The monocentric model in discrete space

Lecture notes #4: EC2410
Matt Turner
Brown University[1]
18 January 2024

In the first lecture, we developed two models of location choice. The first was based on households that were the same. Elaborating this model has given us the monocentric city model and the generalizations that we have studied so far, including the Roback model. The second model we considered in the introduction was based on agents who had heterogenous preferences. With the notion of bid-rent function, we can extend the monocentric city to allow for a small number of classes of agents, as in LeRoy and Sonstelie (1983) or Fujita and Ogawa (1982). However, to really think about people being different from one another, we will want a different modeling framework.

We develop such a framework here. We start with a description of two extreme value distributions, Frechet and Gumbel, and their very useful properties. Next we develop a version of the linear city model based on discrete space and a population of agents whose preferences depend on individual draws from a Frechet distribution. Finally, we discuss a much more general version of this model developed in Heblich et al. (2018).

At heart, these exercises are an elaboration of older techniques for modeling discrete choice problems by imposing extra adding up constraints, e.g., land and labor markets clear, on top of the discrete choice machinery. Given this, it is worth noting two classic books on the discrete choice problem, Anderson et al. (1992) and Train (2009).

A *Extreme value distributions and discrete choice problems*

a The Frechet distribution

Spatial models with heterogenous agents almost always rely heavily on one of two extreme value distributions to describe agent heterogeneity.

The first of these is the Frechet distribution,

$$Pr(z' \leq z) \equiv F(z) = e^{-Tz^{-\epsilon}}, T > 0, \epsilon > 1$$
$$f(z) = T\epsilon z^{-\epsilon-1}e^{-Tz^{-\epsilon}}.$$

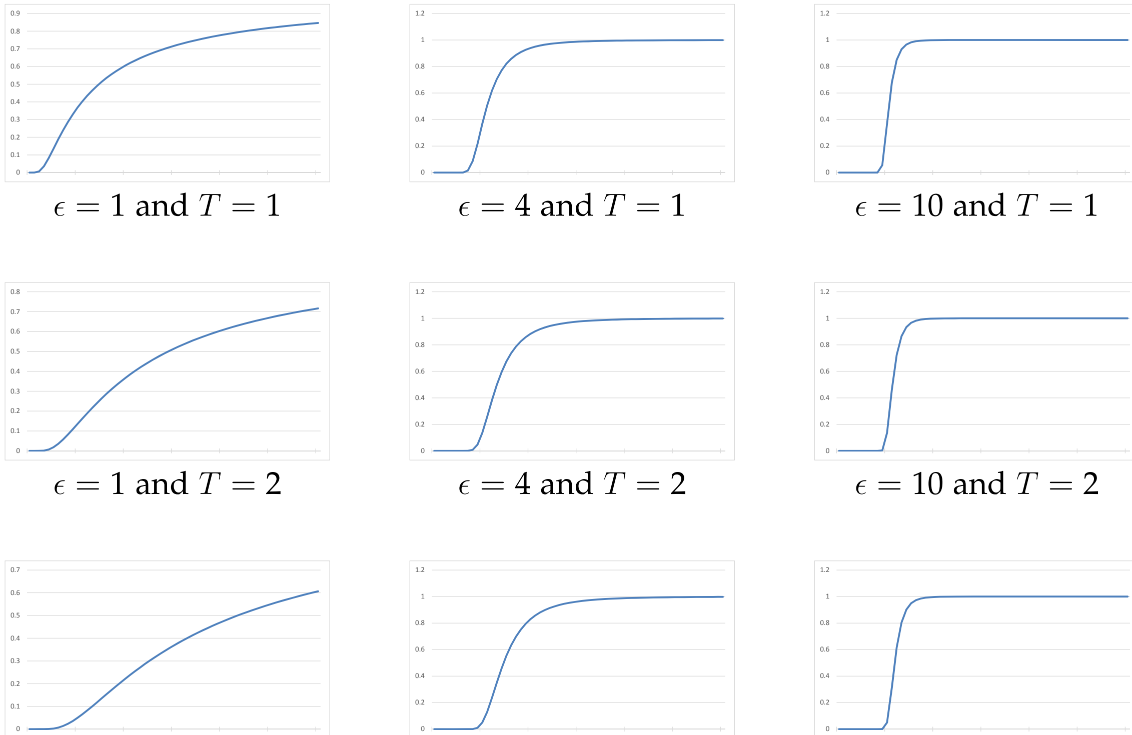
This distribution is governed by two parameters, T, called the level, and ϵ called dispersion. These names are suggestive of 'mean' and 'variance' and are often used in the same spirit. In fact, the mean of z is

$$E(z) = \Gamma\left(\frac{\epsilon-1}{\epsilon}\right)T^{1/\epsilon}.$$

The gamma function in this expression is defined as $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx$ and is a generalization of the factorial operator to the real numbers. In particular, $\Gamma(n) = (n-1)!$ for n an integer.

[1] I am grateful to Lorenzo Aldeco, Jonathan Dingel, Gilles Duranton and Jacques Thisse for helpful comments on earlier versions of these notes.

Figure 1: C.D.F.'s of Frechet distribution



$\epsilon = 1$ and $T = 1$ | $\epsilon = 4$ and $T = 1$ | $\epsilon = 10$ and $T = 1$

$\epsilon = 1$ and $T = 2$ | $\epsilon = 4$ and $T = 2$ | $\epsilon = 10$ and $T = 2$

$\epsilon = 1$ and $T = 3$ | $\epsilon = 4$ and $T = 3$ | $\epsilon = 10$ and $T = 3$

Figure 1 illustrates the CDF of a Frechet distribution for various values of T and ϵ. Note the that probability becomes more concentrated, and outcomes less dispersed, as ϵ decreases. While the mean of F depends on ϵ and the variance on T. With that said, the CDF of F gets flatter as ϵ shrinks and it shifts to the right as T increases, so as a rough way of thinking about the Frechet, thinking of level as mean and dispersion as variance may often be defensible.

The Frechet distribution has the following handy property. Consider two Frechet distributions,

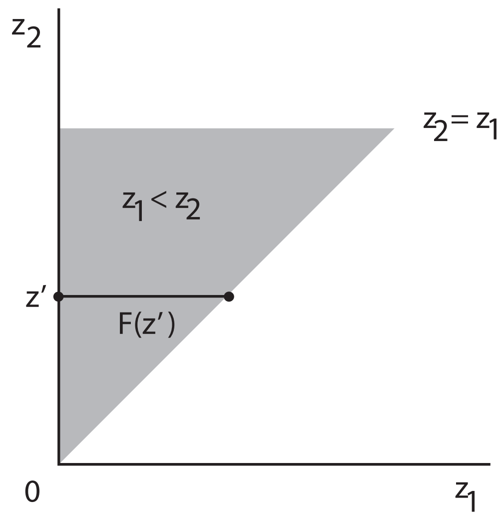
$$F_1(z) = e^{-T_1 z^{-\epsilon}}$$
$$F_2(z) = e^{-T_2 z^{-\epsilon}}.$$

Suppose the two distributions are independent and that we take a draw from each, z_1 and z_2. To motivate our discussion, note that in a spatial model, we expect agents to choose their favorite location from among the available possibilities. If z_1 and z_2 are related to an agent's preferences over two locations, then we will be interested in $Pr(z_1 < z_2)$.

To evaluate this probability, we must integrate the joint p.d.f. over the region indicated in figure 2. The region where $z_1 < z_2$ is shaded light gray. Assuming z_1 and z_2 independent, this integral is

$$Pr(z_1 < z_2) = \int_0^\infty F_1(z) dF_2(z) dz$$

Figure 2: $Pr(z_1 < z_2)$



$$
\begin{aligned}
&= \int_0^\infty e^{-T_1 z^{-\epsilon}} \left[T_2 \epsilon z^{-\epsilon-1} e^{-T_2 z^{-\epsilon}} \right] dz \\
&= \int_0^\infty T_2 \epsilon z^{-\epsilon-1} e^{-(T_1+T_2) z^{-\epsilon}} dz \\
&= \frac{T_1}{T_1+T_2} \int_0^\infty (T_1+T_2) \epsilon z^{-\epsilon-1} e^{-(T_1+T_2) z^{-\epsilon}} dz \\
&= \frac{T_1}{T_1+T_2} \left[e^{-(T_1+T_2) z^{-\epsilon}} \right]_0^\infty \\
&= \frac{T_1}{T_1+T_2} \left[e^{-(T_1+T_2) \frac{1}{z^\epsilon}} \right]_0^\infty \\
&= \frac{T_1}{T_1+T_2} [1-0] \\
&= \frac{T_1}{T_1+T_2}
\end{aligned}
$$

Given this result, suppose we observe the share of draws, π_1, with $z_1 > z_2$, and the share of draws, π_2, with $z_2 > z_1$. Then in a large sample,

$$
\begin{aligned}
\pi_1 &= \frac{T_1}{T_1+T_2} \\
\pi_2 &= \frac{T_2}{T_1+T_2}.
\end{aligned}
$$

This is a system of two equations in two unknowns, but by inspection, we can identify T_1 and T_2 only up to a constant. Given any pair of T's that solves this equation, so does any non-zero scalar multiple. That is, given extreme value draws, I can estimate the level parameters (up to a constant) if all I observe is the share of draws for which outcome 1 is larger than 2, and conversely. Indeed, in this simple example, $\pi_1 = 1 - \pi_2$, so it is enough to observe just one of the two shares. Note that the indeterminacy here is fundamental to discrete choice analysis in which we are just concerned with ranking choices (Train, 2009), and not with cardinal comparisons between them.

This result is central to discrete choice analysis. In absence of an easy closed form solution to the integral that lets us evaluate the likelihood of the first order statistic, we are left with numerical integration. This rapidly becomes intractable as the number of possible alternatives increases. To convince yourself of the simplicity and elegance of this result, try to conduct the corresponding calculation when, for example, z_1 and z_2 are uniformly distributed.

In fact, when we discuss models with transportation or transactions costs, we will not simply be concerned with ranking choices, but with determining whether preferences for one alternative are sufficiently strong to overcome transaction costs. This will sometimes allow us to resolve this problem.

Note that ϵ does not appear in this equation, and so we cannot hope to estimate it from this system alone. This will turn out to be important and we will return to it later.

This result generalizes to a choice over n draws in the natural way. Given n draws $\{z_1,...,z_n\}$, where z_i is distributed Frechet with $F_i(z) = e^{-T_i z^{-\theta}}$, then the probability that $z_i = \max\{z_1,...,z_n\}$ is $\pi_i = \frac{T_i}{\sum_{j=1}^{n} T_j}$. Note that we require that the dispersion parameter be constant across all i, although level parameters can vary.

A large literature is organized around this machinery. Many economic decisions can be framed as 'choose your favorite' from among a discrete choice of alternatives. Examples include the choice be amongst places to live or countries to trade with. The trick is to frame such a 'choose your favorite' problem in such a way that we can trace the share of agents choosing each outcome back to the parameters of a joint Frechet distribution.

b CES demand systems and the Frechet distribution

It is worth pointing out the similarity between the expressions for π_i above to a demand function under constant elasticity of substitution preferences. Here, letting x_i be the demand for good $i \in [01,]$ and p_i be the corresponding price, under common assumptions, we can write the demand for good x_i as $\frac{p_i^{-\sigma}}{\int_0^1 p_j^{1-\sigma} dj}$, where σ is the elasticity of substitution. Note how close is this functional form to the function describing a the shares given above. In particular, if we let $T_i = \rho^{1-\sigma}$ then

$$\begin{aligned}\frac{p_i^{-\sigma}}{\int_0^1 p_j^{1-\sigma} dj} &= T_i^{\frac{-1}{1-\sigma}} \frac{T_i}{\int_0^1 T_j dj} \\ &= T_i^{\frac{-1}{1-\sigma}} \pi_i.\end{aligned}$$

where I've cheated a little bit and assumed that the continuum version of the model leads to the same expression as in the discrete version, with the summation replaced by an integral. It is common to exploit this similarity in the derivation of models to predict, for example, trade shares, e.g. Eaton and Kortum (2002).

c The other extreme value distribution

Over the past few years the study of urban economics has seen the application of models and techniques originally developed to study international trade. Almost all of this work relies on the Frechet distribution to describe the heterogeneity of locations or economic agents.

An older urban economics literature, and more recently Diamond (2016), develops models based on much the same intuition but uses the Gumbel distribution,

$$Pr(z' \leq z) \equiv F(z) = e^{-e^{-z}}. \tag{1}$$

Like the Frechet distribution, the Conditional Logit leads to a particularly simple expression for the share of trials for which a particular draw is the largest of a set of draws.

Given a set of n scalars $\{u_1,...,u_n\}$ and n draws from a Gumbel distribution, $\{z_1,...,z_n\}$, define $v_i = u_i + z_i$. In this case,

$$Pr(v_i = \max\{v_1,...,v_n\}) = \frac{\exp(u_i)}{\sum_{j=1}^{n} \exp(u_j)}. \tag{2}$$

While the formulation of the Gumbel distribution above is pervasive in discrete choice estimation, it is helpful to consider the slightly larger family of densities given by

$$Pr(z' \leq z) \equiv F(z) = e^{-e^{-z/\mu}}. \tag{3}$$

Here, the parameter μ plays much the same role as does the dispersion parameter ϵ in the Frechet. The literature on discrete choice commonly normalizes this parameter to 1. Given this literature's focus on ordinal rankings, this is without loss of generality. However, by considering the more general function we can illustrate the relationship between Frechet and Gumbel densities.

To see this, consider a set of n scalars $\{u_1,...,u_n\}$ and n draws from a Gumbel distribution of the form given in (3), $\{z_1,...,z_n\}$, and define $v_i = u_i + z_i$. In this case,

$$Pr(v_i = \max\{v_1,...,v_n\}) = \frac{\exp(u_i/\mu)}{\sum_{j=1}^{n} \exp(u_j/\mu)}. \tag{4}$$

If we make the change of variable $w_i = \ln v_i$, we have

$$Pr(v_i = \max\{v_1,...,v_n\}) = \frac{w_i^{1/\mu}}{\sum_{j=1}^{n} w_j^{1/\mu}}. \tag{5}$$

This is exactly the expression we get if we start with a Frechet distributed shock. As a practical matter it does not appear to be very important which density we work with. Conventionally, however, trade economists work with the Frechet and everyone else uses Gumbel.

The reliance on extreme value distributions to model individuals' choices is not quite as arbitrary as it seems. There are 'extreme value theorems' that parallel the 'central limit theorem'. Loosely, if x is a random variable that is a maximum over a many draws of some other random variable y, then we expect x to have a Frechet or Gumbel distribution, or a similar third distribution, depending on the characteristics of the distribution on y. See Embrechts et al. (2013) for examples.

d A surprising and useful property of extreme value distributions

Consider x_i distributed Frechet, so

$$F_i(x) = e^{-T_i x^{-\theta}} \tag{6}$$

$$f_i(x) = T_i \theta x^{-\theta-1} e^{-T_i x^{-\theta}} \tag{7}$$

$$E(x) = \Gamma(1-\frac{1}{\theta})T_i^{1/\theta} \tag{8}$$

We really care about the properties of the maximum of a set of Frechet draws. To make things easy, just think about pairs of draws. Define

$$x^* = \max\{x_1, x_2\}, \tag{9}$$

for x_i independent Frechet draws as above.

In this case,

$$Pr(x_1 > x_2) = \frac{T_1}{T_1 + T_2}, \tag{10}$$

and we can calculate the distribution of x^*, $F(x^*)$ as follows:

$$\begin{aligned} F(x) &\equiv Pr(x^* < x) && (11)\\ &= Pr(x_1 < x^* \bigcap x_2 < x^*) && (12)\\ &= Pr(x_1 < x) Pr(x_2 < x) && (13)\\ &= F_1(x) F_2(x) && (14)\\ &= e^{-T_1 x^{-\theta}} e^{-T_2 x^{-\theta}} && (15)\\ &= e^{-(T_1+T_2)x^{-\theta}} && (16) \end{aligned}$$

Therefore F is also Frechet. Hence

$$\begin{aligned} f(x) &= (T_1+T_2)\theta x^{-\theta-1} e^{-(T_1+T_2)x^{-\theta}} && (17)\\ E(x^*) &= \Gamma(1-\frac{1}{\theta})(T_1+T_2)^{1/\theta} && (18) \end{aligned}$$

We are also interested in $E(x^*|x_1 > x_2)$. Evaluating this one is a little trickier. We want to calculate an expectation over the region in quadrant I of the (x_1,x_2) plane that lies to the right of the $x_1 = x_2$ line. The probability of each x_1 is f_1, the weight we assign to it is $F_2(x_1)$. Since the probability mass of the whole region is less than one, we need to scale everything up by the inverse probability of $x_1 > x_2$.

Thus we have,

$$\begin{aligned} E(x^*|x_1 > x_2) &= Pr(x_1 > x_2)^{-1} \int_0^\infty f_1(x_1) F_2(x_1) x_1 dx_1 && (19)\\ &= \frac{T_1+T_2}{T_1} \int_0^\infty \left(T_1 \theta x_1^{-\theta-1} e^{-T_1 x_1^{-\theta}}\right) e^{-T_2 x_1^{-\theta}} x_1 dx_1 && (20)\\ &= \int_0^\infty (T_1+T_2)\theta x_1^{-\theta-1} e^{-(T_1+T_2)x_1^{-\theta}} x_1 dx_1 && (21)\\ &= \int_0^\infty f_*(x_1) x_1 dx_1 && (22)\\ &= E(x^*) && (23) \end{aligned}$$

Thus we have three interesting facts:

1. The expected realization of the maximum of Frechet draws conditional on any particular realization being the best is the same as the unconditional expected realization.

2. The expected realization of the maximum of Frechet draws conditional on any particular realization being the best does not vary with the winner.

3. This one is a little more subtle. Looking at equations (14) and (15), we see that the argument for #2 above works iff the probabilities of each variable being the winner takes the familiar Frechet formula. Thus, 'all expected utilities conditional on a realized choice are equal' is equivalent to 'choice probabilities take the familiar Frechet form'. That is, we have two equivalent ways of implementing the random utility choice problem.

B *The linear city with discrete space and heterogenous preferences*

a Frechet preference shocks

Consider a discrete version of a linear city consisting of three neighborhoods. Index neighborhoods by $i \in \{1,2,3\}$. Let x_i denote a neighborhood's distance from the CBD, with $x_i = i$. That is, the three neighborhoods are, respectively, 1, 2, and 3 units of distance from the center. Suppose that the cost to commute one unit distance is τ.

The city is populated by measure $\overline{L}$ of agents indexed by j. Each agent chooses a neighborhood, pays land rent specific to that neighborhood, R_i and commutes to the center to earn wage w.

The utility of agent j locating at i is

$$u_{ij} = [w - R_i - i\tau]z_{ij}, \tag{24}$$

where z_{ij} is a person and location specific taste shock, drawn from a Frechet distribution, $F(z) = e^{-Tz^{-\epsilon}}$. That is, each agent gets one Frechet shock for each of the three possible locations.

If we solve (24) for z and substitute into the distribution of z, we derive the implied distribution of u, $G(u) = e^{T[w-R_i-i\tau]^{\epsilon}u^{-\epsilon}}$.

Define an equilibrium to occur when each agent chooses his favorite location, and require that all landlords in each location choose the same rent.

To begin, let π_i denote the share of population living at location i. If all agents choose their favorite location, then the probability that any particular agent chooses location 1 is $Pr(u_1 > u_2$ and $u_1 > u_3)$.

Using the results developed above, we have that

$$\begin{aligned} \pi_1 &= \frac{[w - R_1 - 1\tau]^{\epsilon}}{\sum_{k=1}^{3}[w - R_k - k\tau]^{\epsilon}} \\ \pi_2 &= \frac{[w - R_2 - 2\tau]^{\epsilon}}{\sum_{k=1}^{3}[w - R_k - k\tau]^{\epsilon}} \\ \pi_3 &= \frac{[w - R_3 - 3\tau]^{\epsilon}}{\sum_{k=1}^{3}[w - R_k - k\tau]^{\epsilon}}. \end{aligned} \tag{25}$$

Suppose that each location is occupied by exactly one third of the population so that $\pi_i = 1/3$ for all i, and that land rents are not observed. Since the denominator of each of the fractions in (30) is the same, it must be that the numerators are also the same. In turn, this requires that $R_1 - R_2 = \tau$ and $R_2 - R_3 = \tau$. This is the discrete analog of having the land rent gradient decrease at the same rate as commute costs increase.

More generally, with τ unknown, (30) is a system of three equations in 5 unknowns, τ, ϵ and the three R_i. In general, we can solve the system only if we add more information. By analogy to the linear city model, two natural candidates would be to require that R_3 be equal to an exogenous agricultural land rent and that income net of rent and commuting reach some threshold value. This 'calibration' exercise is a much simpler version of what is often done in the recent literature that will be the topic of much of the rest of this section.

What we have done here is, effectively, to solve the analog of the closed city model. We fixed population and then asked what land rent would rationalize an equilibrium in which every person consumed the 'same amount of space'.

There is an important difference between this discrete linear city and a linear city with a continuum of locations. In the continuous model, all agents are identical and in equilibrium all obtain the same level of utility. In the discrete case, all agents within a location have different levels of utility.

This means that calculating welfare in the discrete case is more difficult. We must calculate both land rent and consumer surplus. This parallels the difference between the simple model with homogenous and heterogenous agents that we began with.

In fact, we can calculate the expected utility of an agent living in this city. This expectation is,

$$
\begin{aligned}
E(u) &= \mathbb{E}\left(max_{i\in\{1,2,3\}}[w - R_i - i\tau]^{\epsilon} z\right) \\
&= \Gamma(\frac{\epsilon-1}{\epsilon})\left(\sum_{i\in\{1,2,3\}}[w - R_i - i\tau]^{\epsilon}\right)^{1/\epsilon}.
\end{aligned}
$$

At this point, it makes sense to think about the nature of the equilibrium we have described. Since everyone chooses their favorite location, after realizing shocks and choosing, no one will want to move. This is broadly consistent with the notion of equilibrium we used in the closed monocentric city model. Comparing with the open monocentric city model seems more problematic. In this, if population flows into the city on the basis expected utility, as described above then we should expect that some unlucky agents will realize bad shocks for everywhere, and even if they choose their best option from {1,2,3} will regret their decision to migrate to the city and would prefer the exogenous outside option.

One natural way around this would seem to be to introduce the 'outside option' as another of the choice alternatives, and not as a reservation expected utility level. It is not clear, however, if introducing such an option, with a fixed utility level, requires substantive changes in other parts of the model.

In addition, it is worth pointing out the timing of how prices are determined remains a little vague. Interpreting the model literally, agents are choosing their location on the basis of prices, here land rent, that cannot be determined until after everyone resolves their random choice problem. Thus, it is natural to think about agents as making decisions on the basis of their expectations of prices, and that their ex post welfare depends on realized prices. Implicitly, we are relying on a law of large numbers that says that as the number of agents choosing becomes uncountably large, all ex ante uncertainty about prices goes away.

Finally, notice that we do not observe ϵ, the parameter that governs preference dispersion, and so we have no basis to guess at the utility levels of anyone in our city. It is to the issue of estimating ϵ that we now turn.

b Estimation

Let S_i be the sample analog of π_i. As described above, our discrete city is populated by a continuum of people. In this case, we will have $S_i = \pi_i$.

Suppose that, in addition to observing these shares, we also observe enough other information that the system described by (30) is exactly identified. This will mean that each of the three equations hold identically.

We will discuss 'gravity regressions' at some length below. They are a widely used approach to learning about the dispersion parameter. To derive the analog of a gravity regression in the context of this model, take the logarithm of each of the three equations in (30). Since the denominator is constant across locations, this yields,

$$\begin{aligned} \ln S_i &= -\ln\left(\sum_{k=1}^{3}[w - R_k - k\tau]^{\epsilon}\right) + \epsilon\ln(w - R_i - i\tau) \\ &= A + \epsilon\ln(w - R_i - i\tau), \end{aligned} \tag{26}$$

for $i = 1,2,3$.

This looks like a regression, and it is common to estimate a regression of this form. Such regressions are often called gravity regressions, although we will need to wait to see a more complicated model to understand the rational for this name.

Suppose that we estimate a regression based on this equation. That is,

$$\ln S_i = A + \epsilon\ln(w - R_i - i\tau) + \mu_i. \tag{27}$$

This is the same as (26), except that we have introduced a stochastic error term μ_i and implicitly, an assumption about the relationship between this error and the regressor, e.g. $cov(\mu, \ln(w - R_i - i\tau)) = 0$. In this case, we are able to estimate the dispersion parameter ϵ.

Notice that if we subsequently use this estimated value of ϵ in the system (26) and solve this system exactly, then this introduces a logical inconsistency. In particular, the gravity regression is based on a logarithmic transformation of (26) that is assumed to hold inexactly, while the rest of the model requires that it hold exactly.

Alternatively, suppose that our data describes the location of a finite number of people, say n_i in each location. In this case, we would expect that sampling error will lead realized shares in each location, $S_i = n_i / \sum_j n_j$, to differ slightly from expected values π_i. In this case, we can ask which parameter values are most likely to lead to the observed values of S_i. This leads to a maximum likelihood with likelihood function

$$\mathcal{L} = \Pi_{j=1}^{3}\left(\frac{[w - R_1 - j\tau]^{\epsilon}}{\sum_{k=1}^{3}[w - R_k - k\tau]^{\epsilon}}\right)^{n_i}. \tag{28}$$

c Gumbel preference shocks

We can also describe a discrete linear city model under Gumbel shocks to preferences. Since the formula for calculating 'choose your favorite' under Gumbel shocks revolves around additive shocks (rather than the multiplicative shocks we used for the Frechet case) this requires a slightly different specification of preferences.

Specifically, preserve all of the notation from the discrete linear city model above, but change preferences so that we have an additive Gumbel shock rather than a multiplicative Frechet shock. In this case we have,

$$u_{ij} = [w - R_i - i\tau] + z_{ij}, \tag{29}$$

where z_{ij} is a person and location specific taste shock, drawn from a Gumbel distribution, $F(z) = e^{-e^{-z}}$. That is, each agent gets one Gumbel shock for each of the three possible locations.

As for the Frechet case, let π_i denote the share of population living at location i. If all agents choose their favorite location, then the probability that any particular agent chooses location 1 is $Pr(u_1 > u_2 \text{ and } u_1 > u_3)$.

Given the Gumbel distribution of the z_{ij} we have that,

$$\begin{aligned} \pi_1 &= \frac{e^{w-R_1-1\tau}}{\sum_{k=1}^{3} e^{w-R_k-k\tau}} \\ \pi_2 &= \frac{e^{w-R_2-1\tau}}{\sum_{k=1}^{3} e^{w-R_k-k\tau}} \\ \pi_3 &= \frac{e^{w-R_3-1\tau}}{\sum_{k=1}^{3} e^{w-R_k-k\tau}}. \end{aligned} \tag{30}$$

This formulation of the problem and the corresponding Frechet based formulation are obviously similar and the discussion of that model applies almost without adjustment. There does not seem to be a strong reason to prefer one to the other.

C *Using a general model with discrete space and agent heterogeneity to study the role of the London underground in the economic geography of London*

Among the most robust comparative statics in the various formulations of the monocentric city model is that a city should spread out as transportation costs decrease. We saw this result in the context of the linear city model, the monocentric city model with housing, and in the model of commuting mode choice by LeRoy and Sonstelie (1983). In each of these models, we allow households to choose their location of residence, but work location is fixed outside the model.

Fujita and Ogawa (1982) generalizes to allow both firms and workers to choose their location, and to allow workers to choose both their location of work and their location of residence. In this model we see that if transportation costs are sufficiently high, a 'fully mixed' equilibrium may emerge. That is, if commuting costs are sufficiently high, in equilibrium everyone works where they live and firms and workers are more-or-less evenly distributed, and no household commutes to work. As the cost of commuting falls, given agglomeration forces, we expect to see cities with a monocentric structure, firms at the center and commuting workers arranged around.

Thus, the available theory presents alternative hypotheses about the effect of reductions of transportation costs on urban form. Should we see the dispersion predicted by the moncentric city model, or the centralization of work predicted by Fujita and Ogawa (1982)?

This question has also been the subject of a recent empirical literature. Baum-Snow (2007) counts radial segments of the US interstate highway system emanating from old central cities. That is, highway segments that are well situated to reduce the cost of travelling away from the center of US cities. He then examines the extent to which the number of such radial highways can explain the change from 1950 to 1990 in central city share of metropolitan area population. This was a period when the share of metropolitan area population declined dramatically, and Baum-Snow's estimates suggest that almost the entire decline can be explained by the creation of radial highways. In a more recent working paper, Baum-Snow (2017) also finds the radial interstate highways decentralized employment in US, as well as population.

The research design pioneered by Baum-Snow has been replicated in China by Baum-Snow et al. (2017) and in Europe by Garcia-López et al. (2015). Both studies arrive at qualitatively similar conclusions about ability of highways to spread cities out. Similarly, Gonzalez-Navarro and Turner (2018) examine the effect of subways on rate at which night light decays with distance from a city enter. Consistent with Baum-Snow (2007), they find that night light is more dispersed in cities with more extensive subway networks, and that light becomes more dispersed as the extent of a subway network increases.

In short, this empirical literature would seem to strongly support the predictions of the monocentric city model.

Against this background, the finding in Heblich et al. (2018) is surprising. This paper conducts two main exercises. The first documents changes in the pattern of residential and commercial land use in London around the time that its subway began to operate. They find that London's geography changed in a way that is broadly consistent with the predictions in Fujita and Ogawa (1982). That is, the employment and residence locations became segregated as the center of London became home to an extraordinary concentration of employment, the rest of the city specialized in housing, and people went back and forth between central employment and peripheral residences on the trains.

The second main exercise conducted in Heblich et al. (2018) is to develop a model with which to rationalize the changes in London that appear to follow from the construction of the rail network. This model is a generalization of the discrete space linear city model described above, and is the main topic of this section.

Briefly, Heblich et al. (2018) develop a model with discrete locations, heterogenous freely-mobile households differentiated by Frechet shocks to their preferences, and a production sector governed by perfect competition and free-entry. There are no agglomeration effects, but agents have preferences over consumption and housing and pay for commute distance by foregoing a portion of their daily wage. This is, therefore, an important advance to our ability to think about location choice. In particular, their model describes household choices of work and residence and the location of production in a model that can perfectly replicate a cross-section of behavior in London as an equilibrium.

This is an improvement on the monocentric city model and on Fujita and Ogawa (1982) in a number of dimensions. It is inferior in two regards. First, unlike Fujita and Ogawa (1982) it does not consider agglomeration economies. In fact, the technique developed in Heblich et al. (2018) derives from and is simpler than Ahlfeldt et al. (2015) which does allow for agglomeration economies. Second, the complexity of the model is such that analytic comparative statics are impractical. The entire exercise is largely organized around a single, numerical comparative static: What would London look like without the subway?

With that preamble, the details of the model follow. London consists of R discrete 'boroughs'. Index this set of locations by i if we are referring to work locations and by n if referring to residence locations. Each borough contains floor space, L_n, with price Q_n if used as housing for households, or Q_i if used as an input into production.

Measure H households live in London. Of these, H_n live in borough n and H_i work in borough i. Index households with ω. All households supply one unit of labor inelastically and have an outside utility level $\overline{u}$. Land rent is collected by absentee landlords.

Households derive utility from the consumption of a composite consumption good, with price P, and housing according to a Cobb-Douglas utility function with consumption share α. Conditional on a choice of work location i, household ω has income w_i^*. Thus, the indirect utility function for household ω living in n and working in i is

$$u_{ni} = \frac{w_{ni}^*}{P_n^{\alpha} Q_n^{1-\alpha}}. \quad (31)$$

Next, make the simplifying assumption that $P_n = 1$ for all n. Thus, the composite consumption good is the numeraire and, implicitly, it trades costlessly across boroughs.

Income w_i^* depends on the wage available in borough i, on commuting costs between boroughs n and i, and on a random shock,

$$w_{ni}^* = \frac{z_{ni}(\omega) w_i}{\kappa_{ni}}. \quad (32)$$

Here, w_i is the wage in borough i. $\kappa_{ni} > 1$ is an 'iceberg' commute cost. With $\kappa_{ni} > 1$, fraction $1 - 1/k$ of the households labor 'melts' in the commute: if κ units of labor begin a commute, only 1 unit finishes it. Note that this is different from the additive formulation of commuting costs we have considered so far. This assumption is purely to ease computation. It is common to all of the papers in this literature and is inherited from the trade literature from which this literature is derived, e.g., Eaton and Kortum (2002).

$z_{ni}(\omega)$ is an individual specific shock reflecting some combination of ω's idiosyncratic productivity in i and his taste for the commute between n and i. Each ω will draw one such z for each of the possible pairs of location/residence choices.

Assume that the $z_{ni}(\omega)$ are drawn from a Frechet distribution

$$G_n(z) = e^{-B_n z^{-\epsilon}}. \quad (33)$$

Note that the level parameter varies over residence boroughs only. It will operate as a measure of the 'residential amenities' in borough n. The dispersion parameter, ϵ is common across all residence-workplace pairs.

Substitute (32) into (31) and solve for z. Next substitute in (33)and recall that $P_n = 1$, to derive the Frechet distribution for $u_{ni}(\omega)$,

$$G_{ni}\left(\frac{\kappa_{ni}Q_n^{1-\alpha}}{w_i}u_{ni}\right)$$
$$=\exp\left[-B_n\left(\frac{\kappa_{ni}Q_n^{1-\alpha}}{w_i}u_{ni}\right)^{-\epsilon}\right]$$

Let $\pi_{ni} = \frac{H_{ni}}{H}$ denote the share of households choosing residence n and workplace i. Since we consider a continuum of agents, this share will exactly equal the probability of an agent preferring the workplace residence pair (n,i) to all others. That is,

$$Pr(u_{ni} \geq u_{n'i'}\,\forall n',i' \in R). \tag{34}$$

Recalling the particular properties of the Frechet distribution discussed above, this means that

$$\pi_{ni} = \frac{B_n w_i^{\epsilon}(\kappa_{ni}Q_n^{1-\alpha})^{-\epsilon}}{\sum_{r\in R}\sum_{s\in R} B_r w_s^{\epsilon}(\kappa_{rs}Q_n^{1-\alpha})^{-\epsilon}} \tag{35}$$

Summing across residence locations we get the share of workers in borough i,

$$\begin{aligned}\pi_i &= \frac{H_i}{H} \\ &= \sum_{r\in R}\pi_{ri} \\ &= \sum_{r\in R}\frac{B_r w_r^{\epsilon}(\kappa_{ri}Q_r^{1-\alpha})^{-\epsilon}}{\sum_{r'\in R}\sum_{s\in R} B_{r'} w_s^{\epsilon}(\kappa_{r's}Q_{r'}^{1-\alpha})^{-\epsilon}}\end{aligned} \tag{36}$$

Similarly, summing across workplaces, we get the share of households resident in n,

$$\begin{aligned}\pi_n &= \frac{H_n}{H} \\ &= \sum_{s\in R}\pi_{ns} \\ &= \sum_{s\in R}\frac{B_n w_s^{\epsilon}(\kappa_{ns}Q_n^{1-\alpha})^{-\epsilon}}{\sum_{r\in R}\sum_{s'\in R} B_r w_{s'}^{\epsilon}(\kappa_{rs'}Q_r^{1-\alpha})^{-\epsilon}}\end{aligned} \tag{37}$$

The $2R$ equations described by (36) and (37) correspond loosely to the three equations describing population shares in the three location linear city model. Thus, we should expect them to be important for estimates to calibrate or estimate the model. This turns out to be the case, and the equations are central to the calibration exercise preformed in Heblich et al. (2018).

Note that, despite the stochastic foundations of this model, there is no uncertainty in its predictions. Because we are drawing a from a continuum of households, the fraction of observed households in each neighborhood will exactly match the expected quantity. This eases calculation, but will complicate efforts to calibrate the model. If the model does not fit exactly, there is not source of structural error to appeal to explain the divergence between model and observations.

Again using the properties of the Frechet distribution, we can evaluate the expected utility of a household living in n and working in i,

$$E(u_{ni}) = \delta \left[\sum_{r \in R} \sum_{s \in R} B_r w_s^{\epsilon} (\kappa_{rs} Q_n^{1-\alpha})^{-\epsilon} \right]^{\frac{1}{\epsilon}}$$

where $\delta = \Gamma(\epsilon - 1)/\epsilon$. This calculation is analogous to the corresponding calculation that we performed earlier for the three location linear city, and the same comments about spatial equilibrium and the price mechanism apply.

It is worth noting that we can also use equation (35) to calculate the share of people living in n working in i, for each possible workplace i,

$$\pi_{ni|n} \equiv \frac{H_{ni}}{H_n} \tag{38}$$

$$= \frac{\frac{H_{ni}}{H}}{\frac{H_n}{H}} \tag{39}$$

$$= \frac{\pi_{ni}}{\pi_n} \tag{40}$$

$$= \frac{\pi_{ni}}{\sum_{s \in R} \pi_{ns}}. \tag{41}$$

Substituting from (35) and simplifying, we arrive at

$$\pi_{ni|n} = \frac{\left(\frac{w_i}{\kappa_{ni}}\right)^{\epsilon}}{\sum_{s \in R} \left(\frac{w_s}{\kappa_{ns}}\right)^{\epsilon}} \tag{42}$$

This is a 'gravity equation' for commuting. It provides that most widely used basis for estimating the dispersion parameter, a topic that we will return to later.

We now turn our attention to the production of the composite good.

Each borough produces the tradable composite consumption good from floor space, labor, and an intermediate input, 'services', produced in the borough, and not traded. The intermediate input, in turn, is produced from floor space and labor. Both sectors are constant returns to scale, and both sectors have a borough specific total factor productivity. Equilibrium behavior is determined by perfect competition and free-entry.

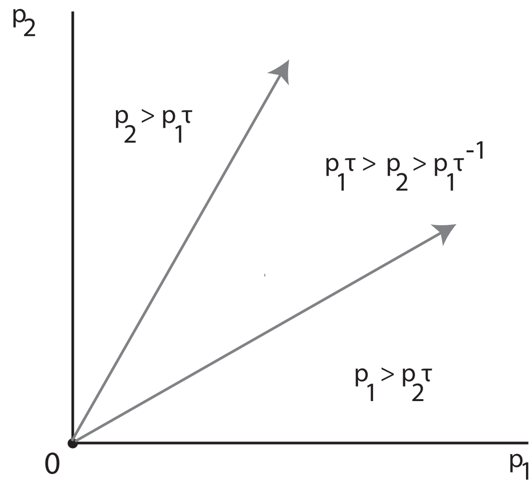
Omitting a good deal of detail, this leads to an equilibrium in which the wage in borough i is given by

$$w_i = A_i^{\frac{1}{\tilde{\beta}}} Q_i^{\frac{1-\tilde{\beta}}{\tilde{\beta}}}, \tag{43}$$

where A_i compounds total factor productivity in the intermediate and final goods sector in borough i, and $\tilde{\beta}$ compounds parameters from intermediate and final goods production functions. As before, Q_i is the price of floor space in borough i.

With this abbreviated description of production in place, closing the model requires only that we specify market clearing conditions. Market clearing requires that two main accounting identities be satisfied.

Figure 3: Ricardian trade with trade costs



The first of these involves balancing the numbers of workers with the number of residents and commuting patterns. Everyone must live somewhere, work somewhere, and commuting flows must synchronize these two distributions.

The second involves ensuring that the land market clears. Firms and residents must pay the same price for floor space, and total payments for floor space by households and firms must match to total rent received by landlords.

With preferences, production and market clearing conditions specified, it is now possible to calibrate the model. This process is conceptually similar to what we described above in the three location discrete analog to the linear city model, albeit considerably more complicated. This leads to all estimates for all of the exogenous parameters of the model. This, in turn, permits the evaluation of counterfactual transportation policies, really numerical comparative statics. In particular, it allows the authors to investigate what London would look like in the absence of its subway. Not too surprisingly, these counterfactuals suggest that London's subway played an important role in determining the economic geography of the city.

D *Ricardian Trade, gravity, and market access*

Ricardian Trade As we saw in the first lecture, heterogeneity of agents is important for understanding the welfare implications of particular patterns of location. More specifically, for the purpose of welfare, the thing that is really important is how different peoples' tastes are from one another. This means that understanding preference dispersion is important.

This same intuition is central to trade. Here, however, it is differences in comparative advantage rather than differences in taste that drive results. In a formal sense, these are similar. We can recast the decision to commute to a far away location as a decision to import labor, and the cost of commuting becomes the cost to transport labor.

Given the importance of the dispersion of preferences or comparative advantage for determining the extent to which people avail themselves of far-away goods or jobs, this section discusses the most common way of learning about such dispersion, the 'gravity regression'.

It is easiest to proceed in the context of an example that describes trade in goods. To begin, consider a world that consists of two locations, $i \in \{1,2\}$. There are a continuum of goods, $c \in [0,1]$. Each location can produce all goods at unit cost $p_i(c)$, where each price is draw from a distribution of prices, $F(p)$. To simplify the problem, both locations draw from the same price distribution. Thus, each good can be produced in two locations at different prices.

Suppose each location is inhabited by a single consumer who consumes one unit of each good inelastically, but tries to minimize expenditure. That is, each consumer buys the good from the cheapest available source.

We can now consider the patterns of trade that emerge as the cost of trade between the two locations varies.

If trade between the two locations is infinitely expensive, each consumer buys all goods locally. If trade between the two locations is free, then each consumer buys each good from the cheapest source. In this case, good c is produced in location 1 if and only if $p_1(c) < p_2(c)$.

Now consider the intermediate case, where trade is possible, but each unit traded incurs cost a pairwise iceberg trade cost, $\tau \geq 1$. This means that in order for one unit of good c produced in location 2 to arrive in location 1, τ units must be shipped from location 2.

With trade costs, each good will have two prices in each location, one each for the domestic and foreign varieties. To keep track of this, let $p_{ij}(c)$ denote the price in location i of good c produced in j. With iceberg trade costs, $p_{12} = \tau p_2$ and $p_{21} = \tau p_1$. Suppose that no shipping costs are incurred for goods produced and consumed locally, so that $p_{11} = p_1$ and $p_{22} = p_2$.

For each particular good, c, the economy receives two price draws, $p_1(c)$ and $p_2(c)$. The consumers in each location buy good c from the cheapest source available to them. This gives rise to three possible cases. If $\tau p_2(c) < p_1(c)$ then $p_{12}(c) < p_{11}(c)$ and it is cheaper for the consumer in location 1 to buy good c from location 2 (and pay the trade cost) than to buy the locally produced variety. Trivially, this requires that the consumer in location 2 buy locally produced c rather than import it from location 1. Thus, if $\tau p_2(c) < p_1(c)$ then all good c consumed in locations 1 and 2 is produced in location 2. Figure 3 illustrates the range of draws for which this outcome obtains as the cone bounded on the left by the y-axis and on the right by the line $p_2 = \tau p_1$.

If $\tau p_2(c) < p_1(c)$ then we get the opposite result. All good c consumed in locations 1 and 2 is produced in location 1. Figure 3 illustrates the range of draws for which this outcome obtains as the cone bounded on the left by the line $p_1 = \tau p_2$ and below by the x-axis.

When $\tau > 1$, it is also possible that neither location has a sufficient comparative advantage in producing c to overcome the resistance of trade costs. This occurs if $\tau p_1(c) > p_2(c) > \frac{p_1(c)}{\tau}$. In this case, both locations produce good c exclusively for local consumption. The region of price draws for which this occurs is the central cone in figure 3.

We are interested to understand the interaction between productivity dispersion, i.e., the variance of F, trade costs, and patterns for trade. In particular, we would like to know whether observations of trade costs and patterns of trade can allow us to learn anything about productivity dispersion.

Consider first the case where trade costs are zero. In this case, the decision about where to buy a good is purely ordinal. If a good is cheaper in location i, buy from i. Because it doesn't matter how much cheaper, the dispersion of prices also doesn't matter.

With trade costs, this is no longer true. In particular, it is natural to suspect that trade increases as the dispersion of productivity increases. This suggests the possibility of learning about the dispersion of productivity by examining the way that the share of traded goods changes with trade costs. This is exactly the logic of 'gravity regressions'.

It is easy to make this intuition precise when prices follow a Frechet distribution. Suppose that both p_1 and p_2 are distributed according to $F(p) = e^{-Tp^{-\epsilon}}$. With iceberg trade costs, $p_{12} = \tau p_1$. It follows that p_{12} is distributed Frechet, with $G(p) = e^{-T\tau^{\epsilon}p^{-\epsilon}}$.

Let π_{12} denote the fraction of goods consumed in location 1 which are produced in region 2. Since a good is produced in region 2 and consumed in region 1 if and only if $p_{12} < p_1$. Recalling earlier results about the Frechet distribution, this means that

$$\begin{aligned} \pi_{12} &= \frac{T}{T\tau^{\epsilon} + T} \\ &= \frac{1}{\tau^{\epsilon} + 1} \end{aligned} \tag{44}$$

Since τ and π_{12} are observed, this is one equation in one unknown, we can solve this expression for ϵ. Thus, the introduction of trade costs makes it possible to learn about the dispersion of prices just from easily observable trade shares and information about transportation costs.

This is precisely the logic behind the widely gravity regression. By exploiting changes in patterns of trade in goods or commuting (trade in labor), as trade costs change but the productivity distribution does not, the gravity regression seeks to identify the dispersion of location specific productivity.

Notice that in a more complicated geography, to create a centralizing force. Implicit in the Ricardian set-up, comparative advantage, productivity shocks, are a feature of a location, not of some mobile factor that happens to be at the location. Moreover, the fact that the support of the Frechet is unbounded means that every location will always trade with every other. This is going to create an incentive for people to locate in the 'center' of the geography. This is where (usually) you will minimize your average trade cost.

Two comments about this. First, the advantage of a central location is a feature of any model of trade in which agents trade with everyone. In particular, this is also true for two for two other common foundations for trade models, 'Armington', and 'monopolistic competition'. Both frameworks, like the Ricardian set-up, require every location to trade with every other. Second, the incentive for centralization looks like an 'agglomeration economy', but it's not really. The centralizing force in trade models is a feature of the location and does not depend the amount of economic activity in the location. In this sense, at least in the Ricardian model, the centralizing force is exogenous. Once you pick your geography and shock, the importance of the centralizing force is determined. This is not how agglomeration forces work in conventional urban models. See Thisse et al. (2021) for more on this.

The Gravity Regression In our discussion of Heblich et al. (2018) we presented the following gravity equation,

$$\pi_{ni|n} = \frac{\left(\frac{w_i}{\kappa_{ni}}\right)^{\epsilon}}{\sum_{s \in R} \left(\frac{w_s}{\kappa_{ns}}\right)^{\epsilon}} \tag{45}$$

This equation motivates a regression equation whose logic is much the same as in the simple example above.

However, this gravity equation is different than the simple example of (44). The gravity equation from Heblich et al. (2018) involves pairwise commutes, trades of labor, across many different boroughs, not just two locations as in the simple example of (44). This means that pairwise commuting between i and j depends on the attractiveness of commuting to k. Thus, the gravity equation that emerges from a model based on Frechet productivity shocks allows for the possibility that commuting between i and j is limited not because people in i don't want to commute to j, but because location k is still more attractive.

In addition, with the addition of many possible origin and destination pairs, the easy analytical solution for the dispersion parameter that is possible in the two location example is no longer possible. Instead, it is common to evaluate the dispersion parameter using a regression.

To see how this works, first take logarithms of both sides of (45) to get

$$\log(\pi_{ni|n}) = \log(\left(\frac{w_i}{\kappa_{ni}}\right)^{\epsilon}) + \log(\sum_{s \in R} \left(\frac{w_s}{\kappa_{ns}}\right)^{\epsilon}). \tag{46}$$

Next, note that the second term does not vary within location n, so we can write this as

$$\log(\pi_{ni|n}) = \epsilon \log(w_i) - \epsilon \log(\kappa_{ni}) + A_n + \mu_{ni}. \tag{47}$$

In this regression, A_n is a location specific constant which reflects the denominator of (45) and μ_{ni} is an error term.

Given this formulation of the gravity equation as a regression, we can regress pairwise commute shares on a location indicator and on location specific wages and recover an estimate of the dispersion parameter as the wage coefficient. This is the most common way to estimate the dispersion parameter.

This requires three comments. First, in spite of the elaborate stochastic structure of the location choice model, there is no way to interpret the error μ as part of an individual or firm decision. It is a purely ad hoc measurement error term. Second, if we include only wages as a regressor, then the implied assumption is that the pairwise trade costs is part of the regression error. This, in turn necessitates an assumption about the orthogonality of location wages and pairwise commute costs. Third, as discussed in the context of the discrete linear city model, there is no guarantee that this approach to estimating ϵ will be logically consistent with a subsequent calibration of the model that relies on this estimate.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the berlin wall. *Econometrica*, 83(6):2127–2189.

Anderson, S. P., de Palma, A., and Thisse, J.-F. (1992). *Discrete choice theory of product differentiation*. MIT Press.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.

Baum-Snow, N. (2017). Urban transport expansions, employment decentralization, and the spatial scope of agglomeration economies. *Processed, University of Toronto*.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017). Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, 106(3):479–524.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

Fujita, M. and Ogawa, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional science and urban economics*, 12(2):161–196.

Garcia-López, M.-À., Holl, A., and Viladecans-Marsal, E. (2015). Suburbanization and highways in spain when the romans and the bourbons still shape its cities. *Journal of Urban Economics*, 85:52–67.

Gonzalez-Navarro, M. and Turner, M. A. (2018). Subways and urban growth: Evidence from earth. *Journal of Urban Economics*, 108:85–106.

Heblich, S., Redding, S. J., and Sturm, D. M. (2018). The making of the modern metropolis: evidence from london. Technical report, National Bureau of Economic Research.

LeRoy, S. F. and Sonstelie, J. (1983). Paradise lost and regained: Transportation innovation, income, and residential location. *Journal of Urban Economics*, 13(1):67–89.

Thisse, J.-F., Turner, M., and Ushchev, P. (2021). A unified theory of cities. Technical report, National Bureau of Economic Research.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.