



Willingness to Travel with Endogenous Distance: Evidence from the Changing Retail Landscape

Yue Cao, Judy Chevalier, Jessie Handbury, Hayden Parsley, Kevin R Williams

February 2026

Spatial data is fundamental to resolving various economic and policy-related questions

Examples:

- Education: Which school to send my children to?
- Healthcare: How do I select which doctors to visit?
- Retail: Where do I want to shop?

We frequently infer an individual's value for a service or amenity by measuring willingness to travel to it

- Locations of services, supplier, or amenities are typically treated as predetermined

Point of the Paper

We argue that measuring willingness to travel requires addressing the endogenous location-setting behavior of suppliers:

- More complex than standard price endogeneity (individual distances)
- Introduce a novel instrument to address distance endogeneity
- Apply it to study consumer value for retail opportunities

Retail upheaval has been characterized as the “retail apocalypse”

- Accelerated movement to e-commerce; welfare gains skew to high-income households
- Shift from traditional B&M retail to a hybrid physical/online retail environment may leave some (low-income) shoppers with reduced access

Contributions of the paper

1. Document new facts on consumer shopping patterns for general merchandise, brick-and-mortar chains

2. Estimate a model of retail demand by income and region

- Use cell phone data covering trips regardless of method of payment choice
- Instrument for distance exploiting preference externalities and chain entry behavior; results in 37% (43%) higher travel costs for income quartile 1 (4)

3. Quantify welfare impacts of long-run changes in retail opportunities

- Welfare per trip to general merchandise stores did not significantly decline from 2010 to 2019
- Ignoring distance endogeneity results in surplus losses (to lower-income households) consistent with the “retail apocalypse” narrative

Data on consumer shopping behavior

- Smartphone data on visits to polygons, home locations, and household income from Precisely PlacelQ (Couture et. al., 2022)
- Focus specifically on visits to general merchandise retailers (NAICS 452)
 - Department stores: e.g., Nordstrom, Sears, Kohls
 - Discount department stores: e.g., Ross, Burlington Coat
 - Warehouse clubs: e.g., BJ's, Sam's Club
 - Supercenters: e.g., Walmart, Target
 - Dollar Stores: Dollar Tree, Family Dollar, Dollar General
 - All identified NAICS 452 establishments included in our analysis, e.g., Von Maur, Boscov's, etc.
- We select 18 of the top 20 CBSAs / 2.7 million smartphones
 - Data cleaning: time in store; delivery?; distance around home, etc.

Advantages and disadvantages of using smartphone data

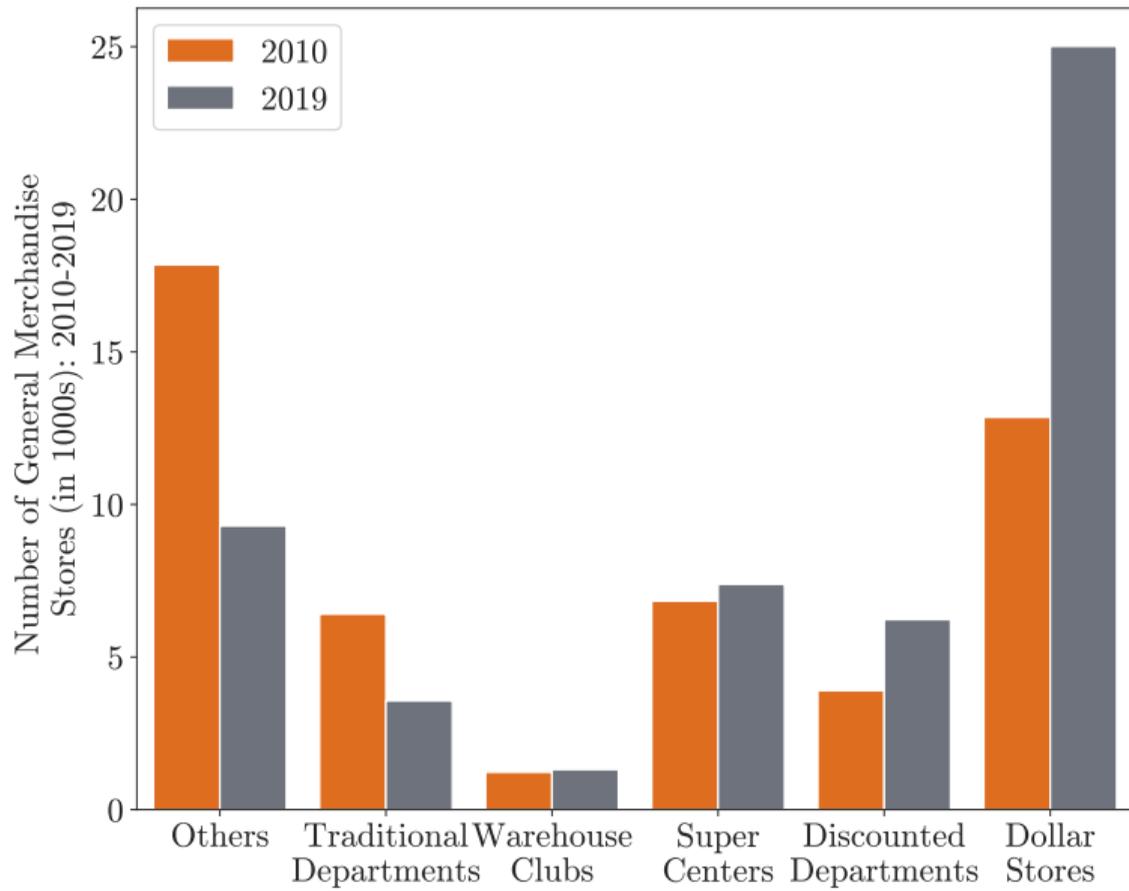
Smartphone data are broadly representative of the US population (Chen and Pope, 2020; Couture et. al., 2022)

- 30 percent of retail transactions were in cash in 2019
- Consumer propensity to have and use credit cards is smaller for low-income households (2019 Survey of Consumer Payment Choice)
- Retail datasets don't cover many general merchandise categories
- Smartphone data agreements permit disclosure of retailer identities

We can't completely get welfare

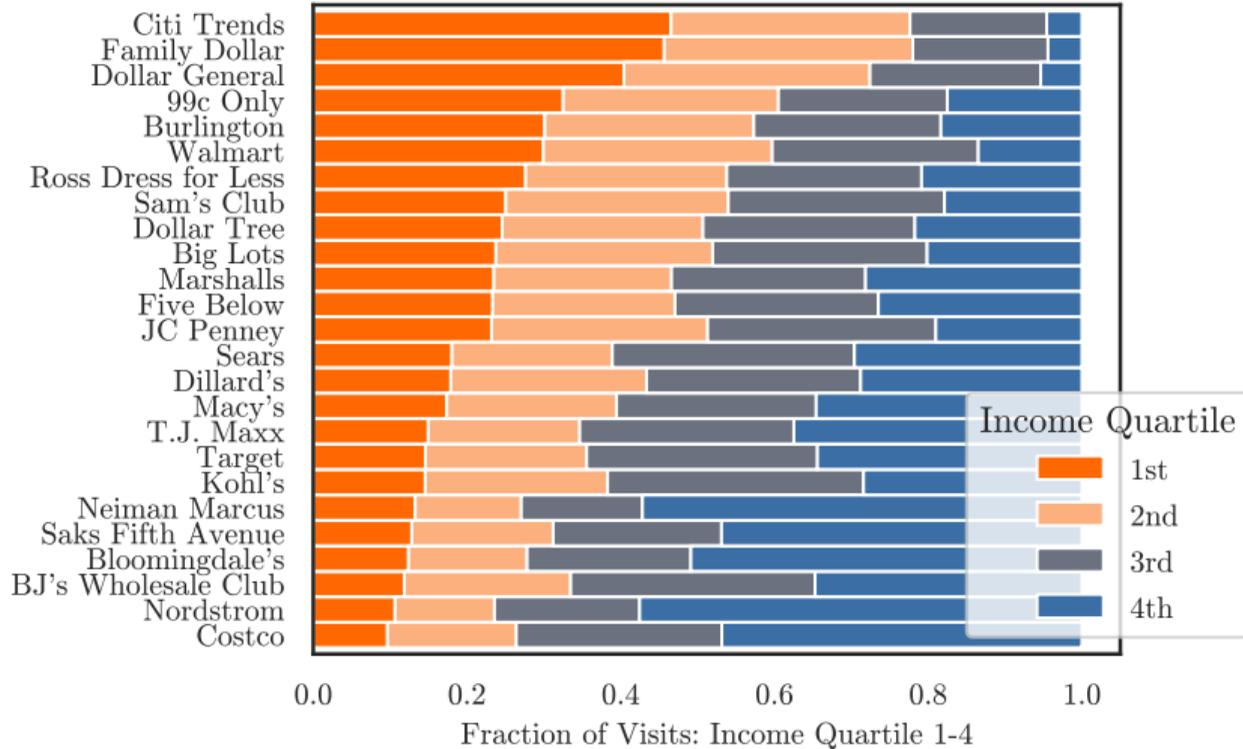
- Data do not cover online behavior
- We can't observe individual purchases, expenditures, or prices
- Data only available recently; data not as useful going forward
- We can only estimate retailer demand in 2019

Fact 1: General merchandisers have grown, led by dollar stores



Fact 2: Income gradient across chains in line with expectations

Visit shares to general merchandisers by income



Model of consumer brick-and-mortar shopping choices

A consumer i gets indirect utility from visiting store j :

$$u_{i,j} = \begin{cases} v_{i,j} \left(x_j, \text{dist}_{i,j}; \beta_{y(i)} \right) + \xi_{i,j} + \varepsilon_{i,j} & j \in J_i^c \leftarrow \text{stores from 25 identified chains} \\ \Gamma(J_i^f) + \varepsilon_{i,0} & j \in J_i^f \leftarrow \text{fringe stores (regional chains)} \end{cases}$$

- $v_{i,j} \left(x_j, \text{dist}_{i,j}; \beta_{y(i)} \right)$ and $\Gamma(J_i^f)$: deterministic value of visiting a chain and fringe store
- $\xi_{i,j}$: unobserved heterogeneous preferences over chains, potentially correlated with distance
- $\varepsilon_{i,j}$: idiosyncratic component of consumer's indirect utility, assumed iid type-1 extreme value

Deterministic utility components for inside chains

If consumer i visits an identified chain, the deterministic part of their utility is:

$$v_{i,j} \left(x_j, \text{dist}_{i,j}; \beta_{y(i)} \right) = \underbrace{\beta_{y(i),j}}_{\text{chain amenity}} + \underbrace{\beta_{y(i)}^{d1} \text{dist}_{i,j} + \beta_{y(i)}^{d2} \text{density}_{i,j}}_{\text{travel cost}}$$

- $\beta_{y(i),j}$: income quartile $y(i)$'s quality assessment of store j
- $\text{dist}_{i,j}$: distance in miles from the consumer i 's home to store j
- $\text{density}_{i,j} = \ln(1 + \# \text{ stores within } 0.1 \text{ mile of store } j \text{ in } i\text{'s choice set})$: trip-chaining adjustment

We collapse j to be a representative store for a given chain (nearly 78% of visits are to nearest outlet), taking mean-weighted distance and density across chain outlets.

Deterministic utility components for fringe stores

If consumer i visits a store in the fringe, the deterministic part of their utility is:

$$\Gamma(J_i^f) = \omega_{y(i)} \log \left(|J_i^f| + 1 \right)$$

- $\omega_{y(i)}$ captures the income-quartile specific taste for fringe stores
- $|J_i^f|$ is the total number of stores available in the fringe for consumer i
- Why not have the outside option be to stay home and use, e.g., BLP (1995)?
41% zeros at zip code-week level

Identification of key demand parameters

Disutility from distance:

- Propensity of individuals within an income quartile to visit more proximate versus more distant stores.

Chain preferences:

- Relative frequency with which consumers visit stores from each chain
- Do they pass a Dollar General to go to a Walmart?
They must like Walmart more!
- Interpret the income-specific taste as the taste for the price-assortment characteristics of the store

But, there is an endogeneity problem

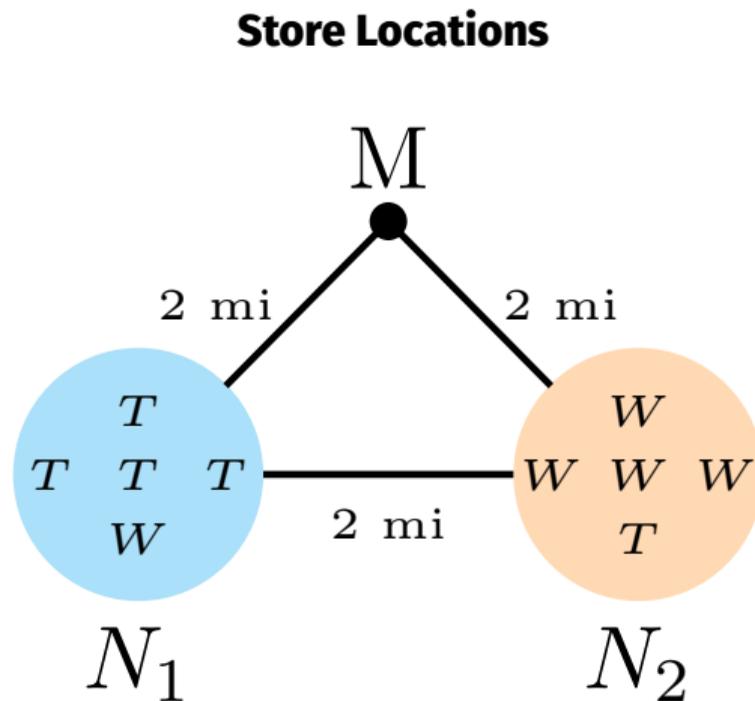
Consider a world where there are “Target” (T) types and “Walmart” types (W)

Suppose that:

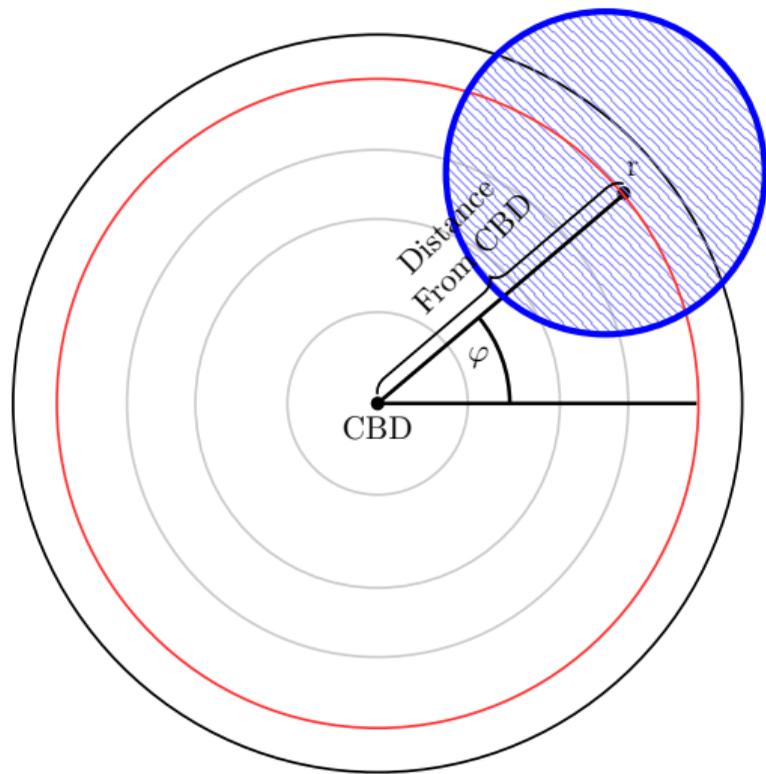
- T will travel 2 miles to a Target, but 0 miles to Walmart
- W will travel 2 miles to a Walmart, but 0 miles to Target
- There is no Walmart nearby

N_2 may be the profit-maximizing Target site, further from T 's.

Strategic siting will generate correlations between $\xi_{i,j}$ and $dist_{i,j}$.
Bias unclear

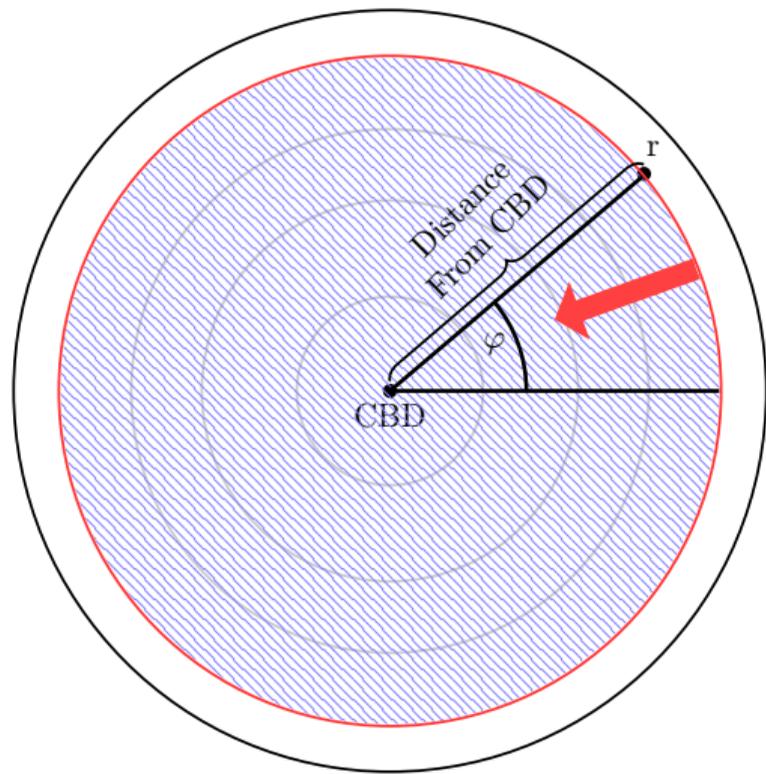


IV Strategy: “Neighborhood Instruments”



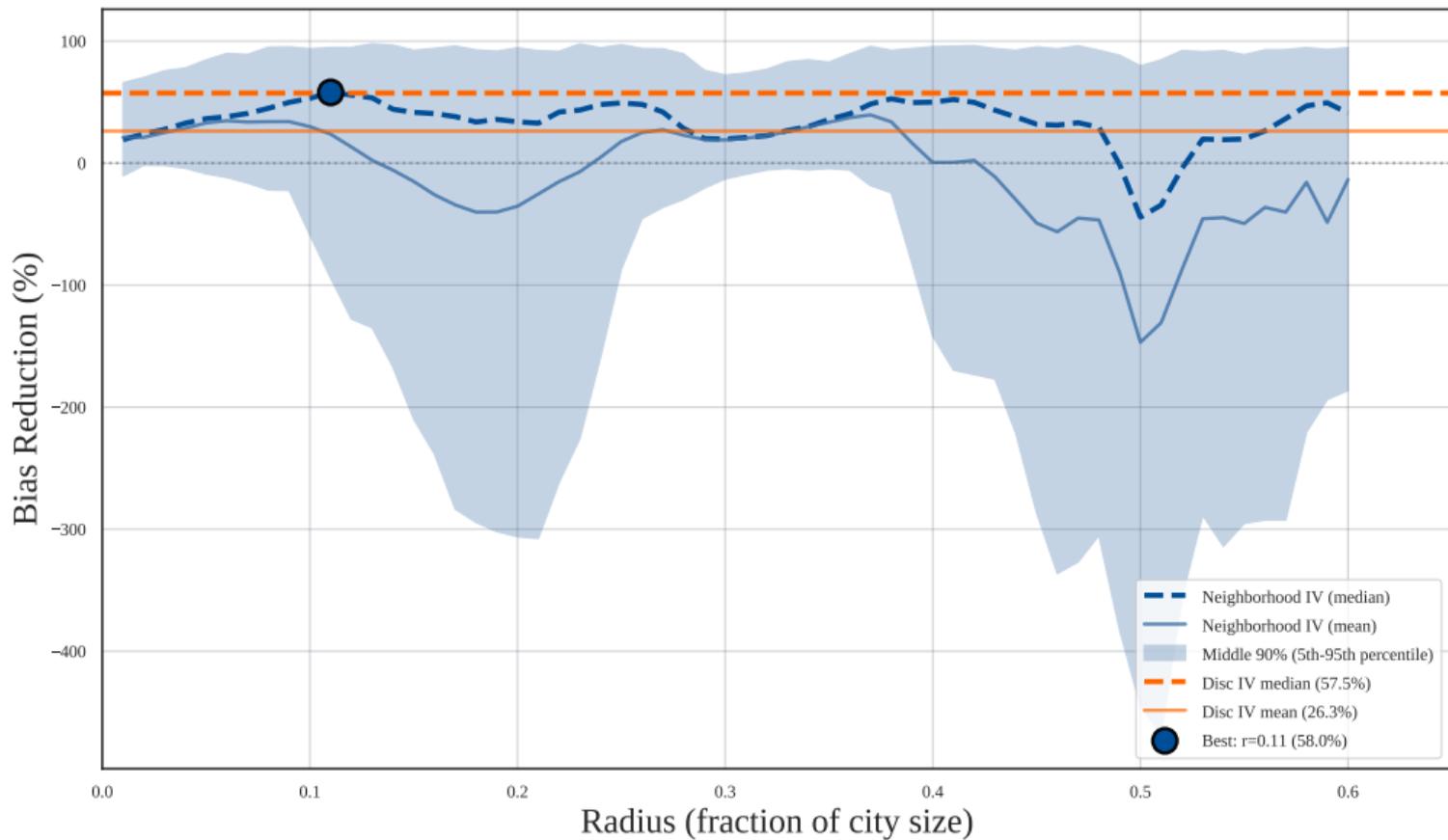
- Use local income composition interacted with chain fixed effects
- Leverages preference externalities: chains locate near concentrations of target income groups
- Intuition: if many low-income households live nearby \rightarrow more likely a Dollar General is close
- Vulnerable to angular sorting on unobservables within income quartiles
- Monte Carlo: performance varies substantially across radius choices

IV Strategy: “Disc Instruments”



- Construct instruments using discs centered on the CBD
- For each consumer at distance r from CBD: Compute share of each income quartile living within radius
- Every consumer on the same ring gets the same instrument values; exploits income gradients in cities
- Mitigates local sorting on unobservables; requires no radial sorting on unobserved tastes (conditional on income)

Monte Carlo Evidence: Disc vs. Neighborhood



Estimates: Averages across CBSAs (Endogenous Distance)

Income Quartile		Inc. 1	Inc. 1 SE	Inc. 2	Inc. 2 SE	Inc. 3	Inc. 3 SE	Inc 4.	Inc. 4 SE
<u>Parameter</u>									
Distance	β^{d1}	-0.444	(0.020)	-0.469	(0.010)	-0.475	(0.009)	-0.540	(0.010)
Density	β^{d2}	0.339	(0.083)	0.206	(0.018)	0.177	(0.014)	0.141	(0.015)
Fringe	ω	1.478	(0.769)	1.354	(0.066)	1.340	(0.023)	1.130	(0.018)
Control Function	ρ	0.181	(0.016)	0.221	(0.007)	0.212	(0.006)	0.242	(0.007)
<u>Summary</u>									
Number of Visits		1,476,820		7,737,705		11,797,999		13,796,397	
Number of Devices		130,157		605,128		886,500		1,132,056	
Avg. First Stage Partial R^2		18.9%		13.5%		11.6%		12.8%	
Avg. First Stage Partial F -stat		237.1		720.4		873.9		1514.0	

Findings

- powerful instruments
- positive travel costs, density, and fringe create value

Estimates: Averages across CBSAs (Exogenous Distance)

Income Quartile		Inc. 1	Inc. 1 SE	Inc. 2	Inc. 2 SE	Inc. 3	Inc. 3 SE	Inc 4.	Inc. 4 SE
<u>Parameter</u>									
Distance	β^{d1}	-0.278	(0.016)	-0.270	(0.006)	-0.274	(0.005)	-0.310	(0.006)
Density	β^{d2}	0.357	(0.074)	0.245	(0.021)	0.193	(0.017)	0.148	(0.016)
Fringe	ω	1.404	(0.799)	1.292	(0.078)	1.248	(0.025)	0.994	(0.017)
<u>Summary</u>									
Number of Visits		1,476,820		7,737,705		11,797,999		13,796,397	
Number of Devices		130,157		605,128		886,500		1,132,056	

Findings

- Significantly lower travel costs
- Analogous to familiar price endogeneity results

Use model to investigate preferences in mile equivalents

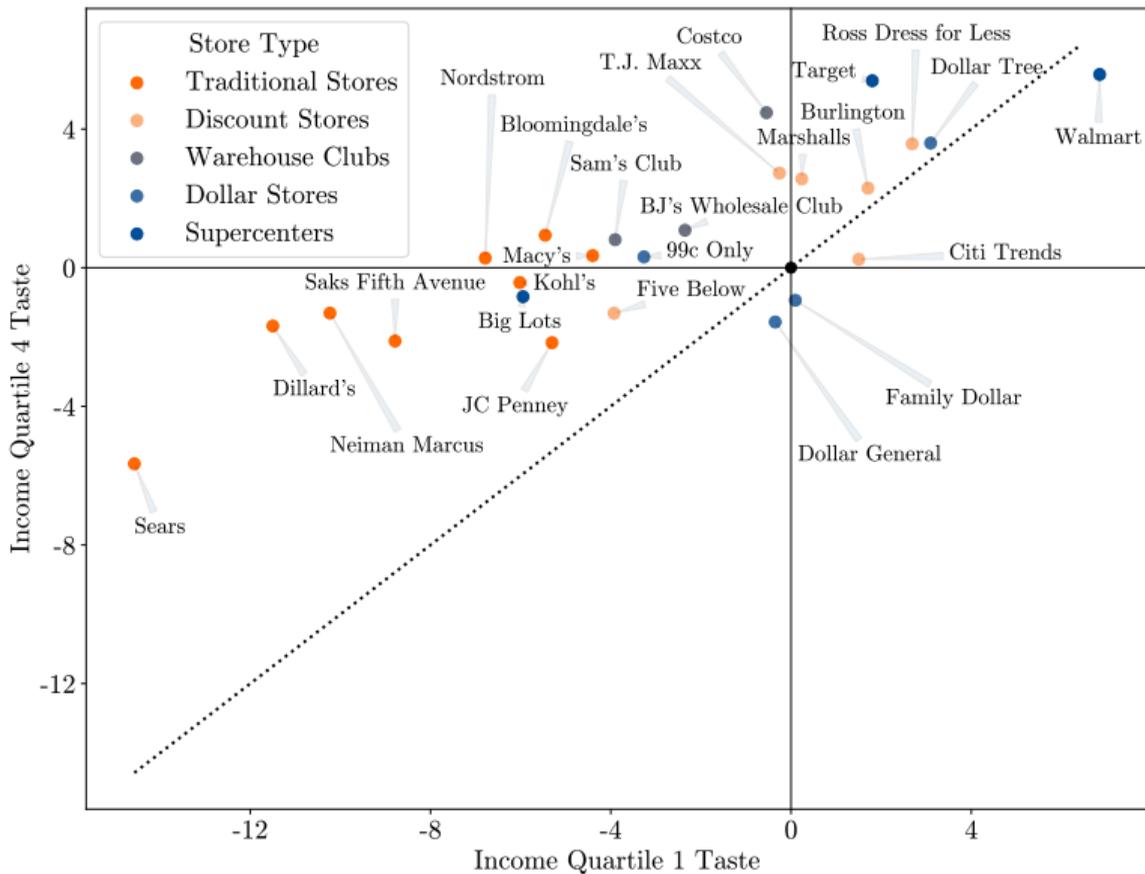
For each chain j , for each income group $y(i)$ we calculate a mile-normalized taste measure:

$$\tilde{\beta}_{y(i),j} = \frac{\beta_{y(i),j} + \beta_{y(i)}^{d2} \text{median}(\text{density}_{.j}) - \omega_{y(i)} \text{median}(\Gamma(j^f))}{|\beta_{y(i)}^{d1}|},$$

where $\text{median}(\text{density}_{.j})$ is the CBSA-specific chain-specific median density

Interpret chain taste coefficients as how much further the representative consumer travel to visit chain j instead of a fringe store

Chain preferences vary predictably across income quartiles



Welfare changes from 2010 to 2019

- Calculate consumer surplus from the store array for 2010 and 2019 by income quartile.
 - Calculate distances from tract centroids to 2010 and 2019 store locations using the Data Axle data.
 - Aggregate across Census tracts using income quartile x tract populations.
 - Use 2019 preferences estimates.
- We keep 2019 population in place
- Our model outputs welfare changes in mile-equivalents

Counterfactual Results

Table: Welfare Change from 2010-2019 by Income Quartile, Distance Endogenous

Income Quartile	1	2	3	4
ΔW	-0.220	-0.139	-0.053	0.115
Conf. Interval	(-0.477, 0.275)	(-0.320, -0.043)	(-0.245, 0.093)	(0.034, 0.231)
Δ Fringe Only	-0.577	-0.400	-0.382	-0.183
Conf. Interval	(-1.851, -0.471)	(-1.049, -0.376)	(-0.387, -0.387)	(-0.246, -0.166)
Δ Chains Only	0.372	0.312	0.283	0.359
Conf. Interval	(0.273, 1.942)	(0.245, 0.414)	(0.160, 0.509)	(0.338, 0.474)

Findings

- Modest, insignificant declines for quartiles 1 and 3
- Modest, significant increases (decreases) for quartile 4 (2)
- All quartiles have benefited from entry

Counterfactual Results

Table: Welfare Change from 2010-2019 by Income Quartile, Distance Exogenous

Income Quartile	1	2	3	4
ΔW	-0.514	-0.459	-0.261	0.061
Conf. Interval	(-1.705, -0.222)	(-0.755, -0.367)	(-0.626, -0.060)	(-0.101, 0.221)
Δ Fringe Only	-0.987	-0.786	-0.584	-0.313
Conf. Interval	(-1.990, -0.837)	(-1.137, -0.780)	(-0.673, -0.591)	(-0.386, -0.315)
Δ Chains Only	0.354	0.388	0.346	0.466
Conf. Interval	(0.038, 0.692)	(0.209, 0.614)	(0.105, 0.653)	(0.413, 0.679)

Findings

- Overstates welfare loss, significant

Wrap up

- IV strategy can be applied in a variety of spatial settings
- Accounting for distance endogeneity results in much higher travel costs
- In terms of the retail apocalypse, cannot conclude that welfare has declined, accounting for endogeneity
- See paper for discussion of:
 - Alternative instrument approaches and addressing additional endogeneity concerns
 - Analysis with travel time instead of distance
 - Additional retail facts and analysis of individual chains
 - Additional counterfactuals, e.g., population moves as well