname{Var}(\eta_c) + \operatorname{Var}(\xi_c) + 2\varepsilon \operatorname{Cov}(\eta_c, \xi_c)\right)$, and,

   $$\operatorname{Cov}(\ln L_{c(i)}, \eta_c) = \frac{\varepsilon}{1-\varepsilon\alpha}\operatorname{Var}(\eta_c) + \frac{1}{1-\varepsilon\alpha}\operatorname{Cov}(\eta_c, \xi_c)$$

   When $\eta_c \perp \xi_c$, we have a nice result. Defining $\lambda = \frac{\operatorname{Var}(\xi_c)}{(\varepsilon^2 \operatorname{Var}(\eta_c) + \operatorname{Var}(\xi_c))}$ to be the supply shock to total noise ratio,

   $$\operatorname{plim}\hat{\alpha} = \lambda\alpha + (1-\lambda)\frac{1}{\varepsilon}.$$

   When supply shocks dominate $\operatorname{Var}(\xi_c) \gg \operatorname{Var}(\eta_c)$ then $\lambda \to 1$ and we pick up the agglomeration elasticity, α. If demand shocks drive variation in the data, then $\lambda \to 0$ and regressing wages on population size recovers the (inverse) supply elasticity, $1/\varepsilon$.

So, what can we do? Early work pioneered by Ciccone and Robert E. Hall (1996) used instrumental variables Z_c like distance to waterways, historical existence of railroads, or historical population as instruments for

$L_{c(i)}$. The idea is that these are essentially long-run supply shifters ξ_c that affect a city's contemporary size without affecting its productivity, $(Z_c \perp \eta_c, e_{i,c(i)})$. These instruments remain popular in settings with only cross-sectional data, and usually hinge on establishing a valid exclusion restriction, as the first stage is usually very strong (more on that, later).

Another popular strategy is to use micro panel data on workers. Suppose we observed workers' location histories in a panel, with time indexed by t, and focused on workers who move cities? Then we could run a panel regression,

$$\ln w_{i,t,c(i,t)} = \alpha \ln L_{c(i,t)} + \mu_i + \rho_{c,t} + u_{i,c(i),t}, \tag{1.2}$$

where μ_i and $\rho_{c,t}$ are individual and city-year fixed effects. The inclusion of these time fixed effects nets out a spurious correlation between city size and wage if both are trending in time. By conditioning on worker effects μ_i, we remove the potential for sorting bias. What this regression does is look within the history of a worker, and ask what happens to their wages when they move cities. Is α in Equation (1.2) identified? We have removed the potential confounder $\text{Cov}(\ln L_c, \mu_i)$ by conditioning on worker effects. What about simultaneity bias? $\rho_{c,t}$ absorbs this too! Glaeser and Maré (2001) pioneered using worker effects to purge sorting bias, and found that much of the urban wage premium evaporates when accounting for sorting (yet a sizable chunk of the premium still remains!). Equation (1.2) looks a lot like the canonical Abowd et al. (1999) worker-firm effects specifications and is an exciting frontier for research: how much of observed place effects are driven by the sorting of productive establishments across space? Do high-wage-earners sort to high wage cities? These types of questions are what recent papers like Card et al. (2025) seek to address.

Now, this is all very encouraging, but we still have an issue. Our regression sample is effectively 'movers,' a very selected population. If effects are heterogeneous (i.e., α_i varied by individuals) than we may measure the 'treatment on the treated,' not the average treatment effect.

Is there a causal effect of place?

What other kinds of experimental designs would allow us to learn about the treatment associated with moving to a city? One idea would be to perform an RCT, assigning the treatment of changing city size at random. This is quite difficult to do, though some work has been done in Bangladesh, providing bus tickets to incentivize

rural-to-urban migration (Bryan, Chowdhury, et al., 2014; Lagakos et al., 2023).

There is, however, substantial quasi-experimental evidence for measuring the causal effect of place and smaller spatial scales than the level of a city. Such studies examine 'neighborhood effects:' the idea that living in a better neighborhood can causally improve one's wellbeing. Early work on neighborhood effects simply accounted for the characteristics of one's community in determining on education and earnings (see, e.g., Datcher, 1982). Such sociologically interesting correlations fail to measure treatment, as there may be spatial sorting of households so that unobserved ability correlates with determinants of neighborhood choice. To get at a causal effect, we need some experimental variation. The most famous experiment on neighborhood effects is the Moving To Opportunity project (MTO). The MTO was a large-scale randomized controlled experiment that took place in the 1990s in U.S. public housing, sponsored by HUD. The treated received housing vouchers that could be used in neighborhoods with low poverty rates, essentially incentivizing moves from high- to low-poverty neighborhoods. Many, many studies have been done on the MTO, but the most important is likely Chetty et al. (2016), who find that the effects moving to better neighborhoods accrue to the young, and occur later in life. Treated children were more likely to graduate college, and have higher earnings later in life, with effects declining in the age at the time of the move.

Similarly, Chyn (2018), studies what happened to those displaced by the demolition of public housing in Chicago during the 1990s. A famous example of this is the razing of Cabrini–Green, a set of mid- and high-rise public housing developments, sandwiched between Old Town and River North in Chicago. It was a area of highly spatially concentrated poverty, known for problems with crime and decaying infrastructure. The project features heavily in the horror film *Candyman* (Dir. Rose, 1992), where it serves as the lair to the ghost of 'Candyman,' a specter of historical racial violence haunting the present. Chicago vacated and demolished the structures in 2002. Chyn studies what happens to children in families that were in demolished projects relative to those that remained in place. He finds that children in relocated households had better later-in-life outcomes.

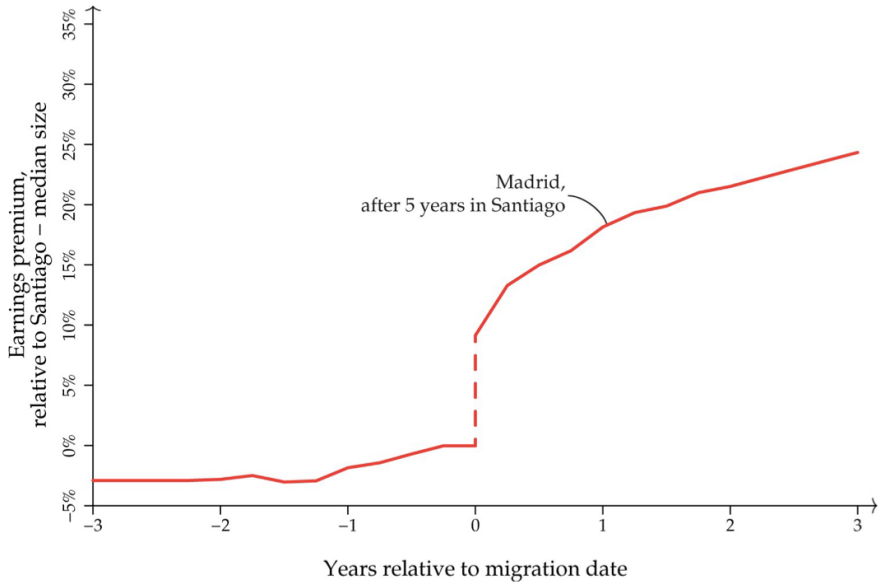

Figure 1.6: Reproduction of Figure 6 from Roca and Puga (2017).

Dynamic versus static effects

That neighborhood effects studies find a causal effect of place that occurs years later demands we ask, is the urban wage premium a static or dynamic effect? That is, is it a premium accrued quickly upon moving to a new city (e.g., a level effect on the wage), or slowly (i.e., that moving to a big city improves the rate of wage growth)?

A key reference to examine static versus dynamic effects is Roca and Puga (2017).[3] They look within worker histories in Spain what happens when workers move from small to large cities. The experiment in the data is visible in Figure 1.6. It tracks the average wage of movers from small to large cities, in percent terms relative to the wage one year prior to the move, a type of event study. Not only is there a spike in wages in moving from a small city (Santiago) to a large one (Madrid), but the rate at which earnings grow steepens, consistent with faster learning in bigger cities.

They find these 'dynamic,' longer-term earnings premia are more elastic with respect to city size than the 'static' premium, suggesting that the wage growth effects of moving to cities are as–if not more–important

[3] Another key reference is Baum-Snow and Pavan (2012). They show that in big cities, workers have steeper experience-earnings profiles, which is consistent with faster learning in bigger cities. They also find that the employer-employee 'match' component of earnings has a steeper experience profile, consistent with better search-and-matching in the labor market.

than the level effect. Figure 1.7 shows both static and dynamic urban wage premia in Spain.

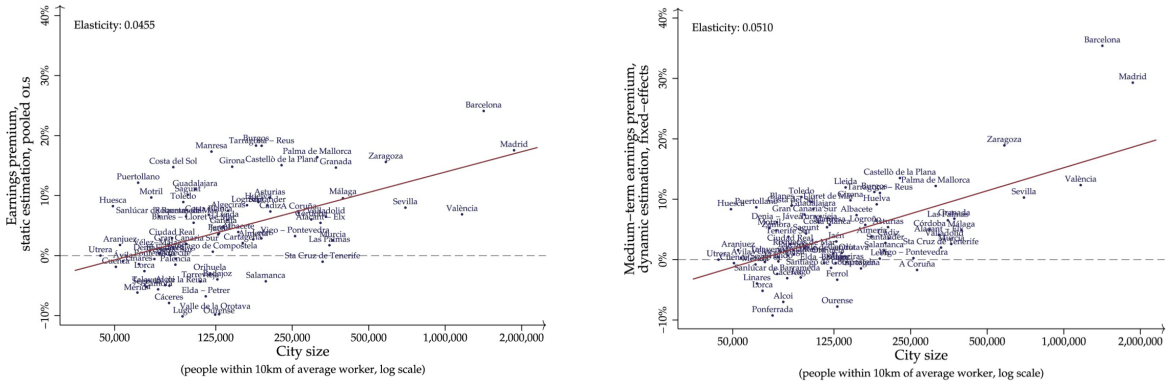

Figure 1.7: Reproduction of Figures 2 and 5 from Roca and Puga (2017). Left: static wage premia vs. city size. Right: dynamic wage premia.

1.3 Developing a theory of cities

To understand the role cities play in shaping economic development, we have to build up a theory of what cities are. First and foremost, cities are places: a large amount of people on a small amount of land. And they are places that exhibit some empirical regularities that help us understand what a city is: they tend to have a downtown, or several, where land is used for production rather than residency, coupled with residential neighborhoods around such employment clusters. The density of construction is bigger in urban core than the periphery, and thins out the farther you get from business activity. Moreover, what people do in cities is commute: they travel to work and consume, rather than engaging in home production and consumption. In other words, what cities do is allow for this decoupling, generating income gains through the specialization of land use.

So, we shall now build up a model of cities. We will begin with von Thünen's model of land use, and then extending the ideas to capture urban form, through the synthesis of work done by Alonso, Muth, and Mills. We will then add back in the ingredients that make cities special: agglomeration and spatial spillovers.

1.3.1 Organizing space: von Thünen and the bid rent curve

I now present a simple model of land use as a warm-up to thinking about the spatial organization of cities: it is the classic von Thünen model.[4] Von Thünen asks us to consider how agricultural land is organized around a small market town, in distance units x from the market. Suppose agricultural products a vary in their value yield per acre, y^a as well as their transport costs τ^a per unit x. Variation in y^a is determined by the physical productivity per unit land and the price a fetches in the market, taken as exogenous. Transport costs vary at the product level: it is expensive to transport fruit and vegetables, which must be refrigerated and handled with care so as to not bruise. Cereals are cheaper to move across space – no refrigeration needed – and cattle are even cheaper: heck, they'll do some of the transportation themselves if you point them in the right direction!

If it were possible, farmers would like to locate all production at $x = 0$ and pay no transport costs, but land is in fixed supply everywhere, and so production cannot concentrate in space. How, then, does a market economy organize production across space? Suppose land is owned by absentee landlords who set rent to that of the highest bidder. Let $r(x)$ denote the bid rent for a unit of land at distance x from the market. A farmer producing good a at x earns profit $\pi^a(x) = y^a - \tau^a x - r(x)$. Perfect competition ensures that $\pi^a(x) \leq 0$. When $\pi^a(x) < 0$, farmers don't produce, so a is only produced if $\pi^a(x) = 0 \implies r(x) = \max_a y^a - \tau^a x$. This is the *outer envelope* of all product a-specific bid rent curves, and is illustrated in Figure 1.8. It embeds a notion of *spatial equilibrium*: profits must everywhere be the same. In the absence of this condition, there could be the opportunity of spatial arbitrage, and therefore this organization would not constitute a competitive equilibrium.

What does this theory imply? First, it shows us how market forces can organize production across space: competition and the price system organize space to maximize output net transportation costs. Second, it tells us that the opportunity cost of land is spatially varying. An alternative form of land use: e.g., the construction of housing at x, would have to pay $r(x)$ for that land, too, implying that e.g., the price of housing is more expensive closer to $x = 0$.

[4] Johann Heinrich von Thuünen was a 19th century German economist.

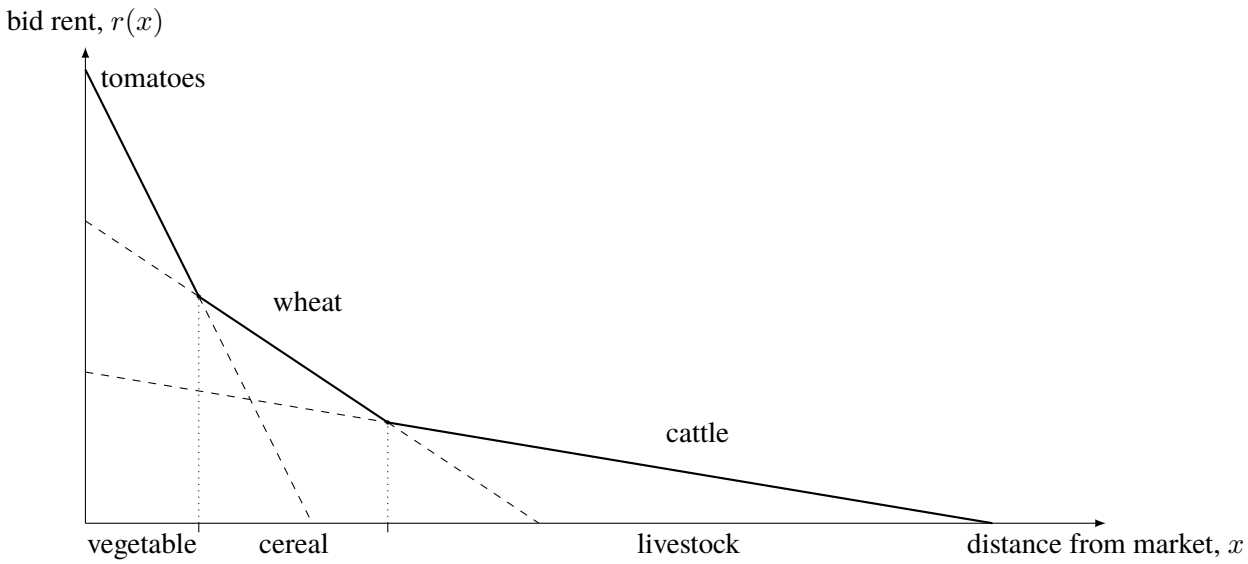

Figure 1.8: Von Thünen's agricultural bid rent curves for different products. Their outer envelope is presented in solid black.

1.3.2 Spatial equilibrium and compensating differentials: Alonso-Muth-Mills

What does land use look like in a city? Figure 1.9 shows land use patterns for Paris, from Duranton and Puga (2015). What's striking is most built up area that is not used for transportation is residential, and displays a striking pattern: It is densest in the urban core and decays in space from the downtown. This type of land use pattern is common to many cities: there is a built up urban core, often replete with mid- or high-rise buildings–a downtown or central business district–and the density of construction decays in distance to that urban core. The canonical land use model to think about these patterns is has grown out of the synthesis of work done by William Alonso (1964, *Location and land use*), Richard Muth (1961, *The spatial structure of the housing market*), and Edwin Mills (1972, *Studies in the structure of the urban economy*), sometimes referred to as simply '*the* standard urban model' (SUM) or '*the* monocentric city model.'

The model setup is as follows. There is a downtown of a city, the 'central business district' (CBD – no, not like that), where people work and shop. They can choose to live at any distance from the CBD, denoted x. This means that the CBD is located at $x = 0$. However, if they choose to live x miles away from the CBD, they must pay τx dollars in gas and transport costs (let alone Uber Eats fees).

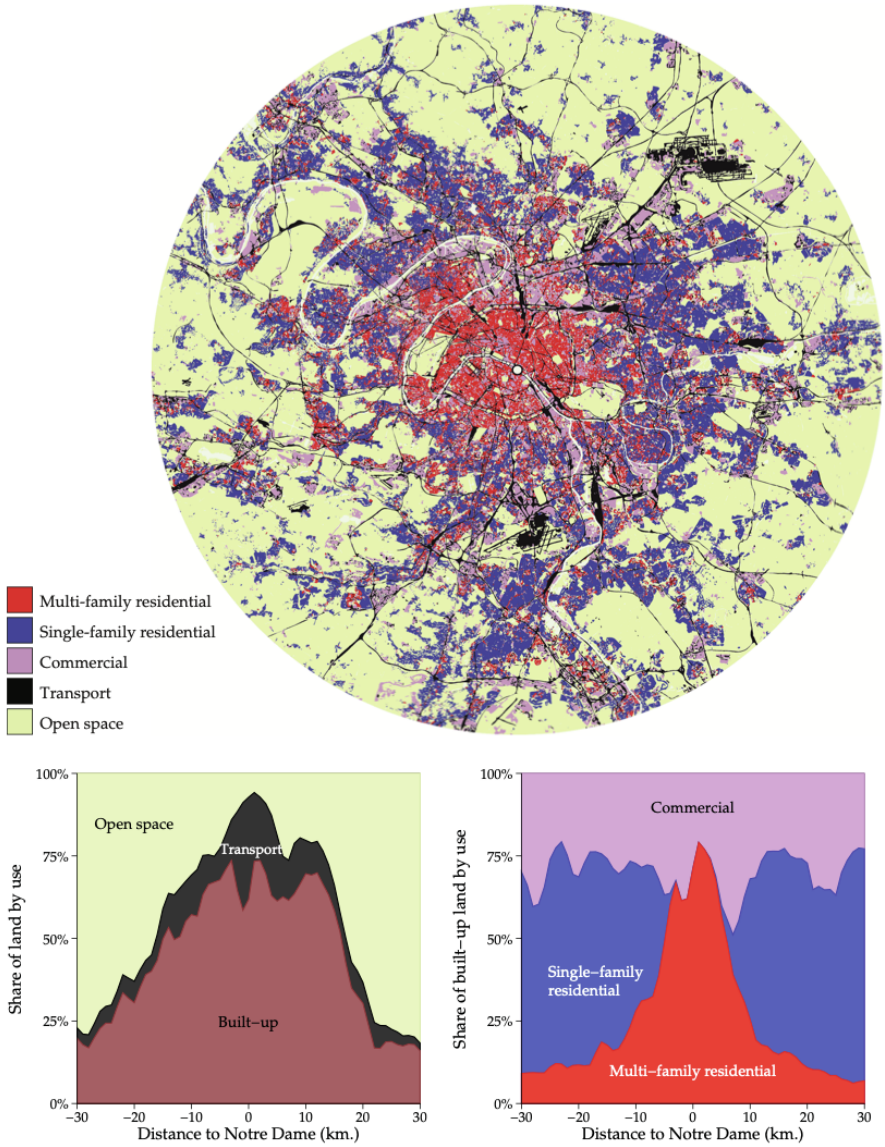

Figure 1.9: Reproduction of Figure 1 from Duranton and Puga (2015).

When they choose to live at x, they consume $h(x)$ units of housing. The price of housing is $r(x)$. Everything else they consume – all consumption goods, c – do not depend on their location, and the price of these will be 1. They are produced in the CBD. One dollar gets you one unit of non-housing consumption c.

Households maximize their utility, $U(c, h)$ subject to the budget constraint $c + r(x)h(x) = w - \tau x$. Their budget constraint varies with x – it *depends on where they choose to live*!

In a spatial equilibrium, utility is everywhere the same. Let's call this level of utility $\bar{U}$, and suppose it

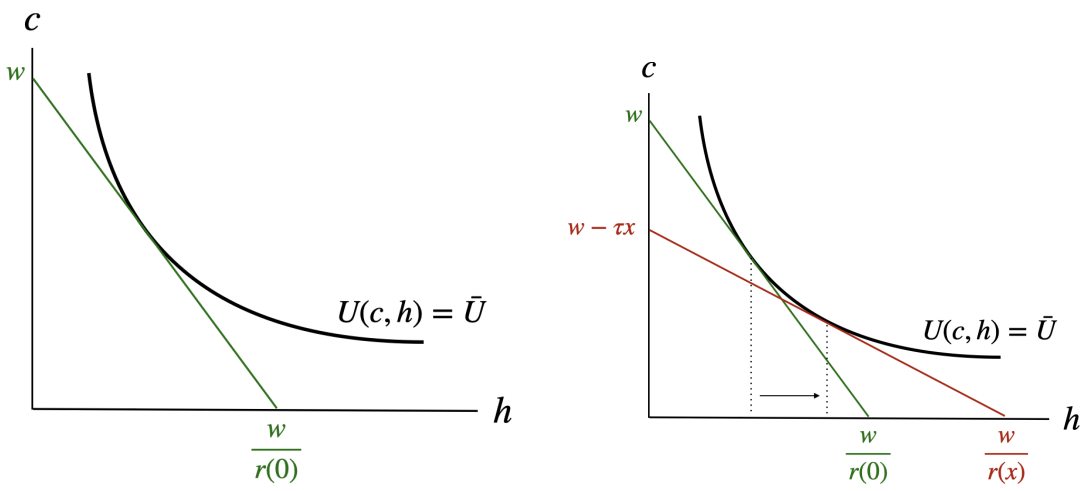

Figure 1.10: Left: Budget constraint at the CBD. Right: Utility maximization at $x = 0$ and x.

is known. Consider someone who lives in the downtown, $x = 0$. Their budget constraint is simply, $c + r(0)h(0) = w$. The price of housing must adjust to achieve spatial equilibrium, so household utility maximization looks like the left panel of Figure 1.10. In the right panel, we move farther from the downtown. A household that wants to live x miles away will have to pay more in commuting, so their income is $w - \tau x$. However, to maintain spatial equilibrium, housing supply *must adjust* to compensate them for living far away. The price of housing must *fall*! As it falls, there is a substitution effect: they will want to consume more housing.

This looks a lot like the structure of cities. Households far away from the downtown pay less in rent, and have bigger houses, too. A high price of housing and limited land incentivizes construction firms in the center to build up, explaining skyscrapers and densely populated downtowns.

What is the bid rent curve in this environment? In spatial equilibrium, it must solve,

$$\underbrace{R(x, \bar{U})}_{\text{bid rent curve}} = \max_{c,h \geq 0} r(x) \quad \text{s.t.} \quad u(c, h) = \bar{U},\ r(x)h + c = w - \tau x$$

that is, it must be the maximum rent that optimizing agents are willing to pay to hold their utility fixed at $\bar{U}$. Let $c^*(h, \bar{U})$ denote the *Hicksian* demand for consumption: the amount of consumption and agent enjoying h units of housing would need to have utility equal $\bar{U}$. Substituting this in to the above expression, we can

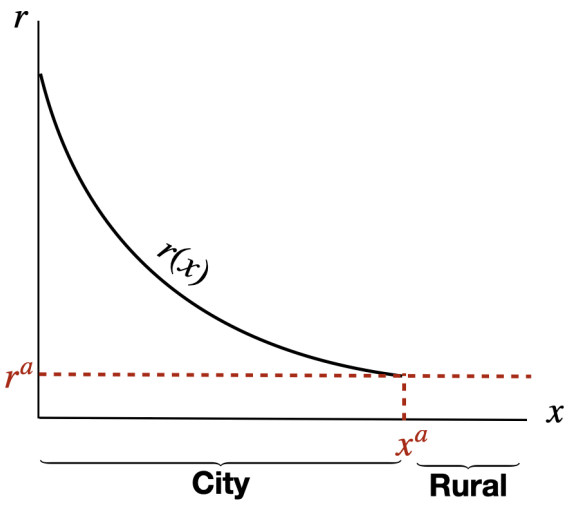

Figure 1.11: The division of land in the Alonso-Muth-Mills framework.

see that,

$$R(x, \bar{U}) = \max_h \frac{w - \tau x - c^*(h, \bar{U})}{h}.$$

Differentiating and applying the envelope theorem reveals the gradient,

$$\frac{dR(x, \bar{U})}{dx} = -\frac{\tau}{h^*(x, \bar{U})} \implies \frac{d \ln R(x, \bar{U})}{d \ln x} = \frac{-\tau x/w}{r(x, \bar{U}) h^*(x, \bar{U})/w}$$

which says that the log slope of $r(x)$ depends on the ratio of commuting cost expenditure share and the housing expenditure share.

When does the city end? As you move farther and farther from the city, to incentivize someone to still work and shop downtown, the rent has to fall dramatically. Eventually, the amount a landlord could make leasing his land for housing will fall below the opportunity cost of that land: using it for, say, farming. The rental rate of land in farming is given by r^a. We will call the fringe of the city x^a. At $r(x^a) = r^a$, a landlord is indifferent to giving up his land for urban use or farming. As shown in Figure 1.11, As we move even farther away, $x > x^a$, all land will be farming or have the housing density equal to that at the fringe of the city.

1.3.3 Spatial spillovers and internal structure: Lucas and Rossi-Hansberg (2002)

We've been talking about urban land use without really making reference to what makes cities engines of economic development. Moreover, we've been assuming the existence of a downtown. What economic forces foster the creation of 'downtowns?' We will now build a model of a city in which an urban 'core' is endogenous to the very forces that make workers in cities productive: agglomeration forces. The model we will study is from Lucas and Rossi-Hansberg (2002), which itself builds on Lucas (2001) and Fujita and Ogawa (1982).

Locations in the city are denoted in polar coordinates (x, ϕ), and we will fix the size of the city to a fixed radius R (so total area is πR^2). All land can be used for residential *or* productive purposes; production *need not take place at* $x = 0$; there is competition for land! We will study symmetric allocations so that at fixed x, all activity across $\phi \in (0, 2\pi)$ is the same.

The key advent here is the addition of a production technology. $\theta(x) \in (0, 1)$ will denote the fraction of land used at x for the production of the numeraire good that's sold to the outside economy at a competitive price. Production per unit land is,

$$y(x) = f(\underbrace{z(x)}_{\text{technology}}, \underbrace{n(x)}_{\text{labor}})$$

where technology is *endogenous*, and rises in the local density of production,

$$z(x) = \delta \int_0^R \int_0^{2\pi} \underbrace{e^{-\delta K(x,s,\phi)}}_{\text{spatial discounting}} \underbrace{s\theta(s,\phi)n(s,\phi)}_{\text{total employment}} d\phi ds$$

where

$$K(x, s, \phi) = \sqrt{x^2 - 2\cos(\phi)xs + s^2}$$

is the Euclidean (chord) distance between two points in a disk. This captures two margins of agglomeration forces: spatial spillovers (productivity at one location 'spills over' to nearby ones) and increasing returns (larger clusters of employees in space increase productivity).

Agents can choose where to live and where to work. Since are only studying symmetric (across ϕ) allocations, agents will only ever want to commute along rays. An agent living at x who works at s will pay a fraction of

their wage $1 - \tau(x, s)$ to commute, where,

$$\tau(x, s) = \exp(-\kappa|x - s|).$$

To close the model, we must fully specify preferences, firm behavior and market clearing. Agents in the model solve,

$$\max_{x,s} U(c, h) \quad \text{s.t.} \quad c + p(x)h \leq \tau(x, s)w(s)$$

There is one unit of housing per location. Firms pay workers their marginal products, and in equilibrium, there can be no spatial arbitrage (utility everywhere must be equalized). The paper specifies tractable functional forms for preferences $U(c, h)$ and technology $f(z, n)$ and the full equilibrium definition, which I'm omitting here for clarity.

The key point is that *technology* is endogenous, and depends on how workers are distributed in space. Workers can commute anywhere. What equilibrium forces do we have here? Workers want to live near high-paying employers to reduce commuting costs. Firms that are more spatially clustered are more productive, drawing in workers. However, land is in fixed supply! Workers and firms compete for land, driving up prices and incentivizing economic activity to spread out over space. Urban form is an equilibrium outcome that balances these centripetal and centrifugal forces such that all agents are optimizing, all markets are clearing, and there's not spatial arbitrage.

Urban form in this model is *self-organizing*. The left panel of Figure 1.12 shows an example of equilibrium land use patterns that can emerge as a solution to the model for particular parameters. Land use is determined by the strength of agglomeration forces: in the right panel, we see the land rent pattern generated by different values of δ. When agglomeration forces disperse slowly, activity is less concentrated in the center of the city, but when agglomeration forces are hyper local, competition for land in the center is fierce, driving rents sky-high!

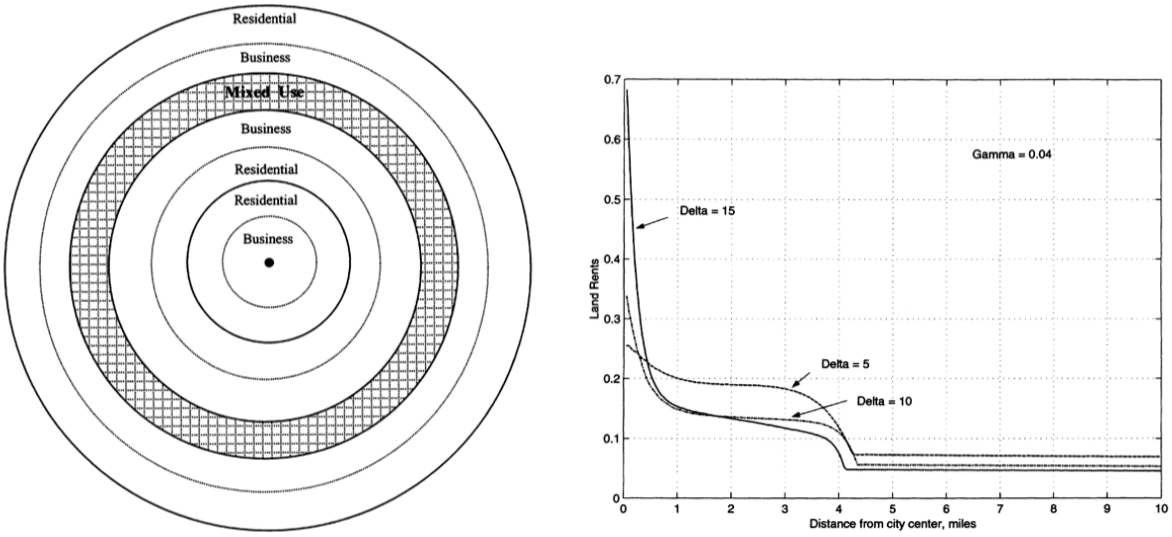

Figure 1.12: Figure 1 (left) and Figure 9 reproduced from Lucas and Rossi-Hansberg (2002).

How can we measure these agglomeration forces?

Evidence abounds that these kinds of agglomeration forces. For example, for fans of *Mad Men*, Arzaghi and Henderson (2008) study hyper-local spillovers in the advertising industry in New York. In the education literature, there is evidence that the productivity of one's college roommate 'spills over' (Sacerdote, 2001). Baum-Snow, Gendron-Carrier, et al. (2024) shows how establishment TFP shocks spill over to neighboring employers at hyper local spatial scales.

In Anisfeld and Rosenthal-Kay (2025), we tried to find an experiment in the data that would allow us to trace out spatial spillovers and identify δ. We studied casinos that opened after competition for a license to operate. We can examine how local neighborhood employment responds to a jobs boom at a treated (license-winning) site, relative to trends in employment across space in a control (almost-winning) neighborhood. The left panel of Figure 1.13 shows the design of this experiment in one city (Philadelphia), though the paper pools evidence across several cities. In the right panel, we see how employment changes across space in reaction to the jobs boom at the casino location, relative to counterfactual land use: employment expands nearby, but draws in labor from more distant locations. Using a structural model similar to Lucas and Rossi-Hansberg (2002), we show that we can infer that productivity spillovers decay rapidly: our estimates imply that doubling productivity at a given location achieves only a 10% increase in productivity at a drivetime of 5 minutes away.

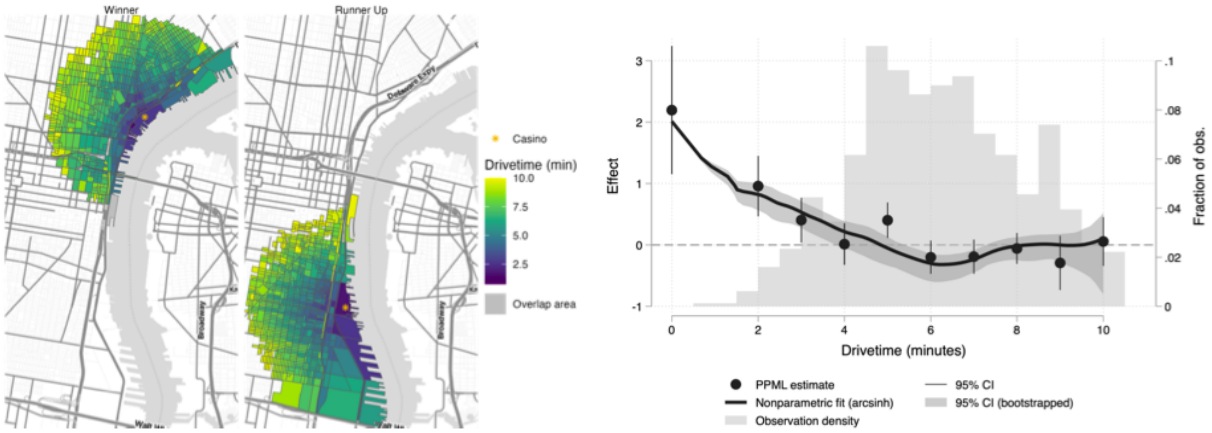

Figure 1.13: Reproduction of Figures 1 and 2 from Anisfeld and Rosenthal-Kay (2025).

References

Abowd, John M, Francis Kramarz, and David N Margolis (1999). "High wage workers and high wage firms". In: *Econometrica* 67.2, pp. 251–333.

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani (2019). "The economic effects of density: A synthesis". In: *Journal of Urban Economics* 111, pp. 93–107.

Alonso, William (1964). *Location and land use: Toward a general theory of land rent*. Harvard university press.

Anisfeld, Ari and Jordan Rosenthal-Kay (2025). "How local is local development? Evidence from casinos". In: *Regional Science and Urban Economics*, p. 104139.

Arzaghi, Mohammad and J Vernon Henderson (2008). "Networking off madison avenue". In: *The Review of Economic Studies* 75.4, pp. 1011–1038.

Baum-Snow, Nathaniel, Nicolas Gendron-Carrier, and Ronni Pavan (2024). "Local productivity spillovers". In: *American Economic Review* 114.4, pp. 1030–1069.

Baum-Snow, Nathaniel and Ronni Pavan (2012). "Understanding the city size wage gap". In: *The Review of economic studies* 79.1, pp. 88–127.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak (2014). "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh". In: *Econometrica* 82.5, pp. 1671–1748.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis (2020). "Cities in the developing world". In: *Annual Review of Economics* 12.1, pp. 273–297.

Card, David, Jesse Rothstein, and Moises Yi (2025). "Location, location, location". In: *American Economic Journal: Applied Economics* 17.1, pp. 297–336.

Chauvin, Juan Pablo et al. (2017). "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States". In: *Journal of Urban Economics* 98, pp. 17–49.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz (2016). "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment". In: *American Economic Review* 106.4, pp. 855–902.

Chyn, Eric (2018). "Moved to opportunity: The long-run effects of public housing demolition on children". In: *American Economic Review* 108.10, pp. 3028–3056.

Ciccone, Antonio and Robert E. Hall (Mar. 1996). "Productivity and the Density of Economic Activity". In: *The American Economic Review* 86.1, pp. 54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2011). "The identification of agglomeration economies". In: *Journal of economic geography* 11.2, pp. 253–266.

Datcher, Linda (1982). "Effects of community and family background on achievement". In: *The review of Economics and Statistics*, pp. 32–41.

Duranton, Gilles and Diego Puga (2004). "Micro-foundations of urban agglomeration economies". In: *Handbook of regional and urban economics*. Vol. 4. Elsevier, pp. 2063–2117.

– (2015). "Urban land use". In: *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 467–560.

Fincher, David (2010). *The Social Network*. United Staes: Sony Pictures Releasing.

Fujita, Masahisa and Hideaki Ogawa (1982). "Multiple equilibria and structural transition of non-monocentric urban configurations". In: *Regional science and urban economics* 12.2, pp. 161–196.

Glaeser, Edward, Hedi D Kallal, et al. (1992). "Growth in cities". In: *Journal of political economy* 100.6, pp. 1126–1152.

Glaeser, Edward and David C Maré (2001). "Cities and skills". In: *Journal of labor economics* 19.2, pp. 316–342.

Hall, Robert E and Charles I Jones (1999). "Why do some countries produce so much more output per worker than others?" In: *The quarterly journal of economics* 114.1, pp. 83–116.

Jacobs, Jane (1969). *The economy of cities*. Vintage.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh (2023). "The welfare effects of encouraging rural–urban migration". In: *Econometrica* 91.3, pp. 803–837.

Lin, Jeffrey (2011). "Technological adaptation, cities, and new work". In: *Review of Economics and Statistics* 93.2, pp. 554–574.

Lucas, Robert E (1988). "On the mechanics of economic development". In: *Journal of monetary economics* 22.1, pp. 3–42.

– (2001). "Externalities and cities". In: *Review of Economic Dynamics* 4.2, pp. 245–274.

Lucas, Robert E and Esteban Rossi-Hansberg (2002). "On the internal structure of cities". In: *Econometrica* 70.4, pp. 1445–1476.

Mills, Edwin S (1972). "Studies in the Structure of the Urban Economy." In.

Muth, Richard F (1961). "The spatial structure of the housing market". In: *Papers of the Regional Science Association*. Vol. 7. 1. Springer, pp. 207–220.

Reggio, Godfrey (1982). *Koyaanisqatsi*. United States: Zoetrope Studios.

Roca, Jorge De La and Diego Puga (2017). "Learning by working in big cities". In: *The Review of Economic Studies* 84.1, pp. 106–142.

Rose, Bernard (1992). *Candyman*. United States: TriStar Pictures.

Sacerdote, Bruce (2001). "Peer effects with random assignment: Results for Dartmouth roommates". In: *The Quarterly journal of economics* 116.2, pp. 681–704.

Lecture 2: Systems of cities and the 'New' economic geography

2.1 Systems of cities, 'New' economic geography

The 'new' economic geography (NEG) is no longer new: it is a tradition primarily pioneered by Paul Krugman that combines modeling tools from industrial organization and international trade to answer spatial questions. NEG was primarily developed in the 1990s, and formalized intuitions dating back to 'high development theory' of the 1950s using tools primarily developed in the 1980s. Krugman was interested in the operation of economic geography forces at a macro scale: what forces shape the population distribution across space, how do 'cities' compete with one another for population, and what determines where cities locate and how they grow? Krugman, along with friends like Masahisa Fujita and Anthony Venables, thought about these questions in a way that was mostly divorced from what more empirical urban economists were doing at the time – studying issues of urban public finance, real estate markets, and transit systems: more in the weeds of how cities worked than taking a stab at the Big Questions with the new toolkit the NEG boys cooked up.

We're going to learn the skeleton of these NEG models, but will start outside of the tradition, with Henderson (1974), which bridges urban economics with a multi-city model that uses many of the key ideas in NEG in a way that's more rooted in the tradition of urban economics.

2.1.1 "The sizes and types of cities" – Henderson (1974)

Henderson (1974) is about the sizes and types of cities. He tries to answer two questions: what determines the size of a given city, and what determines its type (what it specializes in).

Henderson develops an urban model of a system of cities indexed by $i = 1, ..., N$. In the background of his framework, you can think that there is a freely mobile population of L workers that choose in which city to live. Each city offers a different level of utility U_i depending on its size, L_i. Building on the concepts of last lecture, Henderson comes up with a utility function that looks like,

$$U_i(L_i) = \underbrace{aL_i^{\gamma_i}}_{\text{agglomeration benefits}} - \underbrace{bL_i^{\kappa}}_{\text{congestion costs}} . \tag{2.1}$$

This model contains a lot of information: cities differ in type because of differences in the scale elasticity of different goods, γ. Each good can only be produced in one location, so each location has an incentive to produce the good with the largest scale economies, leading to a hierarchy of cities sorted by the scale elasticity of their export industry. Cities with higher γ_i will be bigger in equilibrium.

Second, a spatial equilibrium occurs when,

$$\underbrace{\bar{U} = U_i(L_i)}_{\text{utility is equalized across cities}} , \qquad \underbrace{\sum_i L_i = L}_{\text{everyone lives somewhere}} .$$

In Henderson's original analysis, he takes $\bar{U}$ as given, which means there is an elastic supply of workers to the city: if $U_i(L_i) > \bar{U}$, the population of the city will increase until $U_i = \bar{U}$, and if $U_i(L_i) < \bar{U}$, people will leave the city until utility is again equalized.

With the shape of the utility function given in (2.1), we have that for any city and any level of $\bar{U}$, there are two possible equilibria (excluding the knife-edge case when $\bar{U} = \max_L U_i(L)$): one where the city is too small, and one where it is too large.

Consider the graph of $U_i(L)$ in Figure 2.1, displayed in red, with a peak around $L_1 = 0.3$. If we are at a spatial equilibrium $\bar{U}$, either L_1 is around 0.1, or around 0.6. At 0.1, the equilibrium is unstable: we are on the upwards sloping part of the curve. If someone moves there, it only gets better, incentivizing more people to move there. Around $L_1 = 0.6$, the equilibrium is stable. The marginal agent indifferent between getting $\bar{U}$ and $U_1(L)$ will make U_1 fall if they move there, and so they will want to move back to getting $\bar{U}$ elsewhere immediately.

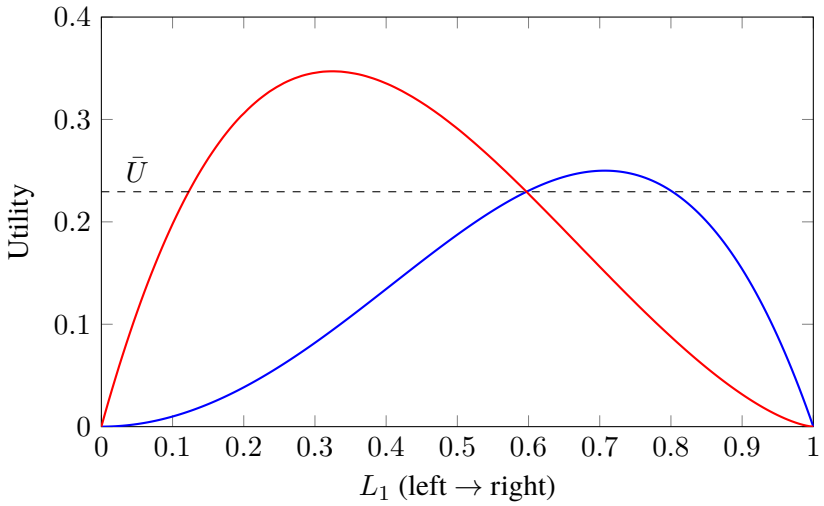

Figure 2.1: Graph of $U_1(L_1)$ in red and $U_2(1 - L_1)$ (as $L_2 = 1 - L_1$, L_2 can be read as moving from right←left) following Henderson (1974).

How is $\bar{U}$ determined? Well, everyone's got to live somewhere! Consider the case where total population is equal to 1, and there are two cities, $L_1 + L_2 = 1$. To satisfy spatial equilibrium,

$$U_1(L_1) = \bar{U},\ U_2(L_2) = \bar{U} \qquad \text{(spatial equilibrium)}$$

$$U_1(L_1) = U_2(L_2) \qquad \text{(substitution)}$$

$$U_1(L_1) = U_2(1 - L_1) \qquad \text{(population constraint)}$$

So, $U_1(L_1) = U_2(1 - L_1)$ solves for the equilibrium size of city 1, L_1, which can then be plugged into $U_1(L_1)$ to get $\bar{U}$.

This is what's going on with the blue curve in Figure 2.1. It plots $U_2(1 - L_1)$, and I assume $\gamma_1 > \gamma_2$. The spatial equilibrium $\bar{U}$ occurs when $U_1 = U_2$ on the plot, which determines the size of cities. Here, both cities are too big! Both would give a higher level of utility with fewer people. But everyone's got to live somewhere.

2.1.2 The NEG toolkit: a primer

Before we see NEG in action, we need to understand the key ingredients of the NEG toolkit. These are: (1) monopolistic competition + love-of-variety preferences; (2) iceberg trade costs (3) free mobility in multi-

region models (usually, two regions or some type of continuous space.[1]

These tools are simply that: tools. Krugman (1997) freely admits to this. These tools allow for a formalization of old ideas: how agglomeration economies work, so-called 'circular causation' (big cities are big because they are big), forwards- and backwards-linkages that generate multipliers, all hallmarks of what was once called 'high development theory.' The New Economic Geographers developed a toolkit that formalized these ideas, making them precise, legible to economists, and developing a somewhat empirical framework that allows for researchers to attempt to falsify some of its hypotheses.

Generating scale economies from monopolistic competition

In the Technical Appendix 2.2.5 I show you how monopolistic competition and 'love of variety' preferences generate scale economies. I use a model from Ethier (1982), which is a version of the Dixit and Stiglitz (1977) model on which Krugman so heavily relied.

Here's the idea of the model: Output Y is produced in a region using intermediate inputs. There are diminishing returns to any input use, but constant returns in the aggregate. There are N firms producing intermediate inputs. When N is larger, output producers use more intermediate inputs and run into decreasing returns less quickly ('love of variety'), generating scale economies, so Y is increasing in N, but producers of Y don't get to choose N. N is where the scale economy comes from.

The exact math to make this work out is a little tricky: We need fixed costs to generate scale economies, but this alone won't generate an externality in the size of a region, or the scale of production. To make these scale economies *external* to the firm – hence making them externalities – we need even more machinery: we will suppose that intermediate input suppliers are the ones paying a fixed cost, and they engage in monopolistic competition. Firms freely enter, competing away all profits. In bigger markets, there are more profits to compete away, so more firms enter.

The equilibrium is governed by two equations: $P = \mu MC$ where μ is a markup from monopolistic competition, and $P = AC = MC + F/X$ which says that under free entry, profits must be zero: price must equal

[1] These tools were combined with tools from the study of dynamic systems: bifurcation analysis, symmetry breaking, and so on. It is a really fun and underused mathematical toolkit that I will not get into in these notes.

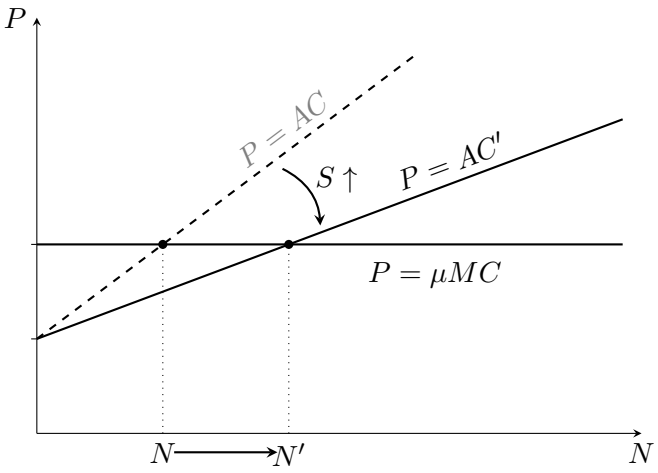

Figure 2.2: Intermediate goods equilibrium in a Dixit-Stiglitz-Krugman model.

average cost. Average cost is equal to marginal cost + the fixed cost F of production divided by output X. With scale economies, these fixed costs can be more easily spread out over sales X. If S is the overall size of a market, and there are N firms competing, then sales are $X = S/N$, and average costs $AC = MC + \frac{F}{S}N$.

This equilibrium is visualized in Figure 2.2. The free entry curve $P = AC$ slopes upwards in N: You need a higher price to support more firms when they are competing. The horizontal line is $P = \mu MC$, representing pricing under monopolistic competition. As the size of the market increases S, at a given price P, more firms N can be supported, so the $P = AC$ curve rotates in as $S \uparrow$, driving the equilibrium number of firms from N to N'. With $N \uparrow \implies Y \uparrow$, there's our scale externality.

Iceberg trade costs

Suppose a good produced in region j and sold in region i has a price p_{ji}. Trade economists have found it useful to write this as $p_{ji} = \tau_{ji} p_j$. These are *icerberg trade costs*, and they multiply the 'free on board' (fob) price p_j. This means that we assume goods transport costs are denominated units of the good being shipped: you pay the truck driver shipping Teslas from Los Angeles to Detroit in Teslas; for every Tesla meant to arrive in Detroit, you have to ship $\tau_{\text{Los Angeles}\to\text{Detroit}}$ out the door. This allows us to talk about space without modeling the transport sector: we can have regions that trade with each other that have some distance from each other. Space affects the size of τ_{ij}. This is a convenient modeling trick.

2.1.3 "Increasing returns and economic geography" – Krugman (1991)

Krugman (1991) asks, without ex-ante differences between regions, what determines the size and composition of regions? This question is very similar to that of Henderson (1974), but Krugman instead focuses on the interplay between regions, and the role transport costs play in shaping the equilibrium.

The model is 'two-by-two' in that there are two industries and two regions, but unlike 'two-by-two' classical trade models (e.g., Heckscher-Ohlin), here we will have two unique ingredients: (1) *trade costs* that link regions together with 'realistic' geography and (2) increasing returns. Our two industries are manufacturing and agriculture, and our two regions are labeled 1 and 2. No matter where the agricultural good is produced, it's the same, and it's freely traded and we'll set that price equal to 1. We'll assume that agricultural workers are immobile, tied to the land they work on. In a Krugman-like move, I'll just say total agricultural income in any region is equal to 1, as well.[2] Manufacturing goods, unlike agriculture, will be region-specific: A manufacturing good produced in region 1 will not be a perfect substitute for a manufacturing good in region 2. This is a metaphor for product differentiation and monopolistic competition, like Fords vs Teslas, and is a common ingredient in Krugman models.

The question of how and where an industrial base might develop depends on the mobility decisions of manufacturing workers. We'll normalize population in shares so manufacturing worker population so that $L_1 + L_2 = 1$. In this model, manufacturing workers enjoy real wages,

$$\omega_i = \frac{\overbrace{w_i}^{\text{nominal wage}}}{\underbrace{P_i}_{\text{local price index}}}.$$

The price index is different in different regions because there are trade costs. We will label these $\tau \geq 1$ and they are iceberg trade costs.

The final key ingredient to the Krugman model is increasing returns. Manufacturing production is subject to agglomeration economies, where,

$$y_i = Z_i L_i, \quad Z_i = L_i^{\alpha}$$

[2] Krugman is far more clever in his normalizations, often generating very nice and misleadingly transparent equations in which one parameter may play two totally different roles.

So α here is our increasing returns parameter: bigger regions are more productive because of agglomeration economies.

So, to recap: we have two regions, both producing agricultural goods. We have a population of manufacturing workers who can choose whether to live in region 1 or region 2. We have increasing returns: the bigger a region gets, the more productive it gets, generating lower prices. These increasing returns are our centripetal, agglomerative force that incentives the clustering of production. Immobile agricultural workers and trade costs generate a centrifugal, dispersive force: these can't move, want to purchase manufacturing goods, and prefer if production occurs in their region so they don't have to pay trade costs, τ. When trade costs are high, production wants to dissipate across space to meet the demand from the immobile workers. When trade costs are low, the need to spread out production evaporates, and uneven spatial development emerges: the formation of a manufacturing hub.

Why does this occur? Our key endogenous variables are population L_i and wages w_i. There's a tight link between wages and prices: prices are marginal costs, $p_i = w_i/Z_i$ and because of increasing returns, prices fall as a region gets bigger: $p_i = w_i L_i^{-\alpha}$. Price indices are an increasing function of prices, so $P_1 = P(p_1, \tau p_2)$ and $P_2 = P(\tau p_1, p_2)$. When you are in region 1, and everyone else is in region 2, your price index is worse $P_1 > P_2$ (higher prices are bad), because the region 2 good is produced at scale and you have to pay transport costs τ to get it.

How do we solve for these two variables? (1) Prices p_i must be such that supply equals demand for each region's manufacturing good ('market clearing') and (2) a spatial equilibrium must hold $\omega_1 = \omega_2$ in which everyone lives somewhere $L_1 + L_2 = 1$.

Figure 2.3 shows what the incentive to migrate to region 1, ω_1/ω_2 looks like as a function of L_1. The left panel is a reproduction of Figure 1 from Krugman (1991), while the right panel is a version I made. Computing power has come a long way since 1991! Equilibria occur when $\omega_1/\omega_2 = 1$. You can see when transport costs are high ($\tau = 4$), there is only one equilibrium: a 50-50 split of the population between the two regions. This equilibrium is stable: putting more people in region 1 makes it worse off (and region 2 better off, incentivizing migration to undo the population change). As transport costs fall ($\tau = 2$), the incentive to agglomerate starts to fight against the incentives to dissipate, and it makes sense to locate more

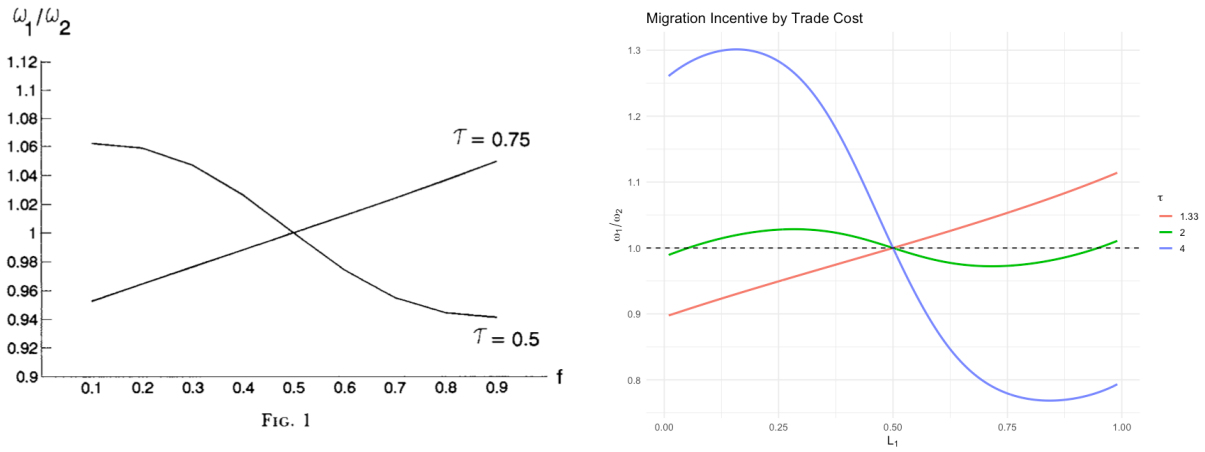

Figure 2.3: Left: Reproduction of Figure 1 from Krugman (1991). f denotes the share of the population in region 1, and Krugman's 'τ' is the inverse of mine. Right: My reproduction (slightly different parametrization) for several values of τ.

manufacturing in one region to take advantage of scale economies. In this regime, there are three equilibria: the even split, which is stable, and two unstable equilibria with L_1 near 0 or 1. As transport costs fall to a low level ($\tau = 1.33$), the even split equilibrium becomes unstable, and we are left only with equilibria in which there is total agglomeration in one of the two regions.

Another classic paper that uses this framework is Krugman and Venables (1995). The title of that paper is not 'History of the World, Part I' – it is titled, rather boringly, 'Globalization and the inequality of nations.' But 'History of the world' is what the original title was, and it is a far better title. They modify the machinery of Krugman (1991) to eliminate free mobility across regions, but allow for free mobility *between sectors* within each region. They ask, where will manufacturing end up, as transport costs fall? Is the emergence of a 'manufacturing core' and 'agricultural periphery' endogenous? It is in this sense that the paper is a history of the world: trade costs have fallen a ton ('globalization') – to what extent is the rise of international inequality – a 'Global North' versus 'Global South' – simply an emergent property of market systems with multiple equilibria?

Their analysis leads to a very famous graph – the *Tomahawk diagram*. It is simply a bifurcation diagram (for those familiar with the study of systems of differential equations). It plots the equilibria levels of L_1, and their stability, as a function of τ. I will do this for our simple Krugman (1991) model.

Figure 2.4 displays the Krugman (1991) tomahawk. When transport costs are high, there is only one equi-

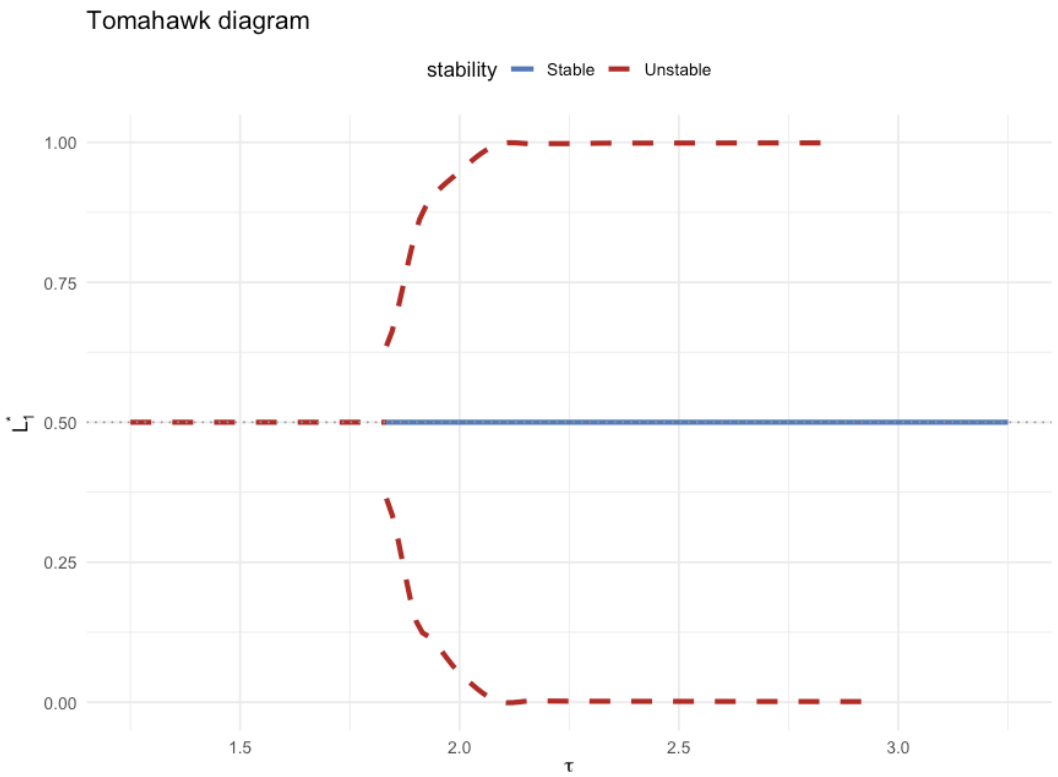

Figure 2.4: The 'Tomahawk' of Krugman and Venables (1995), using my variation of Krugman (1991).

librium, and it is stable: $L_1 = 1/2$. However, as transport costs fall, we are dramatically introduced to two additional near $\{0, 1\}$ equilibria. As transport costs continue to fall, these equilibria collapse until we are left with $L_1 = 1/2$ again as an *unstable* equilibrium.

2.2 Urban growth and the size and spatial distribution of cities

2.2.1 Zipf's law for cities

Zipf's law is that the population size distribution for cities follows a Pareto law with tail parameter $\zeta = 1$. This can be expressed through the relationship,

$$\log \text{rank}_i = r_0 - \zeta \log \text{size}_i$$

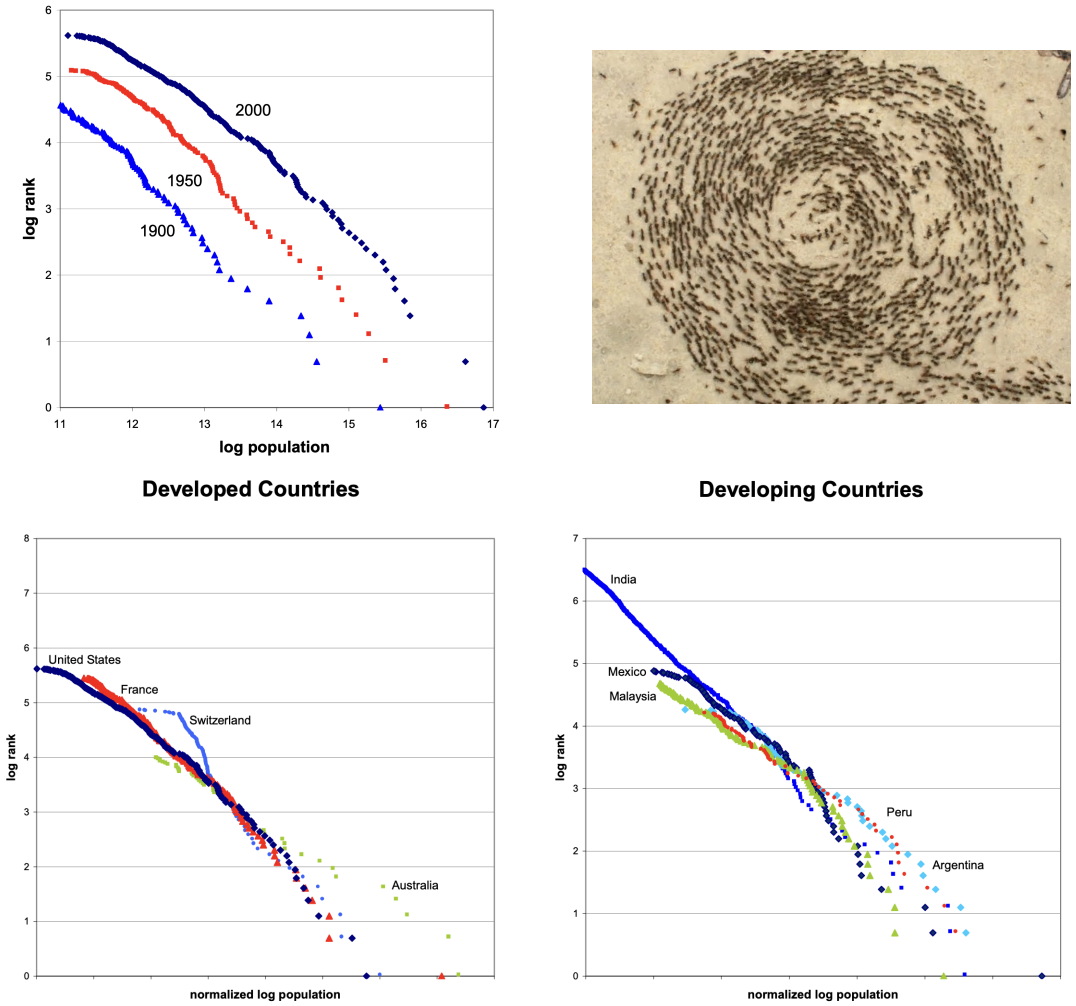

Figure 2.5: Top Left: Zipf's law for US cities, reproduced from Rossi-Hansberg and Wright (2007). Top Right: Ant Mill. Bottom panel: Zipf's law around the world, also from Rossi-Hansberg and Wright (2007).

with which ζ can be estimated via linear regression.[3] Zipf's law, that $\zeta = 1$, does not just hold today, in the U.S. but has appeared to hold across all cities within countries, and across all time. The U.S. 21st century is in the left panel of Figure 2.5.

Now, is ζ always measured as equal to one? No. It varies a bit. And does the size distribution always look perfectly Pareto? Also no. The largest city in many African nations is larger than anticipated ('urban primacy'). Including small villages–of which there are often many of equal size–can flatten the left tail. And of course it depends on measurement: what's a city, what's a nation, and so on. These are mainly excuses to feel less nervous about an eerie empirical regularity. The population size distribution over places across

[3]This is not the MLE; the MLE is called a Hill estimator.

which there is free mobility has basically looked the same, everywhere, forever. This gives some people the willies, as if we've pulled back a curtain onto the spooky forces governing society. Is the universe speaking to us? What are the implications for free will: are we really just ants spiraling in an ant mill (Figure 2.5, right panel)?

Anyway, when do you get Zipf's law for cities? You need some force to both generate a power-law city size distribution, and recover $\zeta = 1$. Let's try a Henderson-like model of city size. Suppose,

$$U_j(L_j) = A_j L_j^{-\kappa}, \quad \underbrace{A_j \sim \text{Pareto}(1, \theta)}_{\text{location fundamentals}}$$

Equilibrium is $U_j = \bar{U}$ and $\sum_j L_j = 1$. Using these, we get,

$$L_i = \frac{A_i^{1/\kappa}}{\sum_j A_j^{1/\kappa}}$$

and using Pareto,

$$L_i \sim \text{Pareto}(1, \theta\kappa)$$

This basically says, when dispersion forces are constant-elasticity, and location fundamentals are Pareto, than city sizes are Pareto. That's not much of an explanation, and rather restrictive; $\theta\kappa = 1$ is arbitrary!

Gabaix (1999) tries to answer this question using a simple dynamic, model where Zipf's law emerges mechanically. To generate Zipf's law, we need two ingredients: random growth (Gibrat's law) and a *reflecting barrier*.

Gibrat's law says that the rate of growth for cities (or firms...) is independent of its size. For example the law of motion for a city size can be,

$$S_{t+1} = \exp(g_t) S_t, \quad g_t \sim N(\mu, \sigma^2).$$

A reflecting barrier in this context simply says that $S_t > s_{min}$, and so any growth that pushes $S_{t+1} < s_{min}$

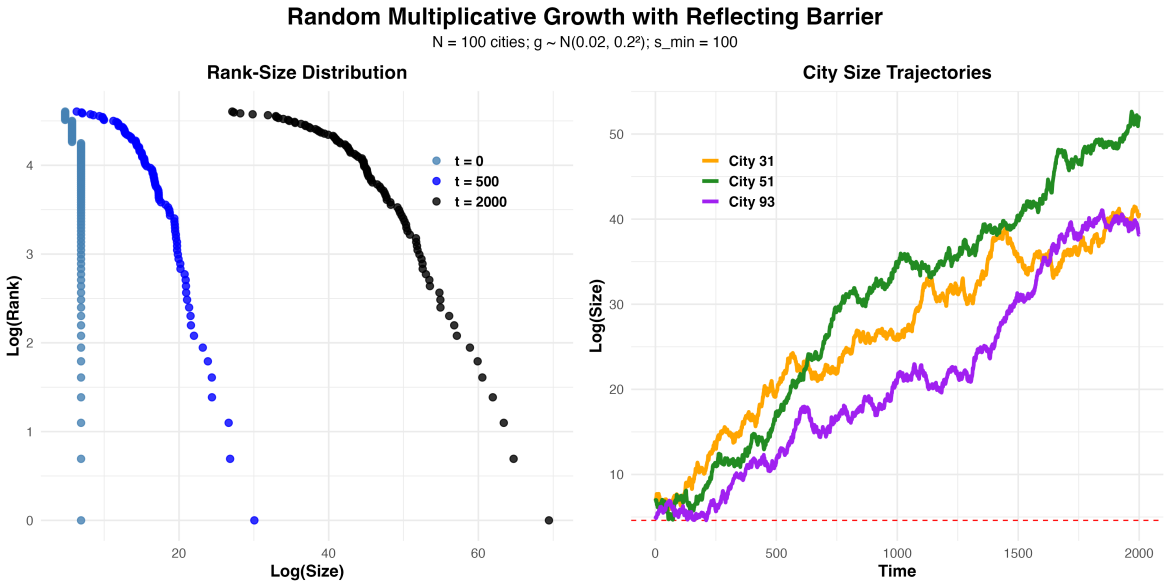

Figure 2.6: Simulating random growth – Gibrat's law with a reflecting barrier. Left: Log rank log size distribution. Right: city size time series.

is 'reflected,' e.g.,

$$S_{t+1} = \begin{cases} \tilde{S}_{t+1} & \text{if} \quad \tilde{S}_{t+1} > S_{min} \\ s_{min} + |\tilde{S}_{t+1} - s_{min}| & \text{if} \quad \tilde{S}_{t+1} \leq s_{min} \end{cases},$$

$$\tilde{S}_{t+1} = \exp(g_t) S_t.$$

Gabaix (1999) proves that in this environment, the stationary distribution of S_t converges to a power law with exponent $\zeta = 1/(1 - s_{min}/\mathbb{E}[S])$, so, when cities are defined as places with a positive population, $\zeta = 1$.

Figure 2.6 simulates this process. As you can see, it generates Zipf's law.

2.2.2 Testing economic geography theories – Davis and Weinstein (2002)

We've now developed two models that explain the city size distribution: *increasing returns*, offset by some congestion forces, and random growth, which can generate Zipf's law. Both have some time series properties: Krugman's increasing returns theory suggests concentration as transport costs fall. Random growth implies city size should follow Gibrat's law.

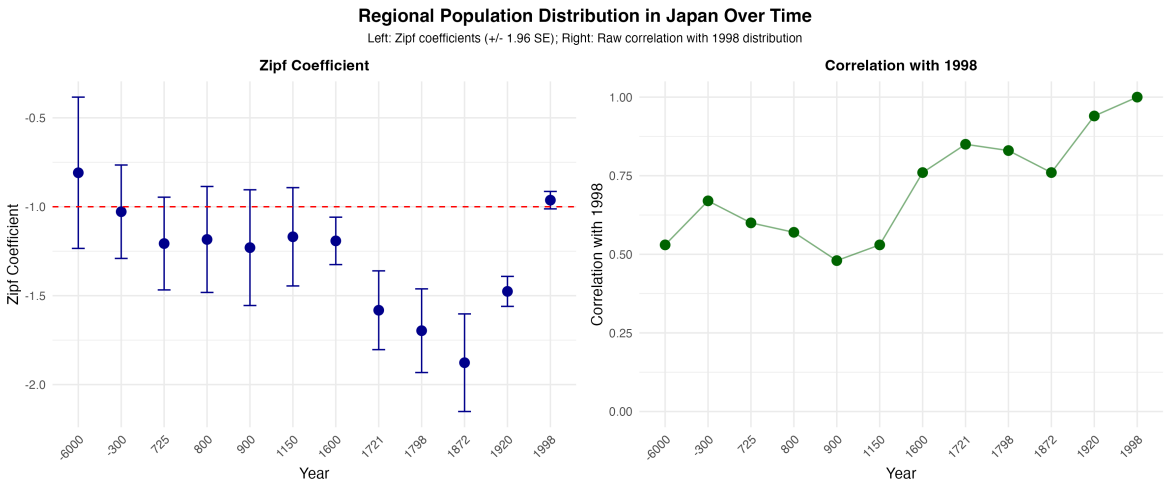

Figure 2.7: Graphical recreation of Table 1 from Davis and Weinstein (2002): Zipf coefficients and correlation with 1998 population distribution over time.

Davis and Weinstein (2002) is a classic and clever paper that tries to test various theories of what determines the city size distribution using some novel data sources. The paper is title 'Bones, Bombs, and Break Points.'

The paper introduces one more simple theory to stress test: what they call 'location fundamentals.' The idea is that *first nature* geography explains the spatial distribution of population. Simply put, some places are better than others: people tend to like coastlines and valleys and protected natural harbors.

What are the Bones? They use archaeological data going back as far as 6000 BC to try and recover statistics about the spatial distribution of population in Japan over time. They find that it accords remarkably well with Zipf's law. The Zipf' coefficient ζ for Japan is plotted over time, from the Stone Age to the modern era, in the left panel of Figure 2.7. It only deviates from during the Edo and Meiji periods. Remarkably, the spatial distribution of population in Japan predicts the population distribution today (Figure 2.7, right panel)!

This is very strong persistence. The correlation with population today remaining strong is points in favor of a location fundamentals theory. However, as transport costs fell under the Tokugawa shogunate and the industrial revolution began, and market size increased as trade with the West began, the population distribution concentrated, consistent with a Krugman-y increasing returns theory (recall our Tomahawk diagram).

How about the bombs? The left panel of 2.8 shows the population time series for Hiroshima and Nagasaki, which the United States bombed with atomic weapons on August 6th and August 9th, 1945, killing over

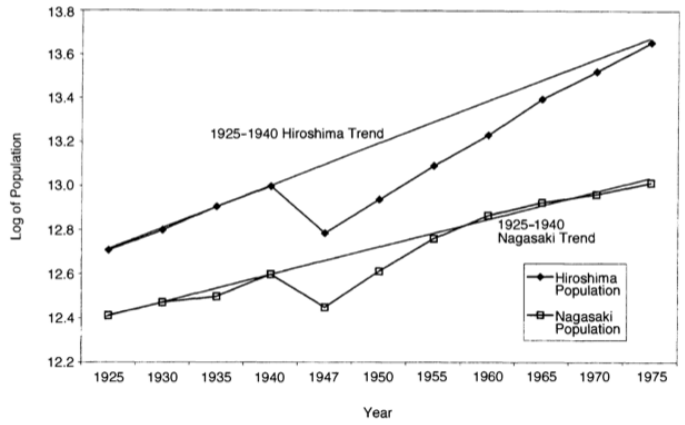

TABLE 4—MATCH BETWEEN THEORIES AND PREDICTIONS

Stylized fact	Increasing returns	Random growth	Locational fundamentals
Large variation in regional densities at all times	−	+	+
Zipf's Law	−	+	+
Rise in variation with Industrial Revolution	+	−	−
Persistence in regional densities	?	?	+
Mean reversion after temporary shocks	?	−	+

Figure 2.8: Reproductions of Figure 2 and Table 4 from Davis and Weinstein (2002).

200,000 people by the year's end, most of whom were civilians. In spite of this needless devastation, both cities rebounded rapidly in terms of population, catching up to their pre-bombing growth trajectories before the end of the century. This does not accord with random growth theory, which predicts that log city size will follow a random walk. This kind of bounceback accords with either a strong location fundamentals story, or an increasing returns story where stable equilibria bifurcate between cities either being large or small, and that the bombing was not enough to push the population of these regions past the 'break point' to another size equilibrium.

The right panel of Figure 2.8 collects the evidence and asks whether it is consistent $(+)$, inconsistent $(-)$, or doesn't make a strong prediction $(?)$, towards our three theories.

2.2.3 Physical geography and the location of U.S. cities – Bleakely and Lin (2012)

Bleakley and Lin (2012)'s paper 'Portage and Path Dependence,' brings in the idea of path dependence for explaining the size distribution of cities.[4] The basic concept is that increasing returns can 'lock in' a city size distribution explained by some historical accident.

They note that many U.S. cities today exist at former *portage* sites: places where there were barriers to water navigation that required travelers to stop, and briefly haul goods and ships over land. Such sites of portage constituted natural resting spots on commercial voyages on U.S. waterways, making for attractive spots to

[4]This idea is not new, but they examine it in data.

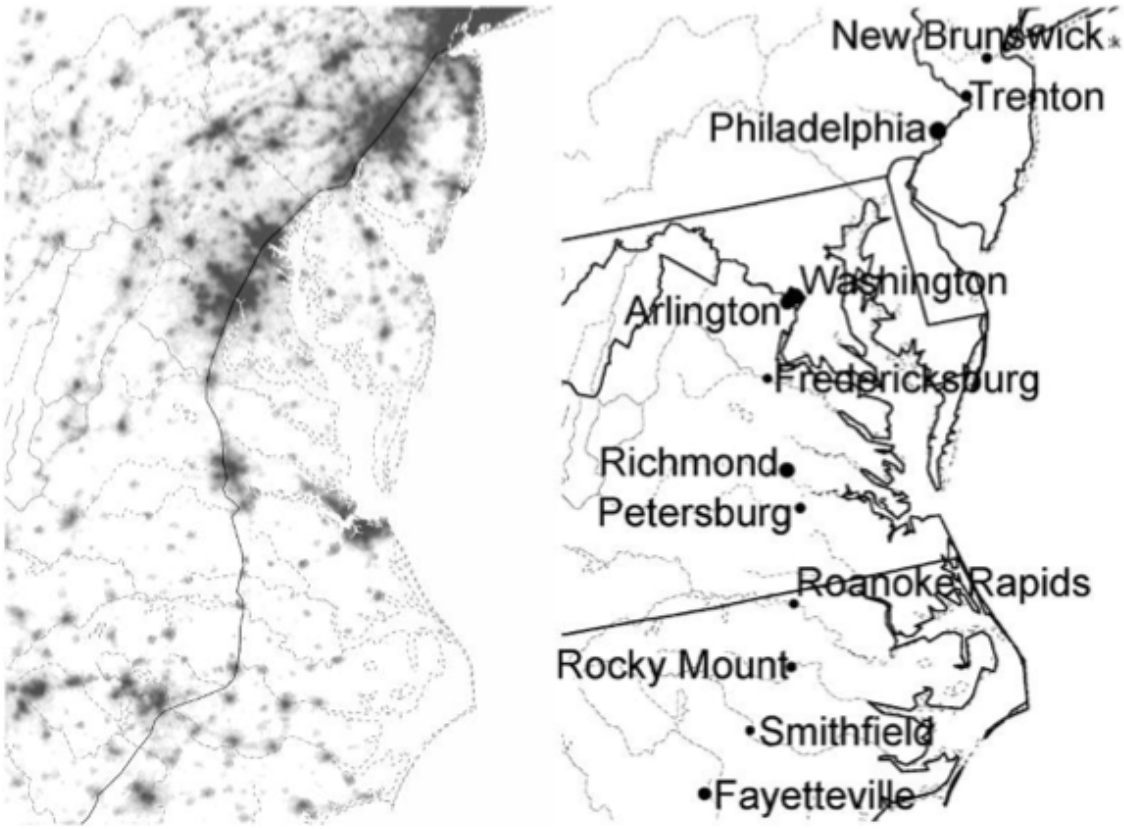

FIGURE IV
Fall-Line Cities from North Carolina to New Jersey

Figure 2.9: Reproduction of Figure 4 from Bleakley and Lin (2012).

exchange goods. They consider the case of the portage site along the falls of James River in present day Virginia. Tobacco was grown throughout Viriginia during the colonial era, and growers downriver of the falls could access the ocean to engage in trade. However, plantations upriver of the falls could not navigate down it, nor could seafaring vessels cross the falls to access these producers. Subsequently, plantations upriver sent their Tobacco on canoe, and made portage at the falls, which became a site of exchange. Trade via waterways was made obsolete by the railroad and car, making this portage site obsolete. Nonetheless, economic activity persisted at this location: it is the site of present day city of Richmond (paraphrased from Bleakley and Lin, 2012, page 588).

The Fall Line is a natural geomorphological feature where bedrock and (softer) sedimentary rock meet, which causes rivers to have sudden changes in elevation, resulting in rapids and waterfalls. Sites of portage

occurred at these locations, where the fall line met waterways. Figure 2.9 shows the fall line and the current spatial distribution of economic activity, as represented by nightlights (left) and place names (right). This Figure is basically the thesis of the whole paper: thought portage sites have no current economic relevance, they generated early agglomerations which locked in the spatial distribution of economic activity – path dependence!

The book *Nature's Metropolis* Cronon (1991) chronicles how these economic geography forces shaped the development of Chicago. The Chicago Portage, linking the Des Plains and Chicago Rivers, through the Valparaiso Moraine was used for hundreds (perhaps thousands) of years by Native Americans to connect Lake Michigan with the Mississippi. In 1673, Kaskaskia tribe members lead French explores Louis Jolliet and Jacques Marquette to the Chicago Portage to form a connection for French Fur Trade between Montreal and New Orleans; by 1755 the 'Port de Checageu' appears on French fur trading maps. In 1803, the U.S. government builds Fort Dearborn to protect the Chicago Portage for commerce. Having established the Chicago portage as *the* place for exchange and commerce on the Lake Michigan-Mississippi route, in 1848 the U.S. builds the I&M canal to link Lake Michigan to the Mississippi (this canal was replaced by the Chicago Sanitary and Ship Canal in 1900). This gave farmers access to eastern markets, allowing agriculture to expand locally. Agriculture storage technology was then developed in Chicago, to store merchandise before shipping it East or downstream the Mississippi, including grain elevators. To manage large volumes, grain elevator operators created a grain-grading system with receipts: a farmer could give a bushel of grade-A grain to the grain elevator, and receive a receipt good for *any* bushel of grade-A grain. Such receipts then could be traded for monetary equivalents, and upon these early commodities futures contracts were built. Thus, the Chicago financial industry was born, which generated financial agglomeration economies in market size and expertise. The operations of the Chicago Board of Trade today are so far removed from the Port de Checageu, but nonetheless influenced by it, making Chicago a great example of the interplay between agglomeration economies and path dependence.

2.2.4 Policy implications: agglomeration externalities and local development – Kline and Moretti (2014)

What are the implications of these increasing returns for place-based policy? If some regions are too small, can't policy help stimulate regional development by encouraging agglomeration economies? What are the costs and benefits associated with doing so? Kline and Moretti (2014b) try and answer these questions. What follows is my own variant of one of the key takeaways from Kline and Moretti (2014b) and Kline and Moretti (2014a).

Whether there's a policy rationale to encourage regional growth when there are agglomeration externalities depends on whether a benevolent social planner would do something differently than what the laissez-faire equilibrium produces. As it turns out, whether there's a role for policy depends on the nature of the agglomeration externalities.

To see this, suppose wages in city i did depend on city size $w_i = \bar{w}_i L_i^{\alpha}$, as in Chapter 1, but holding wages fixed, utility u_i fell at rate $\delta > \alpha$ in city size, and a benevolent social planner wanted to allocate all people $\bar{L}$ to different cities to maximize average utility,

$$\max_{L_i} \sum_i u_i \frac{L_i}{\bar{L}} \quad \text{subject to} \quad \sum_i L_i = \bar{L}$$

as $u_i = \bar{u}_i L_i^{\alpha-\delta}$, the first order condition is,

$$\frac{1}{\bar{L}}(1 + \alpha - \delta)\bar{u}_i L_i^{\alpha-\delta} = \lambda$$

where λ is the lagrange multiplier on $\sum_i L_i = \bar{L}$. Now, compare this to the spatial equilibrium condition:

$$u_i = \bar{U} \implies \bar{u}_i L_i^{\alpha-\delta} = \bar{U}$$

Both $\bar{U}$ and λ are pinned down by $\sum_i L_i = \bar{L}$. These conditions are the same! That means a benevolent social planner would choose the spatial equilibrium *even though there are externalities*. This is clear when

noting that,

$$\lambda = (1 + \alpha - \delta)\bar{U}/\bar{L}$$

which is the shadow value of a person – the marginal benefit to the planner of one more person to put somewhere – is equal to the marginal increase in average utility in the spatial equilibrium condition!

That's right: this is a market failure that needs no government solution. These externalities cancel out in the aggregate!

To justify place-based policy, you need α_i to be heterogeneous. Then the planner's and spatial equilibrium will differ: the planner will want to put more people into higher α_i regions.

References

Bleakley, Hoyt and Jeffrey Lin (2012). "Portage and path dependence". In: *The quarterly journal of economics* 127.2, pp. 587–644.

Cronon, William (1991). *Nature's Metropolis: Chicago and the Great West*. New York: W. W. Norton & Company.

Davis, Donald R and David E Weinstein (2002). "Bones, bombs, and break points: the geography of economic activity". In: *American economic review* 92.5, pp. 1269–1289.

Dixit, Avinash K and Joseph E Stiglitz (1977). "Monopolistic competition and optimum product diversity". In: *The American economic review* 67.3, pp. 297–308.

Ethier, Wilfred J (1982). "National and international returns to scale in the modern theory of international trade". In: *The American Economic Review* 72.3, pp. 389–405.

Gabaix, Xavier (1999). "Zipf's law for cities: an explanation". In: *The Quarterly journal of economics* 114.3, pp. 739–767.

Henderson, J Vernon (1974). "The sizes and types of cities". In: *The American Economic Review* 64.4, pp. 640–656.

Kline, Patrick and Enrico Moretti (2014a). "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority". In: *The Quarterly journal of economics* 129.1, pp. 275–331.

– (2014b). "People, places, and public policy: Some simple welfare economics of local economic development programs". In: *Annu. Rev. Econ.* 6.1, pp. 629–662.

Krugman, Paul (1991). "Increasing returns and economic geography". In: *Journal of political economy* 99.3, pp. 483–499.

– (1997). *Development, geography, and economic theory*. Vol. 6. MIT press.

Krugman, Paul and Anthony J Venables (1995). "Globalization and the Inequality of Nations". In: *The quarterly journal of economics* 110.4, pp. 857–880.

Rossi-Hansberg, Esteban and Mark LJ Wright (2007). "Urban structure and growth". In: *The Review of Economic Studies* 74.2, pp. 597–624.

Technical appendix

2.2.5 Scale externalities from monopolistic competition and free entry

Here I work out the guts of a Dixit and Stiglitz (1977)-Ethier (1982) style model used in Krugman's work.

Suppose a region produces a good Y, by aggregating the output of N intermediate input suppliers x_i with the following production function,

$$Y = \left(\int_0^N x_i^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}},$$

paying price p_i for each input. This generates demand for intermediate inputs:

$$\min_{\{x_i\}} \int_0^N p_i x_i \, di \quad \text{s.t.} \quad Y = \left(\int_0^N x_i^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}} \qquad \text{(cost minimization)}$$

$$p_i = \lambda Y^{1/\sigma} x_i^{-1/\sigma} \qquad \text{(FOC for } x_i\text{)}$$

$$x_i = \left(\frac{p_i}{\lambda} \right)^{-\sigma} Y \qquad \text{(solve for } x_i\text{)}$$

$$P \equiv \lambda = \left(\int_0^N p_i^{1-\sigma} \, di \right)^{\frac{1}{1-\sigma}} \qquad \text{(unit cost / price index)}$$

and therefore the conditional input demand is

$$x_i = \left(\frac{p_i}{P} \right)^{-\sigma} Y.$$

Now suppose each intermediate firm produces output y_i using labor

$$\ell_i = f + y_i,$$

where f is a fixed cost (in labor units). Marginal cost is the wage w, so under monopolistic competition with demand elasticity elasticity σ,

$$p_i = \mu w, \qquad \mu \equiv \frac{\sigma}{\sigma - 1}.$$

In a symmetric equilibrium $p_i = p$ and $x_i = x$ for all i, and goods-market clearing implies $x = y$.

Labor market clearing is

$$L = \int_0^N \ell_i \, di = N(f + y) = N(f + x) \quad \Rightarrow \quad x = \frac{L}{N} - f.$$

Under symmetry, final output is

$$Y = \left(\int_0^N x^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}} = \left(N x^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} = N^{\frac{\sigma}{\sigma-1}} x = N^{\frac{1}{\sigma-1}} (L - Nf).$$

We now solve for N via free entry. Firm profit is revenue minus labor costs:

$$\pi = px - w(f + x) = (p - w)x - wf.$$

Using $p = \mu w$ gives

$$\pi = (\mu - 1)wx - wf = \frac{w}{\sigma - 1}x - wf.$$

Free entry implies $\pi = 0$, hence

$$x = f(\sigma - 1).$$

Combining with $L = N(f + x)$ yields

$$L = N(f + f(\sigma - 1)) = N\sigma f \quad \Rightarrow \quad N = \frac{L}{\sigma f}.$$

Returning to our production function, we have,

$$Y = \left(\frac{L}{\sigma f}\right)^{\frac{1}{\sigma-1}} \left(L - \left(\frac{L}{\sigma f}\right) f\right) = \tilde{\sigma} \underbrace{L^{\frac{1}{\sigma-1}}}_{\text{scale externality}} \times L$$

where $\tilde{\sigma} = (\sigma f)^{\frac{1}{1-\sigma}} (1 - 1/\sigma)$

2.2.6 Complete Krugman (1991)-style model

This appendix details out the simple version of Krugman (1991) I use in the chapter text. The two key ingredients are increasing returns and transportation costs (and a little sleight-of-hand with immobile agricultural workers).

Environment There are two regions, $i \in \{1, 2\}$, and two sectors: agriculture (A) and manufacturing (M). Regions are ex-ante symmetric.

Agriculture Each region has a fixed endowment of 1 unit of the agricultural good, which is freely traded and serves as the numéraire ($p_A = 1$). Agricultural workers are immobile.

Manufacturing Manufacturing workers are mobile across regions. Let L_i denote the mass of manufacturing workers in region i, with $L_1 + L_2 = 1$. Production is subject to an agglomeration externality: a competitive firm in region i takes aggregate productivity $Z_i = a_M L_i^{\alpha}$ as given and operates a constant-returns technology $y = Z_i \cdot \ell$, where ℓ is the firm's own labor input. Because individual firms perceive constant returns, I avoid the machinery of monopolistic competition common to Krugman's work. The parameter $\alpha > 0$ governs the strength of agglomeration economies: larger regions are more productive.

Since firms perceive constant returns, competitive pricing gives

$$w_i = p_i \cdot Z_i = p_i \cdot a_M L_i^{\alpha}$$

and aggregate output in region i is

$$Y_i = Z_i \cdot L_i = a_M L_i^{1+\alpha}.$$

Because firms take the externality as given, the wage bill exhausts revenue: $w_i L_i = p_i Y_i$; there are no profits.

Preferences Consumers have Cobb-Douglas preferences over agriculture and a manufacturing composite, spending a fraction μ on manufactures and $1 - \mu$ on agriculture. The manufacturing composite aggregates goods from the two regions with a CES ('Armington') structure,

$$U(C^A, C^M) = (C_i^A)^{1-\mu}(C^M)^{\mu}, \quad C_i^M = \left(c_{ii}^{\frac{\sigma-1}{\sigma}} + c_{ji}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1$$

where c_{ii} is consumption of the local manufacturing good and c_{ji} is consumption of the import.

Trade costs / geography Shipping goods between regions is subject to iceberg trade costs: to deliver one unit from region j to region $i \neq j$, one must ship $\tau \geq 1$ units. The delivered price of region j's good in region i is therefore τp_j.

Price indices and demand The CES manufacturing price index in region i is

$$P_i = \left(p_i^{1-\sigma} + (\tau p_j)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

Total income in region i is the sum of agricultural endowment income and the manufacturing wage bill,[5]

$$I_i = 1 + w_i L_i.$$

Consumers spend μI_i on the manufacturing composite. CES demand for each origin's good, in quantity received, is

$$c_{ii} = \frac{\mu I_i}{P_i}\left(\frac{P_i}{p_i}\right)^{\sigma}, \qquad c_{ji} = \frac{\mu I_i}{P_i}\left(\frac{P_i}{\tau p_j}\right)^{\sigma}.$$

Market clearing Because τ units must be shipped for each unit received, market clearing for region i's manufacturing good requires that production cover both local consumption and the gross quantity shipped abroad:

$$Y_i = c_{ii} + \tau \cdot c_{ij}$$

where c_{ij} is the quantity of region i's good received in region j. Substituting the demand expressions, the market-clearing conditions for $i = 1, 2$ are

$$p_1 Y_1 = \left(\frac{p_1}{P_1}\right)^{1-\sigma} \mu I_1 + \left(\frac{\tau p_1}{P_2}\right)^{1-\sigma} \mu I_2$$
$$p_2 Y_2 = \left(\frac{p_2}{P_2}\right)^{1-\sigma} \mu I_2 + \left(\frac{\tau p_2}{P_1}\right)^{1-\sigma} \mu I_1.$$

These two equations jointly determine the equilibrium manufacturing prices (p_1, p_2) for any given allocation of workers (L_1, L_2).

Migration The question of how and where an industrial base might develop depends on the mobility decisions of households. Workers have real wages

$$\omega_i = \frac{w_i}{P_i^{\mu}} = \frac{p_i \cdot a_M L_i^{\alpha}}{P_i^{\mu}}$$

where P_i^{μ} is the cost-of-living index under the Cobb-Douglas upper tier.[6] Workers migrate toward the region offering higher real wages. An interior equilibrium with $L_1 \in (0, 1)$ requires $\omega_1 = \omega_2$, while a corner solution with all workers in one region is an equilibrium if the concentrated region offers weakly higher real

[5] Since $w_i L_i = p_i Y_i$, I can equivalently write $I_i = 1 + p_i Y_i = 1 + p_i a_M L_i^{1+\alpha}$.

[6] The agricultural good is the numéraire with price 1, so only the manufacturing price index enters with weight μ.

wages.

Computing the ω_1/ω_2 How do you actually solve this thing? Start with a value of $L_1 \in (0, 1)$ – this implies L_2. With those values, solve the market clearing equations (supply = demand) for equilibrium prices, compute ω_1, ω_2. Equilibria occur when $\omega_1 = \omega_2$.

Lecture 3: Quantitative spatial economics

Let's recap. We've now developed theory of what cities are and how agglomeration economies and urban costs (congestion forces: housing prices and commuting costs) may operate within them, moving from Alonso-Muth-Mills to Lucas and Rossi-Hansberg (2002). We then developed a system-of-cities theory where at a high level, cities were a concept that captured the trade-off between the agglomeration and congestion forces, by way of Henderson (1974). We then asked how geography mattered for determining the spatial distribution of economic activity through a stylized two-sector, two-region model with Krugman (1991).

This is a rich set of ingredients for understanding cities and the spatial distribution of economic activity. At a high-level, I've emphasized how economic activity is linked across space through both direct spatial forces–interregional trade and mobility–and through the spatial equilibrium concept. Thus, these models caution us on how to interpret spatial data, and how to make sense of particular data patterns at a high level. Yet, what we still lack is a *quantitative* theory, something we can take to data and use as both an interpretation and measurement device – how to make sense of regressions using spatial variation, or how to measure things like agglomeration forces – as well as a tool to answer policy-relevant questions by way of counterfactual analysis.

To understand what I mean by this, consider a question like understanding the macroeconomic costs of hurricanes. We can examine, in the data, the effect of hurricanes on economic outcomes by comparing, e.g., regions that were and were not affected by hurricanes. Yet, the spatial theory we've developed tells us this won't capture the true aggregate effect: hurricanes may cause migration, affect local capital stocks and regional trade, and diminish local agglomeration economies. Such effects will 'spill over' to unaffected regions.

We want our theories to work beyond critique: in an interlinked spatial economy, what's the *right* regression to run? What spillovers should we care about, and how should we measure them? How can we interpret those effects and 'scale them up' to understand the macroeconomic impact?

Answering these types of quantitative, general equilibrium questions is the key goal of quantitative spatial economics, which builds 'quantitative spatial models' (QSMs) of cities, regions, and countries. Like New Economic Geography, Quantitative Spatial Economics borrows tools from urban economics, industrial organization, and international trade to build such quantitative models.

3.1 Quantitative spatial models

What are quantitative spatial models? They are general equilibrium models that can easily fit to spatial data. One way to think of them is that they are machines: they take in parameters estimated from data and return prices and quantities. They allow us to do *counterfactual* analysis: if a parameter changed, what would happen to prices? The spatial distribution of population? GDP?

The general equilibrium part is that prices and quantities are *endogenous*: they are determined in some equilibrium. A model consists of behavioral equations (e.g., consumers' demand slopes down, producers' supply slopes up) and adding up rules (supply must equal demand). Combining these defines an equilibrium and gives some system of equations that can be solved on a computer.

The building blocks of most quantitative spatial models is the spatial equilibrium framework referred to as the *Rosen-Roback model* (Roback, 1982, is the canonical reference). Roback's 1982 model had firms, and Rosen carefully thought about hedonics, and what the theory does is pin down determination of wages and rents to identify amenity values. We're going to ignore all that and take a simple, partial equilibrium approach to fix ideas. The basic idea is that money-metric utility of living in some place i, V_i, can be written as,

$$V_i = \underbrace{A_i}_{\text{amenities}} \times \frac{\overbrace{w_i}^{\text{nominal wage}}}{\underbrace{P_i}_{\text{local price index}}}, \qquad \underbrace{V_i = V}_{\text{spatial equilibrium}}$$

so that w_i/P_i is the real wage. This should look fairly similar to the manufacturing workers' location choice

Place	Mechanism	Implication
Anchorage, AK	bad weather $A_i \downarrow$ high prices (costly to import) $P_i \uparrow$	$\Rightarrow$ wages $w_i \uparrow$ (about 10% higher than nat'l average)
New York, NY	great cultural amenities $A_i \uparrow$ productive industries $w_i \uparrow$	$\Rightarrow$ prices $P_i \uparrow$ (rents greater than 2x nat'l average)
Honolulu, HI	average wages $w_i -$ high cost of living $P_i \uparrow$	$\Rightarrow$ infer there are great amenities $A_i \uparrow$ (fact check: true)

Table 3.1: The logic of spatial equilibrium

problem in Krugman (1991), except we've augmented it with amenities, A_i. Amenities are non-market goods like good weather that make some locations more attractive than others. The spatial equilibrium condition is that utility must be everywhere the same. If people are free to move, then there cannot be any spatial arbitrage in equilibrium. Consider the logic of this theory as summarized by Table 3.1. It provides a good description of the world: places with poor amenities and high costs of living must have high wages to compensate, as in Anchorage, Alaska. To maintain spatial equilibrium, the cost of living must adjust upwards for productive, high amenity places like New York City. Amenities in the theory can act as unmeasured compensating differentials, too: consider Honolulu, Hawai'i. Wages are average there, but the cost of living is quite high. People must be paying for something: it surely offers a great quality of life!

While this theory is implausible if taken literally, it provides a useful way to characterize the world. What quantitative models do is *endogenize* the prices w_i and P_i by specifying supply-and-demand in the goods and labor market, and solve for the population distribution across places such that spatial equilibrium holds. The main model we'll work with will have some similarities to Krugman (1991), but we will eliminate the agricultural workers and, for now, the agglomeration forces. However, we will expand to a much more realistic geography of $N > 2$ locations. Expanding the geography will be fruitful for thinking about data and what we learn from empirical work that leverages spatial variation.

3.1.1 Quantitative economic geography – Allen & Arkolakis (2014)

Allen and Arkolakis (2014) present a workhorse economic geography model that is easy to modify, take to the data, and solve on a computer. It's a model that's useful for both thinking through economic logic

('supply and demand... in space!' Allen and Arkolakis, 2023), as well as computing general equilibrium effects associated with some policy.

In the model there are N regions, labeled i or j. There's a mass of agents L who can choose to live anywhere. If they choose to live in i, their utility is V_i,

$$V_i = \underbrace{A_i}_{\text{amenities}} \times \underbrace{\frac{w_i}{P_i}}_{\text{real wage}},$$

they earn the wage w_i by providing labor to i's representative firm, and use that wage to finance the consumption of goods from every location by solving,

$$\max_{c_1,...,c_n} U(c_1, ..., c_n) \quad \text{s.t.} \quad \sum_j p_{ji} c_j \leq w_i$$

the term p_{ji} is the price i pays for a good produced in j. So, this is kind of a trade model, and goods flow $j \to i$. P_i is an *ideal price index*, which is the cost of an optimal bundle of c_j and is a function of p_{ji}.

Let's pause for a second and remind ourselves what an ideal price index is. It's the dollar cost of 'real consumption,' i.e., the dollar price of consumption at some given level of utility $\bar{u}$. This is the expenditure function in demand theory: $P(\bar{u}) = e(p, \bar{u}) = \min_c\{p \cdot c : U(c) \geq \bar{u}\}$. Under homotheticity (i.e., relative demand c_i/c_j depends only on prices, and not the level of utility), $e(p, \bar{u}) = e(p)\bar{u}$. Under homotheticity, when optimizing, $\sum_j p_{ji} c_j = P_i C_i$ where $C_i = U(c_1, ..., c_N)$ is real consumption.

Geography, in this model, shows up inside the price index through transportation costs. Instead of modeling a transport sector that moves goods across space, we make a convenient modeling trick: we assume *iceberg trade costs*, τ_{ji}. These are the same iceberg trade costs we saw in Lecture 2.1.3. This means that we assume goods transport costs are denominated units of the good being shipped: you pay the captain of the ship that takes iPhones from Shenzhen to Seattle in iPhones; for every iPhone meant to arrive in Seattle, you have to ship $\tau_{\text{Shenzhen}\to\text{Seattle}}$ out the door. Therefore $p_{ji} = \tau_{ji} p_j$, where p_j is the price j receives.[1]

[1] This is sometimes called the f.o.b. (free on board) price.

In the original paper, and most quantitative work, $U()$ is a constant elasticity of substitution (CES) function,

$$U(c_1, ..., c_N) = \left(\sum_{j=1}^{N} (c_j)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \implies P_i = \left(\sum_j (\tau_{ji} p_j)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

This function is extremely useful in quantitative modeling, has a rich history, and is not the point of the lecture. What we care about is that with it, we can define P_i. The functional form for P_i captures that the 'cost of living' is different in different places: it depends how accessible cheap goods are in a given location; that is, it captures location i's *market access*.

The representative firm in each i makes a good with the technology,

$$\underbrace{Y_i}_{\text{location } i\text{'s good}} = \underbrace{Z_i}_{\text{productivity}} \times \underbrace{L_i}_{\text{population of } i}$$

The goods are sold at price p_i. Wages reflect marginal products of labor, $w_i = p_i Z_i$.

So, what's endogenous in this model? We need to determine (1) where everyone lives L_i (2) the prices of each good p_i. We also need to determine the average level of utility in the model, V. In equilibrium,

1. $V_i = V$ for all i (spatial equilibrium condition)
2. $\sum_i L_i = L$ (everyone has to live somewhere)
3. $p_i y_i = \sum_j p_{ij} c_j$ (total revenue comes from sales – 'market clearing')

The market clearing one is the trickiest of these: this is a supply = demand condition,

$$\underbrace{p_i y_i}_{\text{region } i\text{'s supply}} = \sum_j \underbrace{p_{ij} c_j}_{\text{region } j\text{'s demand for goods from } i}$$

Now, using the CES demand system, in this model, it turns out this is,

$$p_i y_i = \sum_j \left(\frac{\tau_{ij} p_i}{P_j} \right)^{1-\sigma} p_j y_j$$

This can be rearranged by dividing $p_i^{1-\sigma}$, substituting for y_i and dividing, and then taking logs,

$$\ln p_i = \underbrace{-\frac{1}{\sigma}\left(\ln Z_i + \ln L_i\right)}_{\text{supply effect}} + \underbrace{\ln MA_i}_{\substack{\text{market access /} \\ \text{demand effect}}}$$

The supply effect term says that prices are lower for bigger or more productive regions. This is because these shift the supply curve out; we move down the demand curve for region i's good! The other term is a complicated object called region i's market access, MA_i and it depends on the prices everywhere else, weighted by trade costs τ_{ji}.[2] When region i has good market access, it is near rich consumers, and that is equivalent to shifting out the demand curve for region i's good: we move up the supply curve.

Despite the algebra we can do with them, these equations need to be solved in a computer. So, why bother with this at all?

Let's suppose you have some shock, D_i that affects a region's productivity: for example, D_i might represent a dummy for whether a region was hit by a hurricane. Suppose hurricanes affect productivity,

$$\ln Z_i = \zeta D_i + z_i^0$$

Suppose you had data on wages, and wanted to know the effect of a hurricane on wages. You run a regression,

$$\ln w_i = \beta_0 + \beta_1 D_i + u_i$$

What does our model say β_1 picks up?

$$\begin{aligned}
\ln w_i &= \ln Z_i + \ln p_i && \text{(wages are marginal products)} \\
&= \zeta D_i + z_i^0 + \ln p_i && \text{(definition of } Z_i\text{)} \\
&= \zeta D_i + z_i^0 - \frac{1}{\sigma}(\ln Z_i + \ln L_i) + \ln MA_i && \text{(solution for prices)} \\
&= (1 - 1/\sigma)\zeta D_i + z_i^0 - \frac{1}{\sigma}(\ln L_i) + \ln MA_i && \text{(definition of } Z_i\text{)}
\end{aligned}$$

[2] Formally, it is,

$$\ln MA_i = \frac{1}{\sigma} \ln \left(\sum_j \tau_{ij}^{1-\sigma} \frac{p_j y_j}{P_j^{1-\sigma}} \right)$$

here.

Compare this to our regression,

$$\ln w_i = \underbrace{(1 - 1/\sigma)\zeta}_{\beta_1} D_i + \underbrace{z_i^0 - \frac{1}{\sigma}(\ln L_i) + \ln MA_i}_{\beta_0 + u_i}$$

What does this tell us about identification? It tells us that the basic regression does not identify the causal effect of D_i on Z_i:

1. We need to control for population, L_i. If D_i affects L_i, that could affect the wage, because it would move us along the labor demand curve!

2. We need to control for market access, MA_i. If the shock is spatially correlated, it may also affect neighboring areas, and affect wages by affecting prices through trade and market access!

3. We need to argue for random assignment, $D_i \perp z_i^0$. If regions that get hit by the hurricane $D_i = 1$ are less productive than average, then $Cov(D_i, z_i^0) < 0$, which would bias β_1 downwards.

What does this tell us about the macroeconomic effects of the hurricane? That is, does β_1 tell us the aggregate impact of the hurricane? Certainly not! If we knew σ, we could use our regression to identify ζ, and then imagine a world without a hurricane: that would allow us to benchmark the costs of the hurricane, incorporating spillovers.

The point here is that, (1) we can use this kind of quantitative Rosen-Roback framework to think through what we learn when we use geographic data, and how to identify the parameters we want to identify, and (2) how to go from micro estimates to macro effects, incorporating spillovers.

So, let's do that.

We can control for population L_i and proxy for MA_i fine, but random assignment is hard! Hurricanes often hit coastlines, which are at baseline more productive regions on average, thus $Cov(D_i, z_i^0) > 0$. What's the solution? Panel data and fixed effects. Suppose we observed two time periods $t = 0, 1$, and we believed that amenities and productivities A_{it} and Z_{it} stochastically varied in time. Suppose we estimated,

$$\ln w_{it} = \beta_1 D_{it} + \beta_2 \ln L_i + \beta_3 \ln \widetilde{MA}_{it} + \xi_i + \chi_t + v_{it}$$

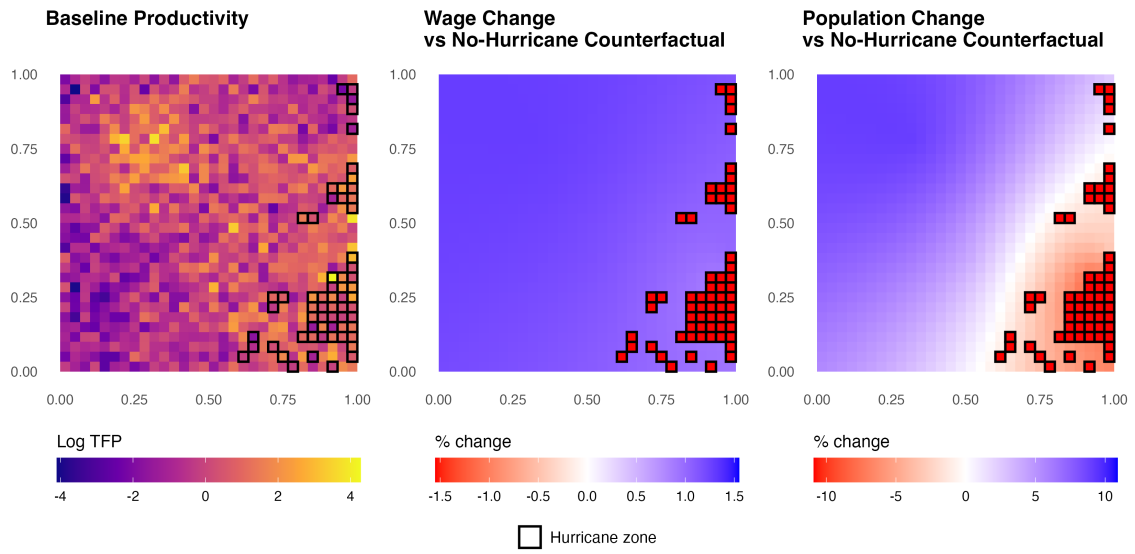

Figure 3.1: Fake data

where at $t = 0$ is before the pre-period, and $D_{it} = 1$ when $t = 1$ and i is treated. This is a difference-in-differences experiment. The unit fixed effect ξ_i absorbs time-invariant features of Z_i—i.e, z_i^0—while χ_t absorbs common growth trends across units. Then it might be fair to assume,

$$D_{it} \perp v_{it} \mid \ln L_{it}, \ln MA_{it}, \xi_i, \chi_t.$$

This is random assignment: it says that once we condition on unit fixed effects and the other endogenous regressors (population and market access), that D_{it} is as good as random. In other words, that random changes in Z_{it} from $t = 0$ to $t = 1$ are not correlated with D_{it} once we net out z_i^0 through ξ_i.

Let's replicate this experiment in fake data. The left panel of Figure 3.1 shows the distribution of $\ln Z_i$ at $t = 0$ in a simulated economy. Each cell is a location, and they are colored by the Z_{i0} value. Black cells are the ones hit by the imaginary hurricane. You can see that they are on the 'eastern seaboard' which is more productive at baseline, baking in $Cov(D_{it}, z_i^0) > 0$.

What happens when we estimate our regression models?

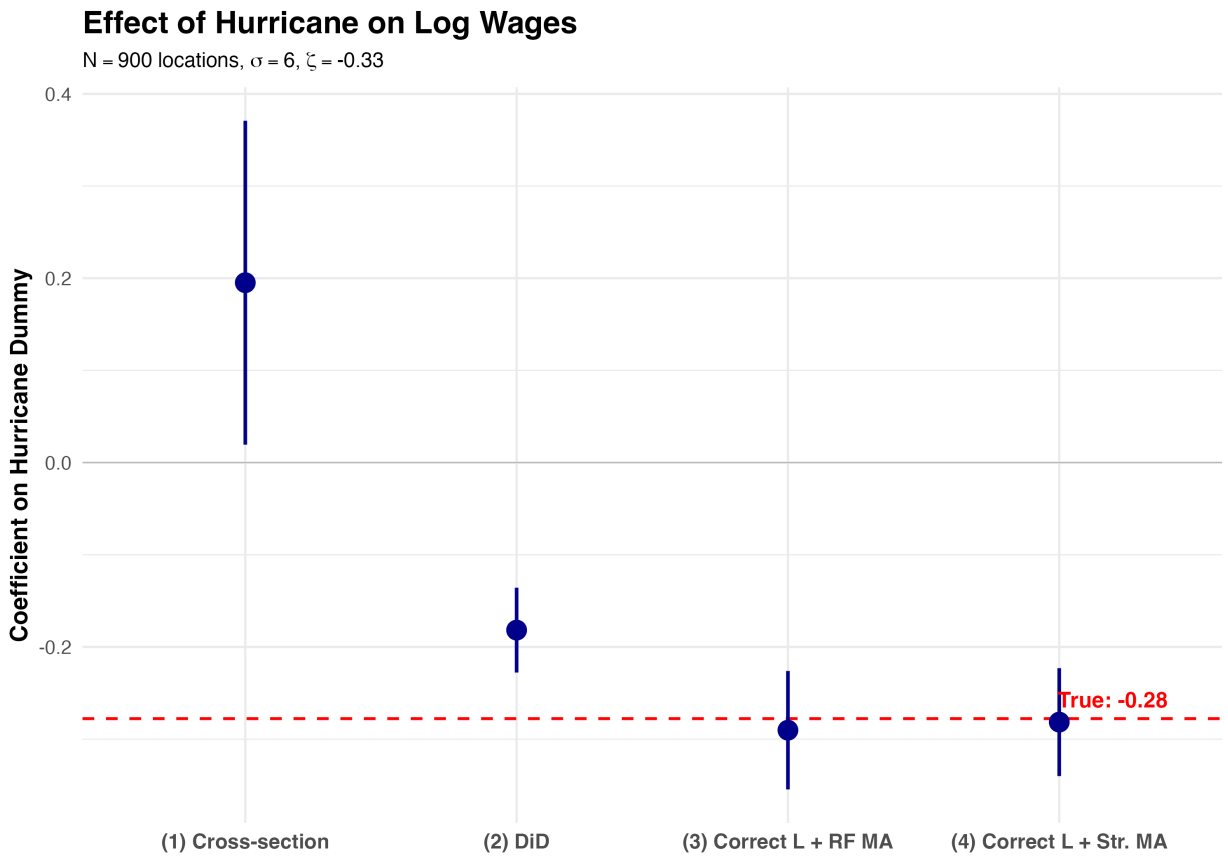

Figure 3.2: Regressions in fake data

In Figure 3.2, I first consider the following regressions,

$$\ln w_{i1} = \beta_1^{CS} D_{i1} + u_{i1} \qquad \text{(1) - Cross section}$$

$$\ln w_{it} = \beta_1^{DID} D_{it} + \xi_i + \chi_t + v_{it} \qquad \text{(2) - DID}$$

$$\ln w_{it} = \beta_1^{MRF} D_{it} + \beta_2 \ln L_{it} + \beta_3 \ln \left(\sum_j \text{dist}_{ij}^{-1} D_{jt} \right) + \xi_i + \chi_t + v_{it} \qquad \text{(3) - Model, MA proxy}$$

$$\ln w_{it} = \beta_1^{MRF} D_{it} + \beta_2 \ln L_{it} + \beta_3 \ln MA_{it} + \xi_i + \chi_t + v_{it} \qquad \text{(4) - True MA control}$$

You can see in the figure that the cross sectional regression is bad and gets the sign wrong because the hurricane is not randomly assigned. The DID estimate gets the sign right, but does not identify the effect of the hurricane on productivity – it's a 'reduced form' object that muddles together three channels: the direct productivity effect (negative), the population reallocation effect via the labor demand curve (positive), and

the market access effect from spatially correlated shocks (ambiguous sign).

Controlling for population and spillovers correctly nails the effect. As it turns out, to control for spillovers, it appears in this kind of model I just need to control for whether nearby regions are shocked, and don't need to control for the correct notion of market access, which requires knowing τ_{ij}, P_i, σ, and so on.

Now, can we simply just take the 'true' treatment effect and scale it up to get the macroeconomic effect of the hurricane? Absolutely not! As Figure 3.1 demonstrates: there are spillovers: wages rise in some places, and the population adapts to the hurricane shock by reallocating across space. We need to account for these *general equilibrium* effects to understand the macroeconomic impact.

To do so, we need to simulate the counterfactual economy in which the hurricane never occurred. This requires actually estimating A_i, Z_i, τ_{ij} and solving for what population and wages would have looked like if not for the shock. So, now a brief diversion.

How do you estimate this kind of model?

Let's think about the model's parameters. We have,

1. A_i amenities, in every location, so N of these parameters
2. Z_i productivities, in every location, so N again,
3. τ_{ij} trade costs, so $N(N-1)$ of these ($\tau_{ii} = 1$), cut that in half if we assume symmetry $\tau_{ij} = \tau_{ji}$,
4. and the parameter σ.

That's a lot of parameters! What data might be useful to estimate them all?

1. We know L_i, the population distribution across space (N observations)
2. w_i nominal wages across space (N)

These will prove useful to get A_i and Z_i. How about τ_{ij}?

There are two directions to take on the τ_{ij}. You can model them as a function of observables like distance – better when you only have some data, or noisy data, or very granular data with lots of zeros – or try to fit them to exactly match trade flows. Let's do it with data. For this, we must rely on the model's *gravity equation*. Using the CES demand system, trade flows—the dollar amount of goods purchased in j from i, X_{ij}—have a gravity form,

$$X_{ij} = \left(\frac{\tau_{ij} p_i}{P_j}\right)^{1-\sigma} w_j L_j$$

Taking logs and using fixed effects,

$$\ln X_{ij} = -(\sigma - 1) \ln \tau_{ij} + \xi_i + \chi_j$$

If we model $\tau_{ij} = \text{distance}_{ij}^{\kappa} v_{ij}$ – i.e., as a function of distance – we can construct $\tau_{ij}^{1-\sigma}$ using estimates from,

$$\ln X_{ij} = \eta \ln \text{distance}_{ij} + \xi_i + \chi_j + \ln v_{ij}$$

where $\eta = -(\sigma - 1)\kappa$. So, this allows us to estimate τ_{ij}.

However, we're still short σ. The usual route here is to use a σ that has been estimated elsewhere, preferably with quasi-experimental variation. For example, if you had tariff data, and included it in the gravity equation, then the coefficient on tariffs would be $(\sigma - 1)$, which allows us to pick up σ. NB: $\sigma - 1$ is often referred to as *trade elasticity*.

How do you get everything else? Well, recall the model is a machine that takes in parameters and returns L_i and w_i. So if you *invert* the model, you can find A_i and Z_i that exactly rationalize the data L_i, w_i as an equilibrium of the model. This is typically done on a computer.

Back to getting the GE effect

Ok, so suppose now we've estimated the model and can construct what the economy would have looked like absent the hurricane. Was this a waste of time? to find out, consider examining what the GDP effect of the hurricane was. We'll do so by looking at real output $Y = \left(\sum_i (Z_i L_i)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, with and without the hurricane.

How would you do this naively, if you just had the DID coefficient, β_1^{DID} =-0.18? Well, to the first order,

$$\frac{dY}{Y} = \sum_i s_i \frac{dZ_i}{Z_i},$$

where s_i is unit i's share of output. You can approximate $s_i = \frac{w_i L_i}{\sum_j w_j L_j}$. At baseline, treated tracts are about 32% of the wage bill, so, assuming $\frac{dZ_i}{Z_i} = \beta_1^{DID}$, you get that -0.18 $\times$ 0.32 ≈ -0.058, or the hurricane reduced real output by 5.8%.

This is not right in general equilibrium, and the DID coefficient is wrong *twice*: one, we know $\beta_1^{DID} \neq \frac{dZ_i}{Z_i}$. Second, we missed the spillovers that reallocate population across space and change wages even in untreated places. When we compute Y_i in the observed and counterfactual (no hurricane) economy, I find (numerically, in this particular example) that the real output effect is only a 2.3% real output loss.

3.1.2 Free mobility? Assessing the gains associated with migration

This last example relied on a model with free mobility. This seems naive, especially for short time horizons like computing the welfare effects of a hurricane. A lot of the damages occur because people cannot easily 'escape' affected regions.

One obvious twist to our framework is adding mobility frictions, violating spatial equilibrium. Free mobility cannot literally be true – we all know there are moving costs – and there can be a large opportunity cost associated with abandoning the accumulated knowledge and social capital that's place-specific. Some people are trapped in bad locations by happenstance of their birth! Do these mobility frictions matter in practice? How can we model them, and what's their consequence for our spatial equilibrium theory?

The evidence suggests that removing migration barriers improves outcomes. Nakamura et al. (2022) studies what happens to movers who lost their house in a volcanic eruption in Iceland. They essentially regress earnings on moving, using exposure to the volcanic eruption as an instrument for moving. Figure 3.3 replicates some key figures from this paper. The treatment effect they estimate shows that 'the gains to moving may be very large for those badly matched to the location they happened to be born in, even if differences in average income are small' (Nakamura et al., 2022). This suggests even with little wage variation across space, there

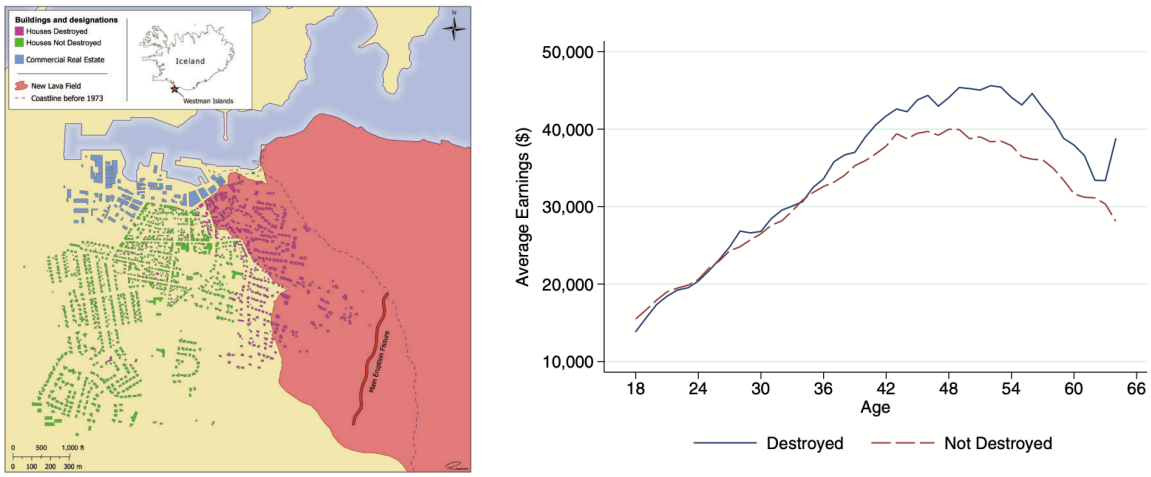

Figure 3.3: Reproductions of Figures 1 and 5 from Nakamura et al. (2022).

are potentially very large mobility barriers that reduce aggregate productivity by keeping people trapped in places where they cannot realize their potential.

How can we measure these migration costs and assess the macroeconomic costs associated with migration barriers? A common approach is to use a model. I'll now lay out the foundations of a *location choice model* (or *discrete choice model*). These models are extremely popular in quantitative spatial economics, industrial organization, and, increasingly, macroeconomics.

We begin with imagining that there are $i = 1, ..., N$ locations each with L_i persons. We can index persons with ω. Agents get utility V_j by living in place j. Agents also enjoy locations for *idiosyncratic* reasons, modeled by a shock $\epsilon_j(\omega)$, so that,

$$\underbrace{V_j(\omega)}_{\text{utility } \omega \text{ gets in j}} = \underbrace{V_j}_{\text{common component}} \times \underbrace{\epsilon_j(\omega)}_{\text{idiosyncratic component}} .$$

Now, suppose there are moving costs μ_{ij} such that for a mover ω in i who enjoys $V_j(\omega)$, a move to j discounts the reward so that they would enjoy $V_j(\omega)/\mu_{ij}$. Then, an agent born in i solves,

$$\max_j \frac{V_j}{\mu_{ij}} \epsilon_j(\omega)$$

What is expected utility, for someone born in i? It is $\mathbb{E}[\max_j V_j/\mu_{ij}\epsilon_j(\omega)]$. This cannot be the same everywhere unless $\mu_{ij} = 1$ for all i, j. A spatial equilibrium cannot hold!

These are complicated objects. Fortunately, there are distributions for which the maximum of a vector of random variables is well-behaved. They are called *max stable* distributions. In spatial work, it is common to use a Frechet distribution, that $\epsilon_j(\omega) \sim \exp\left(-(x/s)^{-\varepsilon}\right)$. The parameter s is the scale, which we will assume is 1, and ε is the shape parameter.

This distribution has nice properties, which are worked out in the Technical Appendix to this chapter,

$$P(\text{choose } j \mid \text{born in } i) = \frac{(V_j/\mu_{ij})^{\varepsilon}}{\sum_k (V_k/\mu_{ik})^{\varepsilon}}$$
$$\mathbb{E}\left[\max_j \frac{V_j}{\mu_{ij}}\epsilon_j(\omega) \mid j \text{ is max}\right] = \Gamma_0 \left(\sum_k (V_k/\mu_{ik})^{\varepsilon}\right)^{1/\varepsilon}$$

where Γ_0 is a constant that depends on ε.

The choice probability formula generates a *migration gravity equation*,

$$\underbrace{L_{ij}}_{\text{people moving from } i \text{ to } j} = \frac{(V_j/\mu_{ij})^{\varepsilon}}{\sum_k (V_k/\mu_{ik})^{\varepsilon}} L_i \implies \ln L_{ij} = -\varepsilon \ln \mu_{ij} + \xi_i + \chi_j$$

thus we can estimate migration costs by assuming they are a function of distance, or other variables.

With these estimates, we could slot them into our hurricane model, and check what the right estimating equation is when there are migration costs, and recompute the counterfactuals.

Including migration costs also allows us to assess the importance of migration barriers. For example, Bryan and Morten (2019) use this framework to assess the welfare gains from reducing migration barriers μ_{ij} in Indonesia, and find that there could be large GDP gains if people could access the top regions, even when accounting for selection on skills.

3.1.3 Revisiting the city – Ahlfeldt, Redding, Sturm, and Wolf (2015)

Ahlfeldt et al. (2015) is a seminal paper. In their paper, a city is a bunch of locations – city blocks – labeled i or j. Workers ω can choose which block to live in i, and which block to work in j. Just like Alonso-Muth-Mills, they trade off commuting costs and housing prices when trying to access high wage locations. They

maximize,

$$\max_{ij} \underbrace{\frac{1}{\delta_{ij}}}_{\substack{\text{commuting costs} \\ \text{from } i \text{ to } j}} \times \underbrace{A_i}_{\text{amenity in } i} \times \frac{\overbrace{w_j}^{\text{wage in } j}}{\underbrace{P_i}_{\text{prices in } i}} \times \underbrace{\epsilon_{ij}(\omega)}_{\substack{\text{idiosyncratic} \\ \text{Fréchet shock}}}$$

In their model, they also bring in ideas from Lucas and Rossi-Hansberg (2002): in particular, they assume amenities A_i and productivities Z_i (which determine wages) have a functional form like,

$$A_i = \bar{A}_i \times \left(\sum_j e^{-\text{distance}_{ij}} L_j \right)^{\kappa}, \quad Z_i = \bar{Z}_i \left(\sum_j e^{-\text{distance}_{ij}} L_j \right)^{\lambda}$$

i.e., amenities and productivities depend on the local population density.

This model also contains a gravity equation:

$$\underbrace{L_{ij}}_{\text{commuters from } i \text{ to } j} = \frac{\left(\delta_{ij}^{-1} A_i w_j / P_i \right)^{\varepsilon}}{\sum_{i',j'} \left(\delta_{i',j'}^{-1} A_{i'} w_{j'} / P_{i'} \right)^{\varepsilon}}.$$

These gravity estimates are a unifying feature of quantitative spatial models—they show how locations are linked together, and reappear in adding up constraints: the workers at j must be equal to the sum of commuters from i to j across all i, and so on, thus revealing things like: residences with better access to jobs should have higher rents, or that employers that are farther away need to pay higher wages to have as large a workforce as a more centrally located employer. These are the kinds of testable insights that allow us to interpret spatial data at various spatial scales, just as we did for the hurricane example.

In Ahlfeldt et al. (2015), they fit their model to Berlin, before and after the construction of the Berlin wall. Figure 3.4 shows how the construction of the Berlin wall shifted economic activity in West Berlin. As the wall cleaved the central business district in two, destroying the economic center of activity. Economic activity had to reorganize in space, pushing the downtown from the Potsdamer Platz and Mitte area towards the Kurfürstendamm. Today, Berlin consequently looks more polycentric, with the Mitte resurfacing but the West Berlin downtown near the present-day Zoo persisting.

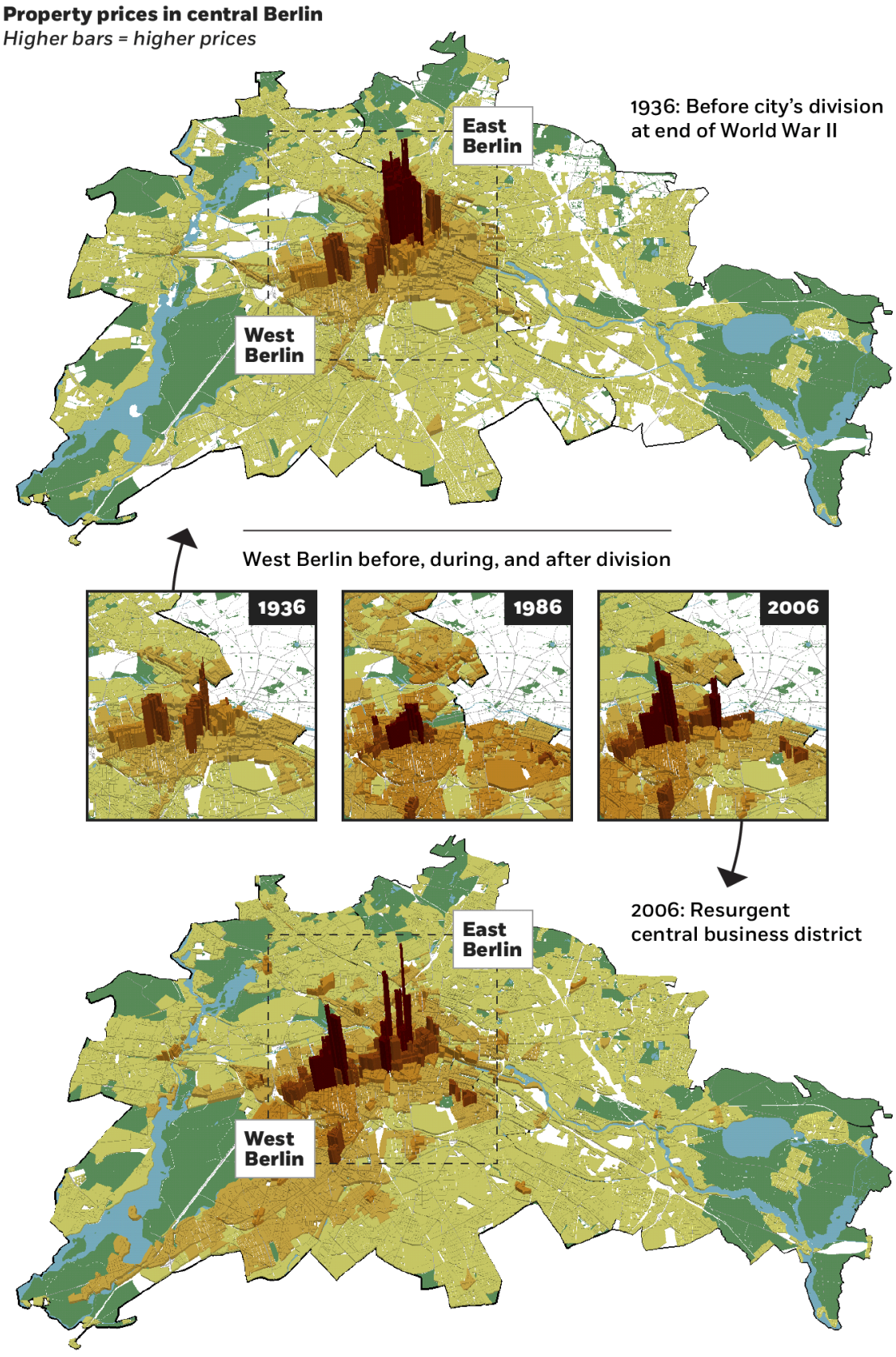

Figure 3.4: Changes in urban spatial structure in Berlin; reproduced from Chicago Booth Review (2018).

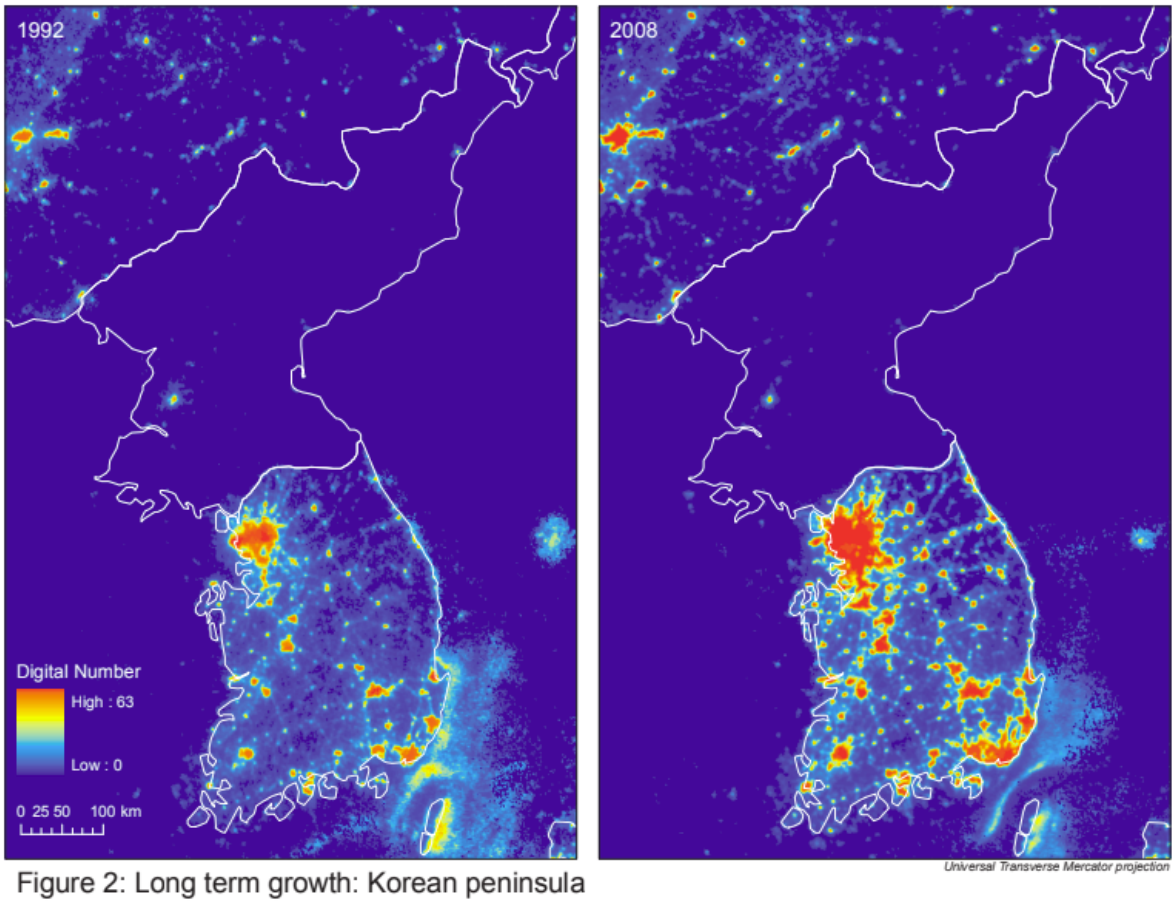

Figure 3.5: Korea from space, 1992 and 2008; reproduced from Henderson et al. (2012).

3.2 Spatial data analysis

Another side of modern, quantitative spatial economics is leveraging spatial data. These include, but are not limited to, using nightlights, other remote-sensed data, cellphone trip data, traffic data, building height data, and beyond.

Nightlights Henderson et al. (2012) popularized using satellite-derived measures of nighttime luminosity to infer economic growth in data poor and subnational levels. The basic idea is given in Figure 3.5. Using nightlights is now a common way to infer local GDP, especially in places where data is poor, like the developing world.

Other data sources and their uses abound, and combining them and using them alongside quantitative spatial

models is an exciting frontier in spatial economics. For example, both Harari (2020) and my job market paper use data on urban form—literally, city shapes and characteristics of their physical infrastructure—to infer measures of city quality. Cellphone data has been used to infer commuting patterns in developing cities (Kreindler and Miyauchi, 2023) or consumption trips (Miyauchi et al., 2025). Google maps queries have been used to analyze traffic in cities around the world (Akbar et al., 2023).

References

Ahlfeldt, Gabriel M et al. (2015). "The economics of density: Evidence from the Berlin Wall". In: *Econometrica* 83.6, pp. 2127–2189.

Akbar, Prottoy A et al. (2023). *The fast, the slow, and the congested: Urban transportation in rich and poor countries*. Tech. rep. National Bureau of Economic Research.

Allen, Treb and Costas Arkolakis (2014). "Trade and the Topography of the Spatial Economy". In: *The Quarterly Journal of Economics* 129.3, pp. 1085–1140.

– (2023). "Economic activity across space: a supply and demand approach". In: *Journal of Economic Perspectives* 37.2, pp. 3–28.

Bryan, Gharad and Melanie Morten (2019). "The aggregate productivity effects of internal migration: Evidence from Indonesia". In: *Journal of Political Economy* 127.5, pp. 2229–2268.

Chicago Booth Review (Aug. 21, 2018). *What the Berlin Wall Can Teach Us about Urban Development*. Accessed: 2026-02-25. URL: https://www.chicagobooth.edu/review/what-berlin-wall-can-teach-us-about-urban-development.

Harari, Mariaflavia (2020). "Cities in bad shape: Urban geometry in India". In: *American Economic Review* 110.8, pp. 2377–2421.

Henderson, J Vernon (1974). "The sizes and types of cities". In: *The American Economic Review* 64.4, pp. 640–656.

Henderson, J Vernon, Adam Storeygard, and David N Weil (2012). "Measuring economic growth from outer space". In: *American economic review* 102.2, pp. 994–1028.

Kreindler, Gabriel E and Yuhei Miyauchi (2023). "Measuring commuting and economic activity inside cities with cell phone records". In: *Review of Economics and Statistics* 105.4, pp. 899–909.

Krugman, Paul (1991). "Increasing returns and economic geography". In: *Journal of political economy* 99.3, pp. 483–499.

Lucas, Robert E and Esteban Rossi-Hansberg (2002). "On the internal structure of cities". In: *Econometrica* 70.4, pp. 1445–1476.

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J Redding (2025). "The economics of spatial mobility: Theory and evidence using smartphone data". In: *The Quarterly Journal of Economics* 140.4, pp. 2507–2570.

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson (2022). "The gift of moving: Intergenerational consequences of a mobility shock". In: *The Review of Economic Studies* 89.3, pp. 1557–1592.

Roback, Jennifer (1982). "Wages, rents, and the quality of life". In: *Journal of political Economy* 90.6, pp. 1257–1278.

Technical appendix

Consider an agent who chooses among n options to maximize

$$\max_{i \in \{1,\ldots,n\}} c_i z_i,$$

where c_i is a systematic component (e.g. productivity, utility, or cost) and z_i is an idiosyncratic draw. To make this tractable, assume $z_i \overset{iid}{\sim}$ Fréchet,

$$F(z) = \exp\left(-z^{-\theta}\right),$$

where $\theta > 0$ governs the dispersion of the idiosyncratic shocks. Then $c_i z_i$ has CDF

$$\Pr[c_i z_i \leq z] = \Pr\left[z_i \leq \frac{z}{c_i}\right] = \exp\left(-\left(\frac{z}{c_i}\right)^{-\theta}\right) = \exp\left(-c_i^{\theta}\, z^{-\theta}\right),$$

so scaled Fréchet draws remain Fréchet.

What's the CDF of the max of these draws, $G()$?

$$\begin{aligned} G(z) &= P\left(\max_i c_i Z_i < z\right) \\ &= \prod_i P\left(c_i Z_i < z\right) \\ &= \prod_i P\left(Z_i < z/c_i\right) \\ &= \prod_i F(z/c_i) \\ &= \exp\left(-\sum_i c_i^\theta z^{-\theta}\right) \end{aligned}$$

The CDF of the maximum is therefore

$$F_{\max}(z) = \exp\Big(-\Phi\, z^{-\theta}\Big), \qquad \Phi \equiv \sum_{i=1}^{n} c_i^\theta.$$

To recover the choice probabilities, we need a little work: First, densities. differentiating $F_i(z) = \exp\big(-c_i^\theta\, z^{-\theta}\big)$ gives the PDF,

$$f_i(z) = \frac{d}{dz} F_i(z) = c_i^\theta\, \theta\, z^{-\theta-1}\, \exp\Big(-c_i^\theta\, z^{-\theta}\Big).$$

Now, we can get the choice probabilities,

$$\begin{aligned} \pi_i &= \int_0^\infty \underbrace{\prod_{j\neq i} F_j(z)}_{\substack{\text{Prob. every other} \\ \text{option is below } z}} \cdot\, f_i(z)\, dz \\ &= \int_0^\infty \exp\left(-\sum_{j\neq i} c_j^\theta\, z^{-\theta}\right) \cdot c_i^\theta\, \theta\, z^{-\theta-1}\, \exp\Big(-c_i^\theta\, z^{-\theta}\Big)\, dz \\ &= \int_0^\infty c_i^\theta\, \theta\, z^{-\theta-1}\, \exp\Big(-\Phi\, z^{-\theta}\Big)\, dz \\ &= \frac{c_i^\theta}{\Phi} \underbrace{\int_0^\infty \Phi\, \theta\, z^{-\theta-1}\, \exp\Big(-\Phi\, z^{-\theta}\Big)\, dz}_{=1 \text{ (density of Fréchet}\left(\Phi^{1/\theta}, \theta\right)\text{integrates to 1)}} \\ &= \frac{c_i^\theta}{\Phi}. \end{aligned}$$

These are the Fréchet choice probabilities. Increasing θ concentrates mass on the highest-c_i option. The elasticity of π_i with respect to c_i is

$$\frac{\partial \log \pi_i}{\partial \log c_i} = \theta \left(1 - \pi_i\right).$$