Discussion of "Consumption Zones"

by

Batch, Bridgman, Dunn, Gholizadeh

Emi Nakamura, UC Berkeley

Goals

- "Commuting zones" designed to model competition in labor markets
  - Geography of consumption potentially totally different
- Develop new unit of analysis for consumption behavior: "Consumption Zones"
  - Clustered by regions in which consumption is most substitutable
  - Use to estimate concentration within ConZ's as opposed to CZ's
- Methodology loosely based on CZ methodology introduced by Tolbert & Sizer (1987) to summarize commuting behavior

Data

- Exciting new dataset from Fiserv
  - Large card transaction intermediary
  - County level data
  - Aggregated/ anonymized by third party (Palantir)
    - Some data suppression to preserve privacy
- Beyond this specific project, gives hint of how big data might be used by BEA, BLS, Census etc. going forward
  - Very important initiative

More on Data

- Cardholder's home location estimated based on the transaction history of the consumer using information on all transactions across all industries in Fiserv database (for card)
  - 15 NAICS industries accounting for 79% of personal consumption spending
  - Excludes E-commerce (poor coverage)
- Every county-industry: estimate share of revenues for establishments in that county coming from consumers *in that county*

Missing Data

- Goal: Generate transaction flows in nearly *all* areas where final goods are sold across the United States
  - Must impute missing flows
- Imputation using "flexible models"
  - e.g. impute flows for missing industry based on observed flows for other industries in the county and various other covariates
  - Important aspect of procedure (perhaps deserves more discussion)

Findings: Home Bias in Consumption

- Approx 68% of expenditures take place in county where individuals reside
  - 87% within 100 mile radius!
- Lots of variation:

Table 1. Distribution of Spending Share From Consumers that Reside in the Same County as the Firm

	Median	10th	25th	75th	90th
Accommodation (NAICS 721)	0.152	0.058	0.104	0.215	0.312
Ambulatory Health Care Services (NAICS 621)	0.760	0.563	0.664	0.873	0.939
Amusement, Gambling, and Recreation Industries (NAICS 713)	0.509	0.233	0.353	0.665	0.788
Building Material and Garden Equipment and Supplies Dealers (NAICS 444)	0.824	0.639	0.738	0.893	0.941

Two Questions

1. Size of CZs vs ConZs
   - Application to concentration
2. Role of Symmetry

Size of CZs vs ConZs

- Paper's Intuition:
  - Because employment has a large effect on income, individuals may be willing to travel far for better job opportunities
  - In contrast, individual consumption items are a small share of an budget, so less likely to travel
  - Suggests that CZs should generally be larger than ConZs
- Is this true when products can be "delivered" cheaply? E.g., Amazon
  - How is "location" of good determined?
  - Key for online sales (which are excluded here)

Application to Concentration

- HHI used as cut-off by DoJ for allowing horizontal mergers
- How do HHI's differ in ConZs vs CZs?
  - HHIs lower in ConZs
  - Largest gap for products with infrequent purchases/wide geography
  - Not obvious (right)?
    - 1235 ConZs vs 810 CZs
    - If ConZs are generally smaller, would have expected *higher* concentration all else equal
    - But result actually goes the opposite way

CZs vs ConZs

- Distinct geographies
  - 50% of ConZs contain two or more CZs
  - Perhaps some differences arise from changes over time (old definitions of CZs?)
  - For very local industries (e.g. grocery) ConZs may actually obscure even smaller consumption zones within counties

Role of Symmetry in ConZs

- Dissimilarity matrix:

$$D_{i,j} = 1 - \frac{f_{i,j} + f_{j,i}}{min(Rev_i, Rev_j)},$$

- f_ij is spending flow from i to j
- Use to group together counties that have a lot of spending flows between them
- Symmetry between buying and selling!

Role of Symmetry in ConZs

- "Dissimilarity" matrix assumes consumption flows both ways
  - Is that always the case?
- Suppose Regions C and D contain interchangeable sellers
  - Herfindahl index should group C/D together since stores compete (I think)
- But what if C purchases only from itself; same for D
  - Would this asymmetry make C and D less likely to be grouped together?