Residential Sorting and Access to Consumption*

Sisi Zhou †

Department of Economics

Purdue University

Click here for the latest version.

Abstract

This paper estimates an equilibrium sorting model that jointly considers residential and consumption choices to evaluate the impact of urban policies aimed at improving household access to consumption amenities. I structurally estimate the model using data from the household travel survey, business information from Google Maps, and housing transaction records in Chicago. Employing the estimated model to evaluate counterfactual policies, I find that the introduction of new grocery stores in underserved areas increases housing prices in nearby neighborhoods. While higher housing costs may promote low-income households to relocate, they experience welfare gains through enhanced access to grocery options. Furthermore, the findings suggest that such policies reduce city-wide income inequality by fostering residential mobility across neighborhoods.

Keywords: Consumption amenities, Residential sorting, Food desert

JEL Code: D12, R21, R23

* The author gratefully acknowledges Farid Farrokhi, Chong Xiang, Ben Zou, and the participants at the 94th Annual Meeting of the Southern Economic Association for their valuable comments. Appreciation is also extended to the Chicago Metropolitan Agency for Planning for providing the CMAP Household Travel Survey Data. All opinions expressed and any errors in this paper are solely those of the author.

†Department of Economics, Daniels School of Business, Purdue University, 403 Mitch Daniels Blvd, West Lafayette, IN 47907, USA. Email: zhou1053@purdue.edu.

1 Introduction

Trips related to consumption play a critical role in individual's daily life and can significantly influence residential housing choices. As shown in Figure 1, majority of trips of households living in metropolitan area are non-commuting trips (grocery shopping, dining, etc.), especially on weekend. Therefore, consumption access is a critical factor in evaluating the welfare benefits associated with specific residential areas, and urban policies aimed at improving household access to consumption amenities have been highlighted in recent urban development plan in many metropolitan area. On the one hand, those policies will improve consumption accessibility, and thus increase residents' welfare and attract more people to move into the neighborhood. On the other hand, areas with premium consumption access are typically associated with higher property values, which may force low-income households to relocate away from the neighborhood. Thus, a framework that jointly modeling the consumption and residential decisions is necessary to evaluate the impact of urban policies that enhance the access to consumption amenities. Existing research have studied the role of commuting costs (Gu et al. (2024); Barwick et al. (2024)) in shaping residential location decisions. However, these studies overlook the influence of non-commuting costs associated with consumption-related trip, which is crucial for evaluating the consumption-oriented urban policy. Leveraging detailed household travel survey data, property transactions, and business information from Google Map, I develop and estimate a residential sorting model with rich household's preference heterogeneity, where consumption access is treated as an endogenous channel for housing choice. I then use the estimated model to evaluate the impact of urban policies aimed at improving household access to consumption amenities on housing prices, residential choices, households' welfare, and city-wide income disparity.

The housing market is defined at the unit of census block. Following the literature, I assume that work locations are determined ex ante. Conditional on the workplace location, households make the house location decision. Conditional on their home locations, they choose the business location for the trips of various consumption categories (e.g. grocery stores and restaurants). The model is a two-stage discrete choice framework: in the first stage, households determine which block to live conditional on workplace location; in the second stage, they choose business locations for consumption trips across various categories (e.g., grocery and restaurants). The residential-choice component follows the methodology specification in Bayer et al. (2007). Preferences are parameterized as functions of house prices, block characteristics, and unobserved attributes. Households' valuation of choice attributes is alloId to vary with their own demographic characteristics. I also include consumption access (determined by price, quality, travel cost to consumption, etc.) as a key determinant in housing decisions. The consumption-choice stage of the model is characterized by a multinational logit model. Conditional on home location, households derives utility from trips from travel cost, price level, review rating, and also the those characteristics interacted with their own demographic characteristics. The two-stage sorting model enables counterfactual simulations to predict equilibrium outcomes under counterfactual regional development policies, specifically examining impacts on consumption patterns, residential location choices, housing prices, and welfare distribution.

I employ three primary datasets: the Chicago CMAP My Daily Travel Survey, Infutor data, and Google Maps Places. To estimate household consumption preferences, I merge trip and household data from the CMAP survey with business data from Google Maps Places. This combined dataset provides demographic information of the individual, home location, and details of each business they visited (category, price level, and rating). For household housing preferences, I use the Infutor dataset, which offers comprehensive records on housing transactions and buyer demographics in Cook County. This dataset includes household income, age, and detailed property characteristics (e.g., address, sale price, and features). Our sample focuses on homeowners of single-family houses, yielding 54,028 observations. I follow a two-step estimation strategy, drawing on prior studies of residential choice (Phaneuf et al. (2008); Barwick et al. (2024)). In the first step, I estimate household preferences for consumption choices using data from the CMAP travel survey and Google Maps. The estimated parameters from this step, along with the set of consumption locations and their attributes, are used to construct consumption-access measures for each category across all blocks within the household's choice set. These consumption-access attributes are then included as household-block specific variables in the housing demand estimation. In the second step, I recover household preferences for housing attributes using simulated maximum likelihood with a nested contraction mapping and instrumental variable (IV) approach. The estimation is based on observed housing transactions from the Infutor dataset. To address the endogeneity of housing prices, I construct an instrument for prices following Bayer et al. (2007) and Gu et al. (2024). The instrument is based on the exogenous characteristics of blocks located between 3 and 5 miles from the target block.

The estimation results confirm that access to consumption is a significant determinant of residential choices. The access to grocery stores and restaurants are both positively correlated with housing choices. In the counterfactual analysis, I analyze the effects of introducing new grocery stores on consumption and residential choices, welfare, and regional inequality through the lens of residential sorting. I assume local government funding facilitates the establishment of full-service grocery stores at the centroid of each identified area with low grocery store access. Two scenarios are assessed: budget stores with low price and low ratings, and premium stores with high price and high ratings. The counterfactual simulation yields four key findings. First, the average housing price within a 0.5-mile radius of new grocery stores increases, but the average housing prices decline between 0.5 mile to 2 mile from the stores. Second, higher-income people are moving into the region, while loIr-income households are relocating away. The average household income increase by \$7061.58 and \$34950.70 within 0.5 miles of new budget stores and premium stores, respectively. Third, even though the low-income residents might be forced to move away from where they currently live due to the increase of house prices, they still experience welfare gains from the opening of new grocery stores for improved access of grocery. Forth, the introduction of new grocery stores in the underserved area drives the movement of households and decrease the income disparity.

This paper contributes to the growing literature on urban consumption by analyzing residents' consumption preferences across multiple sectors using diverse data sources. Prior studies have leveraged household travel surveys (Couture (2016)), credit card transaction data (Agarwal et al. (2017)), and online user-generated content (Davis et al. (2019)) to understand urban consumption patterns. In this paper, to examine demand for grocery stores and restaurants, I construct a novel dataset by combining the travel survey data and user-generated content from Google Maps.

This paper is also related to the residential sorting literature. Existing research emphasizes the role of commuting costs in shaping residential location decisions (Gu et al. (2024); Barwick et al.

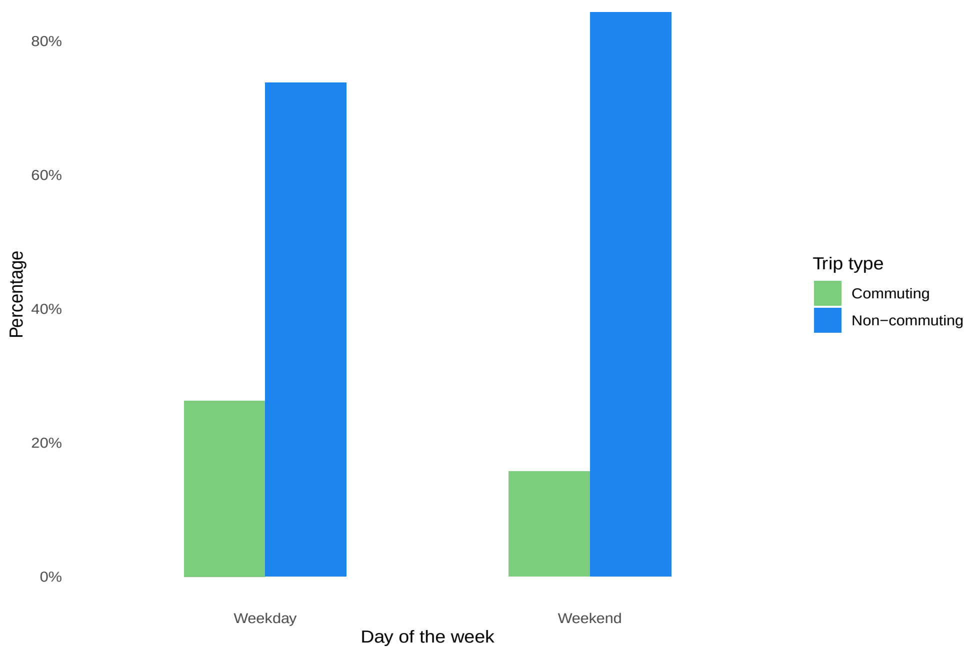
(2024)). However, these studies overlook the influence of non-commuting costs associated with consumption-related travel. Our paper contributes to this literature by integrating consumption behavior into residential sorting models, emphasizing the role of consumption access in location choices. In terms of amenities, sorting models have traditionally focused on households' preferences for exogenous, non-marketed amenities such as air quality and school quality (Bayer et al. (2007); Ferreyra (2007); Epple et al. (2012); Park and Hahm (2023); Kuminoff et al. (2013)), open spaces (Walsh (2007); Phaneuf et al. (2008); Klaiber and Phaneuf (2010); Klaiber and Kuminoff (2014); Timmins and Murdock (2007)),and racial composition (Bajari and Kahn (2005); Hwang (2019)). Public goods, including environmental amenities, have also been studied (Sieg et al. (2004); Almagro and Domínguez-Iino (2024)), primarily within the context of local residential neighborhoods (Bayer and McMillan (2012); Ferreira and Wong (2020); Su (2022)). This paper extends these frameworks by incorporating an additional city-wide consumption access into residential location models, so that residential decisions are shaped not only by exogenous public goods and local amenities but also by endogenous factors such as access to consumption opportunities across the city. I leverage consumption demand estimates to construct a measure of expected city-wide consumption access, which serves as a key determinant in the residential sorting process.

This paper relates to the recent advance using quantitative equilibrium models to study the spatial concentration of economic activity (Miyauchi et al. (2021); Rao (2022); Couture et al. (2024)). Compared to their studies, our model accounts for rich heterogeneity in observed and unobserved preferences by leveraging detailed household-level data. Accounting for such preference heterogeneity is crucial for evaluating the equilibrium responses to regional development policies and provides a more nuanced understanding of urban economic behavior.

Finally, this paper contributes to the literature on the evaluation of urban development policies. Various studies have analyzed the efficiency and effectiveness of urban transportation policies (Gu et al. (2021); Jerch et al. (2021); Kreindler; Yang et al. (2020); Viard and Fu (2015)). Haninger et al. (2017) examines the impact of redeveloping contaminated lands on property values. Gyourko and Molloy (2015) discusses the effect of housing regulation policy. Quigley and Raphael (2005) analyze the effects of land-use regulations on housing prices. Cuffey and Beatty (2022) study policy interventions aimed at improving food access. However, limited research has been conducted on evaluating the equilibrium effects of consumption-related policies. This paper explicitly models household residential location choices and consumption decisions jointly, assessing the impact of these policies on the housing market and regional inequality in equilibrium.

The rest of the paper is structured as follows. Section 2 presensts the model. Section 3 introduces the data and documents the basic consumption pattern of consumption in Chicago. Section 4 details the estimation strategy. Section 5 presents and discusses the estimation results. Section 6 lay outs the simulation algorithm and results of counterfactual analysis. Section 7 concludes.

Figure 1: Percentage of trips by day of the week



Note: This figure shows the percentage of commuting and non-commuting trips on weekday and weekend. The data is from My Daily Travel Survey conducted by the Chicago Metropolitan Agency for Planning (CMAP) in 2019.

2 Model

In this section, I propose a sorting model with two-stage household decision process (Phaneuf et al. (2008); Barwick et al. (2024)). In the residential-choice stage, conditioning on work locations, a household decides where to live. In the consumption-choice stage, given the residential location decision, household decides where to make consumption trips such as dining at restaurants and shopping for groceries.

2.1 Housing demand

I develop a discrete choice model of residential location, building on the framework established in the residential sorting literature (Bayer et al. (2007)). In this model, household preferences over housing options are expressed as a function of both observed and unobserved housing attributes, as Ill as household-specific characteristics. The choice unit is defined at the census block level, where all properties within a given block are assumed to share identical block-level attributes and neighborhood amenities. Following the literature, I assume that work locations are determined ex ante, and households select their residential location conditional on their work location. Additionally, households act as price-takers, with housing prices and block characteristics treated as exogenous to the household's decision-making process.

The utility for household i choosing block j is specified as

$$\max_{j \in J_i} U_{ij} = \alpha_i p_j + \boldsymbol{X}_j \boldsymbol{\beta}_i + \underbrace{\sum_s \phi_{is} EV_{ijs}(\tau_{is})}_{\text{a function of non-commuting cost}} + \xi_j + \varepsilon_{ij} \tag{1}$$

where J_i is the choice set for household i, p_j is the log price per square foot. $\mathbf{X}_j$ is the observed block attributes that do not vary across households, such as average commuting-time, average living area size, average number of bedrooms, average number of bathrooms, green-area ratio, and number of schools. $EV_{ijs}(\tau_{is})$ is the expected consumption utility for household i derived from the optimal consumption choices across $s \in \{restaurant, grocery\}$ sectors. $EV_{ijs}(\tau_{is})$ is a function of τ_{is}, which is travel time from block centroid to all potential consumption locations in sector s. $EV_{ijs}(\tau_{is})$ characterizes block j's attractiveness in terms of household i's consumption benefit and cost. The error term is decomposed into two components: a block-specific term, ξ_j, and a household-block specific term ε_{ij}. ξ_j captures the unobserved quality of the block j. ε_{ij} is an i.i.d. error term with the type I extreme value distribution that reflects unobserved preferences over each housing choice.

Household preferences over price and block attributes are denoted as α_i and $\boldsymbol{\beta}_i$, which consists of a household-specific component and a population average. The household-specific price coefficient α_i is specified as:

$$\alpha_i = \bar{\alpha}_p + \mathbf{z}_i' \boldsymbol{\alpha}_p$$

The household heterogeneous preferences over block attributes are denoted as $\boldsymbol{\beta}_i$. For each element l in $\boldsymbol{\beta}_i$:

$$\beta_{il} = \bar{\beta}_l + \mathbf{z}_i' \boldsymbol{\beta}_l$$

where $\mathbf{z}_i$ is a vector of household characteristics including demeaned income and demeaned age.

The consumption-access preference ϕ_{is} $(s = 1, ..., S)$ are characterized by random coefficients with multivariate normal distribution:

$$\boldsymbol{\Phi}_i \sim MVN(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$$

2.2 House Supply

I assume heterogeneous elasticity for housing supply. The aggregate housing supply in block j is specified as

$$ln(S_j) = c + \rho_j ln(p_j) \tag{2}$$

1% increase in housing price p_j is associated with ρ_j % increase in housing supply.

2.3 Choice of Consumption

For each category s, conditional on the housing location choice j, the indirect utility for a visit by household i to consumption location m is characterized as the multinomial logit model (I suppress sector s to ease exposition):

$$\max_{m \in M_i} U_{ijm} = \mathbf{w}_{im} \boldsymbol{\eta} + \lambda \tau_{ijm} + \nu_{ijm} \tag{3}$$

where $\mathbf{w}_{im}$ includes observed characteristics of consumption location choice like price level, google review rating, price level interacted with household income, and google review rating interacted with household income. τ_{ijm} is the log driving time from home block centroid to the consumption location. ν_{ijm} denotes the unobserved preference over consumption choice with i.i.d type I extreme value distribution.

The ex ante expected consumption utility is defined as (before the realization of preference shocks) is defined as:

$$EV_{ij} = \mathbb{E}_{\nu_{ijm}}\left(\max_{m \in M_i} U_{ijm}\right) \tag{4}$$

3 Data and Summary Statistics

3.1 Policy Background

To estimate the household's consumption and housing preference, I leverage the micro-level data from household travel survey and property transaction data in Cook county, where the city of Chicago is located. With a population of approximately 2.7 million, Chicago is the third-most populous city in the United States and serves as a central economic hub in the MidIst. Figure A1 presents the map of townships within Cook County. The downtown area of Chicago, located in the Chicago Ist township along the shore of Lake Michigan, serves as the primary hub of economic activity. The city's housing market and demographic composition have been shaped by historical systemic discrimination, leading to pronounced income and racial segregation. Ialthier and white residents are concentrated on the North Side, while loIr-income communities of color predominantly reside on the South and Ist Sides. This spatial segregation is mirrored in the uneven distribution of consumption businesses across the city, with high-income areas near downtown and along Lake Michigan offering diverse consumption opportunities, while low-income neighborhoods face food deserts and limited access to quality retail and dining options.

The local government has implemented regional development plans aimed at revitalizing underserved neighborhoods and reducing spatial and economic inequalities within the city. ON TO 2050, a comprehensive regional strategy announced by the Chicago Metropolitan Agency for Planning, outlines a framework for fostering sustainable prosperity across Metropolitan Chicago by addressing the factors that attract individuals to specific communities. The plan aims at improving access to food, particularly in designated food deserts. The City of Chicago allocated $5.5 million to support the construction of a full-service grocery store on the Near Ist Side. More recently, a $5 million Good Food Fund was established to support food entrepreneurs in communities facing inequitable access to food.

3.2 Data

I use three primary datasets for estimation: the Chicago Metropolitan Agency for Planning (CMAP) My Daily Travel survey, Infutor data, and Google Maps Places. To estimate households' consumption preferences, I construct a novel dataset by merging trip-level data from the CMAP survey with business characteristics from Google Maps Places. For each consumption trip, I observe the demographic information and home location of individuals, along with detailed attributes of the businesses they visit, including category, price level, and user ratings. To estimate households' housing preferences, I use the Infutor dataset, which contains comprehensive records of property transactions and buyer demographics in Cook County.

Chicago CMAP My Daily Travel Survey The CMAP survey captures detailed information on commuting (e.g., work, school) and non-commuting trips (e.g., grocery shopping, park visits) taken by households across northeastern Illinois. The latest survey, completed in 2019, collected seven days of GPS-based travel data through a smartphone app. The dataset provides household-level demographics (income, size) and individual-level characteristics (race, gender, education), along with the start and end times and geolocation coordinates of all stops, including residences, workplaces, and consumption sites. Out of the 12391 households surveyed, 4391 Cook County households with positive income are included in our sample, comprising 10086 individuals.

Google Maps Places The CMAP survey records the locations visited by participants but lacks detailed business attributes. I supplement this with data from the Google Maps Places API by sending the geocoordinates of each stop to retrieve additional information, such as the business category (e.g., grocery store, restaurant), price level, user ratings, and number of reviews.

Infutor Data The Infutor dataset provides demographic details for residents of Cook County, including income and household composition, along with property characteristics such as address, sale price, living area, and the number of bedrooms and bathrooms. I focus on homeowners residing in single-family houses, yielding a sample of 54028 households. Residential choice is defined at the census block level, and our analysis covers 28568 blocks with recorded transactions between January 2010 and December 2019.

Other I approximate travel costs using the fastest driving time between locations, calculated via the Open Source Routing Machine (OSRM), which returns the duration and distance of optimal routes. In addition, I use OpenStreetMap (OSM) data to construct the green-area ratio for each census block.

3.3 Summary Statistics

Table 1 and Figure 2 summarize the data used for consumption preference estimation. Table 1 provides the summary statistics of households' demographic information from CMAP My Daily Travel Survey. Figure 2 shows the distribution of grocery stores and restaurants the households visited during the survey week. Grocery stores are generally dispersed across the region, with some clustering near downtown, whereas restaurants are densely concentrated in the downtown area. For grocery stores, the sample includes warehouse clubs and supercenters (e.g., Walmart) and supermarkets (e.g., JeIl-Osco), excluding convenience stores. For restaurants, I include full-service restaurants (e.g., Lou Malnati's Pizzeria), limited-service restaurants (e.g., McDonald's), and snack and nonalcoholic beverage bars (e.g., Starbucks). During the survey week, 269 grocery stores and 2027 restaurants Ire visited at least once, each with corresponding price and rating data from Google Maps.

Table 2 reports summary statistics on the dataset used for housing preference estimation. Panel A provides demographic information on homeowners of single-family houses, who are on average 41 years old, with a mean household income of $95297. Panel B summarizes housing attributes at

Table 1: Summary Statistics of Consumption

	Mean	SD
Panel A: Travel Survey		
Age	33.39	19.04
Income	94598.05	43757.75
Panel B: Grocery store		
$$	0.43	0.50
$$$ bin	0.06	0.24
Rating	4.15	0.27
Panel C: Restaurant		
$$ bin	0.45	0.50
$$$ bin	0.03	0.17
$$$$ bin	0.004	0.07
Rating	4.11	0.41

the census block level.

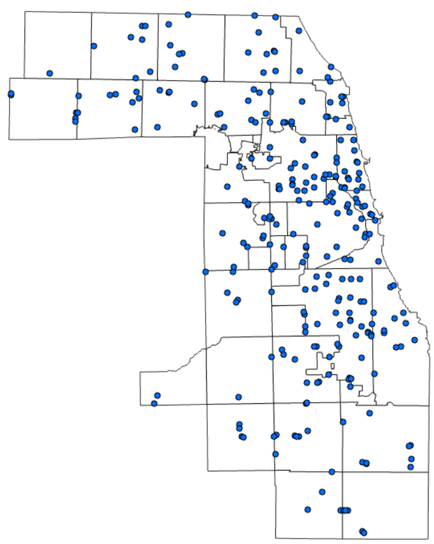
Table 2: Summary Statistics of Housing

	Mean	SD
Panel A: Households	N = 54028	
Age	41.11	12.10
Income (2019 $)	95297	35933.96
Panel B: Block Attributes	N = 28568	
price per sf (2019 $)	213.22	132.95
living area	1830.35	854.47
# of bedrooms	3.48	1.05
# of bathrooms	2.25	0.97
property age	63.91	29.55
green area ratio	0.02	0.10
# of school	2.87	2.18

3.4 Consumption Pattern

Using the dataset for consumption, I describe the consumption pattern of residents. For both grocery stores and restaurants, ”low” price is defined as prices below the 25th percentile, and ”high” price as those above the 75th percentile. Similarly, ”low” and ”high” quality are defined using the 25th and 75th percentiles of user ratings within each category. Figure 3 displays heat maps of travel times from home to grocery stores visited during the survey week. Panel (a) shows travel times to low-price grocery stores, defined as those with a $ price level. Panel (b) shows travel times to high-quality grocery stores, defined as those with an average rating of at least 4.3.

Figure 2: Distribution of consumption locations



(a) Grocery

(b) Restaurant

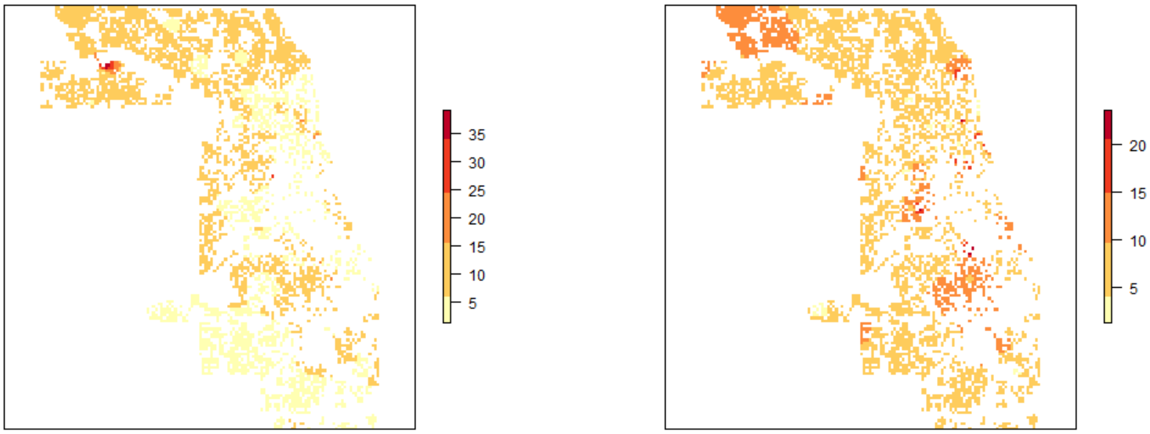
Note: This figure displays the distribution of grocery stores and restaurants that recorded at least one visit during the CMAP My Daily Travel Survey week and have available price and rating information on Google Maps. There are 269 grocery stores and 2027 restaurants.

Residents in the northern and southeastern parts of Chicago tend to travel longer to access low-price grocery stores, while downtown residents have faster access. Residents in the southeast also face longer travel times to reach high-quality grocery stores. Figure 4 shows the heat map of the travel time from home to the low-price ($ price bin), high-price ($$ price bin), and high-quality restaurants (rated 4.4 or above). Overall, residents in urban areas experience shorter travel times to high-end, fine-dining establishments. In contrast, travel times to loIr-priced restaurants are relatively similar across regions, suggesting accessibility is less affected by location for these establishments.

4 Estimation

I follow a two-step estimation strategy, drawing on prior studies of residential choice (Phaneuf et al. (2008); Barwick et al. (2024)). In the first step, I estimate household preferences for

Figure 3: Travel time (min) to grocery stores

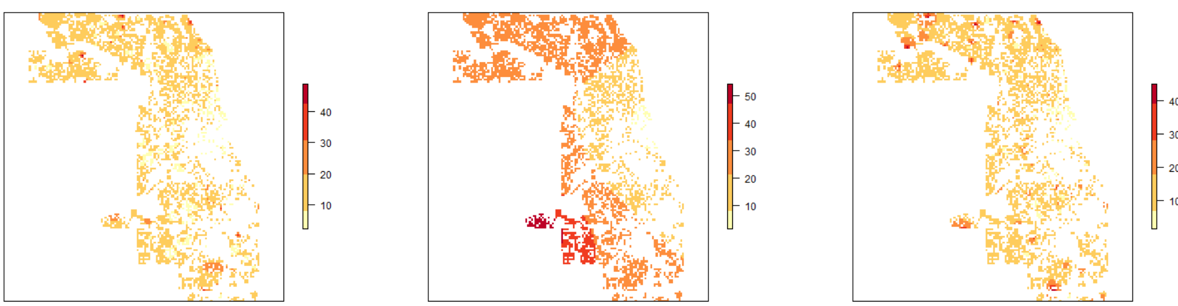


(a) Low price

(b) High quality

Note: The figure displays the travel time from home to the grocery stores which the households visit during the survey week. Panel (a) shows the travel time from home to low-price ($ price bin) grocery stores. Panel (b) shows the travel time from home to high-quality (≥ 4.3) grocery stores.

Figure 4: Travel time to restaurants



(a) Low price

(b) High price

(c) High quality

Note: The figure displays the travel time from home to the restaurants which the households visit during the survey week. Panel (a), (b), and (c) is the travel time from home to the low-price ($ price bin), high-price ($$ price bin), and high-quality restaurants (≥ 4.4), respectively.

consumption choices using data from the CMAP travel survey and Google Maps. The estimated parameters from this step, along with the set of consumption locations and their attributes, are used to construct a consumption-access measure, which is the expected utility from consumption, for each category across all blocks within the household's choice set. These consumption-access attributes are then included as household-block specific variables in the housing demand estimation. In the second step, I recover household preferences for housing attributes using simulated maximum likelihood with a nested contraction mapping and instrumental variable (IV) approach. The estimation is based on observed housing transactions from the Infutor dataset. While I control for a rich set of housing and block characteristics, housing prices may still be correlated with unobserved block-specific attributes. To address potential endogeneity, I construct an instrument for log prices following Bayer et al. (2007) and Gu et al. (2024). The instrument is based on the exogenous characteristics of blocks located between 3 and 5 miles from the target block. The key identifying assumption is that the characteristics of these distant blocks influence local prices through housing market equilibrium but do not directly affect household utility.

4.1 Estimation of Consumption Choice

The parameters of the consumption choices are estimated by maximum likelihood estimation using Chicago household travel survey combined with google map data.

The probability of household i living in j choosing consumption alternative m is given by:

$$P_{ijm}(\boldsymbol{\eta},\lambda) = \frac{\exp\left(\mathbf{w}_{im}\boldsymbol{\eta} + \lambda\tau_{ijm}\right)}{\sum_{k=1}^{M}\exp\left(\mathbf{w}_{ik}\boldsymbol{\eta} + \lambda\tau_{ijk}\right)} \tag{5}$$

The log likelihood function is given by:

$$lnL(\boldsymbol{\eta},\lambda) = \sum_{i}\sum_{j}\mathbb{I}_{ij}ln\left(P_{ijm}(\boldsymbol{\eta},\lambda)\right) \tag{6}$$

where

$$\mathbb{I}_{ij} = \begin{cases} 1 & \text{if } i \text{ chooses } j \\ 0 & \text{otherwise} \end{cases}$$

The estimated parameters from this step, combined with the set of consumption locations and their attributes, are used to construct a measure of consumption access for every block within a household's choice set in the residential choice stage, defined as the expected utility from consumption in each category . These consumption-access measures are then incorporated as household-block-specific variables in the housing demand estimation. The underlying assumption is that, after controlling for locational and demographic differences, the preferences derived from travel surveys are representative of the consumption preferences of homeowners observed in the Infutor data.

4.2 Estimation of Housing Demand

Estimating of the housing demand model follows the procedure developed by Berry et al. (1995), using the household demographic information and housing transaction records from Infutor data. I rewrite the indirect utility function in Equation (1) as a sum of household-specific utility μ_{ij}, population-average utility δ_j, and preference shock ε_{ij}, as shown in Equation (7). The unobserved housing attributes ξ_j is absorbed by δ_j.

$$U_{ij} = \mu_{ij} + \delta_j + \varepsilon_{ij} \tag{7}$$

where μ_{ij} and δ_j are

$$\mu_{ij}(\theta_1) = z_i' \alpha_p p_j + \boldsymbol{X}_j \boldsymbol{\beta} \boldsymbol{z}_i + \sum_s \phi_{is} EV_{ijs} \tag{8}$$

$$\delta_j(\theta_2) = \bar{\alpha}_p p_j + \boldsymbol{X}_j \bar{\boldsymbol{\beta}} + \xi_j \tag{9}$$

I first estimate $\theta_1 = \{\alpha_p, \boldsymbol{\beta}, \phi_{is}\}$ using simulated MLE with a nesting contraction mapping (to invert δ_j). In the second step, I estimate $\theta_2 = \{\bar{\alpha}_p, \bar{\boldsymbol{\beta}}\}$ by regressing the population-average utility δ_j on price p_j using instrument variables to obtain unbiased estimates of the price coefficients.

The probability of household i choosing blocks j is calculated as the average of D = 250 Halton draws:

$$P_{ij}(\theta_1, \delta_j) = \frac{1}{D} \sum_{d=1}^{D} \frac{\exp\left(\mu_{dj} + \delta_j\right)}{\sum_{k=1}^{J} \exp\left(\mu_{dk} + \delta_k\right)} \tag{10}$$

The log likelihood function is given by:

$$lnL(\theta_1, \delta_j) = \sum_i \sum_j \mathbb{I}_{ij} ln\left(P_{ij}(\theta_1, \delta_j)\right) \tag{11}$$

where

$$\mathbb{I}_{ij} = \begin{cases} 1 & \text{if } i \text{ chooses } j \\ 0 & \text{otherwise} \end{cases}$$

I search θ_1 and δ_j to maximize the log-likelihood function. Standard errors are obtained by bootstrapping the estimation 500 times. Once θ_1 is estimated and δ_j is inverted in the first step, the second step runs 2SLS regression to estimate $\theta_2 = \{\bar{\alpha}_p, \bar{\boldsymbol{\beta}}\}$ according to Equation (9). Appendix Section A provides more details about the estimation procedure.

Large Choice Set Issue The high dimensionality of the complete choice sets across blocks introduces a substantial computational burden. To address this issue, I implement a choice set reduction procedure similar to that of Gu et al. (2024). Specifically, I randomly select a subset of 50 potential consumption locations for each household. The choice set for each household includes the actual observed choice and 49 other locations, sampled randomly from the remaining available alternatives with equal probability. McFadden (1977) demonstrates that under random utility models, estimation procedures using choice-based sampling can asymptotically yield consistent parameter estimates.

Endogeneity of House Price The housing price p_j could be correlated with the unobserved the block attributes ξ_j. Following Bayer et al. (2007) and Gu et al. (2024), I construct an instrumental variable for the housing price at the block level. The instrument is derived using characteristics of neighboring blocks located within a 3-5 mile radius. These neighboring blocks are selected to be close enough to influence the housing price of the block of interest through market competition, yet distant enough that their characteristics remain uncorrelated with the unobserved attributes of the block. Specifically, I use the average of exogenous attributes from these neighboring blocks to construct the instrument. I first regress the log of housing prices on these exogenous attributes across the sample and then predict a price for each block based on the estimated coefficients and the average characteristics of the neighboring blocks. This predicted price serves as the instrument for the housing price of the block.

Commuting Cost The Infutor dataset does not provide information on household workplace locations. To estimate commuting costs at the block level, I utilize data from the Chicago CMAP My Daily Travel Survey. Commuting costs are approximated by the driving time between home and work locations. I estimate the driving time between each home-work pair for each individual who live in Cook County and have a fixed workplace. To predict the commuting cost for each block, I interpolate the commuting cost at the block centroid based on inverse distance weighted (IDW) interpolation in Equation (12).

$$Z(x_0) = \frac{\sum_{i=1}^{N} Z(x_k) \cdot w_k}{\sum_{i=1}^{N} w_k} \tag{12}$$

where $Z(x_0)$ is the predicted commuting cost at the block centroid x_0, $Z(x_k)$ is the commuting cost at house location x_k, w_k is the weight for point x_k, calculated as $w_i = \frac{1}{d(x_k,x_0)^2}$. $d(x_i,x_0)$ is the distance between x_k and x_0. This interpolation strategy assumes that commuting costs are more similar among geographically proximate locations compared to those that are farther apart.

4.3 Housing Supply Elasticity

The housing supply elasticity ρ_j of block is constructed based on the data from Baum-Snow and Han (2024). They perform a comprehensive neighborhood-level analysis of housing supply and estimate housing unit supply elasticities at the census tract level. Given these tract-level elasticities, I compute the elasticity for each block j as a weighted average of the elasticities from intersecting tracts. The weight is determined by the area of the polygon that overlaps between each block and its intersecting tracts. The block-level housing supply elasticity ρ_j is specified as:

$$\rho_j = \sum_{k=1}^{K} \frac{A_{jk}}{A_k} \times \phi_k \tag{13}$$

where A_k is the area of tract k, ϕ_k is the housing supply elasticity of tract k, and A_{jk} represents the area of the polygon overlapping between block j and tract k.

5 Estimation Results

5.1 Choice of Consumption

Table 3 reports the results the estimates for consumption preference over log(travel time), price dummy, rating, the interaction between income and price dummy, and the interaction between rating and price dummy. For grocery shopping, the estimate of log(travel time) is -2.716, indicates that spatial frictions substantially influence individuals' grocery store choices. Households are less likely to visit the grocery store that further away from their home locations in terms of automobile travel time. To illustrate, consider two hypothetical grocery stores with identical characteristics except for driving time from home. If one store is 15 minutes away and the other 30 minutes away by car, an individual is over six times more likely to visit the closer store ($2^{2.716} = 6.57$). The estimates are significantly negative for the the price dummy, and positive for the interaction between income and price dummy. It indicates that, overall, individuals experience disutility from higher prices and loIr ratings. However, as income and ratings increase, individuals become less sensitive to price levels. The coefficient for rating and interaction between rating and price dummy are all significantly positive, which implies that households gain positive utility from the quality of grocery stores, and as the price level gets higher, households have a greater preference over high quality. For restaurant, the estimate of log(travel time) is -2.466. To interpret it, suppose restaurant B takes twice the travel time compared to an identical restaurant A, the individual is over five times more likely to choose the nearer option ($2^{2.466} = 5.53$). Overall, households prefer restaurants with low price, holding all else equal. The estimate for rating is negative. Possible reasoning for this counter-intuitive estimate could be that customers might have a negative perception of restaurants with very high ratings, assuming they are overpriced or overrated, leading to a negative attitude towards the high-rating options. Moreover, due to the limitation of data, some restaurant attributes like wait time are unobserved. A high-rating restaurant usually require longer wait time, so that individuals might avoid dining at these high-rating options.

5.2 Choice of Housing

Table 4 displays the estimates of the nonlinear parameters for housing choice. Two model specifications are examined: one excluding and the other including expected utility (EV) terms related to consumption access. The results show that high-income and younger households exhibit loIr sensitivity to house prices. Additionally, the EV terms for both grocery stores and restaurants are significantly positive, indicating the importance of proximity to these amenities in housing decisions. The parameters of demographic interactions in Model 2, which is our preferred specification, indicate how the preference over housing attributes vary across the age and income of households.

With the mean utility of each block, δ_j, inverted from the MLE procedure, I treat δ_j as the dependent variable and run a linear regression to estimate the population-average utility conditional on Model 2 specification from Table 4. The linear parameter in Table 5 represents the marginal utility of housing attribute for an average household. The OLS estimate of the price coefficient is likely biased toward zero due to endogeneity. To address this, I employ an instrumental variable (IV) approach, which corrects for the endogeneity of housing prices and results in a larger magnitude for the price disutility estimate. To interpret the estimates of linear parameter, I can compare them with the mean coefficient associated with log price. For example, the parameter of log property age over log price is 0.1, which indicates that a one log point increase in property age is equivalent to an increase in price by 0.1 log points. Notice that while some attributes may be subject to multicollinearity or omitted variable bias, their inclusion serves to gauge the validity of the instrumental variable.

Table 3: MLE estimates for consumption

	(1) Grocery	(2) Restaurant
log(travel time)	-2.716	-2.466
	(0.005)	(0.002)
$$ bin	-1.558	-0.068
	(0.124)	(0.040)
$$$ bin	-26.009	-5.934
	(0.944)	(0.311)
$$$$ bin		3.572
		(2.290)
rating	1.135	-0.364
	(0.017)	(0.005)
$$ bin × income	0.135	0.080
	(0.002)	(0.001)
$$$ bin × income	0.254	0.259
	(0.006)	(0.003)
$$$$ bin × income		0.134
		(0.005)
$$ bin × rating	0.251	-0.102
	(0.029)	(0.009)
$$$ bin × rating	5.314	0.773
	(0.216)	(0.069)
$$$$ bin × rating		-0.850
		(0.499)

Table 4: Housing Demand - Nonlinear Parameter

	(1) Model 1	(2) Model 2
Demographic Interactions		
ln price per sqft × age	-0.114	-0.120
	(0.041)	(0.038)
ln price per sqft × income	3.074	2.522
	(0.046)	(0.051)
ln living area × age	1.172	1.169
	(0.080)	(0.078)
ln living area × income	2.479	2.097
	(0.065)	(0.072)
ln # of bedroom × age	-1.072	-1.081
	(0.107)	(0.107)
ln # of bedroom × income	-0.826	-0.746
	(0.075)	(0.082)
ln # of bathroom × age	0.219	0.220
	(0.068)	(0.065)
ln # of bathroom × income	0.260	0.192
	(0.045)	(0.045)
ln property age × age	-0.425	-0.422
	(0.033)	(0.033)
ln property age × income	-0.294	-0.259
	(0.035)	(0.036)
ln green-area-ratio × age	0.010	0.010
	(0.005)	(0.004)
ln green-area-ratio × income	0.020	0.007
	(0.003)	(0.003)
ln # of schools × age	-0.042	-0.042
	(0.007)	(0.007)
ln # of schools × income	-0.068	-0.058
	(0.006)	(0.006)
ln commuting-time × age	0.104	0.100
	(0.066)	(0.061)
ln commuting-time × income	1.496	1.535
	(0.052)	(0.056)
Random Coefficient		
$EV_{grocery}$		2.248
		(0.070)
$EV_{restaurant}$		3.297
		(0.247)

Table 5: Housing Demand - Linear Parameter

	(1) OLS	(2) IV
ln price per sqft	-2.760	-11.927
	(0.508)	(1.609)
ln living area	-1.856	-2.477
	(0.468)	(1.105)
ln # of bedroom	-0.097	-3.493
	(0.269)	(0.738)
ln # of bathroom	0.293	2.676
	(0.177)	(0.568)
ln property age	-1.241	-1.194
	(0.292)	(0.374)
ln green-area-ratio	0.060	0.139
	(0.023)	(0.030)
ln # of schools	-0.409	-0.334
	(0.061)	(0.066)
ln commuting-time	8.256	4.315
	(0.756)	(0.834)
constant	9.841	76.521
	(5.465)	(16.327)

5.3 Robustness Check

5.3.1 Random Coefficient for Consumption

The baseline specification in Equation (3) allows for heterogeneous preferences for characteristics of consumption location by interacting these characteristics with household features. To account for unobserved heterogeneity in how people value different attributes of the choices, I allow for random coefficients by assuming the heterogeneous parameters follow the multivariate normal distribution.

$$\max_{m \in M_i} U_{ijm} = \mathbf{w}_{im}\boldsymbol{\eta}_i + \lambda_i \tau_{ijm} + \nu_{ijm} \tag{14}$$

where

$$\begin{bmatrix} \boldsymbol{\eta}_i \\ \lambda_i \end{bmatrix} = \boldsymbol{\Theta}_i \sim MVN(\boldsymbol{\Theta}, \boldsymbol{\Omega})$$

Section A.1 in the appendix presents details of estimation of random coefficient for consumption. Table A1 presents the MLE estimates for consumption specified by random coefficient model.

5.3.2 Alternative Sampling Size

In the baseline model, I reduce the dimension of the large choice set for housing by randomly select 50 choices from the full choice for each household. I conduct the robustness check by estimating the housing demand preference using sample size of 45, 40, and 35. It yields very similar results, as shown in Table A2.

6 Counterfactual Analysis

Over 9% of the population in the Cook County region resides in food deserts—areas characterized by limited access to nearby stores that provide fresh and nutritious food. Figure 5a illustrates regions with low access to large supermarkets, highlighted in red. In response to this disparity, local government initiatives have focused on regional development policies designed to revitalize underserved neighborhoods and address spatial and economic inequities within the city. For instance, the City of Chicago allocated \$5.5 million to establish a full-service grocery store on the Near Ist Side. More recently, a \$5 million "Good Food Fund" was launched to support food entrepreneurs operating in communities with limited access to affordable, healthy food options.

In the counterfactual analysis, I evaluate the impact of regional development policies intended to improve grocery access. Specifically, I analyze the effects of introducing new grocery stores on consumption and residential choices, welfare, and regional inequality through the lens of residential sorting. In this analysis, I assume local government funding facilitates the establishment of full-service grocery stores at the centroid of each identified area with low grocery store access. In Figure 5b, red stars indicate the proposed locations of hypothetical new grocery stores. Two types of stores are assessed: budget stores, defined as those in the \$ price category with a rating of 4 (in the 25th percentile of existing store ratings), and premium stores, defined as those in the $$$ price category with a rating of 4.5 (in the 90th percentile of existing store ratings).

Willingness to Pay The opening of these new grocery stores expands households' choice sets within the grocery sector, thereby altering the consumption access level associated with their housing choices. I incorporate the counterfactual grocery stores into the households' choice sets and simulate the expected utility (EV) term based on the new grocery options available. I then estimate households' willingness to pay (WTP) for the grocery store expansion policies. Using the updated EV term, I calculate the change in housing prices per square foot that would maintain household utility at a constant level. The results show that there is no significant difference between WTP for budget stores and premium stores. On average, households are willing to pay $17.36 for the introduction of new budget stores, and are willing to pay $17.41 for the premium stores. After taking into account that people prefer grocery stores with high quality while get disutility from high price level, household have a slight preference on premium stores over budget stores. Households living in the northern regions, who have a higher income, have higher WTP, while the households in the south with loIr income express loIr WTP (Figure A5 plots the distribution WTP).

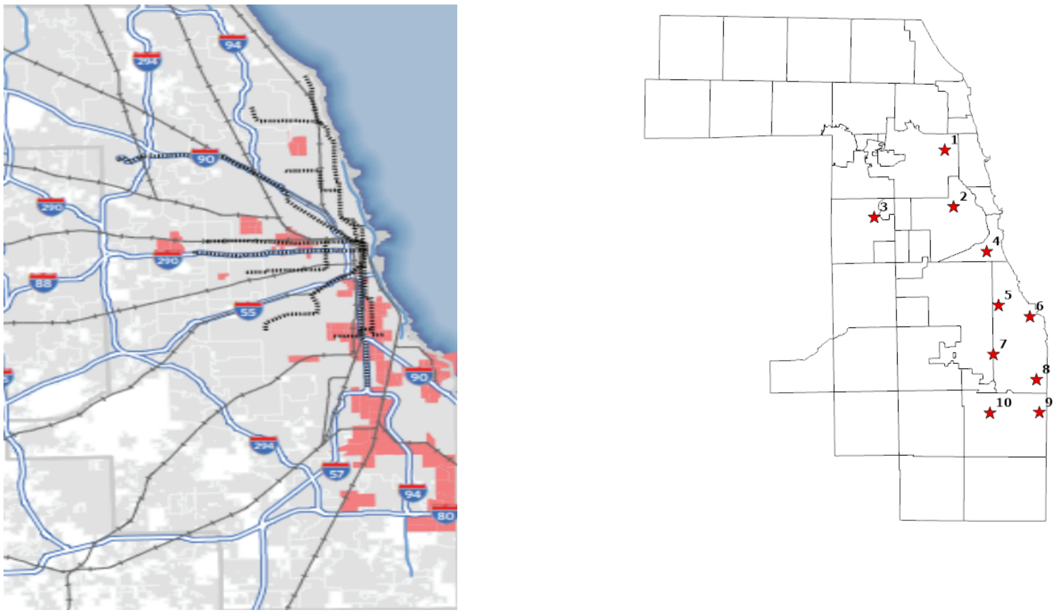
6.1 Simulation New Equilibrium

I further allow residents to reoptimize their consumption and residential location choices jointly and simulate the equilibrium outcomes under the counterfactual scenario. Given the updated EV'_{ijs}, I search for a new housing price vector $\boldsymbol{p}'$ that clears the housing market with housing demand equal to housing supply

$$\sum_{i \in C^{-1}(j)} \frac{1}{D} \sum_{d=1}^{D} \frac{\exp\left(\alpha_i p'_j + \mathbf{X}_j \boldsymbol{\beta}_i + + \sum_s \phi_{is} EV'_{ijs} + \xi_j\right)}{\sum_{r \in C(i)} \exp\left(\alpha_i p'_r + \mathbf{X}_r \boldsymbol{\beta}_i + + \sum_s \phi_{is} EV'_{irs} + \xi_r\right)} = S_j, \forall j \in J \tag{15}$$

The left-hand-side is the simulated demand for neighborhood j and right-hand-side is the housing supply. Suppose the price vector is p^t in iteration t. The aggregate housing supply in block j is specified as $ln(S_j) = c + \rho_j ln(p_j)$. In the counterfactual, I allow the house supply responds to the price change according to

Figure 5: Counterfactual analysis of grocery stores



(a) Low-access to supermarket

(b) Counterfactual grocery stores

Note: Panel (a) shows the areas with low access to large supermarkets in Cook County (source: https://cmap.illinois.gov/). Panel (b) plots the new grocery stores introduced in the counterfactual analysis, labeled from 1 to 10.

$$\frac{S_j^t - S_j^0}{S_j^0} = \rho_j * \left(\frac{p_j^t - p_j^0}{p_j^0}\right) \tag{16}$$

I update the price vector by

$$\boldsymbol{p}_j^{t+1} = \boldsymbol{p}_j^t + \left[\log\left(D_j^t\left(\boldsymbol{p}_j^t\right)\right) - \log\left(S_j^t\right)\right]/a \tag{17}$$

where S_j^t is the housing supply derived from Equation (16), $D_j^t\left(\boldsymbol{p}_j^t\right)$ is the predicted demand in interation t, specified by the left-hand-side of Equation (15), a is a pre-set constant that controls the step size of each iteration. I iterate the housing market conditions until the price vector converges

$$\left\|\boldsymbol{p}^{t+1} - \boldsymbol{p}^t\right\| < e_{\text{tol}} \tag{18}$$

6.2 Counterfactual Results

Table 6 shows the changes of average housing price per square foot, share of population, average household welfare, and average household income in both magnitude and percentage within the 0.5-mile, 1-mile, and 2-mile radius surrounding the new grocery stores under the counterfactual scenarios.

Table 6: Counterfactual result summary

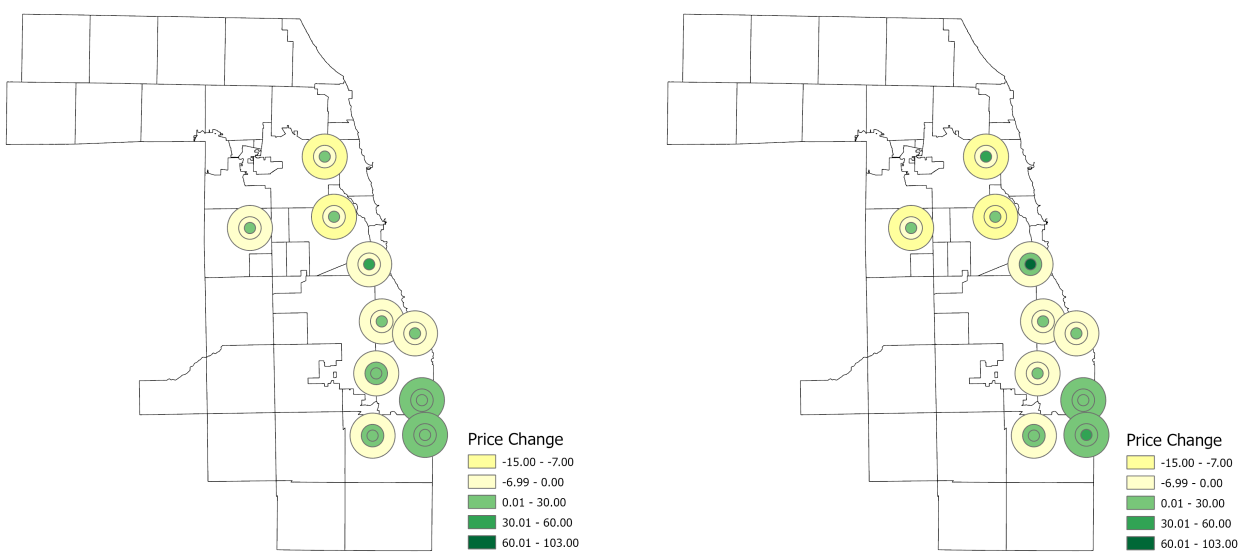
Budget Store		0-0.5 mile	0.5-1 mile	1-2 mile
Housing Price	Δ	19.99	-1.28	-5.57
	$\%\Delta$	11.83%	-0.67%	-2.48%
Share	Δ	2.56×10^{-5}	2.52×10^{-5}	5.16×10^{-5}
	$\%\Delta$	0.35%	0.11%	0.05%
Welfare	Δ	0.42	0.41	0.40
	$\%\Delta$	15.59%	8.31%	4.22%
Income	Δ	7061.58	7262.20	4591.51
	$\%\Delta$	10.27%	9.84%	5.42%
Premium Store		0-0.5 mile	0.5-1 mile	1-2 mile
Housing Price	Δ	36.46	-0.53	-9.11
	$\%\Delta$	21.57%	0.28%	-4.05%
Share	Δ	4.33×10^{-5}	4.51×10^{-5}	9.57×10^{-5}
	$\%\Delta$	0.59%	0.20%	0.10%
Welfare	Δ	0.69	0.68	0.68
	$\%\Delta$	25.78%	13.90%	7.27%
Income	Δ	34950.70	17703.91	6989.87
	$\%\Delta$	50.82%	23.98%	8.25%

Note: This table shows the changes in the simulated counterfactual equilibrium, with the introduction of new grocery stores. Δ is the change in magnitude, $\%\Delta$ is the change in percentage.

6.2.1 Change of housing price

As shown in Table 6, the average housing price per square foot within a 0.5-mile radius of new grocery stores increases, while housing prices decline as the distance from the stores increases. Intuitively, households derive greater utility from proximity to grocery stores, making nearby locations more desirable and thereby driving up housing prices within a 0.5-mile radius. Specifically, housing prices increase by \$19.99 per square foot for properties near budget stores and by \$36.46 per square foot for properties near premium stores, suggesting a higher willingness-to-pay (WTP) among households for proximity to premium stores. Figure 6 further illustrates the spatial heterogeneity of change in housing prices per square foot within the 0.5-mile, 1-mile, and 2-mile radius surrounding each of the new budget and premium grocery stores. It is shown that southern region experiences a larger increase in housing prices than the northern region. Several factors likely drive this regional difference. First, the southern region has feIr existing grocery stores, implying that the introduction of new stores attracts greater consumer demand relative to the north where stores are denser. Second, average housing prices are initially loIr in the southern region, providing greater potential for price appreciation as grocery access improves and begins to approximate that of the northern region. Combining those two factors, the introduction of new grocery stores have a more pronounced impact on housing demand and prices in the southern region.

Figure 6: Price Change in the Counterfactual



(a) Budget Store (b) Premium Store

Note: This figure shows the change of average housing prices per square foot within the 0.5-mile, 1-mile, and 2-mile radius surrounding the new grocery stores in the counterfactual equilibrium.

6.2.2 Resorting of the residents

Following the introduction of new grocery stores, households reallocate across regions to adjust to the updated access to consumption options and changes in housing prices. For areas experiencing an increase in population share, there is a net inflow of residents, whereas blocks with a declining share experience a net outflow. Two forces influence this movement. First, new grocery stores attract residents seeking convenient access to grocery shopping, leading to an inflow of population. However, this influx drives up housing prices, forcing loIr-income households who previously resided in the area to relocate further away to mitigate the disutility associated with higher housing costs. As shown in Table 6, the share of people who choose to live near the the new grocery increase by only a slight magnitude. For example, it increases by only 0.35% within 0.5 miles of budget stores. It indicates that the inflow and outflow of population almost balance with each other. Figure A6 further illustrates the spatial variation of population dynamics around new grocery stores locations.

Households generally prefer housing located in proximity to grocery stores, which provides convenient access to food and daily necessities. Areas with abundant consumption options are typically associated with higher property values, as individuals are willing to pay a premium for such accessibility. This suggests a trade-off between access to consumption and housing costs - while residents in high-amenity communities benefit from superior consumption access, they face higher housing expenditures. Conversely, suburban residents generally face loIr housing costs for comparable property types but may incur longer travel times to reach consumption hubs. This dynamic results in income-based sorting, with higher-income households clustering in areas with superior consumption accessibility, while loIr-income households are more likely to reside in communities with loIr housing premiums but inferior consumption accessibility. The notable increase in average household income in areas surrounding these new stores confirms this resorting pattern of households after the introduction of new grocery stores. For example, the average household income increase by $7061.58 within 0.5 miles of new budget stores. Given the substantial income increase and a relatively stable overall population size, it indicates that higher-income people are moving into the region, while loIr-income households are relocating away. The increase of average income is more pronounced around premium stores than budget stores. This pattern aligns with the intuition that households prefer high-quality stores, and high-income households are less sensitive to the premium prices associated with these stores. Consequently, areas near premium stores attract more high-income households, while loIr-income households are more likely to be displaced from these areas compared to those around budget stores.

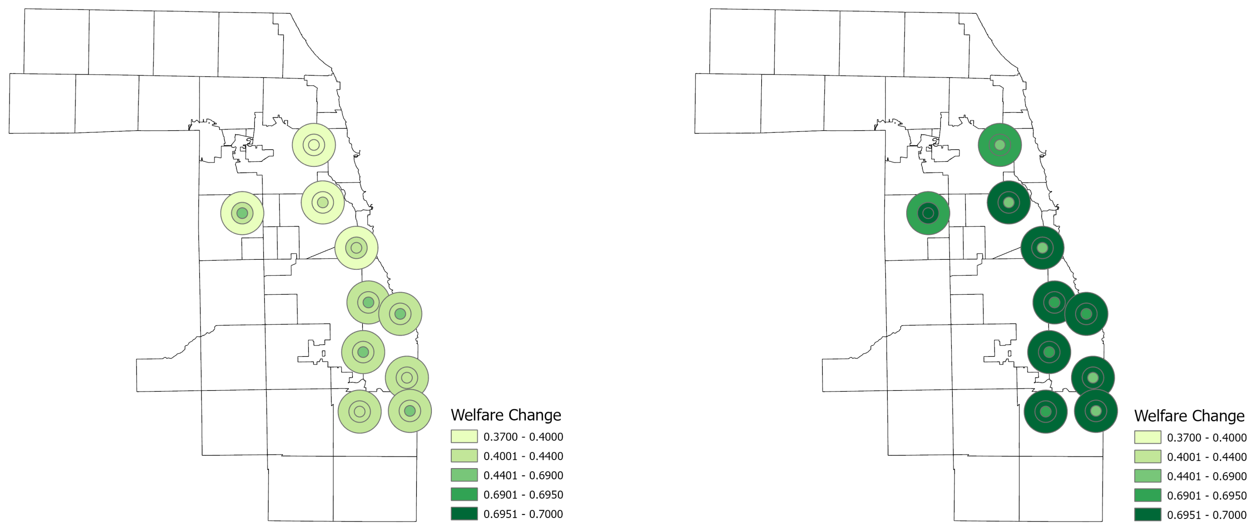
6.2.3 Change of welfare

Table 6 shows that, for residents who currently live within the 2-mile radius of new grocery stores, their average welfare increases after the introduction of new grocery stores. However, as discussed, the current residents near the new grocery stores locations are likely to relocate to other locations in response to increased housing prices associated with the introduction of new stores, especially for those with loIr income. It indicates that even though the current residents might be forced to move away from where they currently live due to the increase of house prices, they still benefit from the opening of new grocery stores for improved city-wide grocery access.

I further examine the spatial variation in the welfare changes among households who currently live within a 2-mile radius of the new grocery store locations. As shown in Figure 7, households in the southern region experience larger welfare gains than those in the north, consistent with the loIr initial density of grocery stores in the south, where new stores have a greater positive impact on household utility.

Figure 8 and Figure 9 present the welfare changes segmented by income groups following the introduction of new budget stores and premium stores, respectively. Specifically, it compares the welfare effects on high-income households (at or above the 90th income percentile) and low-income households (at or below the 10th income percentile). The results indicate that low-income households experience greater welfare increases than high-income households. This pattern aligns with the fact that low-income households are more likely to live in areas with low housing prices, which often are also the areas with limited grocery access. Consequently, the introduction of new grocery stores yields particularly strong welfare gains for low-income households.

Figure 7: welfare Change in the Counterfactual

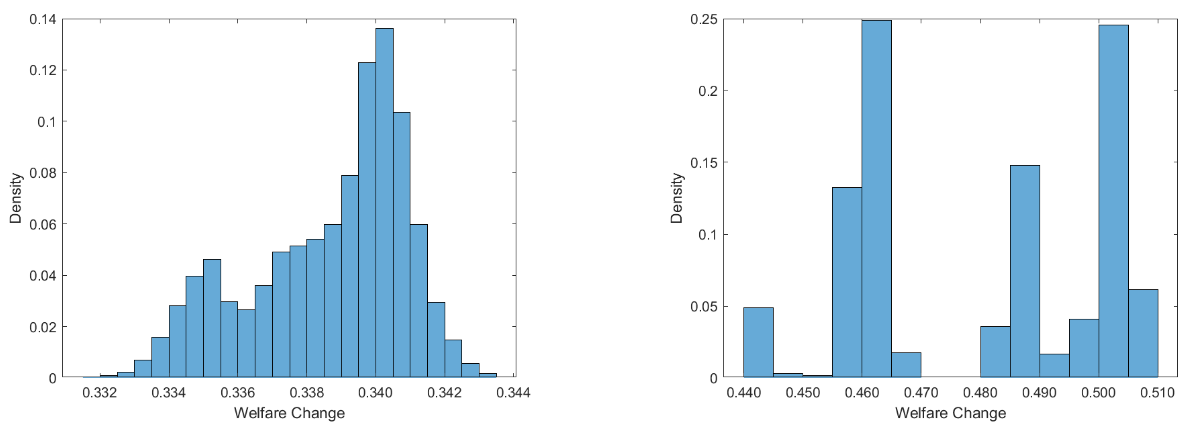


(a) Budget Store

(b) Premium Store

Note: This figure shows the welfare changes for households who previously lived within the 0.5-mile, 1-mile, and 2-mile radius surrounding the new grocery stores in the counterfactual equilibrium.

Figure 8: welfare Change by Income Group with New Budget Store

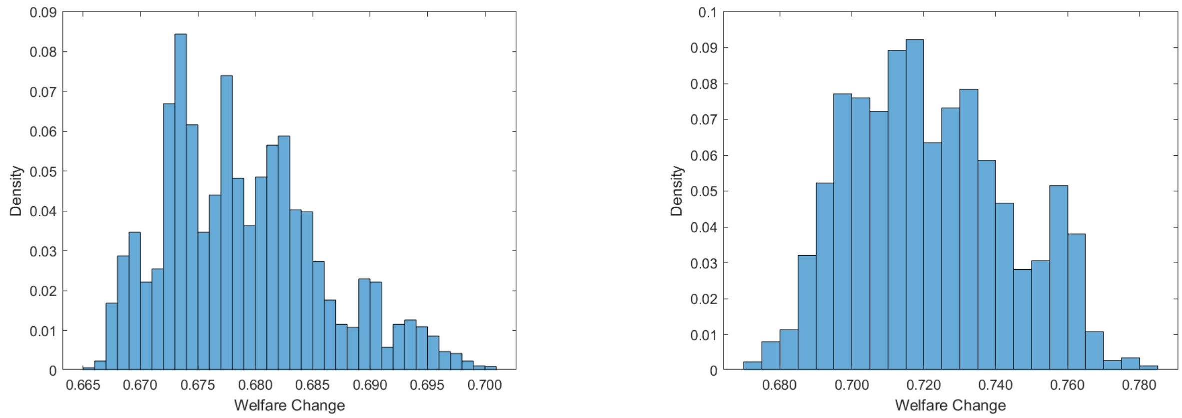


(a) High-income

(b) Low-income

Note: This figure shows the welfare changes in magnitude following the introduction of new budget grocery stores, segmented by income groups. High-income households are those with income at or above the 90th percentile. Low-income households are those with income at or below the 10th percentile.

Figure 9: welfare Change by Income Group with New Premium Store



(a) High-income (b) Low-income

Note: This figure shows the welfare changes in magnitude following the introduction of new premium grocery stores, segmented by income groups. High-income households are those with income at or above the 90th percentile. Low-income households are those with income at or below the 10th percentile.

6.2.4 Income inequality

I use Gini index to measure the income inequality before and after the introduction of counterfactual grocery stores. The Gini index, given by Equation (19), measures the extent to which the distribution of income within an economy deviates from a perfectly equal distribution. A Gini index of 0 represents perfect equality, while an index of 1 implies perfect inequality.

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \mu} \tag{19}$$

where n is the number of blocks, x_i andx_j are the income values of blocks i and j. μ is the mean of the income values. The calculations of the Gini index of income before and after the introduction of new grocery stores indicate a decline in the index from 0.2132 to 0.1304 for budget stores and to 0.1335 for premium stores. This reduction can be attributed to the relocation of high-income households closer to the new grocery stores, which are located in areas with mainly low-income households. Moreover, the outflow of high-income households from their previous neighborhoods reduces housing prices in those areas, attracting low-income households to move in. To further investigate the regional inequality, I compute the Gini index for each municipal township, as illustrated in Figure A1. Compare Figure A9 to Figure A8, I can see that the income inequality decrease significantly in the townships where the new grocery stores are located. This finding suggests that the introduction of grocery stores in areas with limited food access encourages household movement across regions, thereby reducing income disparity.

7 Conclusion

This paper examines the impact of consumption-oriented regional development policies on residential sorting, housing prices, welfare distribution, and income inequality. Leveraging detailed household travel survey data, property transactions, and business information from Google Map, I develop and estimate an equilibrium sorting model that jointly considers residential and consumption choices to evaluate the impact of urban policies aimed at improving household access to consumption amenities. Our findings demonstrate that access to consumption plays a significant role in residential decision-making. Households exhibit a willingness to pay a premium for proximity to consumption amenities. The counterfactual analysis reveals that policies aimed at expanding grocery store access in underserved areas can significantly influence local housing prices and welfare outcomes. The introduction of new grocery stores raises housing prices in nearby areas and induces household reallocation across regions. Though the low-income residents might be forced to move away from where they currently live due to the increase of house prices, they still experience welfare gains from the opening of new grocery stores for improved access of grocery. The results also indicate that these policies can contribute to reducing regional income inequality. This study contributes to the urban consumption and residential sorting literature by demonstrating the importance of city-wide consumption access as an endogenous factor in housing choices. By incorporating rich heterogeneity in household preferences, I offer a more comprehensive understanding of the equilibrium responses to regional development policies. Future research could further explore the role of other channels, such as labor market dynamics and business supply adjustments, in shaping the housing market and spatial distribution of economic activity.

References

S. Agarwal, J. B. Jensen, and F. Monte. The geography of consumption. *Available at SSRN 3002231*, 2017.

M. Almagro and T. Domínguez-Iino. Location sorting and endogenous amenities: Evidence from amsterdam. Technical report, National Bureau of Economic Research, 2024.

P. Bajari and M. E. Kahn. Estimating housing demand with an application to explaining racial segregation in cities. *Journal of business & economic statistics*, 23(1):20–33, 2005.

P. J. Barwick, S. Li, A. Waxman, J. Wu, and T. Xia. Efficiency and equity impacts of urban transportation policies with equilibrium sorting. *American Economic Review*, 114(10):3161–3205, 2024.

N. Baum-Snow and L. Han. The microgeography of housing supply. *Journal of Political Economy*, 132(6):1897–1946, 2024.

P. Bayer and R. McMillan. Tiebout sorting and neighborhood stratification. *Journal of Public Economics*, 96(11-12):1129–1143, 2012.

P. Bayer, F. Ferreira, and R. McMillan. A unified framework for measuring preferences for schools and neighborhoods. *Journal of political economy*, 115(4):588–638, 2007.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63 (4):841–890, 1995.

V. Couture. Valuing the consumption benefits of urban density. *University of California, Berkeley, Working Paper*, 2016.

V. Couture, C. Gaubert, J. Handbury, and E. Hurst. Income growth and the distributional effects of urban spatial sorting. *Review of Economic Studies*, 91(2):858–898, 2024.

J. Cuffey and T. K. Beatty. Effects of competing food desert policies on store format choice among snap participants. *American Journal of Agricultural Economics*, 104(4):1485–1511, 2022.

D. R. Davis, J. I. Dingel, J. Monras, and E. Morales. How segregated is urban consumption? *Journal of Political Economy*, 127(4):1684–1738, 2019.

D. Epple, R. Romano, and H. Sieg. The intergenerational conflict over the provision of public education. *Journal of Public Economics*, 96(3-4):255–268, 2012.

F. V. Ferreira and M. Wong. Estimating preferences for neighborhood amenities under imperfect information. Technical report, National Bureau of Economic Research, 2020.

M. M. Ferreyra. Estimating the effects of private school vouchers in multidistrict economies. *American Economic Review*, 97(3):789–817, 2007.

Y. Gu, C. Jiang, J. Zhang, and B. Zou. Subways and road congestion. *American Economic Journal: Applied Economics*, 13(2):83–115, 2021.

Y. Gu, N. Guo, J. Wu, and B. Zou. Home location choices and the gender commute gap. *Journal of Human Resources*, 59(2):545–575, 2024.

J. Gyourko and R. Molloy. Regulation and housing supply. In *Handbook of regional and urban economics*, volume 5, pages 1289–1337. Elsevier, 2015.

K. Haninger, L. Ma, and C. Timmins. The value of brownfield remediation. *Journal of the Association of Environmental and Resource Economists*, 4(1):197–241, 2017.

Y. Hwang. An estimable general-equilibrium structural model of immigrants' neighborhood sorting and social integration. Technical report, mimeo, 2019.

R. Jerch, P. J. Barwick, S. Li, and J. Wu. Road rationing policies and housing markets. *SSRN Journal. https://doi. org/10.2139/ssrn*, 3766254, 2021.

H. A. Klaiber and N. V. Kuminoff. Equilibrium sorting models of land use and residential choice. *The Oxford handbook of land economics*, pages 352–379, 2014.

H. A. Klaiber and D. J. Phaneuf. Valuing open space in a residential sorting model of the twin cities. *Journal of environmental economics and management*, 60(2):57–77, 2010.

G. E. Kreindler. The welfare effect of road congestion pricing: Experimental evidence and equilibrium implications.

N. V. Kuminoff, V. K. Smith, and C. Timmins. The new economics of equilibrium sorting and policy evaluation using housing markets. *Journal of economic literature*, 51(4):1007–1062, 2013.

D. McFadden. Modelling the choice of residential location. 1977.

Y. Miyauchi, K. Nakajima, and S. J. Redding. Consumption access and agglomeration: evidence from smartphone data. 2021.

M. Park and D. W. Hahm. Location choice, commuting, and school choice. *Commuting, and School Choice (July 24, 2023)*, 2023.

D. J. Phaneuf, V. K. Smith, R. B. Palmquist, and J. C. Pope. Integrating property value and local recreation models to value ecosystem services in urban watersheds. *Land Economics*, 84(3): 361–381, 2008.

J. M. Quigley and S. Raphael. Regulation and the high cost of housing in california. *American Economic Review*, 95(2):323–328, 2005.

D. Rao. The role of environmental amenities in the urban economy: Evidence from a spatial general equilibrium approach. *HKUST Business School Research Paper*, (2022-092), 2022.

H. Sieg, V. K. Smith, H. S. Banzhaf, and R. Walsh. Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International economic review*, 45(4): 1047–1077, 2004.

Y. Su. The rising value of time and the origin of urban gentrification. *American Economic Journal: Economic Policy*, 14(1):402–439, 2022.

C. Timmins and J. Murdock. A revealed preference approach to the measurement of congestion in travel cost models. *Journal of Environmental Economics and management*, 53(2):230–249, 2007.

V. B. Viard and S. Fu. The effect of beijing's driving restrictions on pollution and economic activity. *Journal of Public Economics*, 125:98–115, 2015.

R. Walsh. Endogenous open space amenities in a locational equilibrium. *Journal of urban Economics*, 61(2):319–344, 2007.

J. Yang, A.-O. Purevjav, and S. Li. The marginal cost of traffic congestion and road pricing: evidence from a natural experiment in beijing. *American Economic Journal: Economic Policy*, 12(1):418–453, 2020.

Appendix

A Estimation Details

A.1 Estimating Consumption Choice (Random Coefficient Model)

The parameters of the consumption choices using random coefficient model are estimated by simulated maximum likelihood estimation. Given the D Halton draws for each choice $m = 1, 2, ..., M$ for each household i, the log likelihood function is specified as:

$$
\begin{aligned}
LL &= \ln \left\{ \prod_{i=1}^{N} \left[\int_{\boldsymbol{\Theta}_i} \left(\prod_{m=1}^{M} P_{ijm|\boldsymbol{\Theta}_i}^{\delta_{ijm}} \right) f(\boldsymbol{\Theta}_i)\, d\boldsymbol{\Theta}_i \right] \right\} \\
&\approx \sum_{i=1}^{N} \ln \left\{ \frac{1}{D} \sum_{d=1}^{D} \left(\prod_{m=1}^{M} P_{ijm|\boldsymbol{\Theta}_d}^{\delta_{ijm}} \right) \right\}
\end{aligned}
$$

where

$$
P_{ijm|\Theta_d} = \frac{\exp\left(\mathbf{w}_{im}\boldsymbol{\eta}_i + \lambda_i \tau_{ijm}\right)}{\sum_{r=1}^{M} \exp\left(\mathbf{w}_{ir}\boldsymbol{\eta}_i + \lambda_i \tau_{ijr}\right)}
$$

and

$$
\delta_{ijm} = \begin{cases} 1 & \text{if } i \text{ chooses } m \\ 0 & \text{otherwise} \end{cases}
$$

A.2 Estimating of Housing Demand

Estimation of the housing demand follows the two-step procedure developed by Berry et al. (1995). I rewrite the indirect utility function as:

$$
\begin{aligned}
U_{ij} &= \alpha_i p_j + \mathbf{X}_j \boldsymbol{\beta}_i + \sum_s \phi_{is} EV_{ijs}(\tau_{js}) + \xi_j + \varepsilon_{ij} && (20) \\
&= (\bar{\alpha}_p + z_i' \boldsymbol{\alpha}_p) p_j + \mathbf{X}_j (\bar{\boldsymbol{\beta}} + \boldsymbol{\beta} z_i) + \sum_s \phi_{is} EV_{ijs}(\tau_{js}) + \xi_j + \varepsilon_{ij} && (21)
\end{aligned}
$$

$$U_{ij} = \mu_{ij}(\theta_1) + \delta_j(\theta_2) + \varepsilon_{ij} \tag{22}$$

$$\mu_{ij}(\theta_1) = z_i' \alpha_p p_j + \boldsymbol{X}_j \boldsymbol{\beta} \boldsymbol{z}_i + \sum_s \phi_{is} EV_{ijs} \tag{23}$$

$$\delta_j(\theta_2) = \bar{\alpha}_p p_j + \boldsymbol{X}_j \bar{\boldsymbol{\beta}} + \xi_j \tag{24}$$

where $\theta_1 = \{\alpha_p, \boldsymbol{\beta}, \gamma_i, \phi_{is}\}$ and $\theta_2 = \{\bar{\alpha}_p, \bar{\boldsymbol{\beta}}\}$. Notice that $\boldsymbol{\Phi}_i \sim MVN(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$. In the first step, I estimate $\Gamma = \{\alpha_p, \boldsymbol{\beta}, \Theta, \Sigma\}$ by maximum likelihood estimator (MLE). The estimation steps are:

1. Make D Halton draws for each choice $j = 1, 2, ..., J$ for each household i.

2. Pick starting values for Γ.

   (a) Pick starting values for δ_j, denoted as $\delta_j^{(0)}$

   (b) With the knowledge of δ_j and Γ, compute $S_j^{(0)} = \sum_{i \in C^{-1}(j)} \frac{1}{D} \sum_{d=1}^{D} w_i \frac{\exp\left(\delta_j^{(0)} + \mu_{dj}\right)}{\sum_{k=1}^{J} \exp\left(\delta_k^{(0)} + \mu_{dk}\right)}$

   (c) Compute $\delta_j^{(1)} = \delta_j^{(0)} + ln(s_j) - ln(S_j^{(0)})$, where s_j is the observed demand.

   (d) Iterate $2(a)$ to $2(c)$ till the contraction mapping $\delta_j^{(n+1)} = \delta_j^{(n)} + \ln(s_j) - \ln\left(S_j^{(n)}\right)$ converges. Let $\delta_j = \delta_j^{(n+1)}$

3. With the knowledge of Γ and converged δ_j, construct the log likelihood function
   $\ln(L) = \sum_{i=1}^{N} w_i \ln\left[\frac{1}{D}\left(\sum_{d=1}^{D} \prod_{j=0}^{J} P_{ij|d}^{I_{ij}}\right)\right]$

4. Pick value of Γ to maximize $\ln(L)$.

5. With the values of δ_j at the converged values of Γ, run 2SLS to recover $\theta_2 = \{\bar{\alpha}_p, \bar{\boldsymbol{\beta}}\}$. The instrument variable for price is constructed based on the exogenous characteristics of blocks that are located between 3 and 5 miles from a given block.

B Table

Table A1: MLE estimates for consumption with random coefficient

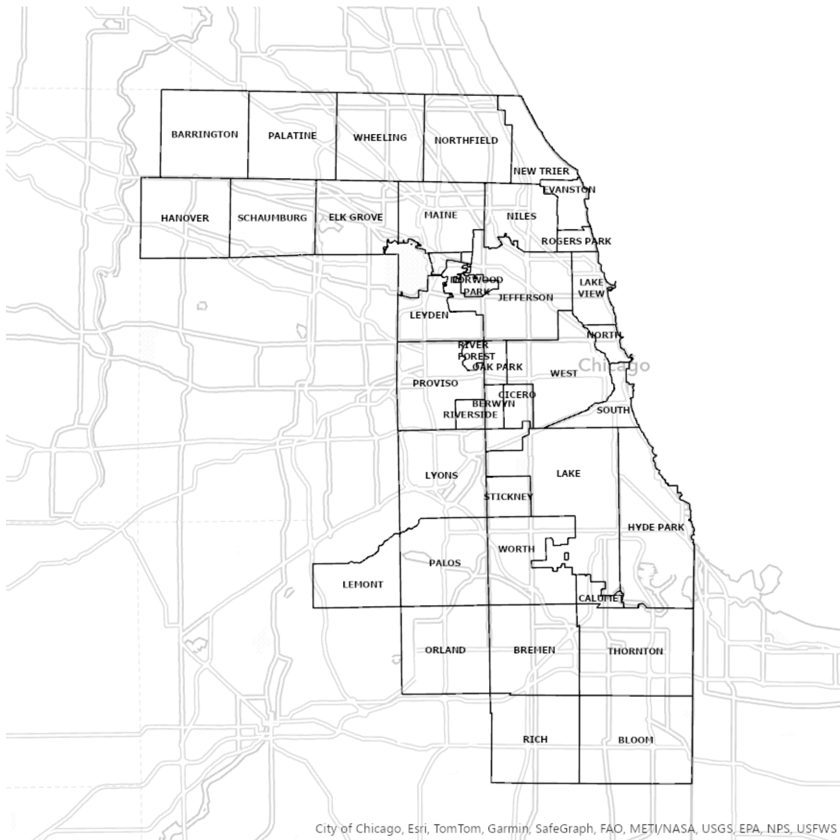
	(1) Grocery	(2) Restaurant
log(travel time)	-3.097	-2.723
	(0.011)	(0.004)
$$ bin	-2.358	0.760
	(0.129)	(0.041)
$$$ bin	-2.863	-3.640
	(1.041)	(0.286)
$$$$ bin		1.841
		(1.951)
rating	1.516	-0.337
	(0.021)	(0.005)
$$ bin × income	0.156	0.083
	(0.002)	(0.001)
$$$ bin × income	0.296	0.262
	(0.007)	(0.003)
$$$$ bin × income		0.146
		(0.005)
$$ bin × rating	0.430	-0.313
	(0.031)	(0.009)
$$$ bin × rating	-0.124	0.256
	(0.242)	(0.064)
$$$$ bin × rating		-0.485
		(0.425)

Table A2: Housing demand - nonlinear parameter for different size of random choice set

	(1) SS = 45	(2) SS = 40	(3) SS = 35
Demographic Interactions			
ln price per sqft × age	-0.127	-0.129	-0.127
	(0.038)	(0.042)	(0.037)
ln price per sqft × income	2.521	2.533	2.549
	(0.057)	(0.041)	(0.043)
ln living area × age	1.161	1.186	1.220
	(0.074)	(0.076)	(0.074)
ln living area × income	2.118	2.130	2.132
	(0.064)	(0.043)	(0.069)
ln # of bedroom × age	-1.072	-1.107	-1.141
	(0.090)	(0.112)	(0.101)
ln # of bedroom × income	-0.759	-0.784	-0.762
	(0.083)	(0.080)	(0.063)
ln # of bathroom × age	0.211	0.210	0.218
	(0.075)	(0.063)	(0.060)
ln # of bathroom × income	0.178	0.180	0.199
	(0.038)	(0.049)	(0.044)
ln property age × age	-0.422	-0.426	-0.418
	(0.039)	(0.023)	(0.033)
ln property age × income	-0.268	-0.256	-0.260
	(0.037)	(0.035)	(0.038)
ln green-area-ratio × age	0.009	0.009	0.009
	(0.005)	(0.005)	(0.005)
ln green-area-ratio × income	0.006	0.007	0.006
	(0.003)	(0.003)	(0.003)
ln # of schools × age	-0.043	-0.040	-0.041
	(0.007)	(0.007)	(0.007)
ln # of schools × income	-0.057	-0.058	-0.057
	(0.005)	(0.007)	(0.005)
ln commuting-time × age	0.096	0.107	0.096
	(0.078)	(0.074)	(0.072)
ln commuting-time × income	1.523	1.541	1.540
	(0.047)	(0.066)	(0.058)
Random Coefficient			
$EV_{grocery}$	2.250	2.245	2.240
	(0.065)	(0.068)	(0.052)
$EV_{restaurant}$	3.305	3.391	3.292
	(0.240)	(0.214)	(0.279)

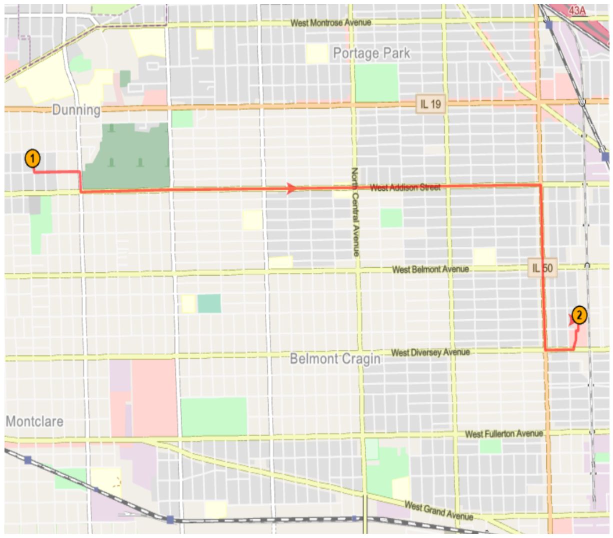
C Figure

Figure A1: Cook County (city of Chicago) townships



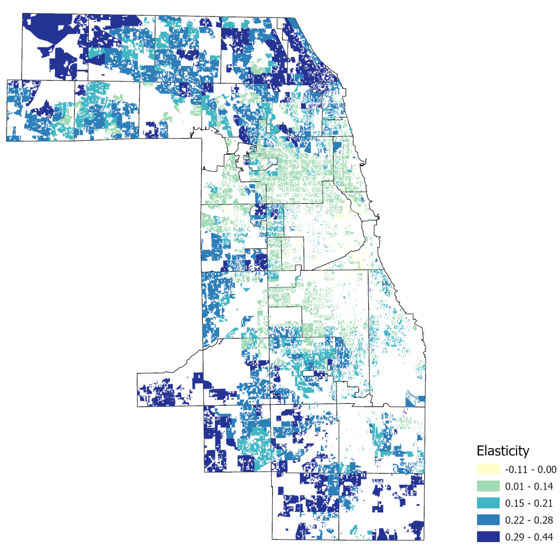
Note: This shows the map of townships of Cook county, where the city of Chicago seats. It is located in the state of Illinois on the southIstern shore of Lake Michigan.

Figure A2: Fastest driving route from stop 1 to 2



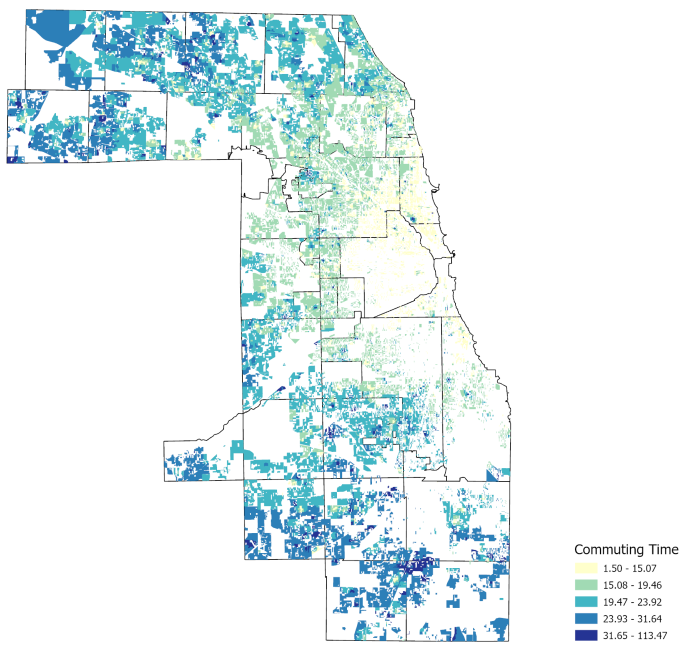
Note: This figure illustrates how I calculate the travel time and distance from location 1 to location 2. It is the time and distance along the fastest route between location 1 and 2, produced by Open source routing machine (OSRM). The travel time and distance from 1 to 2 and from 2 to 1 can be asymmetric.

Figure A3: Block housing supply elasticity



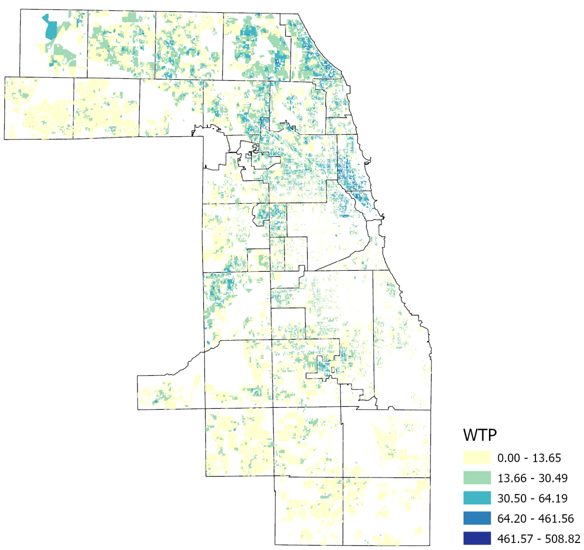
Note: This figure presents the housing supply elasticity at the block level, constructed based on the housing supply elasticity at the tract level from Baum-Snow and Han (2024).

Figure A4: Interpolated block commuting time

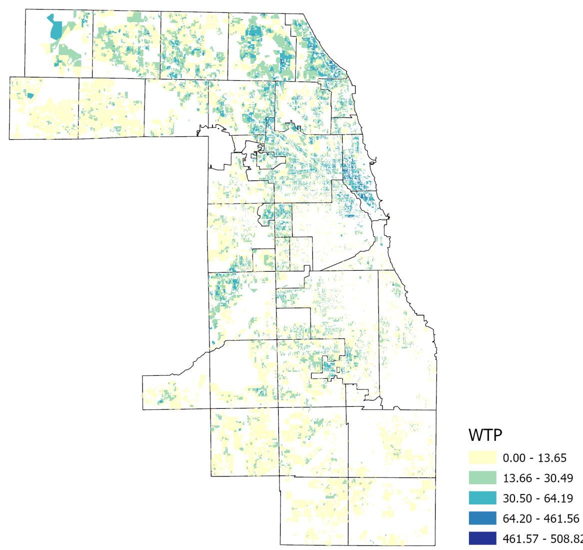


Note: This figure presents the commuting time at the block level, interpolated based on the commuting travel flow from CMAP My Daily Travel Survey in 2019.

Figure A5: Willingness to Pay



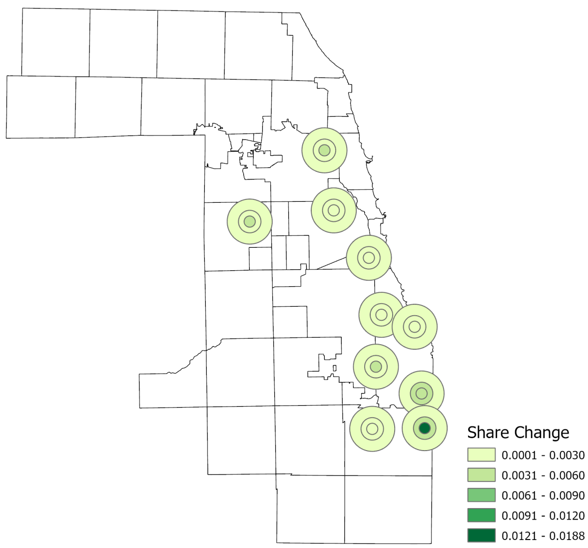
(a) Budget Store



(b) Premium Store

Note: This figure shows households' willingness to pay (WTP) for the grocery store expansion policy, constructed by searching for changes in housing prices per square foot that would keep households' utility constant.

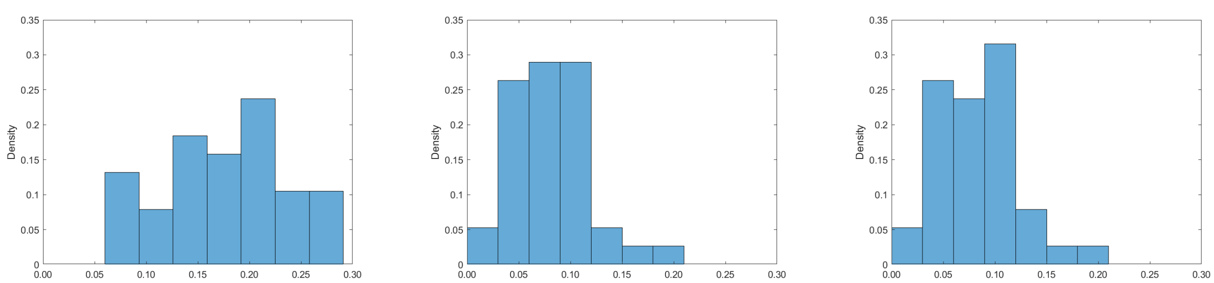
Figure A6: Resorting of Residents



(a) Budget Store (b) Premium Store

Note: This figure illustrates the dynamic of residents flow within the 0.5-mile, 1-mile, and 2-mile radius surrounding the new grocery stores in the counterfactual equilibrium. There is a net inflow if the share change is positive, and there is a net outflow if the share change is negative.

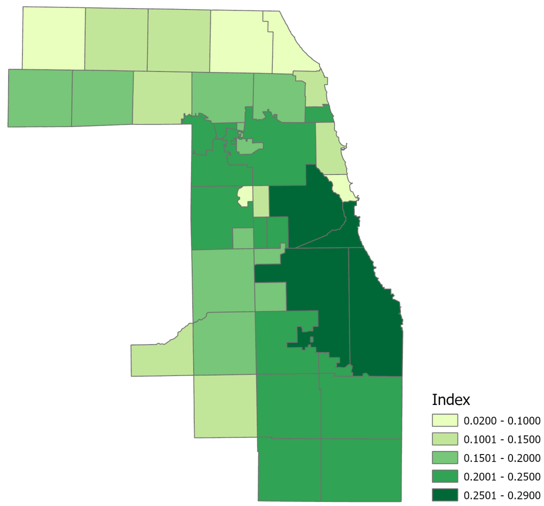
Figure A7: Income Gini Index



(a) Observed (b) Budget Store (c) Premium Store

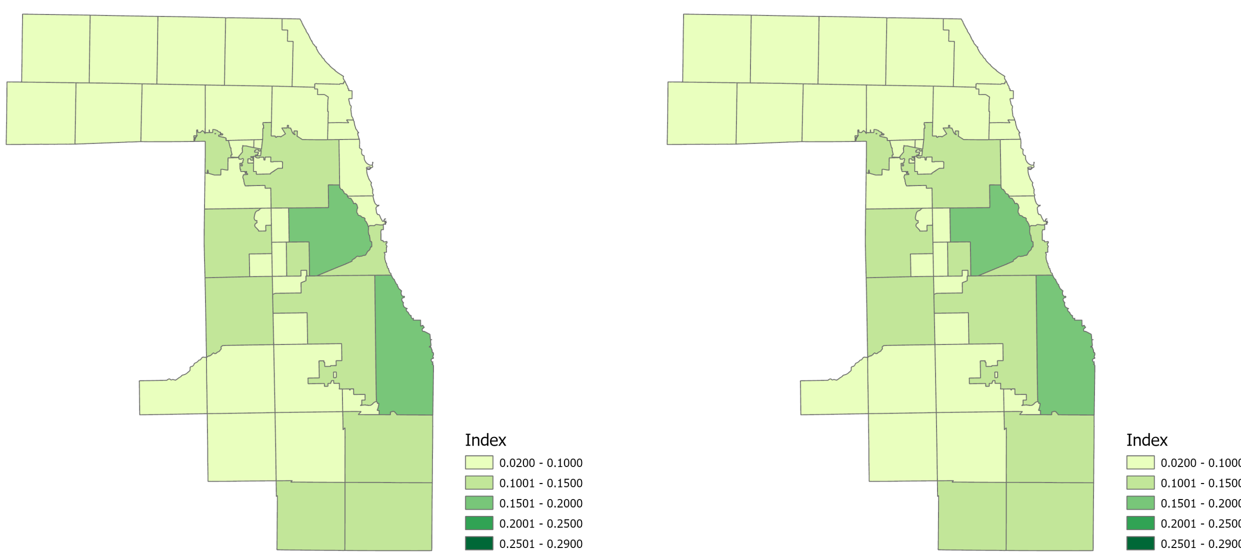
Note: This figure displays the distribution of Income Gini Index by municipal township before and after the introduction of counterfactual grocery stores.

Figure A8: Income Gini Index (Observed)



Note: This figure presents the observed Gini index for income by municipal township. The index is calculated according to Equation (19).

Figure A9: Income Gini Index in the Counterfactual



(a) Budget Store (b) Premium Store

Note: This figure presents the Gini index for income by municipal township in the counterfactual. The index is calculated according to Equation (19).