EC1410 Topic #5.5

The Monocentric City Model with Heterogenous Agents

Matthew A. Turner
Brown University
Spring 2023

(Updated January 23, 2023)

Outline

Why worry about heterogeneity?

- The monocentric city model is not very good for thinking about heterogeneity. We can describe a few types of agents, e.g., rich and poor, but not more. But we think that heterogeneity is of first order importance for many economic questions.
- A central concern of many econometric exercises is whether unobserved person level characteristics are correlated with the outcome of interest. The monocentric city model can't speak to this issue because there is no household level heterogeneity. This can create a problem if we want to use the monocentric city model to inform econometric exercises.

- Unobserved person level heterogeneity is likely to be important economically. If all of the unobservably productive people decide to live in San Francisco, this has important implications for policy, e.g., trying to create a tech hub in North Dakota.
- It turns out that allowing agents to be heterogenous also has important implications for welfare, too.

So, we want to try to figure out what the monocentric city model looks like when there are a continuum of different types of people.

Heterogeneity and discrete space

To allow for a continuum of agent types, we need to assume that space consists of discrete locations, rather than a continuum. This raises two issues.

- How do we describe distance between a bunch of discrete places? Can we just use a coordinate to describe distance from the center as we did for the monocentric city model?
- How do we think about the household optimization problem? If the set of locations is no longer a continuum, calculus stops working: calculus is about finding the top of nice smooth hill.

The discrete choice problem

Suppose there are two locations, $i = 1, 2$ and a single household that obtains utility V_1 and V_2 from the two locations.

This household's discrete choice problem is

$$max\{V_1, V_2\}$$

This is easy. Choose your favorite. This replaces the calculus problem we solved with continuous space.

If households make a choice in each location, then the discrete choice only partially replaces the calculus problem. For example, if households choose housing and consumption once they choose a location, then the household problem is

$$max\{\max_{c_1,h_1} u(c_1, h_1), \max_{c_2,h_2} u(c_2, h_2)\}$$

The logic of this problem is the same if households choose among many discrete alternatives (the case we're interested in) rather than two, there is just more notation.

Now suppose there are many households indexed by j.(Formally, we want a continuum of households. In the jargon, 'a measure' of households, which means a 'length' of households – this is to get around some of the stranger properties of the real numbers.)

Denote household j's value of location i by V_{ij}. Suppose that for each location, each V_{ij} has a systematic or common component, u_i, and an idiosyncratic component that is particular to each household, ε_{ij}, with $V_{1j} = u_1\varepsilon_{1j}$ and $V_{2j} = u_2\varepsilon_{2j}$ for all households j.

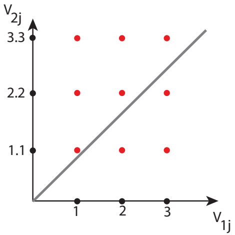
Example: Suppose that $\varepsilon_{ij} \in \{\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}\}$ with each ε_{ij} is equally likely. This means that there are nine different types of households and each type makes up 1/9th of the population.

If $\{\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}\} = \{1, 2, 3\}$ and that $u_1 = 1$ and $u_2 = 1.1$. Then $V_{1j} \in \{1 \times 1, 1 \times 2, 1 \times 3\} = \{1, 2, 3\}$ and $V_{2j} \in \{1.1, 2.2, 3.3\}$. Each household has a valuation for each location (V_{1j}, V_{2j}) drawn from these sets.

Note that location 2 is 'better' than location 1 in the sense that given equal idiosyncratic values for the two locations, households always choose location 2.

To solve our model, we're going to need to be able to figure out the share of agents who choose location 1 and 2. That is, the share s_1 with $V_{1j} = max\{V_{1j}, V_{2j}\}$ and s_2 with $V_{2j} = max\{V_{1j}, V_{2j}\}$. NB: Since $s_1 + s_2 = 1$, its enough to find just one of them.

To do this, consider the following graph,



The pair of payoffs for each of the nine types is a coordinate in (V_{1j}, V_{2j}) space. We want the share of households with $V_{2j} > V_{1j}$.

$V_{2j} > V_{1j}$ holds if an only if we are to the left of the gray line where $V_{1j} = V_{2j}$. Since each dot is equally likely, this occurs 6 times in 9, or for 2/3 of the households. So 2/3 of households choose 2, and 1/3 choose 1.

If let $f(1j, 2j) = 1/9$ denote the share of households of each type (V_{1j}, V_{2j}), then we can write sum all of the households with types left of the gray line in the figure as,

$$
\begin{aligned}
s_2 &= \sum_{j=1}^{3} f(11, 2j) + \sum_{j=2}^{3} f(12, 2j) + \sum_{j=3}^{3} f(13, 2j) \\
&= \sum_{k=1}^{3} \left[\sum_{j=k}^{3} f(1k, 2j) \right] = \sum_{k=1}^{3} \left[\sum_{j=k}^{3} 1/9 \right]
\end{aligned}
$$

The second line is just a compact and conventional rewrite of the first.

We've stated this outcome in terms of a toy example where each household can have only three idiosyncratic values for each location. Nothing prevents us from allowing households to have more values for each location, or from allowing the different dots/types to have different probabilities/shares, i.e., $f(1j, 2j) \in (0, 1)$.

If each household has N possible valuations for each location (so we have N^2 types of households), then the share of households choosing location 2 is

$$s_2 = \sum_{k=1}^{N} \left[\sum_{j=k}^{N} f(1k, 2j) \right]$$

We can also allow for a continuum of possible valuations for each household for each locations. In this case, the ε_{ij} can take a continuum of values for each location and we need to use integration rather than summation,

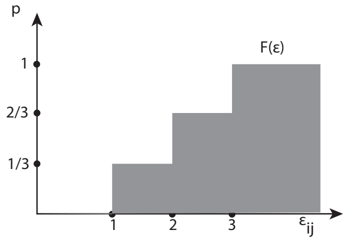
$$s_2 = \int_0^\infty \int_k^\infty f(1k, 2j) djdk,$$

It turns out that evaluating this integral is hard analytically and numerically. The share of households choosing any particular alternative in a discrete choice problem is hard to evaluate analytically once you move away from toy examples. Because integration is hard for computers, it is also hard to evaluate numerically.

Extreme Value Distributions

But there is a special case. If the ε_{ij} follow an 'extreme value distribution', then the shares of households choosing each outcome has an easy analytic solution.

In our simple example, ε_{ij} takes the values 1,2,3 with equal probability. Another way to represent this is with a probability distribution function that reports the share of realizations of ε_{ij} that are less than any given value.



In the jargon, this function is a 'Cumulative Distribution function (CDF)' or 'Probability Distribution Function' for ε. We can write it analytically as

$$F(\varepsilon) = \begin{cases} 0 & \text{if} \quad \varepsilon < 1 \\ 1/3 & \text{if} \quad 1 \leq \varepsilon < 2 \\ 2/3 & \text{if} \quad 2 \leq \varepsilon < 3 \\ 1 & \text{if} \quad 3 \leq \varepsilon \end{cases}$$

F satisfies $Prob(\varepsilon < \overline{\varepsilon}) = F(\overline{\varepsilon})$ or equivalently, the share of households with $\varepsilon < \overline{\varepsilon}$ is $F(\overline{\varepsilon})$.

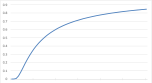
There are two main extreme value distriutions, ‘Gumbel’ and ‘Frechet’. They behave similarly. I will talk about Frechet.

If ε is determined by a Frechet distribution, then its Cumulative Distribution function is

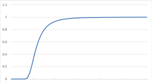
$$F(\varepsilon) = e^{-T\varepsilon^{-\theta}}.$$

This distribution is governed by two parameters, T, ‘level’, and θ, ‘dispersion’. These names are suggestive of ‘mean’ and ‘variance’ and are often used in the same spirit (The Frechet distribution is usually defined for $T > 0, \theta > 1$.)

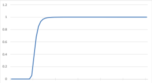
This is what F looks like for a few values of T and θ



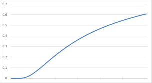
$\theta = 1$ and $T = 1$



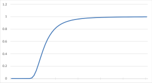
$\theta = 4$ and $T = 1$



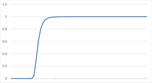
$\theta = 10$ and $T = 1$



$\theta = 1$ and $T = 3$



$\theta = 4$ and $T = 3$



$\theta = 10$ and $T = 3$

Note that as θ increases F goes to step function, which means that all agents have the same ε and there is no heterogeneity.

Big Theorem

Suppose that households choose among N discrete locations. For each location $i = 1, ...N$, household j receives payoff $\varepsilon_{ij} u_i$, and ε_{ij} is drawn from a Frechet distribution, $F(\varepsilon) = e^{-T\varepsilon^{-\theta}}$.

Then the share of household such that

$$V_{ij} = max\{V_{1j}, V_{2j}, ..., V_{Nj}\}$$

is

$$s_i = \frac{u_i^{-\theta}}{\sum_{k=1}^{N} u_k^{-\theta}}.$$

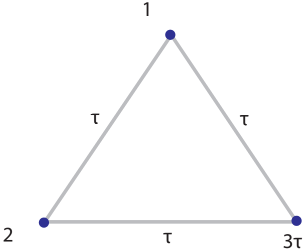
If individual heterogeneity is described by multiplicative Frechet shocks, the shares of households that choose each location are a simple formula of the systematic part of households' payoffs at each location.

This is a big step towards a model of cities with a continuum of types of agents.

Distance in Discrete Space

Given three locations, $i = 1, 2, 3$, there are three possible pairs, $ii' \in \{12, 23, 31\}$. Denote the travel cost to go between any two locations as $\tau_{ii'}$ (implicitly travel from 1 to 2 is the same as from 2 to 1).

Example 1: Suppose our three locations are the vertices of an equilateral triangle and that the cost to travel a unit distance is constant. What are the τ_{ij}'s?

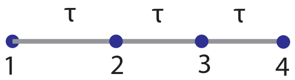


We often assume that costs enter the problem multiplicatively. This is called ‘iceberg transportation costs’. The idea is that if you have τ_{ij} units of value in location *i* and you transport it to *j*, some of it melts, and you are left with 1 unit of value in *j*. This requires that $\tau > 1$ unless you are from *i* to *i* and $\tau = 1$.

This means that we can write a matrix of travel costs for this set of three locations as

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{bmatrix} = \begin{bmatrix} 1 & \tau & \tau \\ \tau & 1 & \tau \\ \tau & \tau & 1 \end{bmatrix}$$

Example 2: Consider four equally spaced locations on a line.



If the cost to travel between adjacent locations is τ, then the cost to travel two units of distance is τ^2 and so on.

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} & \tau_{14} \\ \tau_{21} & \tau_{22} & \tau_{23} & \tau_{24} \\ \tau_{31} & \tau_{32} & \tau_{33} & \tau_{34} \\ \tau_{41} & \tau_{42} & \tau_{43} & \tau_{44} \end{bmatrix} = \begin{bmatrix} 1 & \tau & \tau^2 & \tau^3 \\ \tau & 1 & \tau & \tau^2 \\ \tau^2 & \tau & 1 & \tau \\ \tau^3 & \tau^2 & \tau & 1 \end{bmatrix}$$

Example 3: Given a map of a city with N, e.g., census tracts, one can describe travel costs in the city by calculating the travel costs between each of the possible pairs of census tracts. This results in an empirically founded $N \times N$ matrix of iceberg commute costs. This means that we can use the discrete choice technology to describe choices over real cities, not the stylized examples we've considered before.

A Discrete city with heterogeneous agents

Consider a discrete linear city with three neighborhoods $i \in \{1, 2, 3\}$. Let x_i denote a neighborhood's distance from the CBD, with $x_1 = 1$, $x_2 = 2$, $x_3 = 3$.

The cost to commute one unit distance is τ.

The city is populated by households indexed by j. Each household chooses a neighborhood i, pays land rent R_i, and commutes to the center, at location 0, to earn wage w.

A household's utility is $V_{ij} = c_i z_{ij}$ where c_i is consumption and z_{ij} is a household and location specific valuation. All z_{ij} are drawn from a Frechet distribution, $F(z) = e^{-Tz^{-\epsilon}}$.

A household budget is $w = c_i + R_i + i\tau$, so $c_i = w - R_i - i\tau$. Thus,

$$V_{ij} = [w - R_i - i\tau]z_{ij},$$

and households make the discrete choice

$$max\{V_{1j}, V_{2j}, V_{3j}\}.$$

Applying the *Big Theorem* we have

$$\begin{aligned}
s_1 &= \frac{[w - R_1 - 1\tau]^\epsilon}{\sum_{k=1}^{3}[w - R_k - k\tau]^\epsilon} \\
s_2 &= \frac{[w - R_2 - 2\tau]^\epsilon}{\sum_{k=1}^{3}[w - R_k - k\tau]^\epsilon} \\
s_3 &= \frac{[w - R_3 - 3\tau]^\epsilon}{\sum_{k=1}^{3}[w - R_k - k\tau]^\epsilon}.
\end{aligned}$$

This is three equations in 8 unknowns $\{s_1, s_2, s_3, R_1, R_2, R_3, \varepsilon, \tau\}$, so we can't solve them without more information (probably – they are non-linear equations).

Suppose that (in the spirit of the continuous space monocentric city model) each location is occupied by exactly one third of the population, so that $s_i = 1/3$, and the R_i are not observed.

Then

$$\frac{1}{3} = \frac{[w - R_1 - 1\tau]^\epsilon}{\sum_{k=1}^{3}[w - R_k - k\tau]^\epsilon}$$
$$\frac{1}{3} = \frac{[w - R_2 - 2\tau]^\epsilon}{\sum_{k=1}^{3}[w - R_k - k\tau]^\epsilon}$$
$$\frac{1}{3} = \frac{[w - R_3 - 3\tau]^\epsilon}{\sum_{k=1}^{3}[w - R_k - k\tau]^\epsilon}$$

Then, because the denominators are all the same,

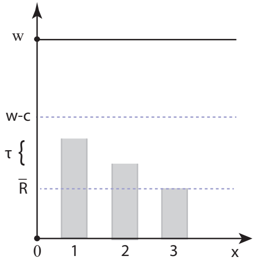
$$[w - R_1 - 1\tau] = [w - R_2 - 2\tau] = [w - R_3 - 3\tau].$$

Which implies that $R_1 - R_2 = \tau$ and $R_2 - R_3 = \tau$.

That is, the land rent gradient decreases at the same rate as commute costs increase, just like the continuous version of the model.

Suppose we also require that land rent at $x = 3$ be equal to the observed agricultural land rent, $\overline{R}$ and that τ is known.

Then we have $(R_1, R_2, R_3) = (\overline{R} + 2\tau, \overline{R} + \tau, \overline{R})$, and this lets us solve for consumption $(c_1, c_2, c_3) = (\overline{c}, \overline{c}, \overline{c})$, just as in the monocentric city model.



Comments:

- Suppose that instead of observing $\overline{R}$, I observed $\overline{c}$? What if I observe both?
- Note that there is an implicit location 0 where people work, but don't live.

- Note that transportation costs are additive here, not multiplicative as above. What is the transportation cost matrix? Did I use all of it?
- This model is better suited to empirical applications than the model with continuous space. If I observed different shares, or if the locations were not evenly spaced, I could still have solved the problem. Much of the current research activity in urban economics revolves around calibrating or estimating more complicated versions of this model (more to follow).

Welfare

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 because they have different z_{ij}'s.

This means that calculating welfare in the discrete case is more difficult. We must calculate both land rent and consumer surplus.

A second big theorem lets us calculate the expected/average utility of a household living in this city. This expectation is,

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

where the 'Gamma function', $\Gamma\left(\frac{\epsilon-1}{\epsilon}\right)$ is a generalization of the factorial operator '!' to the real numbers; $\Gamma(n) = n!$ for counting number n.

This does not account for land rent. Figuring out how to evaluate welfare in these models is an open question.

London and the Tube, 1866-1921

Heblich et al. (2020) is an early application of the discrete cities model. They use it to understand how the construction of the London Underground reorganized the economic geography of London.

London was one of the first cities in the world to build a subway system. Construction began in the 1860’s and a substantial network was in place by the 1920’s.

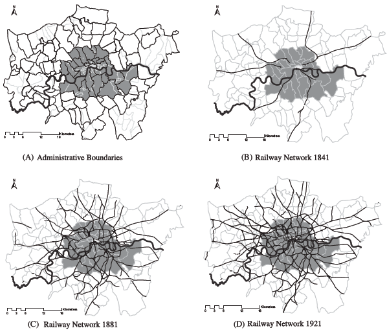


FIGURE I
Administrative Boundaries and the Railway Network over Time

Heblich et al. (2020)

- The ‘City of London’ is the central part of Greater London. During the time that subways were constructed, population intake City of London declined by about 80% while the population of Greater London increased by a factor of about 6.

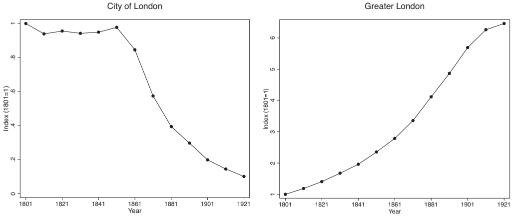


FIGURE II

Population Indexes over Time (City of London and Greater London, $1801 = 1$)

At the same time, employment in the City of London skyrocketed, and the value of central land as a share of the total decreased and then increased.

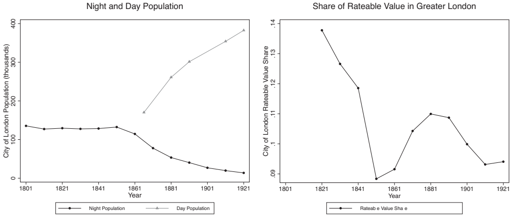


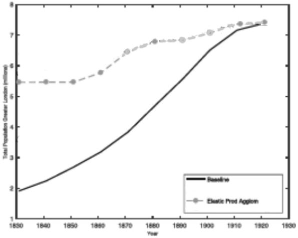
FIGURE III

City of London Day and Night Population and Rateable Value Share over Time

- This suggests that the subway facilitated a separation of work and residence, and increased concentration of central employment, likely with productivity increases resulting from agglomeration economies.
- Heblich et al. (2020) develop a discrete city model to see if this hypothesis is consistent with the economic forces in this model. This model is different from everything we have studied in an important way. It lets people choose where to work and where to live, instead of just where to live. This means that the model also makes predictions about where production occurs. This is an important advance (pioneered in Ahlfeldt et al. (2015).

- Commuting costs are described by a matrix of pairwise iceberg commuting costs between each of about 55 buroughs that make up greater London. This means that the transportation cost matrix is 55×55.
- Iceberg costs, τ_{ij} are estimated using GIS software and guesses about the speed of travel by foot and subway.
- Households chose a workplace-residence pair, and get utility $z_{ij}\frac{u_{ij}}{\tau_{ij}}$ where i is residence ad j is workplace. Note that commute costs are multiplicative. The z_{ij} is a household specific valuation of workplace-residence pair ij drawn from a Frechet distribution. Thus, the household problem involves a discrete choice over 55×55 alternative workplace-residence pairs.

- Households choose housing and consumption for each *ij*, so each of the 55×55 possibilities also involves solving a calculus problem, and can involve a different wage, if wages are different in different workplaces.
- Going through the details of the estimation are beyond the scope of this class
- Once the model is estimated, it can be used to evaluate comparative statics, here called ‘counterfactuals’, such as ‘what would London look like if it had not built the subway?’
- Here is one of their main results. The dark line is ‘with the subway’, the light one is ‘without’. This shows that the city grows much faster with the subway than in the counterfactual case where it did not build the subway (???? FIX)


Baseline
Elastic Prod Agglom
Year

Conclusion I

- We can extend the basic monocentric city model to an environment with discrete space and a continuum of types of households.
- This leads to formulations of the monocentric city model in which we can think about the role of heterogeneity in a more sophticated way, an in particular in a way that is rich enough to allow us to think about issues of omitted variable bias that are so prominent in most econometric analyses.
- It also leads to model which can immediately serve as a basis for calibration exercises. This allows us to, if we believe the models, to evaluate real world counterfactuals, like 'what would happen if London did not have a subway?'

Conclusion II

- This framework depends crucially on the assumption that individual heterogenity is described by an extreme value distribution. This is not assumption for which there is much evidence one way or the other.

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.

Heblich, S., Redding, S. J., and Sturm, D. M. (2020). The making of the modern metropolis: evidence from london. *The Quarterly Journal of Economics*, 135(4):2059–2133.