Coding Assignment

REAL9440 QSMs for Urban Economics
Chris Severen

August 26, 2025
DUE September 12, 2025

Preliminaries

Use R, Python, or Julia. You may use Stata for data wrangling and estimation (Tasks 1-5), but the remaining tasks will be challenging to This flow data will be relatively large, please ensure you have sufficient storage space and RAM. I will provide a way to upload your replication package.

Each of you should code your own assignment. It may be useful to discuss what you do with your classmates, but I highly recommend attempting to code everything yourself.

Tasks

Select an county or group of counties to use as your primary analysis zone. Note this in your submission.

1. Obtain bilateral commuting flow data from either LEHD LODES or a recent CTPP for your analysis zone, along with supporting data on the locations of the tracts or blocks underlying the data. Aggregate the commuting flow data to the tract level if not already at the tract level.

2. Calculate the distance between all pairs of locations. [Bonus: Calculate travel times between all pairs of locations.]

3. Using a linear model, estimate:

$$\ln(N_{ij}) = \theta_i + \lambda_j + \kappa d_{ij} + e_{ij}, \tag{1}$$

   where N_{ij} is the flow from residential location i to workplace j, d_{ij} is distance or travel time, and κ is the semi-elasticity of travel time. In your code, note the estimate of κ. Retain estimates of the fixed effects θ_i and λ_j.

4. Make sure to extend your sample to include all ij observations, including those with zero flows. Then, using PPML, estimate:

$$\ln(\mathbb{E}[N_{ij}]) = \theta_i + \lambda_j + \kappa d_{ij}. \tag{2}$$

In your code, note the estimate of κ and comment on whether the estimates between Equation (1) and Equation (2) differ, and if so, why? Retain estimates of the fixed effects θ_i and λ_j.

5. Revisit the estimation in Tasks 3-4. Your model presumes that the distance (or time) traveled is zero. Suppose that you believe this is incorrect and could induce bias. Brainstorm and implement at least two feasible ways to correct for this.

6. Compare the fixed effects estimated from the different models in Tasks 3-5.

7. Using $\epsilon = 3.2$, your estimate of κ from Task 4, and $\Theta_i = \exp(\hat{\theta}_i)$ and $\Lambda_j = \exp(\hat{\lambda}_j)$ from Task 4, calculate Residential and Workplace Market Access terms, defined respectively as:

$$\Phi_{Ri} = \sum_s \Lambda_s^\epsilon e^{-\kappa\epsilon d_{is}} \tag{3}$$

$$\Phi_{Wj} = \sum_r \Theta_r^\epsilon e^{-\kappa\epsilon d_{rj}}. \tag{4}$$

8. Using $\epsilon = 3.2$, your estimate of κ from Task 4, and $N_{Ri} = \sum_s N_{is}$ and $N_{Wj} = \sum_r N_{rj}$, write a script that solves for $\{\Phi_{Ri}, \Phi_{Wj}\}_{\forall i,j}$ as the solution to the fixed point problem represented by the system of equations:

$$\Phi_{Ri} = \sum_s e^{-\kappa\epsilon d_{is}} \frac{N_s}{\Phi_{Ws}} \tag{5}$$

$$\Phi_{Wj} = \sum_r e^{-\kappa\epsilon d_{rj}} \frac{N_r}{\Phi_{Rr}}. \tag{6}$$

9. Compare the Market Access terms created in Task 7 with those in Task 8. Are there any differences?

10. Section 5 and Appendix F of Brinkman & Lin 2024 detail how to recover productivity and amenity terms, as well as wages and housing prices, in an urban QSM from residential population, workplace population, travel costs, and land area. Make the simplifying assumption that land area is the same for all tracts. Write code that finds the productivity, amenity, and housing price terms using your data and estimates of κ, maintaining the assumption that $\epsilon = 3.2$. Produce a map of the housing price estimates. Do these seem reasonable?

Submission

Submit (1) a PDF document in which you report and discuss the results of Tasks 3-9, and (2) a replication package including a readme file that enables me to generate your results.