The Effect of Sprawl Development, Population Density, and Transportation Costs on Food Access

Stephanie Schauder

June 2019

Abstract

This paper explores the effect of transportation systems on urban food access in the United States. Since the development of the interstate highways system, transportation costs were reduced and many cities developed in a sprawling pattern characterized by low population density and car dependence. I hypothesize that this change caused the decline of neighborhood grocery stores, and local bodegas selling fresh food. I look at how changes in transportation patterns affected the food desert status of all urban census tracts in the United States. I use a difference in differences model as well as an instrumental variables model to attempt to understand if results are consistent given both approaches to addressing endogeneity. The results suggest that sprawl development and car dependence are consistently associated with food desert status; however, the magnitude of the effect is modest.

1 Introduction

Food security relies not only on having access to a subsistence number of calories, but to affordable and nutritious food (U.S. Department of Agriculture, 2018). The topic of food access in the United States has been widely studied and discussed. On the one hand, people living closer to fresh food eat more fresh food (Bodor et al., 2008; Zenk et al., 2009) and have lower rates of obesity (Michimi and Wimberly, 2010; Lopez, 2007; Schafft et al., 2009). Yet, research has shown that simply increasing the supply of grocery stores has a limited effect on food purchasing choices and nutrition (Dubowitz et al., 2015; Cummins et al., 2014; Handbury et al., 2015). A case study of Seacroft, UK finds that the introduction of a supermarket increases diet diversity amongst those with the worst diversity score, but does little to improve healthy eating overall (Freire and Rudkin, 2019). A review of food access literature finds significant heterogeneity in food consumption between households only a small part of which can be explained by access to grocery stores (Ver Ploeg and Wilde, 2018).

However, access to grocery stores has significant implications beyond its propensity or lack thereof to encourage healthy food consumption. The food justice movement was born of the notion that everyone should have access to healthy, culturally appropriate food (Powell et al., 2007). However, there are three times as many supermarkets in wealthy neighborhoods as in lower-income neighborhoods, and four times as many supermarkets in white neighborhoods as in predominately black neighborhoods (Morland et al., 2002). Income has a strong positive association 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. Not only is traveling further to shop an inconvenience but also growing up in a food environment where high fat/rewarding foods are prevalent can shape dietary behaviors for years to come (Teegarden et al., 2009). In fact, it has been proposed that the reason the food desert literature has found unimpressive results from improving the food environment is not because such changes are not important, 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).

The city environment in the United States has changed substantially over the last century with the creation of the interstate highway system. Before cars became the predominant travel method, people congregated in towns and cities which provided necessary retail goods. However, highways reduced transportation time significantly, allowing people to live further from the central business districts. The resulting shift to a low density development pattern, known as urban sprawl, created the suburban environment (Baum-Snow, 2007). As a result, food purchasing habits and preferences changed, and instead of small local grocery stores, supercenters developed (Ross, 2016). 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 in wealthy areas where the demand is highest. 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).

One study of London, Ontario found that an increase in sprawl development and car dependence between 1961 and 2005 led to a decline from 45% to 18% in the number of census tracts with ”easy access” to a grocery store (Larsen and Gilliland, 2008). In fact this would not be the first time that the retail sector responded to a change in transportation patterns. A look back to economic history reveals that rapid electrification of the Boston streetcar caused a 5% decrease in small businesses as transportation costs decreased and location became less important (You, 2017). In a city where people walk or use public transportation regularly, there is higher demand to have a grocery store in one's neighborhood because the costs of traveling to the grocery store are higher. 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). I hypothesize that transportation patterns of cities affect grocery store location. I expect that in sprawling car-dependent cities, there will be a rise in large supercenters, but fewer grocery stores overall and worse food access, than in cities that rely on public transportation and walking.

From an international perspective, it is important 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 global south and asian countries. As a result, increasing concentration in the supermarket industry has developed which threatens local small retailers and farmers who cannot sell the large supermarket chains (Reardon and Gulati, 2008). Increasing 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). The changes in food environment that developing countries are experiencing now have already happened in the United States, and we 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.

However, 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 the people who live there. To deal with endogenous variables, I use two models: a difference in differences model and an instrumental variables model. Both of these methods have drawbacks, as I will discuss below, but consistent results across methods increases credibility. I discuss the theoretical background in section 2 below, followed by the methodology (section 3) and the empirical application (section 4). The results are represented in section 5 and the policy implications are discussed in section 6. Section 7 concludes.

2 Theory

I hypothesize that a sprawl development pattern is an essential contributing factor to the evolution of food deserts. A sprawl development pattern is consistent with relatively low population density, and transportation systems which were planned for cars. This fundamentally changes the way that people shop.

Suppose consumers maximize utility $u = U(F, T, X)$ subject to a budget constraint where F is food, T is per mile transportation effort, and X is all other goods. Food comes from grocery stores, and transportation effort is required to travel to a grocery store. Utility is increasing in food and because food is necessary, as $F \to 0$, $u \to -\infty$. Utility is decreasing in T, transportation effort, because of the loss in happiness from traveling to the store. Now suppose that consumers are heterogeneous in that some have cars and some do not. Those who have cars have transportation effort function $f_1(distance) = T$ which is linearly increasing in distance with a relatively low slope. Those who do not have cars have transportation effort function $f_2(distance) = T$ which is exponentially increasing in distance capturing the fact that using public transit or carrying groceries for a mile or more is much more difficult than driving.

Suppose we fix T at a number t_1 for both those who have cars and those who do not. t_1 could represent the mean amount of effort most people are willing to put into traveling to the grocery store (ex: most people are not willing to walk more than 2 miles to a grocery store). In Figures 1 and 2, the dark squares are neighborhoods, while the lighter circles are the radius that residents of that neighborhood would be able to travel to a grocery store with fixed t_1. Figure 1 contains residents that don't have cars, while figure 2 contains residents that do have cars.

Grocery stores are profit maximizing and they choose location and size. In figure 1 we would expect at least one grocery store to locate in each neighborhood because residents are not able to travel outside of their neighborhood to shop. To see why this is true, suppose a grocery store did not locate in one neighborhood. Residents in that neighborhood have no other food options in this model, so a grocery store could enter the neighborhood, charge very high prices and still receive business. However, in figure 2, residents have far more mobility. Grocery stores could locate in the center of the space and vie for business from all four of the neighborhoods. In figure 2, I would expect there to be fewer stores than in figure 1 because firms would capitalize on economies of scale.

If everyone were to drive a car (including children, the elderly, and disabled people) urban food access would not be a problem because in most cities there are generally grocery stores within a comfortable driving distance. If no one drove a car, I posit that there would also be no food access issues because stores would locate within walking distance of consumers as they did before the advent of the vehicle.

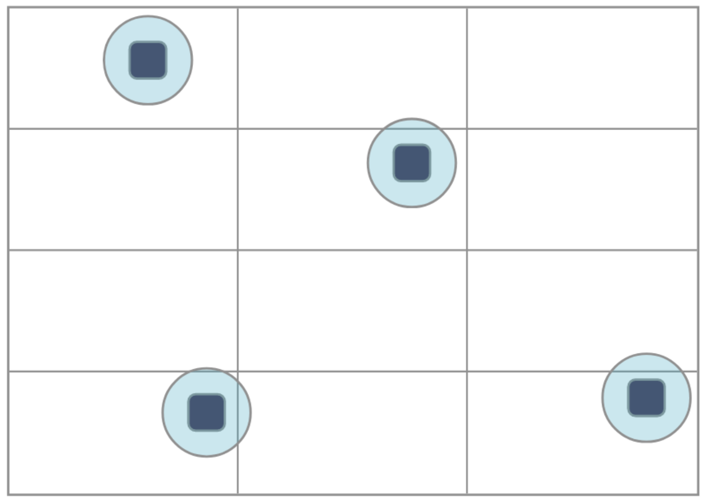

Figure 1: No Residents Own Cars (Dark blue squares are neighborhoods, light blue circles are the radius that residents are willing to travel)

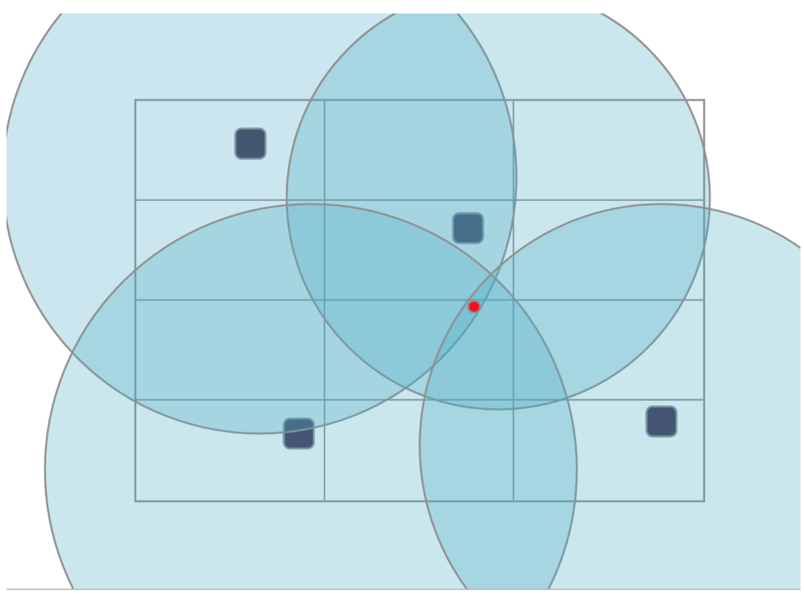

Figure 2: All Residents Own Cars (Dark blue squares are neighborhoods, light blue circles are the radius that residents are willing to travel, red dot is the intersection)

The problem arises when the majority of people drive and a minority do not. The grocery stores locate in a way to serve the majority and the left over minority have food access problems. Therefore, I test empirically whether population density and transportation patterns affect the probability of living in a low-access census tract.

3 Methodology

To test my hypothesis I will look at the effect of transportation patterns and population density on the probability of being in a census tract with low food access (LA). I will first use an Ordinary Least Squares (OLS) specification. However, a plain OLS regression of the effect of the aforementioned explanatory variables on the outcome, could lead to biased results because of endogenous variables. Census tracts that have well developed transportation systems may be wealthier and thus have more money to support grocery stores. Additionally, the people who choose to live in census tracts that are not car dependent may have different food and shopping preferences than those in less car dependent areas. Finally, there could be reverse causality, if a census tract has frequent small grocery stores, people may be more likely to go to those grocery stores without using a car. In order to try to ascertain the effect of transportation and density on the development of food access problems, I will also employ an instrumental variables (IV) and a difference in differences (DD) approach. I use both methods because they each have different advantages, and if there is truly an effect it should be robust to different specifications.

3.1 Ordinary Least Squares

The OLS equation I will use is as follows:

$$LAtract_{ct} = \beta_0 + \beta_1 Explanatory_{ct} + \beta_2 X_{ct} + \epsilon_{ct} \tag{1}$$

Where c is the census tract, and t is the time. *Explanatory* is a continuous explanatory variable. Depending on the specification, *Explanatory* is: the percentage of people in the census tract who commute by 1) driving, 2) public transit, 3) biking, and 4) walking, or the 5) population density of the census tract. X_{ct} is a group of control variables (year, median income, population, education, race/ethnicity, and county fixed effects). The outcome variable, $LAtract_{ct}$ is a dichotomous variable which equals 1 if the census tract is LA and 0 if it is not. Standard errors are robust and clustered at the county level.

The coefficient of interest is β_1. If there were no omitted variables or reverse causality, the interpretation of $beta_1$ would be the effect of the explanatory variables on probability being a LA tract. However, as mentioned above, there are reasons to doubt that the OLS assumptions hold. Therefore I employ the IV and DD approaches below.

3.2 Instrumental Variables

In this section, I use an instrumental variables approach to address the endogeneity mentioned above inherent in assessing the impact of car dependence and population density on the food environment.

Because car dependence and population density are likely correlated with unobservables which affect the food environment, it is necessary to find an exogenous reason why some American cities are characterized by urban sprawl while others are densely populated and have many non-car transportation options.

The interstate highway system has led to a decline in center city population and the growth of suburbanization (Baum-Snow, 2007). Cities that developed and grew primarily after the development of the interstate highway system became more car dependent and were more likely to develop into urban sprawl because of the reduction in transportation costs. In fact, newer cities like Houston have been shown to consume up to 40% more gasoline than older cities like New York reflecting their increased dependence on cars (Newman and Kenworthy, 1989).

The instrument I propose is the number of old homes in a city. A believable instrument should be correlated with the endogenous regressors, but not correlated with the error term conditional on the other covariates. House age is a proxy for the age of the city. Cities that have older houses may also have older transportation patterns, this is why I expect house age to be correlated with car dependence and population density. Conditional on transportation patterns, population density, income, race, education, and geographic fixed effects, I would not expect the age of the house to be correlated with the likelihood to be in a food desert. If anything, I would expect to find older homes in poorer areas. If this is not completely controlled for by income, race, and education variables, it would bias the results downward making these estimates more conservative.

The two stage least squares approach is as follows, where equation 2 is the second stage and equation 3 is the first stage.

$$LAtract_{ct} = \beta_0 + \beta_1 \widehat{Explanatory}_{ct} + \beta_2 X_{ct} + \epsilon_{ct} \tag{2}$$

$$Explanatory_{ct} = \alpha_0 + \alpha_1 houseage_c + \alpha_2 X_{ct} + \nu_{ct} \tag{3}$$

I use $houseage_c$ (measure of the number of old houses in the city) and the census tract characteristics to predict the explanatory variable (transportation patterns or population density). $houseage_c$ is available in 2010 in the dataset. The second stage in equation 2 is identical to equation 1 above except that instead of using actual explanatory data, I use the predicted explanatory data from equation 3. I can test the strength of the instrument with an F-test of the predictive power of the instrument. However, I cannot test if the instrument is correlated with the error term from equation 1. Thus the validity of the these results is contingent on satisfying the exclusion restriction. One might argue for instance that house age is not truly exogenous and could be correlated with factors in the error term that affect food desert status. Because this debate can never be resolved empirically, I choose to include a difference in differences in specification to address the key question through a different method.

3.3 Difference in Differences

For the DD specification, I will look at the effect of a change population density or transportation patterns on a change in the probability of being in a food desert. The DD equation is as follows:

$$LAtract_{ct} = \beta_0 + \beta_1 Explanatory_{ct} + \beta_2 Explanatory_{ct} * Post_t + \beta_3 * Post_t + \beta_4 X_{ct} + \epsilon_{ct} \tag{4}$$

Where c is the census tract, and t is the time. $LAtract_{ct}$, $Explanatory$, and X_{ct}[1] are defined the same as above. I have two years of data so $Post = 0$ in the year 2010 and $Post = 1$ in the year 2015. This variable is the same as $Year = 2015$, I call it $Post$ in this regression to emphasize that we are looking for changes that occurred to commuting patterns and population density between 2010 and 2015 that would cause a change in the likelihood of being a food desert. The coefficient of interest is β_2, the interaction between the explanatory variable and the year. Therefore, the interpretation of the average treatment effect is the effect of a change in population density or a change transportation patterns on a change in the probability of being in a food desert.

Ideally I would like to compare each census tract to an identical counterfactual in an alternative universe that only differs in its propensity to sprawl and in food access. However, because such data does not exist, I use the DD approach to try to create a counterfactual. I compare each census tract to itself at another point in time. My identifying assumption is that if none of the census tracts had changed their transportation patterns or population density, the propensity to increase or decrease in likelihood of being a food desert should not differ between the census tracts that actually did become more or less car dependent. In other words, the transportation patterns and population density were the only things that changed which could be correlated with the outcome.

It is important to note that while the DD and IV approach use the same data, the DD approach looks at changes in transportation patterns over the scale of 5 years, while the IV approach allows for a much larger time frame

4 Empirical Application

I use a panel data set at the census tract level, for the years 2010 and 2015. The data on food access comes from Food Access Research Atlas which is compiled by the U.S. Department of Agriculture (USDA) Economic Research Service (Ver Ploeg et al., 2017). This dataset has a variable for census tracts with low access to grocery stores (which I will refer to as LA). I limited my sample to urban census tracts (a variable in the dataset), because the definition of low-access is different for rural and urban tracts. Out of the original 73,057 census tracts, this leaves me with 55,234 tracts.

The traditional definition of a food desert is a census tract that both has low access to grocery stores and is low income (which I refer to as LILA). The LILA variable is constructed by multiplying two indicator variables, a variable for lack of access to healthy food, and a variable indicating the census tract is low-income. As my main outcome variable, I choose to use only the low food access indicator (LA). This means that I included tracts with both high and low income. The reason for this choice is that including the low-income indicator in the outcome 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. Therefore, 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. For those interested, I include results using the low-income, low-access traditional food desert measure in the appendix.

[1] year is not controlled for because $Post$ is a dummy variable that is completely colinear with year

I use the American Community Survey (ACS) for the explanatory and control variables in the years 2010 and 2015 (U.S. Census Bureau, 2016). The ACS has estimates at the census tract level for a wide range of population and housing characteristics. The five key explanatory variables I examine are 1) population density and four variables capturing aspects of commuting patterns. The four variables are the percentage of people in the census tract that 2) drive alone, 3) use public transit, 4) walk, or 5) bike. The other control variables I use are median income, education, and race/ethnicity.

An urban tract that is characterized by sprawl will have a lower population density, lower levels, of transit ridership, walking, and biking, and higher levels driving. I used the Census TIGER/Line shapefiles to obtain the land area of each census tract and construct the population density (U.S. Census Bureau, 2017). It is important to note that I also control for population. I do this because I want to test the effect of density in particular not population as a whole on food access.

The 2010 ACS also collected data on the number of houses built in each decade, which I accessed from the database, IPUMS NHGIS (Manson et al., 2017). I use the number of houses built before 1939 as my instrument. A large number of of houses still standing from 1939 suggests that the city may maintain features and the development pattern of the pre-highway era.

I present the summary statistics for the data in Table 1. The first section contains the outcome variables, the second contains the explanatory variables, the third section is the instrument, and the remaining variables are control variables.

Table 1: Summary Statistics

Variable Type	Variable	Mean	Std. Dev.	Min.	Max.	N
Outcome	LA at 1 mile	0.449	0.497	0	1	110468
Variables	LILA at 1 mile	0.14	0.347	0	1	110468
Explanatory	Percent Drove Alone	73.667	16.671	0	100	109993
Variables	Percent Used Public Transit	6.941	12.999	0	100	109993
	Percent Walked	3.375	6.389	0	100	109993
	Percent Biked	0.697	1.836	0	100	109993
	Population Density	0.003	0.005	0	0.201	110466
Instrument	House built before 1939	254.47	363.575	0	5753	110468
Control	population	4336.23	2014.177	0	53812	110468
Variables	Year=2015	0.5	0.5	0	1	110468
	Median Income	57843.288	29653.977	3271	249194	109630
	Percent less than 9th grade education	6.416	7.623	0	100	110138
	Percent 9th-12th grade education	8.447	6.402	0	100	110138
	Percent high school graduate	26.78	10.69	0	100	110138
	Percent Some College	20.727	6.737	0	100	110138
	Percent Associates Degree	7.475	3.519	0	100	110138
	Percent Bachelors Degree	18.661	10.759	0	100	110138
	Percent Graduate Degree	11.495	10.185	0	100	110138
	Percent Black	15.699	24.16	0	100	109883
	Percent Native American	0.588	2.201	0	100	109883
	Percent Asian	5.023	9.029	0	100	109883
	Percent Pacific Islander	0.145	0.965	0	71.2	109883
	Percent White	71.977	26.356	0	100	109883
	Percent Hispanic	14.722	20.407	0	100	109883

Nearly half of urban census tracts (45%) have low access (LA tracts) to grocery stores. 14% of census tracts are both low-income and have low access to grocery stores (LILA tracts). In the average census tract, around 74% of people drive alone and around 10% use public transit, bicycling, or walking. Around 15% of people in the average census tract use another mode of transit such as carpooling or rideshare. The population density is measured in persons per square meter of land area. *Year=2015* is a dummy variable equal to 0 in 2010 and 1 in 2015. *House built before 1939* records the number of houses in the census tract built before 1939. *Median Income* is measured in dollars.

Figure 3 compares the means of the five explanatory variables in high and low access areas. As expected the unconditional mean percentage of people driving alone is higher in low access areas, and the unconditional mean for all the remaining variables, (transit, walking, biking, and population density) which are indicators of higher density more walkable cities, is higher in high access areas.

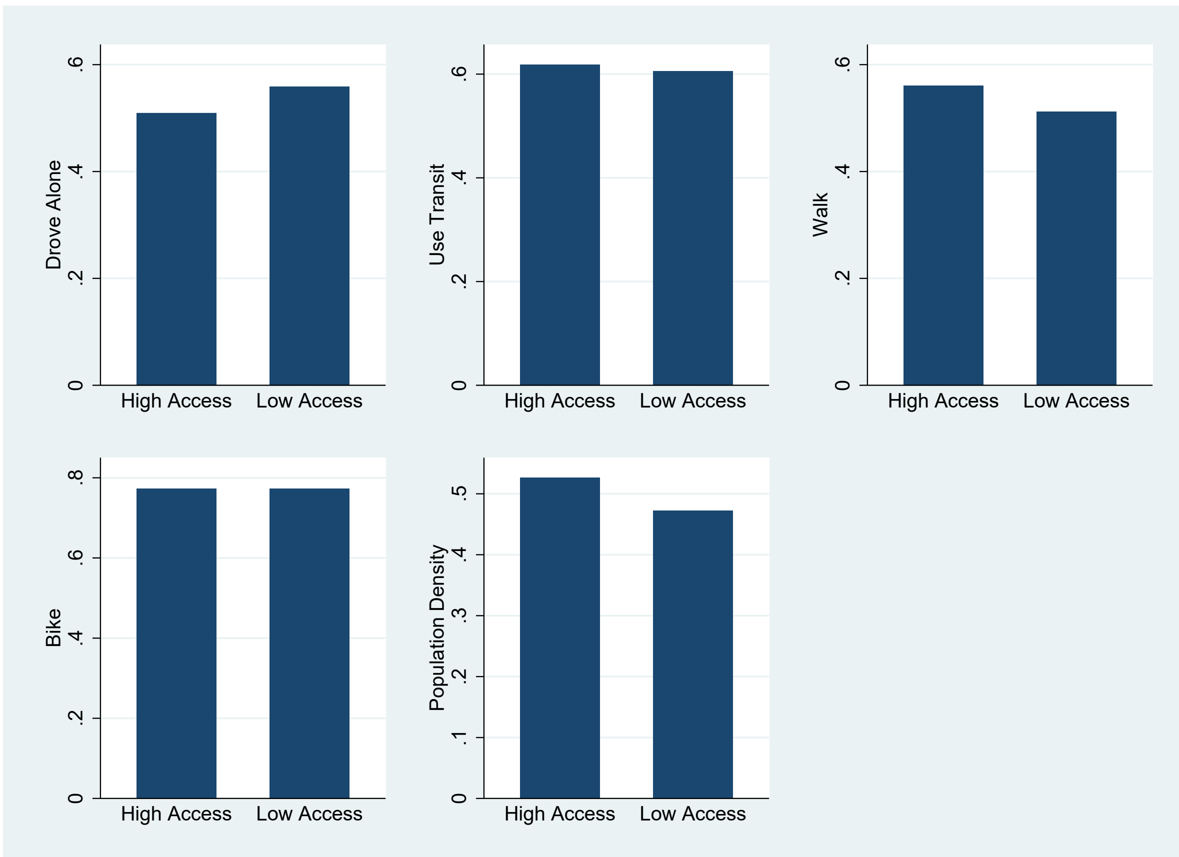

Figure 3: These graphs show the average of each of the key explanatory variables in high access and low access areas (measured using the 1 mile definition)

5 Results

Table 2 presents the OLS results. As expected, the percentage of people who drove alone was positively associated with low-food-access, and the percentage of people who used public transit, walked, or biked, and the population density were associated with high food access. However, the magnitude of these effects was small. For example, a 10 percentage point increase in the percent of people driving alone increases the probability of being in a LA census tract by 4%, and a one standard deviation increase (.005) in persons per square meter, decreases the probability for being in an LA tract by 8.5%. However, as discussed above these results may be biased due to endogeneity.

Table 2: OLS Results: Impact of Explanatory Variables on the Probability of being in a LA Tract

	(1)	(2)	(3)	(4)	(5)
Percent Drove Alone	0.004***				
	(0.000)				
Percent Used Public Transit		-0.006***			
		(0.001)			
Percent Walked			-0.005***		
			(0.000)		
Percent Biked				-0.017***	
				(0.002)	
Population Density					-17.005***
					(3.628)
Year=2015	-0.012***	-0.011***	-0.012***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
population	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education Variables	X	X	X	X	X
Race/Ethnicity	X	X	X	X	X
Income	X	X	X	X	X
Number of Observations	109619	109619	109619	109619	109625
Dependent Variable Mean	0.449	0.449	0.449	0.449	0.449

Note: County fixed effects are included in all regressions. Standard errors clustered at the county level are in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3 shows the IV results. The IV results have the same direction as the OLS results above however they are much stronger. For example, a 1 percentage point increase the percentage of people who commute by bike is associated with a 28% decrease in the probability of being in a food desert. In the OLS model a 10 percentage point increase in the number of people driving alone increased the likelyhood of the tract being a food desert by 4%. In this IV model, the likelyhood is increased by 22% for the same change in driving. The f-statistics for the first stage regressions are greater than 10 indicating that the correlation between house age and LA tracts is strong.

Table 3: Instrument Results: Impact of Explanatory Variables on the Probability of being in a LA Tract

	(1)	(2)	(3)	(4)	(5)
Percent Drove Alone	0.022***				
	(0.003)				
Percent Used Public Transit		-0.034***			
		(0.006)			
Percent Walked			-0.102***		
			(0.018)		
Percent Biked				-0.284***	
				(0.043)	
Population Density					-56.407***
					(16.652)
Year=2015	-0.010***	-0.004	0.006	0.037***	-0.003
	(0.003)	(0.004)	(0.005)	(0.008)	(0.004)
population	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education Variables	X	X	X	X	X
Race/Ethnicity	X	X	X	X	X
Income	X	X	X	X	X
Number of Observations	109619	109619	109619	109619	109625
Dependent Variable Mean	0.449	0.449	0.449	0.449	0.449
First Stage Fstat	108.173	54.026	43.605	37.829	15.619

Note: County fixed effects are included in all regressions. House age is used as an instrument. Standard errors clustered at the county level are in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4 shows the DD regressions. The variables of interest are the five explanatory variables interacted with the year 2015. For example a 10 percentage point increase in the number of people driving alone between the years 2010 and 2015 is associated with a .4% increase in the probability of the particular tract being a food desert. With the exception of population density which is insignificant, the direction of the DD results is consistent with both the OLS and the Instrumental variables results although the magnitude for the DD results is quite small. This could be because we are observing a change over only five years, which is a short time to change the structure of a city. As a result the changes in commuting patterns and population density were likely modest.

Table 4: DD Results: Impact of Explanatory Variables on the Probability of being in a LA Tract

	(1)	(2)	(3)	(4)	(5)
Drove Alone X 2015	0.0004***				
	(0.000)				
Percent Drove Alone	0.0041***				
	(0.000)				
Use Transit X 2015		-0.0002**			
		(0.000)			
Percent Used Public Transit		-0.0059***			
		(0.001)			
Walk X 2015			-0.0007**		
			(0.000)		
Percent Walked			-0.0042***		
			(0.000)		
Bike X 2015				-0.0024**	
				(0.001)	
Percent Biked				-0.0158***	
				(0.002)	
Population Density X 2015					0.2477
					(0.306)
Population Density					-17.1347***
					(3.725)
Year=2015	-0.0406***	-0.0095***	-0.0094***	-0.0080***	-0.0105***
	(0.008)	(0.003)	(0.003)	(0.002)	(0.002)
Population	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education Variables	X	X	X	X	X
Race/Ethnicity	X	X	X	X	X
Income	X	X	X	X	X
Number of Observations	109619	109619	109619	109619	109625
Dependent Variable Mean	0.449	0.449	0.449	0.449	0.449

Note: County fixed effects are included in all regressions. House age is used as an instrument. Standard errors clustered at the county level are in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

The results for the models using LILA tracts as an outcome variable are presented in the appendix. As discussed above, these models are likely less accurate because income may be directly associated with transportation patterns. Overall the results support my hypothesis that car dependence and low population density are associated with food deserts. However, it is important to note that the significance of these results is not very meaningful due to the large number of observations. The strongest support for my hypothesis is found not in the significance of individual results, but rather in the consistency of the direction of impact across specifications and different measures of sprawl development.

6 Policy Implications

Interstate highways and automobiles have afforded numerous gains to society including improvements in efficiency, connectivity, and access to trade. Food access problems arise from a reduction in transportation costs which make the grocery market competitive on aspects other than location, and in some cases driving small local grocery stores out of business. Clearly, the fact that cars reduce transportation costs is good for society. However, it also has costs and these costs are often not paid for by those who drive. Car usage is subsidized through free roads, free parking, and cities that are built specifically to favor cars. The most successful policies would seek not to eliminate car transportation but to accurately price it.

In cities that are already quite car dependent, taxes on car ownership and usage have been proposed to mitigate the harmful effects of traffic, pollution, and carbon emissions (Barter, 2005; Hayashi et al., 2001; Rogan et al., 2011). Such taxes should additionally take into account the effect of car dependence on food access. Alternatively, city officials could take measures to stop subsidizing car usage by using public money to fund infrastructure that only benefits cars, or to divert some of this money to infrastructure that benefits pedestrians, bicyclists, and transit users. In most developing cities, car dependence is not yet a problem. City planners can encourage a development pattern that prioritizes density and public transportation. Instead of building highways, developers could focus on a robust public transportation system, high speed trains, safe bike lines, or sidewalk connectivity (a great example of such development is the construction of TransMilenio in Bogotá, Colombia (Hidalgo and Sandoval, 2002)).

In order to add food access to a cost benefit analysis of public transportation projects or sprawl reduction strategies. Researchers will need to accurately quantify the effect of sprawl development on grocery store location and density. This study is an attempt to do that but it has limitations in that the time frame of the difference in differences analysis is quite short, and there are only two years of data. Those who have access to data on grocery store location for longer time periods may be able to more accurately understand how car dependence affects food access. Such research would be very useful as modernization changes the grocery landscape worldwide.

7 Conclusion

This study seeks to understand and quantify the effect of sprawl development on food access. This is a difficult question to study because it is hard to disentangle the effects of the built environment from income and preferences. I attempt to do this through using IV and DD techniques. I present OLS results for completeness but these are likely biased.

Overall the results suggest that a sprawl development pattern (higher levels of driving, lower levels of transit, biking, walking, and population density) is associated with lower food access. The OLS results showed that sprawling areas were associated with lower food access. The IV results showed that by using the variation in house age (a proxy for the age of the city) to predict sprawl, more sprawling areas were still, in general, associated with lower food access. The difference in difference results exploited small changes in sprawl over a 5 year period, to show that places that became more prone to sprawl developed more food deserts and vice versa. Therefore, infrastructure changes on both a short and longterm basis had an effect on the propensity of a tract to have low access to healthy food.

However, there are several limitations to this study. The effects in the difference in differences study where quite small, and I believe this is because of the short observation period. It would be beneficial to have more years of data to explore this question over a longer time period. In particular, it would be helpful to have case studies of cities that made a change to infrastructure or planning in a way that increased or decreased sprawl to see if this has an effect on the food environment. Additionally, while the instrument captures randomness in transportation patterns in that some cities were largely constructed before the interstate highway system, it is an imperfect measure, and a better instrument may provide more accurate results. Despite these limitations, the results are very consistent and warrant further investigation.

References

Barter, P. A. (2005, nov). A Vehicle Quota Integrated with Road Usage Pricing: A Mechanism to Complete the Phase-Out of High Fixed Vehicle Taxes in Singapore. *Transport Policy 12*(6), 525–536.

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.

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.

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.

Freire, T. and S. Rudkin (2019, feb). Healthy food diversity and supermarket interventions: Evidence from the Seacroft Intervention Study. *Food Policy 83*, 125–138.

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.

Hayashi, Y., H. Kato, and R. V. R. Teodoro (2001, mar). A Model System for the Assessment of the Effects of Car and Fuel Green Taxes on CO2 Emission. *Transportation Research Part D: Transport and Environment 6*(2), 123–139.

Hidalgo, D. and E. Sandoval (2002). TransMilenio: A high capacity–low cost bus rapid transit system developed for Bogotá, Colombia. *... of the Tenth International CODATU Conference*.

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.

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.

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

Manson, S., J. Schroeder, D. Van Riper, and S. Ruggles (2017). PUMS National Historical Geographic Information System: Version 12.0 [Database]. Technical report, Minneapolis: University of Minnesota.

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.

Newman, P. W. C. and J. R. Kenworthy (1989). Gasoline Consumption and Cities Cities with a Global Survey. *Journal of the American Planning Association 55*(1), 24–37.

Powell, L. M., S. Slater, D. Mirtcheva, Y. Bao, and F. J. Chaloupka (2007, mar). Food store availability and neighborhood characteristics in the United States. *Preventive Medicine 44*(3), 189–195.

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

Rogan, F., E. Dennehy, H. Daly, M. Howley, and B. P. Ó Gallachóir (2011, aug). Impacts of an Emission Based Private Car Taxation Policy – First Year Ex-Post Analysis. *Transportation Research Part A: Policy and Practice 45*(7), 583–597.

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.

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.

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

U.S. Census Bureau (2017). 2010 TIGER/Line® Shapefiles.

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

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

Ver Ploeg, M. and P. E. Wilde (2018, aug). How do food retail choices vary within and between food retail environments? *Food Policy 79*, 300–308.

You, W. (2017). The Economics of Speed: The Electrification of the Streetcar System and the Decline of Mom-and-Pop Stores in Boston, 1885-1905.

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.

8 Appendix

Table 5: OLS Results: Impact of Explanatory Variables on the Probability of being in a LILA Tract

	(1)	(2)	(3)	(4)	(5)
Percent Drove Alone	0.001***				
	(0.000)				
Percent Used Public Transit		-0.003***			
		(0.000)			
Percent Walked			-0.000		
			(0.000)		
Percent Biked				-0.003***	
				(0.001)	
Population Density					-7.507***
					(2.131)
Year=2015	0.003*	0.004**	0.003*	0.004**	0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
population	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education Variables	X	X	X	X	X
Race/Ethnicity	X	X	X	X	X
Income	X	X	X	X	X
Number of Observations	109619	109619	109619	109619	109625
Dependent Variable Mean	0.14	0.14	0.14	0.14	0.14

Note: County fixed effects are included in all regressions. Standard errors clustered at the county level are in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Table 6: Instrument Results: Impact of Explanatory Variables on the Probability of being in a LILA Tract

	(1)	(2)	(3)	(4)	(5)
Percent Drove Alone	0.002***				
	(0.001)				
Percent Used Public Transit		-0.004***			
		(0.001)			
Percent Walked			-0.011**		
			(0.005)		
Percent Biked				-0.030***	
				(0.009)	
Population Density					-5.982*
					(3.103)
Year=2015	0.003**	0.004**	0.005***	0.008***	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
population	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education Variables	X	X	X	X	X
Race/Ethnicity	X	X	X	X	X
Income	X	X	X	X	X
Number of Observations	109619	109619	109619	109619	109625
Dependent Variable Mean	0.449	0.449	0.449	0.449	0.449
First Stage Fstat	108.173	54.026	43.605	37.829	15.619

Note: County fixed effects are included in all regressions. House age is used as an instrument. Standard errors clustered at the county level are in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7: DD Results: Impact of Explanatory Variables on the Probability of being in a LILA Tract

	(1)	(2)	(3)	(4)	(5)
Drove Alone X 2015	0.0001				
	(0.000)				
Percent Drove Alone	0.0011***				
	(0.000)				
Use Transit X 2015		0.0002**			
		(0.000)			
Percent Used Public Transit		-0.0031***			
		(0.000)			
Walk X 2015			-0.0001		
			(0.000)		
Percent Walked			-0.0002		
			(0.000)		
Bike X 2015				0.0004	
				(0.001)	
Percent Biked				-0.0035***	
				(0.001)	
Population Density X 2015					0.3170
					(0.194)
Population Density					-7.6720***
					(2.170)
Year=2015	-0.0071	0.0027	0.0035*	0.0033**	0.0034*
	(0.007)	(0.002)	(0.002)	(0.002)	(0.002)
Population	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education Variables	X	X	X	X	X
Race/Ethnicity	X	X	X	X	X
Income	X	X	X	X	X
Number of Observations	109619	109619	109619	109619	109625
Dependent Variable Mean	0.14	0.14	0.14	0.14	0.14

Note: County fixed effects are included in all regressions. House age is used as an instrument. Standard errors clustered at the county level are in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.