The Effect of Sprawl Development on Grocery Store Location and Food Access

Stephanie Schauder

December 27, 2019

Acknowledgements: I am very grateful for the advice, support, and feedback of my committee members, Miguel Gómez, Nicolas Ziebarth, and Oliver Gao in every step of this process. Additionally, this paper would not have been possible without the research of Nathaniel Baum-Snow and his willingness to make his data publicly available. I deeply appreciate Shyam Gouri Suresh who encouraged me to pursue this idea 6 years ago, and who has supported and encouraged my work ever since. I would also like to thank the many Cornell graduate students for their feedback and willingness to discuss ideas, in particular Abhishek Ananth, Sylvia Blom, and James Elwell.

Abstract

A large body of literature shows that poor and minority neighborhoods have fewer full-service grocery stores, but the determinants of this lack of food access are poorly understood. This paper explores the effect of sprawl development on access to fresh, healthy, and affordable food in the United States. I investigate to what extent the grocery landscape, and therefore food access, changed as a result of sprawl. The construction of the interstate highway system reduced transportation costs, causing many cities to develop in a sprawling pattern characterized by low population density and car dependence. Simultaneously, the grocery landscape began to change and small grocery stores that sold fresh food were driven out of business in favor of large supercenters. This paper first develops a theoretical model for consumer utility in the presence of heterogeneous transportation costs for grocery shopping that depend on travel mode. I employ an agent-based simulation to model this preference structure using data on the grocery environment in 3,140 U.S. counties. This simulation generates the testable hypothesis that car dependence causes a reduction in grocery stores and worsens food access. Finally, this paper uses an instrumental variables (IV) model to test this hypothesis with data from 239 major U.S. cities. The IV model finds that cities with greater sprawl development and car dependence are more likely to have food access problems. Specifically, a 1 percentage point reduction in the number of people driving to work alone is associated with a 1.4 percentage point reduction in the percentage of census tracts with low access to food, and a 1 percentage point increase in the number of people who walk reduces the percentage of tracts with low food access by 5.2 percentage points.

Contents

1 Introduction **4**

2 Background: The Highway System and Sprawl **8**

3 Theoretical Model **9**

3.1 Consumers 10

3.1.1 Optimal Quantity of Food 11

3.2 Grocery Stores 12

4 Agent-Based Simulation **12**

4.1 Data 13

4.2 Simulation Lifecycle 14

4.3 Parameterization 16

4.4 Results 17

5 IV Model **18**

5.1 Functional Form 19

5.2 Data 20

5.3 Results 22

6 Discussion and Policy Implications **24**

7 Conclusion **26**

1 Introduction

The interstate highway system fundamentally changed American transportation. Before cars became the predominant travel mode, people congregated in towns and cities, which provided necessary retail goods. However, highways reduced transportation time significantly. This allowed people to live further from the central business districts, and segregated cities as streets became more difficult to navigate for those without a car (Lutz, 2014). The resulting shift to a low density/car dependent development pattern, known as urban sprawl, created the suburban environment (Baum-Snow, 2007). Because, residents were now willing to travel further distances to shop due to reduced transportation costs, grocery stores found it profit maximizing to build fewer but larger stores, and supercenters developed (Ross, 2016). Oligopoly power in the grocery retail market has increased significantly since the 1970's, and in less populated areas, the industry is especially concentrated (Franklin and Cotterill, 1993). A case study of London, Ontario found that an increase in sprawl development and car dependence between 1961 and 2005 was associated with a decline from 45% to 18% in the number of census tracts with "easy access" to a grocery store (Larsen and Gilliland, 2008). At the national level, the number of grocery retailers per zip code in low-income areas decreased from 3.24 to 0.71 between 1970 and 1990, while grocery retailers in all zip codes decreased from 3.25 to 1.79 during the same time period (Thibodeaux, 2016).

In this paper, I first develop a theoretical model to understand how sprawl might affect grocery store location. The theoretical model explores shopping behavior in the presence of heterogeneous transportation costs. Then I investigate the implications of this model with an agent-based simulation using data on actual grocery store density, transportation patterns, and population in United States counties. The simulation generates a testable hypothesis, which I evaluate with an instrumental variables (IV) model.

Food security relies on having access to not only a subsistence number of calories, but also to affordable and nutritious food (U.S. Department of Agriculture, 2018). The Food, Conservation, and Energy Act of 2008 (or the "Farm Bill") defined a food desert[1] as, "an area in the United States with limited access to affordable and nutritious food, particularly such an area composed of predominantly lower-income neighborhoods and communities" (110th Congress, 2008). The topic of food access in the United States has been widely studied and discussed. People living closer to fresh food eat more fresh food (Bodor

[1]The term food desert has been rebranded as an "area with low food access." I use the terms interchangeably.

et al., 2008; Zenk et al., 2009) and have lower rates of obesity (Michimi and Wimberly, 2010; Lopez, 2007; Schafft et al., 2009). Lack of access to healthy food is also associated with kidney disease and hypertension (Suarez et al., 2015).

However, a growing body of research finds that improving food access alone does not necessarily translate to better nutrition (Dubowitz et al., 2015; Cummins et al., 2014; Handbury et al., 2015; Lee, 2012; Allcott et al., 2019). Even when healthy food is available, there are still many obstacles to consuming it such as price, preparation time, and tastes and preferences. It is possible that access to healthy food may be more of a necessary condition than sufficient condition for improving diet. Research suggests that nutritional interventions are more effective for people who have better access to healthy food (Wedick et al., 2015). It has also been proposed that the reason the food desert literature has found unimpressive results from improving the food environment is not because such changes are unimportant, but rather because the long-term persistence of food patterns since childhood is so strong that it is difficult to overcome them in a short time frame (Hammond et al., 2012). Not only is traveling further to shop an inconvenience but also growing up in a food environment where high fat and rewarding foods are prevalent can shape dietary behaviors for years to come (Teegarden et al., 2009).

Beyond nutrition, food deserts exacerbate race and class disparities in quality of life. There are three times as many supermarkets in wealthy neighborhoods as in low-income neighborhoods, and four times as many supermarkets in white neighborhoods as in predominately black neighborhoods (Morland et al., 2002). Food deserts often have lower quality food than non-food deserts, decreasing the desirability of eating fresh food (Hendrickson et al., 2006; Zenk et al., 2006). Income is correlated with car ownership (Dargay, 2001) meaning low-income households and racial minorities often have the least access to transportation, and are least equipped to live in food deserts, making shopping highly difficult. The sociology literature has extensively documented the institutional race and class barriers to healthy, affordable, and culturally appropriate food, and how this affects quality of life (Alkon and Norgaard, 2009; Hossfeld et al., 2017). The food justice movement is evidence that easy access to healthy food is something that people value for its own sake (Alkon and Norgaard, 2009).

Bitler and Haider (2011) remark that, while much research focuses on the existence, importance, or health effects of food deserts, relatively few studies try to understand why food deserts may arise in the first place. Since 2011, a few researchers have attempted to answer this question, primarily via agent-based models because agent-based modeling has been identified as an effective way to study food access (Li et al., 2016).

Agent-based models are useful for exploring the implications of changes to the food environment because they help policy makers understand assumptions about interventions and explore the effects of complicated interactions between variables (Chalabi and Lorenc, 2013). Blok et al. (2015) and Auchincloss et al. (2011) both use agent-based frameworks to look at the effects of residential segregation on food access. Blok et al. (2015) find that residential segregation has the greatest impact on inequalities in food access followed by prices and education. Auchincloss et al. (2011) find that when income segregation was reduced, inexpensive but healthy food stores moved into low-income neighborhoods and diet was greatly improved. Koh et al. (2019) develop an agent-based model to study food security. They find that increasing income and providing transportation options would improve food access. Widener et al. (2013) use an agent-based model to show that the most effective way to increase fruit and vegetable consumption in low-income populations is to increase grocery shopping frequency. In fact, research shows that those who use public transit, walked, or biked had more frequent and smaller grocery trips than those who used cars (Jiao et al., 2016).

In the agent-based simulation used in this paper, consumers choose where to shop and how much to purchase to maximize utility and grocery stores choose where to locate, and how much to charge to maximize profit. Because sprawl is characterized by "an almost total *reliance upon the automobile*," consumers differ in that some have cars and some do not in proportion to the number of people that drive in each county (Burchell et al., 1998). Those who have cars receive linear disutility from traveling to the grocery store, while those who do not have cars have exponential disutitly from traveling. The simulation is able to explain 83% of the variation in the number of grocery stores between counties. The model finds that decreasing the percentage of people who drive alone by 10% increases the number of grocery stores per 1,000 people by 0.08 (32% of the mean).

The simulation suggests that car dependence causes a decrease in grocery stores. I test this result empirically with an instrumental variables (IV) model. Studying the effect of car dependence and sprawl on food access is inherently difficult because living in a particular area is a choice and cities are shaped by the preferences and desires of their inhabitants. The instrument I use is 1947 planned highway rays. This is the same identification strategy as Baum-Snow (2007), and hinges on the fact that the 1947 highway plan is both correlated with measures of sprawl development and only affects food access through its effect on sprawl development. I find that measures of car dependence have significant negative effects on food access, while measures of lack of car dependence have significant positive effects. For example, a 1 percentage point decrease in the percentage of people driving and walking is associated with a 1.4 percentage point reduction and 5.2 percentage point increase, respectively, in the percentage of census tracts which have low access to food.

This simulation contributes to the growing body of literature which uses agent-based modeling to understand food access. This study is the first to use data from over 3,000 United States counties to simulate the effect of sprawl development on food access. Both in the simulation and in the data, the number of grocery stores was negatively correlated with commuting patterns. However, interestingly the correlation was slightly higher in real life than it was in the simulation. This suggests that in a reduced form model, commuting patterns are probably endogenous since they were associated with food access above and beyond what the simulation predicts, and thus justifies the use of the IV model.

The IV model contributes to the literature on food access by proposing a possible mechanism for how food deserts develop thus filling the knowledge gap noted by Bitler and Haider (2011). It is important to understand the implications of sprawl development so that city planners can make informed decisions when shaping the growth of rapidly expanding cities. Additionally, from an international perspective, it is useful to understand how food deserts have evolved in the United States. The global “supermarket revolution” began in the 1900’s and characterizes the spread of large chain grocery retailers to the global south and Asian countries. As a result, the supermarket industry has become increasingly concentrated, which threatens local small retailers and farmers who cannot sell the large supermarket chains (Reardon and Gulati, 2008). An increase in car ownership, among other factors, is implicated in the spread of supermarkets, and a study of Nairobi, Kenya finds that supermarkets disproportionately benefit the wealthy and do little to resolve problems of food insecurity (Berger and van Helvoirt, 2018). Many of these cities are not yet sprawling and car dependent and still have time to make changes.

The rest of the paper proceeds as follows: Section 2 provides provides background information on the development on the highway system and why its development affected food access. Section 3 discuses the theoretical model. Section 4 describes the agent-based simulation and presents the results. Section 5 discusses the IV model and presents the results. Section 6 provides a discussion of the results and corresponding policy implications. Section 7 concludes.

2 Background: The Highway System and Sprawl

In the late 1930's the U.S. federal government became interested in building a national system of roads. The first legislation to this effect was the Federal Aid Highway Act of 1944, which instructed the federal government to form a plan for a national highway system (Weber, 2011). It is difficult to determine if a shift to low-population density car dependent cities *caused* the decline of local grocery stores because there are many confounding factors. As early as 1956, easing congestion became a goal of the interstate highway system (Weingroff, 2017). This means that cities with a large population may have more highways, or that cities with worse public transit networks may have required more highways. Additionally, by the 1960's, city residents began to protest the construction of highways through their neighborhoods. White educated communities were most successful at diverting highway construction, and communities that lacked political power, usually minority neighborhoods, were more likely have their protests ignored (Brinkman and Lin, 2019). Race and education are both associated with food deserts so this is a clear confounding factor. Furthermore, it is not hard to imagine that local commuting patterns would be reflective of local culture and that those who value walking and public transportation may also value healthy food.

The ideal randomized control trial to test the effect of sprawl development on food access would entail having a set of identical cities. In half of the cities, developers would be instructed to invest all transportation funds in building highways, parking lots, and all infrastructure which favors automobiles. In the other half of the cities, the developers would be instructed to invest all transportation funds in a robust public transportation system, walkable streets, and infrastructure for bicyclists. Then 40 or 50 years later, researchers could evaluate food access in each set of cities and determine how each development strategy affected the location, size, and prices of grocery stores. However for obvious reasons, such an experiment would never be feasible or practical. Therefore, the next best option is to find an exogenous reason why some cities would have developed in a more car dependent manner than others.

Before the construction of the interstate highway system, cities were not car dependent because the infrastructure did not exist. Baum-Snow (2007) uses a 1947 interstate highway map as an instrument for actual highways built in order to determine if the construction of highways caused suburbanization. He notes," The validity of the 1947 plan as an instrument depends on the fact that the portion of the system in the plan was designed to facilitate trade and national defense, not to facilitate metropolitan area development" (Baum-Snow, 2007). I choose to use the same 1947 highway map as an instrument for two indicators of car dependence (current highways and the percentage of people who drive) and two indicators of lack of car dependence (the percentage of people who use public transportation and the percentage who walk in the city).

The 1947 highway map shown is shown in Figure 1. This map depicts the location of the highways which were proposed in 1947, but may not have actually been built (Bureau of Public Roads, 1955). In constructing this map, the key considerations were: connecting major cities, expediting the shipment of goods, and facilitating the movement of the military; improving traffic for local commuting was not a listed consideration for the 1947 plan although reducing congestion soon became important and was noted as a concern as soon as 1956 (Baum-Snow, 2007; Weingroff, 2017). Therefore cities in the 1947 plan, received highways based on proximity to other major cities not necessarily on the needs or desires of the city itself.

3 Theoretical Model

I first develop a theoretical model to explore why sprawl development and car dependence would affect food access. The key assumption of this model that distinguishes it from a standard general equilibrium model, is that transportation to a grocery store has a heterogeneous effect on consumer utility depending on car ownership.

3.1 Consumers

Suppose consumer i in time t maximizes its utility function $U_{it} = f(F_{itg}, X_{itg}, K_{ig})$, which is a modified Cobb-Douglas function where:

$$U^*_{it} = \max_{g \in G} \left[\max_{F_{itg}, X_{itg} \in [0,\infty)} \left[(F_{itg})^f (X_{itg})^{1-f} - K_{ig} \right] \right], \tag{1}$$

subject to the constraint

$$I = P_{gt} F_{itg} + X_{itg}. \tag{2}$$

Where G is the set of all grocery stores, and F_{itg} is the quantity of food purchased by consumer i from grocery store g in time t. f is the percentage of income spent on food. X_{itg} is the amount of all other goods consumer i consumes if it goes to grocery store g in time t.[2] I is the consumer's income, which is constant across consumers and over time. P_{gt} is the price of food F at grocery store g in time t. The price of X is normalized to 1. K_{ig} is the negative utility that consumer i experiences from traveling to grocery store g. Where:

$$K_{ig} = \begin{cases} Y * distance_{ig} & \text{if the consumer drives} \\ W * (Z^{distance_{ig}}) & \text{if the consumer walks} \end{cases} \tag{3}$$

K_{ig} is a piecewise function which allows the negative utility from traveling to the grocery store to differ depending on if the consumer drives or walks. Walkers have an exponential function because the desire to walk to a grocery store decreases very steeply with distance due to the physical difficulty of walking long distances with groceries. Drivers have a linear function because traveling to the grocery store does not become increasing difficult with distance. *distance* is the euclidean distance between the position of consumer i and grocery store g. Y, W, and Z are parameters. Neither consumers nor grocery stores can move, so K_{ig} does not vary over time.

[2]Note that X_{itg} is not purchased from a grocery store. X_{itg} is simply the allocation consumers make for other necessary goods. The magnitude of X_{itg} varies depending on the grocery store the consumer selects only because the amount of money the consumer has leftover will vary by grocery store

I intentionally choose to model transportation effort as a reduction in utility rather than as a cost in the budget constraint. If transportation effort were modeled as a cost to the budget constraint, the distance between consumers and the grocery stores they frequent would affect how much money they spend at the grocery store. This is not a realistic assumption. Including an additive term in the utility function discourages the person from traveling far distances to shop but does not change the amount of food the consumer will purchase as a result of travel.

3.1.1 Optimal Quantity of Food

U^*_{itg} represents the utility consumer i gets in time t from its utility maximizing bundle (F^*_{itg}, X^*_{itg}) if it frequents grocery store g. Where:

$$U^*_{igt} = \max_{F_{igt}, X_{itg} \in [0,\infty)} \Big[(F_{itg})^f (X_{itg})^{1-f} - K_{ig} \Big] \tag{4}$$

Note that equation 4 is different from equation 1 because U^*_{itg} is the maximum utility consumer i can receive from grocery store g in time t, while U^*_{it} is the maximum utility consumer i can receive from *any* grocery store in time t. Suppose G is the set of all grocery stores. For consumer i there will be a set $U_{it} = \Big\{ U^*_{itg} \mid g \in G \Big\}$. U^*_{it} is the maximum value in this set.

The grocery store which maximizes the consumer's utility is denoted $g*$. The amount of food purchased at $g*$ is F^*_{itg*}. Because the utility function is a Cobb Douglas function with an additive term at the end, maximizing utility yields the Cobb Douglas solution:

$$F^*_{it} = \frac{f * I}{P_{tg*}} \tag{5}$$

However, in order to find the value of F^*_{it}, g* must first be identified.

3.2 Grocery Stores

Grocery stores have perfect information and know how many consumers will choose to shop at their store at any given time, and how much they will purchase there. Therefore, stores can perfectly predict how much of F they will sell and only stock this amount. Existing stores can only control how much they will charge for F. At certain points in the simulation, stores have the opportunity to open or close. When a new firm opens, it is able to choose the location where it will make the highest profit given current prices. However, existing stores are never able to change location.

Each grocery store g seeks to maximize profit in the current time period t. The profit function is $\Pi_{tg} = TR_{tg} - TC_{tg}$, where Π_{tg} is profit, TR_{tg} is total revenue, and TC_{tg} is total cost for store g in period t. The total cost to the firm is $TC_{tg} = AVC * F_{tg} + FC$, where FC is fixed cost, and the average variable cost, AVC, is equal to a constant and therefore equal to marginal cost, MC. MC is the cost to the store to stock each unit of F. The total revenue is $TR_{tg} = P_{tg} * F_{tg}$. All stores sell F, and each store g chooses a price P_{tg} in period t to charge for F. F_{tg} is the quantity of F sold by store g in period t.

G is the set of all grocery stores in the simulation, but the number of elements of G is dynamic until an equilibrium is achieved. After n periods, if a firm is making a negative accounting profit ($\Pi_{t=n} < 0$) it closes and is removed from the model. If a firm is making a positive economic profit (as opposed to the accounting profit), defined as $\Pi_{t=n} \geq E$, a new grocery store enters the model to compete with it.[3] The new firm chooses the profit maximizing location.

4 Agent-Based Simulation

To explore the implications of this theoretical model, I use an agent-based model to simulate the formation of grocery stores in each of the 3,140 United States counties. The county is represented by a 10 x 10 grid. Agents are consumers and grocery stores. The number of consumers in the model is the population of the county. Consumers solve the utility maximization problem described in Equations 1, 2, and 3 above.

[3] The reason there is a period where $0 \leq \Pi_{t=n} < E$ where firms neither enter not exit the market as a result of this firms profit, is to model the situation where firms are not losing enough money to go out of business but also not generating enough profit to attract additional competitors.

Consumers are assigned to either drive or walk corresponding to statistics on commuting patters for each county. Grocery stores maximize profit as described above.

The simulation predicts the number of grocery stores that should be in each county given the assumptions of the theoretical model. I compare this predicted number stores to the actual number of grocery stores in each county. However, the goal of the simulation is not to be as accurate as possible, but rather to understand what the assumptions of the theoretical model reveal about the relationship between car dependence and food access. For that reason, I do very limited calibration of parameters because I do not want to force the simulation to conform to the actual data. I do evaluate how well the simulation predicts the number of grocery stores in real life, but this is not a target in and of itself.

In the simulation, the only variables that distinguish one county from another are commuting patterns and population. Therefore any effect of these variables on grocery store density is causal, within the simulation. This does not mean that sprawl causally affects food access in real life, but it demonstrates a mechanism whereby sprawl could affect food access. The agent-based simulation yields the hypothesis that sprawl development reduces food access, and this hypothesis can be tested with the IV model.

4.1 Data

The percentage of people who commute by driving alone in a county corresponds to the percentage of consumers that drive in the simulation. The remainder of the consumers are assumed to walk. The data on driving patterns and population come from the American Community Survey (ACS) (U.S. Census Bureau, 2016). The number of grocery stores per county and grocery store density comes from the Food Access Research Atlas (FARA) a dataset and interactive tool created by the U.S. Department of Agriculture Economic Research Service (Ver Ploeg et al., 2017). The county level was used because this is the smallest Census denomination that was available in all datasets. All data come from the year 2014.

Table 1 shows the summary statics for data used in the simulation. The outcomes of interest are "Number of grocery stores" and "Grocery stores per 1,000 people" (henceforth referred to as GSPP). The data contain 3,140 counties out of the original 3,142.[4] The average county has about 98,322 residents and

[4] Two counties were exclude due to missing data.

79% of residents commute by driving alone. However, there are counties where as few as 6% of residents commute by driving alone. The huge variation in the percentage of people who drive between counties is beneficial because the simulation is run in a wide variety of environments. The average GSPP is 0.25. The average number of grocery stores per county is 21.

Parameterizing K_{ig} in Equation 3 is difficult because it is hard to evaluate how much negative utility consumers receive from walking or driving to a grocery store. To attempt to answer this question, I use data from the National Household Food Acquisition and Purchase Survey (FoodAPS), also compiled by the ERS (U.S. Department of Agriculture Economic Research Service, 2017). FoodAPS is a nationally representative dataset which records detailed information at the household level about food purchases and consumption. FoodAPS records how survey participants get to the grocery store (walking or driving), and the distance they travel to get there. I use this data to find the parameters (W, Y, and Z) for equation 3. These parameters do not differ by county because the number of people in each county is too small.

Table 2 presents the FoodAPS data. There were 4,151 people who drove in the sample and 267 people who walked in the sample. Of the people that drive to the grocery store, the mean length of trip was 4.7 miles. Of the people who walk to the grocery store, the mean length of trip was 0.5 miles. The longest anyone was willing to drive to the grocery store was 156 miles and the longest anyone was willing to walk was 1.4 miles.

4.2 Simulation Lifecycle

The simulation is essentially a function where the inputs are the percentage of people that drive in every county and the county population. The output is the predicted number of grocery stores per county. Figure 2 shows that a 2 X 3,410 vector with the population and the percentage of people who drive in each county is fed into the simulation. The simulation runs separately for each county and yields a 1 X 3,140 vector which is the predicted number of stores in each county. The GSSP is calculated after the fact by dividing predicted grocery stores by population in thousands.

For each county in the simulation, a 10 X 10 grid is initialized and populated with the number of consumers equal to the population of that county. These people are randomly distributed across the 100 locations provided in the 10 x 10 grid.[5] The initial number of firms is set to equal the mean GSPP, 0.25. These initial firms are located randomly within the grid. However as the simulation runs, firms close or open depending on how much profit they make. The initial price all firms charge is marginal cost. For 15 periods, the original firms update prices to reflect the profit maximizing price given last period's prices. Each firm simultaneously completes an exhaustive search between the prices $MC - 1$ and $MC + 20$ for the profit maximizing price given the prices of each of the other firms in the previous period. Then profit is calculated for each firm. At the end of the 15 periods, firms are given the chance to open and close. If a firm is making a negative profit it closes and is removed from the model. If a firm is making not just an economic profit but an accounting profit, defined as $\Pi + E$, another firm enters to compete and searches the entire grid for the profit maximizing location and locates there. If all the firms in the model close, a new firm is added in the profit maximizing location.

Therefore, after firms are given the chance to close or open, there may be a different number of firms than the original number. Alternatively, the same number of firms could exist but they may be different firms and thus located in different locations (if an equal number of firms close and open). While existing firms can never change location, new firms locate in the profit maximizing location and are then "stuck" in that position for the rest of the simulation. A new firm can locate in the same spot as an existing firm if this is optimal.

Next, the current set of firms (the original firms plus any new firms that opened and minus any firms that closed) updates prices for 15 periods, just as before, after which firms are either added or eliminated again. This cycle continues 10 times for a total of 150 periods.[6] In this way, the original firms will only exist at the end of the simulation if they were located in a very profitable location. Figure 3 shows a visual representation of the lifecycle of this model. The simulation was run 50 times, and the output of all 50 rounds was averaged for each county.

[5] Due to computing time, if the population is greater than 80,000 people, the county is divided into equal sections of no more than 80,000 people. The model is run for a representative section and the predicted number of stores in that county is scaled up by the population.

[6] Ideally, the cycle would continue until equilibrium (i.e. the number and position of firms no longer changes). However, do to computing time this was not possible. Through experimentation I determined that the quantity of firms does not change significantly after 150 periods.

4.3 Parameterization

To parameterize the model, I chose values that most correspond to the reality of food access in the United States. I is is set to \$59,000 which is the average household income for the counties in the model. f is set to .15, to match the amount of income the average family spends on food (Tuttle and Kuhns, 2016). MC is chosen to be 20. Because the unit of food is arbitrary, the actual cost of the food is irrelevant. Because of the structure of the Cobb-Douglas function, consumers will always spend 15% of income on food. FC and the additional amount necessary to make an economic profit, E, were calibrated to minimize squared error in predicted grocery stores using a 5% sample of counties. These values were: $FC = \$3,000,000$ and $E = \$3,000,000$. It is important to note that this calibration only affects scale and not dispersion in the number of grocery stores which is the main concern. For example, lower values of FC and E would result in a similar correlation between predicted and real grocery store numbers, but the absolute number of grocery stores would be different.

The dispersion in the number of grocery stores is entirely affected by the parameters in the transportation effort function. The parameters that determine K_{ig} (W, Y, and Z) are determined using the FoodAPS data described above. If the customer drives a car, K_{ig} is a linear function of distance and if the customer walks, K_{ig} is exponential function of distance. These functional forms were assumptions that I made based on the theory explained above and are not based on data. To find the optimal values of W, Y, and Z, I assume that disutility from traveling is the same at the mean driving distance and the mean walking distance for drivers and walkers respectively. I also assume disutility at the maximum driving distance and maximum walking distance are the same. Given these two conditions I found the exponential and linear functions which best fit the data. The corresponding parameters are: $W = 0.58$, $Y = 0.67$, and $Z = 39.5$. These parameters were decided before running the simulation, and they were *not* calibrated to the grocery store data. This is because if the parameters were calibrated to the grocery store data, the difference in grocery stores between highly car dependent and less car dependent areas could be caused by reasons other than car dependence.

4.4 Results

The simulation predicts the number of grocery stores that should exist in every county based *only* on population and the percentage of people who drive. Table 3 shows that the correlation between the percentage of people driving and the predicted number of stores and GSPP is -0.20 and -0.37 respectively. This means that given the assumptions about transportation costs and the utility function, driving has a negative effect on the quantity of grocery stores. It is important to understand that this is not an association but rather a causal relationship within the simulation. Figure 4 graphically shows the relationship between predicted GSPP, and the percentage of people who drive in a city. The red line is the best fit line. Table 4 shows that a 1 percentage point increase in the percentage of people who drive causes a 0.008 reduction in GSPP which is 3.5% of the mean.

Another interesting result from Table 3 is that the correlation between the number of stores and percent driving and the correlation between GSPP and percent driving are more negative in the actual data than in the predicted data (-0.27 vs -0.20 and -0.48 and -0.37 respectively). This indicates that the relationship between commuting patterns and grocery stores is actually stronger in the real data than it is in the simulation. This result reveals that there are likely reasons for the decrease in grocery stores in car dependent areas other than high transportation costs. This was not unexpected, and confirms the necessity of addressing the endogeneity in the reduced form analysis by finding an instrument. Non-car dependent areas may have more grocery stores because they are also wealthier or because preferences are different in these areas. This agent-based simulation reveals that higher transportation costs explains some of the variation in grocery store density between sprawling and non-sprawling areas, but not all of it.

Table 5 evaluates how well the predicted GSPP and number of grocery stores match the actual GSPP and number of grocery stores. The root mean squared error (RMSE) for the number of grocery stores is higher than the mean absolute error (MAE) indicating that the predicted number of grocery stores is not accurate for extreme values. After eliminating the 13 counties that have more than 500 grocery stores, the RMSE for the number of grocery stores drops to 13.3 (about half the average number of stores). The RMSE for GSPP remains the same even when excluding the 13 outliers, indicating that the population weighted number of grocery stores is still very accurate for extreme values.

The adjusted R^2 for the number of stores is .83 indicating the simulation is able to explain 83% of the variation in the number of stores between counties. However the number of stores takes into account both population and commuting patterns. When looking at the predictive power of commuting patterns alone (GSPP), we see that the percentage of people driving and the assumptions of the model are able to explain 12% of the variation in GSPP between counties. The objective of the simulation was not to predict these values as accurately as possible. I could have included many more variables and assumptions which would have lead to greater predictive power. However, doing this would make it difficult to isolate the predictive power of sprawl development. By only allowing population and transportation patterns to vary, I can observe how these variables alone affect the number of grocery stores.

5 IV Model

The theoretical model proposes a mechanism through which sprawl development may affect food access. Next, the agent-based simulation employs the theoretical model and generates the hypothesis that sprawl development negatively affects food access. The goal of the IV model is to evaluate this hypothesis. To do this, I look at the effect of indicators of sprawl development on food access using the 1947 highway plan as an instrument.

In order to identify the local average treatment effect, it is imperitive to satisfy the following conditions as outlined by Angrist and Pischke (2008) which I reframe to fit this context:

1. The instrument (1947 planned highways rays) has a causal effect on measures of car dependence
2. The 1947 highway plan affects food access *only* through its affect on car dependence
3. If the presense of more highways in the 1947 highway plan is causally related to higher levels of car dependence, then there can be no cities for which the presense of a greater number of highways in the 1947 highway plan causes the city to have lower levels of car dependence (No defiers)

Condition 1 is easily verified by looking at the first stage F-statistic. The F-statistic should be larger than 10 if the instrument has a causal effect on the measure of car dependence. The second two conditions must be motivated by reasoning. As discussed in Section 2, the validity of the second condition relies on the fact that the 1947 plan allocated highways based on proximity to other major cities. The goal was to facilitate travel between cities more than to provide transportation within them. Therefore, the original highway plan should not affect food access outside of its affect on future commuting patterns. The third condition says that if the presense of more highways in the 1947 plan nudges some cities to actually construct more highways, there can be no cities for which the presence of more highways in the 1947 plan would cause it to have less highways than if it was not nudged by the plan. It is difficult to think of a reason why the presense of more highways in the 1947 plan would have a negative effect on the actual number of highways in given city, so condition 3 is a reasonable assumption.

The primary measure of the extent of the interstate highway system in a particular city is a "ray." Baum-Snow (2007) defines as ray as:

> "A segment of road that connects the central business district (CBD) of the central city with the region outside the central city. If a highway passes through the central city, it counts as two rays whereas if a highway terminates in or near the central city it counts as only one. Rays must pass within one mile of the CBD and serve a significant area of the MSA outside the 1950-definition central city to be counted. Highways that split at or near the border of the 1950-definition central city count as multiple rays. Two highways that pass within one mile of the CBD and converge count as only one ray in the direction of convergence (Baum-Snow, 2007)."

Baum-Snow (2007) collects data on the number of rays in the 1947 planned highway map, and data on the number of rays from current highways.

5.1 Functional Form

The two-stage least squares model is as follows:

$$\widehat{Sprawl} = \alpha_0 + \alpha_1 PlannedRays + \alpha_2 X + \nu \tag{6}$$

$$Outcome = \beta_0 + \beta_1 \widehat{Sprawl} + \beta_2 X + \epsilon \tag{7}$$

Equation 6 is the first stage regression, and Equation 7 is the second stage. *PlannedRays* is 1947 planned highway rays. *Sprawl* represents one of following four variables: (1) Actual highway rays, (2) percentage of people who drive alone, (3) percentage of people who use public transit, or (4) percentage of people who walk. The number of highways and people who drove alone are indicators or car dependence and sprawl, while the percentage of people who use transit or walk are indicators of a lack of sprawl and car dependence. X is a vector of control variables. The unit of analysis for all variables is the metropolitan statistical area (MSA).

5.2 Data

The dataset from the IV model contains data from several different sources. A portion of the data comes from Nathaniel Baum-Snow, who generously posted the data from Baum-Snow (2007) online. The rest of the data is from the U.S. Census Bureau and the U.S. Department of Agriculture Economic Research Service (ERS). The smallest denomination at which the data is available is the MSA.

The two variables I use from Baum-Snow (2007) are planned highway rays from 1947 and actual highway rays from 1999. These variables were collected from a manual examination of old highway maps and a dataset called PR-511. The PR-511 dataset was mandated by the 1956 Interstate Highway Act, which required states to record the completion of federally funded highways (Baum-Snow, 2007). The remainder of my data is from the year 2010 because this is the earliest year for which food access data is available. Although Baum-Snow's data on current highways is from 1999, this is a good approximation of 2010 highways since highway milage only increased by 0.3% during this time (U.S. Department of Transportation, 2019).

To construct measures of food access, I use 2010 data from the FARA as I did in the agent-based simulation (Ver Ploeg et al., 2017). However, instead of measuring the number of grocery stores, I use variables that directly measure food access, because this is a more accurate description of the food environment. I could not use these variables in the simulation because the definition of food access is relatively complicated and did not lend itself to be a reasonable outcome of the simulation.

The FARA has a variable for census tracts with low-access to grocery stores. In an urban area such as an MSA, a low-access food environment is defined as a census tract "where a significant number (at least 500 people) or share (at least 33 percent) of the population is greater than 1.0 mile from the nearest supermarket (Ver Ploeg and Rhone, 2017)." I aggregate these data to construct two measures of food access at the MSA level. The first measure is the percentage of census tracts in the MSA that are considered low-access. The second measure is the per capita number of census tracts in the MSA that are considered low-access. While in practice these variables are very similar, the second variable takes into account that census tracts in some cities could be larger than census tracts in other cities.

I use the American Community Survey (ACS) for the additional explanatory and control variables in 2010 (U.S. Census Bureau, 2016). In addition to the effect of current highways on sprawl development, I look at the effect of current commuting patterns on sprawl development. The ACS includes the percentage of people who drive to work alone, use public transportation, and walk. These are three of the explanatory variables in equations 6 and 7 above. The other control variables I use are population, median income, education, race, and ethnicity. I obtain all the variables from the ACS at the census tract level. I then aggregate them to the MSA level using code from Baum-Snow (2007) to maintain consistency over time.

I present the summary statistics for the data used in the reduced form analysis in Table 6. The table contains the outcome variables, the instrument, the explanatory variables, and the control variables. On average, about half of the census tracts in a given MSA are considered low-access. If this number seems high, it is because the usual definition of a food desert is a tract that is both low-income and low-access. The reason I choose to include low-access tracts of all income levels is because including the low-income indicator in the definition of a food desert may bias the results. Good access to public transportation and walkable streets could be associated with the wealth of census tract since such investments are paid for by taxes. Cities with reduced car usage would automatically be less likely to be in a food desert if food deserts are defined using the low-income indicator. Therefore, the low-access indicator is more appropriate.

To measure sprawl development, I look for indicators of car dependence. A city with a large number of highway rays or a large percentage of people who drive is likely car dependent. Conversely, a city with larger numbers of people walking and using public transportation is less likely to be car dependent and sprawling simply because such modes of travel are possible. Table 6 also shows that the number of planned highway rays in a MSA ranges from 0 to 7 with a mean of 2.12. The number of actual highway rays ranges from 0 to 15 with a mean of 3.41. 79% of people drove alone to work, 2% used public transportation and 3% walked in the average city. The remaining 16% either drove in a car with more than 1 person, used a ride share, worked from home, or used some other means. The variables like population and median income have maximum and minimum values which are not whole numbers because they are a population weighted average of the values from each census tract within the MSA. The three education variables are mutually exclusive and collectively exhaustive so graduate education was omitted arbitrarily in the regressions.

5.3 Results

Table 7 presents four regressions, each one containing one of the indicators of car dependence as an explanatory variable. The explanatory variable is regressed on the number of low-access census tracts in the MSA, using the instrument, 1947 planned highway rays in each regression. An instrument is considered weak if the first stage F-statistic is less than 10 (Stock et al., 2002). The F-statistic for actual highway rays in 1999 is highly significant at 360.56. The F-statics for the other measures of commuting patterns are much lower. The public transportation variable (F-stat=9.87) is on the cusp of being considered a weak instrument (likely because so few people use public transportation) so these results should be interpreted with caution. The F-statistics for driving (19.42) and walking (13.84) are large enough for planned rays to not be considered a weak instrument for these variables. Although planned highway rays is a better instrument for actual highway rays than it is for other commuting patterns, actual data on commuting patterns is a more direct measure of car dependence than the number of highway rays.

The regression results in Table 7 indicate that building one additional highway ray in an MSA leads to a 1.14 percentage point increase in the number of tracts in the MSA which have low access to food. The other commuting variables also support the hypothesis that sprawl development leads to low food access. Increasing the percentage of commuters who drive alone by one percentage point corresponds to a 1.37 percentage point increase in the percentage of low-access tracts. Increasing the percentage of people using public transit and walking corresponds to 3.36 and 5.18 percentage point decreases in the percentage of low-access tracts respectively. Results for actual highway rays and driving alone are significant at the 5% level while results for using public transit and walking are significant at the 10% level. The sample size is relatively small in all regressions (239 MSA's).

Table 8 presents the same regressions as Table 7; however, the outcome variable is the number of low-access census tracts per million people. The purpose of this set of regressions is to understand how robust the outcome variable is to a different measure of food access. Census tracts do not necessarily have the same number of people, so while the first outcome variable reveals how many census tracts in a city have low food access, this outcome variable is more indicative of how many *people* live in areas of low food access. The results are similar indicating the regressions are robust to different measures of food access. The first stage results are the same in both tables because the first stage equation is the same for the regressions reported in both tables. The only difference between these two tables is the outcome variable. The magnitude of the results in Table 7 is about half that of the results in Table 8. This is simply because mean of the *percentage of tracts in the MSA which are low-access* is about half the mean of the *number of low-access tracts per million people.*

It is helpful to look at both the means and the ranges of the explanatory variables when trying to understand the magnitude of the effects presented in Tables 7 and 8). A one ray increase in highway rays is a considerable increase since the average number of rays is 3.4 and the maximum number is 15. This is associated with only a 1.1 percentage point increase in the percent of low-access tracts. However, a 1 percentage point increase in the percentage of people who are driving alone causes a 1.4 percentage point increase in the percent of low-access tracts. Given, that the mean percentage of people who drive is 79%, it is completely reasonable to expect that local transportation policies could change the percentage of people who drive by at least 1% with much less effort than adding or removing a highway. Similarly, while the average percentage of people using public transportation and walking is very low, the range is very large, so incentivizing walking or using transit could influence this value by at least 1%. Therefore targeting a change in commuting patterns would likely have a larger effect on food access than targeting a change in highways.

6 Discussion and Policy Implications

The theoretical model proposes a mechanism for how car dependence leads to fewer grocery stores. Given the model's assumptions about the behavior of the consumers and firms, the simulation shows that a larger percentage of drivers leads to fewer grocery stores. The goal of the simulation was not to predict the number of grocery stores as accurately as possible, but rather to use as little information as possible to understand how transportation patterns specifically, affect food access. For this reason I did not calibrate the transportation effort function to maximize the accuracy of the data, but rather used parameters derived from the FoodAPS data. I did not try different specifications because this could lend itself to choosing the specification that justified my prior beliefs. Instead, I used theory to decided on a functional form and a way to parameterize the transportation effort function before running the simulation. Even still, the simulation performs well, explaining 83% of the variation in the number of grocery stores. However, most of the explanatory power comes from differences in population. Transportation patterns and the assumptions of the model explain 12% of the variation in grocery stores per 1,000 people.

The reduced form results show that indicators of sprawl (highway rays and driving alone) are associated with lower food access, and the variables that are indicative of less car dependence (public transit and walking) are associated with higher food access. The magnitude of the effect of highway rays is much smaller than the effect of commuting patterns. For example: An increase in highway rays by 1 ray and an increase in the percentage of people driving by on percentage point have approximately the same effect on the percentage of tracts which are low-access. However, a 1 ray increase is 1/3 of the mean number of rays, while a 1 percentage point increase in drivers is 1/80 of the mean percentage of drivers. Therefore, an increase in 1 ray is much larger than an increase in 1 percentage point in the percentage of drivers. The difference in the magnitude of the effects could be attributed to the fact that commuting patterns are likely more indicative of sprawl than the number of highways because commuting patterns measure car dependence directly.

There are several limitations to this study. It would be useful to have data on the development of transportation systems over time and on grocery store density and location in the same years. Then the instrument (planned highways) could be used and its effect could be observed over multiple years.

It would also be useful to be able to verify that the instrument is uncorrelated with the error term conditional on the regressors. Finding another instrument would help to verify this claim. Another limitation is that the agent-based simulation uses assumptions about dislike for traveling to the grocery store that are based on how far people travel to a grocery store in real life. In terms of available data, the FoodAPS dataset is very useful for understanding how far shoppers are willing to travel depending on their mode of transportation. However, it is not a perfect measure of disutility from traveling, and disutility from traveling may vary from city to city. Also the agent-based simulation, while informative in demonstrating the mechanism whereby sprawl can affect grocery store location, cannot actually show causality in the real world.

Policy makers may be able to improve food access and equity by reducing car dependence. Sprawl development disproportionately hurts those who do not have cars. Not only is it more difficult to walk to a supermarket when the grocery industry (in response to the preferences of the majority) developed large supercenters driving neighborhood chains out of business, but it is also more difficult to walk an equivalent distance in a car dependent city than it is in a less car dependent city because the roads are not designed for pedestrians. Reducing car dependence may not only produce a larger quantity of stores, but also make existing stores easier to access. Furthermore, it is difficult to travel anywhere (work, doctor, etc.) without a car in a city that was built to revolve around them. Such an environment worsens inequality by making certain resources less accessible to the non-driving population.

The changes in the food environment that developing countries are experiencing now have already happened in the United States, and leaders can learn from their consequences. As consumers worldwide have more disposable income, global car ownership is expected to double by 2040 (Smith, 2016). If increasing car dependence does lead to the development of supermarkets and lower reliance on local food retailers, this could be harmful to the poor who do not have good access to transportation. Understanding the impacts of car dependence on food access and grocery store location is important for developers worldwide looking to maintain the stability of their food systems. If it is possible to plan and develop cities in way that all residents have sufficient access to healthy food, this could greatly improve quality of life, especially for the minority who lack transportation. This analysis, and any studies that explore this topic further, could be included in a cost benefit analysis for expanding public transportation systems, funding pedestrian and cyclist infrastructure, and other policies that reduce urban sprawl.

7 Conclusion

This paper studies the ways that sprawl development and car dependence influence grocery store location and food access. The theoretical model suggests that differences in transportation disutility between drivers and walkers may cause sprawl development to affect food access. In the simulation, consumers behave as the theoretical model suggests they would. The simulation shows that given the assumptions of the theoretical model, car dependence decreases food access. The IV model evaluates this hypothesis using data on actual United States cities. The IV model finds evidence that sprawl development does negatively impact food access.

This paper contributes to the literature on food deserts by providing an explanation of how food deserts form. Traditionally, research has focused on the effects of food access on nutrition, rather than attempting to understand why food access has become a problem. In addition to the fact that dietary patterns do not change quickly (as noted in the Introduction), the model illuminates another reason why some studies find that food deserts do not have a strong influence on diet. If the government subsidizes the construction of a new grocery store in a car dependent city, the commuting time will drop among those who normally pick up groceries with a car, but their shopping patterns are unlikely to change significantly. However, the minority of people who do not have good transportation to a grocery store, may see a significant improvement in quality of life and nutrition.

Additionally, this paper uses an instrument that could be used in future research to assess the impact of food deserts on health. Because food deserts are more likely to form in cities that developed in a car dependent manner, one could use a panel dataset of transportation patterns, food access, and nutrition, in combination with this instrument, to better understand how nutrition and food access changed over time in response to transportation patterns.

Subsequent work should seek to understand if the benefits of the current sprawling supercenter landscape, such as lower prices and enormous variety, outweigh the costs of reducing food access and quality of life for those in food deserts. In a future paper, I plan to investigate how growing up in a food desert affects preferences over the lifespan and explore how long it takes a person to develop new preferences after moving. To do this I would like to use a panel dataset on grocery store location, sprawl development, and food choices or health.

Overall the results suggest that a sprawl development pattern is associated with lower food access. Continuing this research is important because as cities continue to grow worldwide, it is important for city planners to understand the implications of different development patterns. This work suggests that reducing car dependence could increase grocery store density, improving food access.

References

110th Congress (2008). Food, Conservation, and Energy Act of 2008.

Alkon, A. H. and K. M. Norgaard (2009, aug). Breaking the Food Chains: An Investigation of Food Justice Activism*. *Sociological Inquiry 79*(3), 289–305.

Allcott, H., R. Diamond, J.-P. Dubé, J. Handbury, I. Rahkovsky, and M. Schnell (2019, nov). Food Deserts and the Causes of Nutritional Inequality. *The Quarterly Journal of Economics 134*(4), 1793–1844.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, New Jersey: Princeton University Press.

Auchincloss, A. H., R. L. Riolo, D. G. Brown, J. Cook, and A. V. Diez Roux (2011, mar). An Agent-Based Model of Income Inequalities in Diet in the Context of Residential Segregation. *American Journal of Preventive Medicine 40*(3), 303–311.

Baum-Snow, N. (2007). Did Highways Cause Suburbanization? *The Quarterly Journal of Economics 122*(2), 775–805.

Berger, M. and B. van Helvoirt (2018, aug). Ensuring Food Secure Cities – Retail Modernization and Policy Implications in Nairobi, Kenya. *Food Policy 79*, 12–22.

Bitler, M. and S. J. Haider (2011). An Economic View of Food Deserts in the United States. *Journal of Policy Analysis and Management 30*(1), 153–176.

Blok, D. J., S. J. De Vlas, R. Bakker, and F. J. Van Lenthe (2015). Reducing Income Inequalities in Food Consumption: Explorations with an Agent-Based Model. *American Journal of Preventive Medicine*.

Bodor, J. N., D. Rose, T. A. Farley, C. Swalm, and S. K. Scott (2008, apr). Neighbourhood Fruit and Vegetable Availability and Consumption: the Role of Small Food Stores in an Urban Environment. *Public Health Nutrition 11*(04), 413–420.

Brinkman, J. and J. Lin (2019). Freeway Revolts! *Federal Reserve Bank of Philadelphia Working Papers*.

Burchell, R. W., N. A. Shad, D. Listokin, H. Phillips, A. Downs, S. Seskin, J. S. Davis, T. Moore, D. Helton, and M. Gall (1998). The Cost of Sprawl-Revisited. Technical report, Transit Cooperative Research Program, Washington D.C.

Bureau of Public Roads (1955). *General Location of National System of Interstate Highways.*

Chalabi, Z. and T. Lorenc (2013, nov). Using Agent-Based Models to Inform Evaluation of Complex Interventions: Examples from the Built Environment. *Preventive Medicine 57*(5), 434–435.

Cummins, S., E. Flint, and S. A. Matthews (2014, feb). New Neighborhood Grocery Store Increased Awareness of Food Access but did not Alter Dietary Habits or Obesity. *Health affairs (Project Hope) 33*(2), 283–91.

Dargay, J. M. (2001, nov). The Effect of Income on Car Ownership: Evidence of Asymmetry. *Transportation Research Part A: Policy and Practice 35*(9), 807–821.

Dubowitz, T., M. Ghosh-Dastidar, D. A. Cohen, R. Beckman, E. D. Steiner, G. P. Hunter, K. R. Florez, C. Huang, C. A. Vaughan, J. C. Sloan, S. N. Zenk, S. Cummins, and R. L. Collins (2015, nov). Diet and Perceptions Change with Supermarket Introduction in a Food Desert, but not Because of Supermarket Use. *Health Affairs 34*(11), 1858–1868.

Franklin, A. W. and R. W. Cotterill (1993). An Analysis of Local Market Concentration Levels and Trends in the U.S. Grocery Retailing Industry. Technical report, Food Marketing Policy Center: University of Connecticut.

Hammond, R. A., J. T. Ornstein, L. K. Fellows, L. Dubé, R. Levitan, and A. Dagher (2012, oct). A Model of Food Reward Learning with Dynamic Reward Exposure. *Frontiers in Computational Neuroscience 6.*

Handbury, J., I. Rahkovsky, and M. Schnell (2015, apr). Is the Focus on Food Deserts Fruitless? Retail Access and Food Purchases Across the Socioeconomic Spectrum. Technical report, National Bureau of Economic Research, Cambridge, MA.

Hendrickson, D., C. Smith, and N. Eikenberry (2006, oct). Fruit and Vegetable Access in Four Low-Income Food Deserts Communities in Minnesota. *Agriculture and Human Values 23*(3), 371–383.

Hossfeld, L., E. B. Kelly, E. O'Donnell, and J. Waity (2017, nov). Food Sovereignty, Food Access, and the Local Food Movement in Southeastern North Carolina. *Humanity & Society 41*(4), 446–460.

Jiao, J., A. V. Moudon, and A. Drewnowski (2016). Does Urban Form Influence Grocery Shopping

Frequency? A study from Seattle Washington, USA. *International Journal of Retail & Distribution Management 44*, 923–939.

Koh, K., R. Reno, and A. Hyder (2019, apr). Examining Disparities in Food Accessibility Among Households in Columbus, Ohio: An Agent-Based Model. *Food Security 11*(2), 317–331.

Larsen, K. and J. Gilliland (2008, apr). Mapping the Evolution of 'Food Deserts' in a Canadian City: Supermarket Accessibility in London, Ontario, 1961–2005. *International Journal of Health Geographics 7*(1), 16.

Lee, H. (2012). The Role of Local Food Availability in Explaining Obesity Risk Among Young School-Aged Children. *Social Science & Medicine 74*, 1193–1203.

Li, Y., J. Berenson, A. Gutiérrez, and J. A. Pagán (2016, dec). Leveraging the Food Environment in Obesity Prevention: the Promise of Systems Science and Agent-Based Modeling. *Current Nutrition Reports 5*(4), 245–254.

Lopez, R. P. (2007, aug). Neighborhood Risk Factors for Obesity. *Obesity 15*(8), 2111–2119.

Lutz, C. (2014). Cars and Transport: The Car-Made City. In *A Companion to Urban Anthropology.*

Michimi, A. and M. C. Wimberly (2010). Associations of Supermarket Accessibility with Obesity and Fruit and Vegetable Consumption in the Conterminous United States. *International Journal of Health Geographics*.

Morland, K., S. Wing, A. Diez Roux, and C. Poole (2002, jan). Neighborhood Characteristics Associated with the Location of Food Stores and Food Service Places. *American journal of preventive medicine 22*(1), 23–9.

Reardon, T. and A. Gulati (2008). The Supermarket Revolution in Developing Countries: Policies for "Competitiveness with Inclusiveness". *IFPRI Policy Brief 2* (June).

Ross, A. (2016). The Surprising Way a Supermarket Changed the World.

Schafft, K. A., E. B. Jensen, and C. C. Hinrichs (2009). Food Deserts and Overweight Schoolchildren: Evidence from Pennsylvania. *Rural Sociology 72*(2), 153–177.

Smith, M. N. (2016). The Number of Cars Worldwide is Set to Double by 2040.

Stock, J. H., M. Yogo, T. R. We, A. Hall, J. Hausman, T. Hayakawa, G. Judge, W. Newey, and J. Wright (2002). Testing for Weak Instruments in Linear IV Regression. *National Bureau of Economic Rsearch*.

Suarez, J. J., T. Isakova, C. A. Anderson, L. E. Boulware, M. Wolf, and J. J. Scialla (2015, dec). Food Access, Chronic Kidney Disease, and Hypertension in the U.S. *American Journal of Preventive Medicine 49*(6), 912–920.

Teegarden, S., A. Scott, and T. Bale (2009, sep). Early Life Exposure to a High Fat Diet Promotes Long-Term Changes in Dietary Preferences and Central Reward Signaling. *Neuroscience 162*(4), 924–932.

Thibodeaux, J. (2016). A Historical Era of Food Deserts: Changes in the Correlates of Urban Supermarket Location, 1970-1990. *Social Currents 3*(2), 186–203.

Tuttle, C. and A. Kuhns (2016). Percent of Income Spent on Food Falls as Income Rises. *Amber Waves*.

U.S. Census Bureau (2016). American Community Survey. Technical report.

U.S. Department of Agriculture (2018). Definitions of Food Security.

U.S. Department of Agriculture Economic Research Service (2017). FoodAPS National Household Food Acquisition and Purchase Survey.

U.S. Department of Transportation (2019). 2010 Conditions and Performance.

Ver Ploeg, M., V. Breneman, and A. Rhone (2017). USDA ERS - Food Access Research Atlas.

Ver Ploeg, M. and A. Rhone (2017). Documentation.

Weber, J. (2011). “America’s New Design for Living:” The Interstate Highway System and the Spatial Transformation of the U.S. In *Engineering Earth*, pp. 553–567. Dordrecht: Springer Netherlands.

Wedick, N. M., Y. Ma, B. C. Olendzki, E. Procter-Gray, J. Cheng, K. J. Kane, I. S. Ockene, S. L. Pagoto, T. G. Land, and W. Li (2015, mar). Access to Healthy Food Stores Modifies Effect of a Dietary Intervention. *American Journal of Preventive Medicine 48*(3), 309–317.

Weingroff, R. F. (2017). Original Intent: Purpose of the Interstate System - Interstate System - Highway History - Federal Highway Administration.

Widener, M. J., S. S. Metcalf, and Y. Bar-Yam (2013, jun). Agent-based Modeling of Policies to Improve Urban Food Access for Low-Income Populations. *Applied Geography 40*, 1–10.

Zenk, S. N., L. L. Lachance, A. J. Schulz, G. Mentz, S. Kannan, and W. Ridella (2009, mar). Neighborhood Retail Food Environment and Fruit and Vegetable Intake in a Multiethnic Urban Population. *American Journal of Health Promotion 23*(4), 255–264.

Zenk, S. N., A. J. Schulz, B. A. Israel, S. A. James, S. Bao, and M. L. Wilson (2006). Fruit and Vegetable Access Differs by Community Racial Composition and Socioeconomic Position in Detroit, Michigan. *Ethnicity & disease 16*(1), 275–80.

Table 1: Summary Statistics: Agent-Based Model

Variable	Mean	Std. Dev.	Min.	Max.	N
County Population	98321.5	313037.5	82	9818605	3,140
Percent Drove Alone	79.553	7.551	5.982	97.206	3,140
Grocery Stores per 1,000 People (GSPP)	0.253	0.223	0	3.149	3,140
Number of Grocery Stores	21.005	90.406	0	2429	3,140

Note: Data on county population and the percentage of people who drive come from the American Community Survey (U.S. Census Bureau, 2016). The data on number of grocery stores and GSPP come from the Food Access Research Atlas (Ver Ploeg et al., 2017). The level of analysis is the county level for the year 2010.

Table 2: Summary Statistics: FoodAPS Data

Variable	Mean	Std. Dev.	Min.	Max.	N
Distance Drivers Travel to Grocery Store	4.692	6.632	0.014	156.08	4,151
Distance Walkers Travel to Grocery Store	0.516	0.297	0.027	1.405	267

Note: Data on distance to grocery store by travel mode come from the FoodAPS dataset, and is used to calibrate equation 3 (U.S. Department of Agriculture Economic Research Service, 2017).

Table 3: Correlation Matrix

	Percent Driving
Num. Stores (Actual)	-0.268
Num. Stores (Predicted)	-0.204
GSPP (Actual)	-0.475
GSPP (Predicted)	-0.369

Note: Num. Stores (Actual) and GSPP (Actual) refer to the actual number of stores and actual grocery stores per 1,000 people reported in each county as available from the Food Access Research Atlas (Ver Ploeg et al., 2017). Num. Stores (Predicted) and GSPP (Predicted) refer to the number of stores and grocery stores per 1,000 people predicted for each county by the simulation. The correlation coefficients between these four variables and the percentage of people driving in each county (available from the American Community Survey (U.S. Census Bureau, 2016)) is reported in the table.

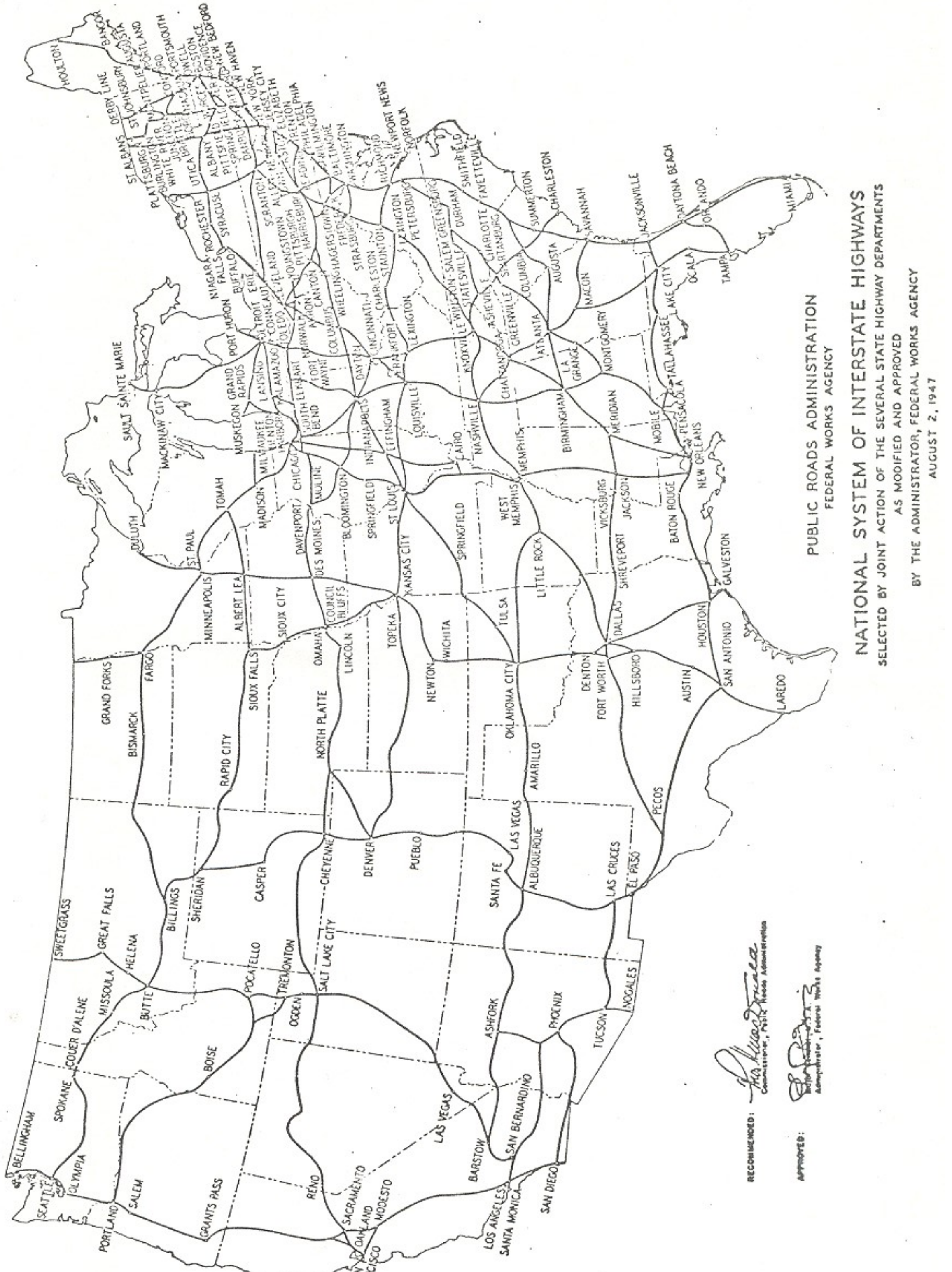

Figure 1: The 1947 planned interstate highway map available from: Bureau of Public Roads (1955)

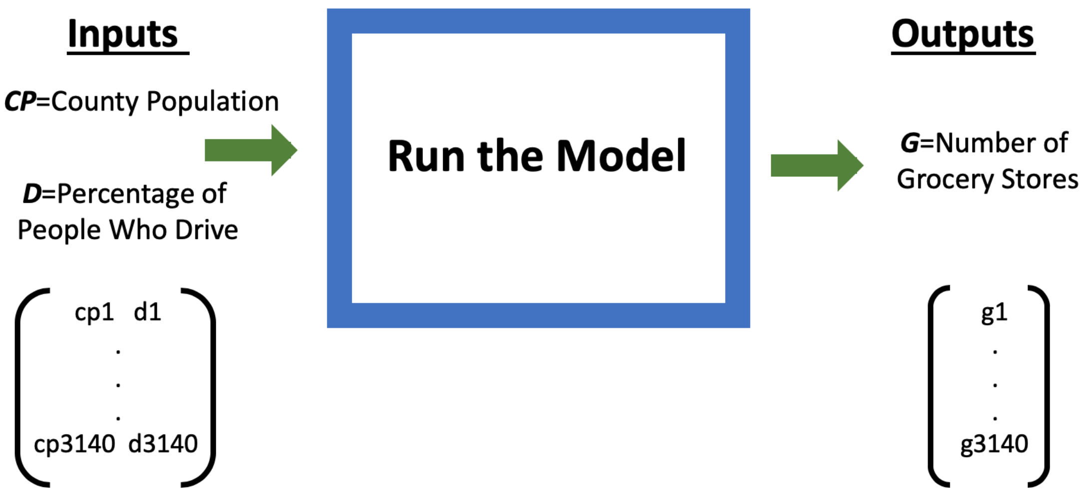

Figure 2: County population and the percentage of people who drive for each of the 3,140 counties is the input to the simulation. The output is the predicted number of grocery stores in that county. The predicted number of grocery stores and county population is used to calculate the predicted grocery stores per 1,000 people (GSPP)

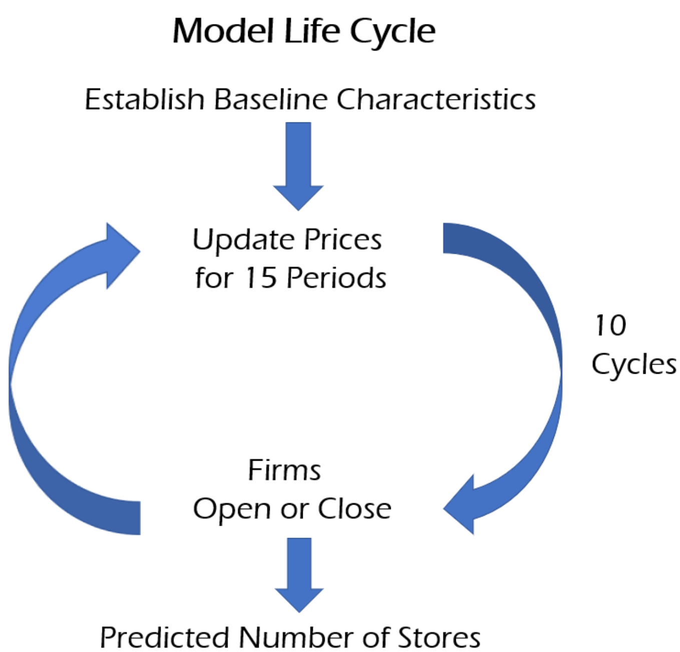

Figure 3: County population and driving statistics are used to establish baseline characteristics. Consumers identify the grocery store which maximizes utility and how much they will purchase there. Grocery stores have complete information and using the knowledge of where consumers will shop, they update prices for 15 periods. At the end of 15 periods stores have the opportunity to open or close. This cycle continues for 10 periods.

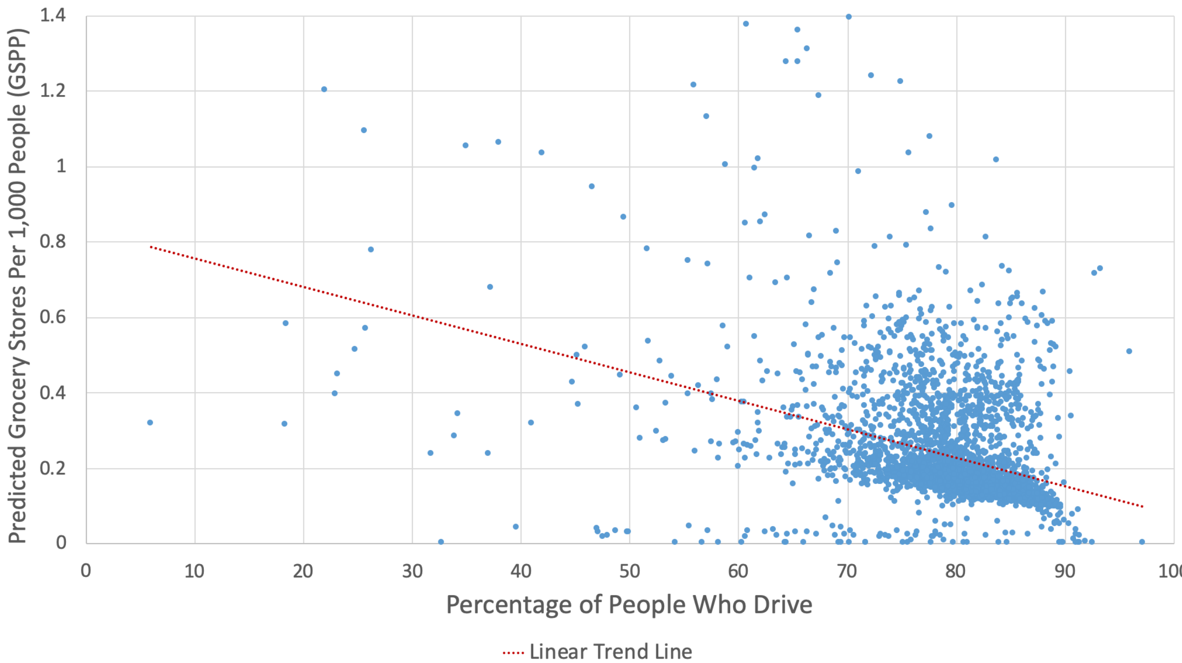

Figure 4: Scatter plot of number of grocery stores Per 1,000 people (GSPP) predicted by the simulation vs. percentage of people who drive (available from the American Community Survey (U.S. Census Bureau, 2016). The red line is the best fit line.

Table 4: Simple Regression of Percentage of Drivers on Predicted Grocery Stores per 1,000 People (GSPP)

Percent of People Who Drive Alone	-0.008***
	(0.0003)
Constant	0.8315***
	0.0271
Mean of Predicted GSPP	0.230

Note: The table shows the output of a simple regression of the percentage of people who drive alone per county (available from the American Community Survey (U.S. Census Bureau, 2016)) on the number of grocery stores per 1,000 people predicted for each county by the simulation. The regression yields the slope and y-intercept for the best fit line in Figure 4. Standard errors are in parentheses: *** indicates $p<0.01$.

Table 5: Evaluation of Goodness of Fit

		Root Mean Squared Error	**Mean Absolute Error**	**Adjusted $\mathbf{R}^2$**
Full Sample	**GSPP**	0.223	0.119	0.122
	Num. Stores	44.052	6.609	0.830
Excluding Extreme Values	**GSPP**	0.223	0.120	0.126
	Num. Stores	13.285	4.657	0.897

Note: This table presents several measures of goodness of fit for the variables GSPP (grocery stores per 1,000 people) and Num. Stores (number of stores). Predicted GSPP and Num. Stores come from the simulation. Actual values of GSPP and Num. Stores come from the Food Access Research Atlas (Ver Ploeg et al., 2017). The full sample includes all 3,140 counties. The sample excluding extreme values contains 3,127 counties, and excludes all counties which have more than 500 grocery stores.

Table 6: Summary Statistics: Reduced Form Model

Variable Type	Variable	Mean	Std. Dev.	Min.	Max.	N	Data Source
Outcome	Percent of Tracts in MSA which are Low-Access	48.093	11.818	13.626	78.419	304	Ver Ploeg et al. (2017)
	Number of Low-Access Tracts per Million People	96.281	24.485	28.975	185.68	304	Ver Ploeg et al. (2017)
Instrument	1947 Planned Highway Rays	2.117	1.49	0	7	239	Baum-Snow (2007)
Explanatory	Actual Highway Rays	3.410	2.357	0	15	239	Baum-Snow (2007)
	Percent Drove Alone	79.053	4.705	49.778	86.150	304	U.S. Census Bureau (2016)
	Percent Used Public Transit	1.945	2.639	0.066	31.07	304	U.S. Census Bureau (2016)
	Percent Walked	2.853	1.692	0.673	15.367	304	U.S. Census Bureau (2016)
Control	Population in Millions	5042.933	18588.961	286.603	295299.75	304	U.S. Census Bureau (2016)
	Median Income	52667.266	9465.535	32411.538	94815.863	304	U.S. Census Bureau (2016)
	High School Education or Less	42.81	8.291	20.558	65.236	304	U.S. Census Bureau (2016)
	Undergraduate Education	47.156	5.706	30.238	65.098	304	U.S. Census Bureau (2016)
	Graduate Education	10.034	4.14	3.769	26.449	304	U.S. Census Bureau (2016)
	Percent Black	10.631	10.569	0.272	52.006	304	U.S. Census Bureau (2016)
	Percent Hispanic	9.891	14.076	0.474	94.586	304	U.S. Census Bureau (2016)

Note: Data on the outcome variables (Percent of Tracts in MSA which are Low-Access and Number of Low-Access Tracts per Million People) were constructed by aggregating the low-access tract indicator variable from the Food Access Research Atlas from the year 2010 to the MSA level (Ver Ploeg et al., 2017). Data on number of highways per MSA in 1947 and highways per MSA in 1999 come from (Baum-Snow, 2007). Data on the percentage of people who drive, walk and use public transportation and data on all the control variables come from the American Community Survey for the year 2010 (U.S. Census Bureau, 2016). These data were originally collected at the census tract level and aggregated to the MSA level.

Table 7: IV: Impact of Car Dependence on the Percentage of Tracts in the MSA which are Low-Access

	(1)	(2)	(3)	(4)
Actual Highway Rays	1.138**			
	(0.536)			
Percent Drove Alone		1.366**		
		(0.681)		
Percent Used Public Transit			-3.357*	
			(1.928)	
Percent Walked				-5.183*
				(2.688)
Population in Millions	-0.000***	-0.000	0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Median Income	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
High School Education or Less	0.504**	-0.306	-0.244	-1.048
	(0.246)	(0.440)	(0.462)	(0.821)
Undergraduate Education	1.109***	0.314	0.138	-0.615
	(0.306)	(0.454)	(0.588)	(0.894)
Percent Black	0.259***	0.206***	0.241***	0.026
	(0.067)	(0.078)	(0.071)	(0.153)
Percent Hispanic	0.064	0.152**	-0.003	-0.039
	(0.047)	(0.070)	(0.064)	(0.065)
Constant	-26.714	-64.120**	49.049	154.803
	(27.038)	(31.845)	(47.966)	(96.086)
Sample Size	239	239	239	239
First Stage F-stat	360.563	19.424	9.865	13.843
Dependent Variable Mean	48.093	48.093	48.093	48.093
Explanatory Variable Mean	3.410	79.058	1.945	2.853
Explanatory Variable Range	0-15	49.8-86.1	0.06-31	0.67-15.37

Note: This table presents results from individual regressions of the four explanatory variables (Actual Highway Rays, Percent Drove Alone, Percent Used Transit, and Percent Walked) on the outcome, Percentage of Tracts in the MSA which are Low-Access. 1947 Planned Highway Rays was used as an instrument for the respective explanatory variable in each regression. Standard errors clustered at the MSA level are in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 8: IV: Impact of Car Dependence on the Number of Low-Access Tracts per Million People

	(1)	(2)	(3)	(4)
Actual Highway Rays	2.356**			
	(1.046)			
Percent Drove Alone		2.829**		
		(1.384)		
Percent Used Public Transit			-6.951*	
			(4.103)	
Percent Walked				-10.733*
				(5.515)
Population in Millions	-0.001***	-0.000	0.001	-0.001**
	(0.000)	(0.000)	(0.001)	(0.000)
Median Income	-0.000	-0.000	0.000	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
High School Education or Less	0.783	-0.894	-0.766	-2.430
	(0.551)	(0.937)	(1.013)	(1.687)
Undergraduate Education	1.695**	0.049	-0.317	-1.875
	(0.714)	(1.003)	(1.294)	(1.853)
Percent Black	0.339***	0.230	0.302**	-0.143
	(0.130)	(0.156)	(0.143)	(0.305)
Percent Hispanic	-0.028	0.153	-0.167	-0.242*
	(0.132)	(0.182)	(0.156)	(0.146)
Constant	-4.111	-81.572	152.781	371.778*
	(59.724)	(67.097)	(103.959)	(196.304)
Sample Size	239	239	239	239
First Stage F-stat	360.563	19.424	9.865	13.843
Dependent Variable Mean	96.281	96.281	96.281	96.281
Explanatory Variable Mean	3.410	79.058	1.945	2.853
Explanatory Variable Range	0-15	49.8-86.1	0.06-31	0.67-15.37

Note: This table presents results from individual regressions of the four explanatory variables (Actual Highway Rays, Percent Drove Alone, Percent Used Transit, and Percent Walked) on the outcome, Number of Low-Access Tracts per Million People. 1947 Planned Highway Rays was used as an instrument for the respective explanatory variable in each regression. Standard errors clustered at the MSA level are in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.