Inequalities in Food Access in Milwaukee County, Wisconsin

Rebecca Rate

with Faculty Advisor Dr. Soren Hauge

BSA 500: Senior Seminar

Dr. Fan Zhang

April 26, 2019

Abstract

Debate continues over “food deserts”, or disparities in food access, which disproportionately affect low-income and minority communities. This study adds to the quantitative empirical analysis of these disparities, informed by economic theory. Data for neighborhoods of Milwaukee County in 2016 are used to test whether, controlling for other economic factors, there remains a statistically and economically significant difference in access to various types of food retailers among neighborhoods of different racial compositions. The process is modelled on an empirical analysis done in Erie County, New York in 2008, with some modifications. Travel times from retailers to neighborhoods (census block groups) measured in a geographic information system provide counts of retailers accessible to each neighborhood. These are used for estimation of inter-neighborhood inequalities with Gini coefficients and incidence rate ratios with Poisson, negative binomial, zero-inflated Poisson and semi-parametric Poisson regressions. The findings show that there are far fewer large supermarkets accessible to neighborhoods that are predominantly black, compared to those that are predominantly white. By contrast, there are a greater number of smaller grocery stores, by all travel modes -- driving, walking and bicycling. Results are mixed for convenience and variety stores, but the statistically significant results show more convenience and variety stores in neighborhoods that are predominantly black. If further research confirms higher costs and lesser variety of healthy foods at smaller grocery stores, convenience and variety stores, these results will have troubling implications for public health in the context of a metropolitan area with high levels of racial segregation.

Inequalities in Food Access in Milwaukee County, WI

I. Introduction

The Centers for Disease Control and Prevention (2017) define food deserts as "areas that lack access to affordable fruits, vegetables, whole grains, low-fat milk, and other foods that make up a full and healthy diet." The USDA also includes access to a vehicle and travel distances to food retailers (Ploeg & Rhone, 2016). As the definition of food deserts becomes more complicated, so do the arguments about their existence. For this study, the issue of defining this term is abandoned. Instead, this study will test for statistically and economically significant *differences* [disparities] in access to various types of food retailers in neighborhoods with different racial compositions, as disparities in food access have been linked to the racial composition of the area, using data from Milwaukee County, Wisconsin in 2016. During this time, the city of Milwaukee, which is the major metropolitan area in Milwaukee County, was identified as the most racially segregated city in the United States (Bayatpour, 2016). If there is a link between food access and race, the unusually high racial segregation in this area should lead to more significant results.

In the first part of this study, Gini coefficients are used to test the null hypothesis that there is no difference in the number of each type of food retailer located within the geographical boundaries of each census block group (CBG). In this test, the null hypothesis of no difference is rejected. There is evidence of disparities in food access at the CBG level in Milwaukee County. The next question is whether these differences can be explained by differences in economic variables. For example, in areas with lower population and/or lower income one would expect to see fewer retail locations. Also, because of economies of scale, one would expect to see fewer large retail locations serving a larger area. When markets diverge from these expected behaviors, convenient assumptions may be made to explain away unexpected market behavior. For example, it is easy to say that there are fewer retail locations than expected because of heterogeneity of preferences. In other words, people in these neighborhoods prefer not to eat fresh fruits and vegetables, therefore suppliers have fled the area. It is equally easy to assume that food suppliers choose not to locate in these areas because they prefer *not* to conduct business in areas that are *not* predominantly white. However, both of these assumptions are harmful and should not be made without statistically significant evidence. This study adds to the limited quantitative empirical analysis of this topic.

In the second part of this study, Poisson regression is used to test the null hypothesis that differences in neighborhood racial composition *cannot* explain difference in food access -- both within the geographical boundaries of CBGs and within a 5-minute walk, bike-ride, and drive from the CBG centroid --*when controlling for economic factors*. To test theories about the correlation between race and food access, a racial indicator is added to this analysis. In these regressions, the null hypothesis is also rejected. There are far <u>fewer</u> large supermarkets accessible to CBGs that are predominantly black, compared to those that are predominantly white. By contrast, there are a <u>greater</u> number of smaller grocery stores accessible by all travel modes in predominantly black CBGs. Results are mixed for convenience and variety stores, but the statistically significant results show <u>more</u> convenience and variety stores in CBGs that are predominantly black.

In the third part of this study, the Poisson regression model is tested using the goodness of fit Chi-square test. Poisson regression assumes that the variance of the independent variable(s) equals its mean. It is highly unlikely that this assumption is true and there is a problem with overdispersion. The negative binomial regression model is designed to be used with over-dispersed count data and includes an extra parameter to model this. Half of the Poisson regressions failed the overdispersion test, and all were repeated using the negative binomial regression model. This model produced similar result usually with smaller confidence intervals. Sometimes, excess zeros in a data set can be mistaken for overdispersion. To account for this, a zero-inflated Poisson regression model was also used to evaluate the same data, while accounting for the high number of data points equal to zero. The regressions that contain high numbers of counts equal to zero produced results that are quite different from the original results. The most different result was for supermarkets, the results changed from far fewer locations in predominantly black neighborhoods with statistical significance to more locations but without significance. Finally, a semi-parametric Poisson model, which is specifically designed for use when there is an unknown regression relationship, was used to evaluate the same data. The results followed the same pattern indicating fewer supermarkets and restaurants in predominantly black and racially-mixed neighborhoods and more small grocery and convenience stores.

Although these results show statistically significant evidence of differences [disparities] in access to types of food retail locations, they do not clearly indicate that these areas are "food deserts". If further research confirms higher costs and less variety of healthy foods at smaller grocery, convenience and variety stores, then the conclusion can be reached that these results have troubling implications for public health in the context of a metropolitan area with high levels of racial segregation like Milwaukee County. However, if smaller food retail locations can provide healthy and ethnically appropriate foods with shorter travel times at reasonable prices, then the greater number of locations would indicate the opposite of a food desert. Further empirical research is needed in this area and would be useful to inform urban planning and policy decisions in the future.

The rest of this paper is structured as follows: Section two provides a brief literature review. Section three explains the empirical models used to measure inequalities in the distribution of food retailers and disparities in food access controlling for economic variables. This section also describes the data sources and the process to create the dataset used in the empirical models. Section four presents the results. Section five provides a discussion of the results. Section six concludes.

II. Literature Review

Economic Theory and the Food Retail Market

This study uses a reduced form model that combines the effects of supply and demand on the food retail market. The literature on this topic provides a detailed explanation of how specific issues of supply and demand affect this market. Although an attempt to separate the effects will not be made in this study, understanding how the economic variables that are controlled for in the model affect supply and demand separately will help to interpret the results.

Input costs affect the supply side of the food retail market and the theories of economies of scale and scope provide a framework to understand how/why input costs vary. In recent decades, few, large, high-quality chain stores have started to dominate this market. The theory of economies of scale explains why food prices are higher at smaller stores than at large supermarkets. Having one large firm, instead of many small firms allows that supplier to distribute fixed costs across more customers and offer lower prices. Johnson et al. (1996) found evidence of this in Milwaukee County, WI and Caspi et al. (2017) found this to be true both inside and outside of areas designated "food deserts". The theory of economies of scope explains why smaller stores have a limited variety of foods. Economies of scope are cost advantages that result when firms provide a variety of products to the same consumers, rather than specializing in the production or delivery of a single product or service. Johnson and Caspi et al. both show evidence of a correlation between store size and variety of food offered. Verschay (2009) finds that this is not exclusively an urban phenomenon. In rural Clark County, WI, small stores also show evidence of limited availability of fresh produce with prices that vary significantly between summer and winter. This explains why policies concerning food deserts have focused on bringing large supermarkets with lower prices and larger varieties to underserved areas (Protect the Harvest, 2016).

The theory of economies of agglomeration provides additional insight into the issue of costs in low-income food desert areas. Economies of agglomeration explains that similar firms locate close together in order to take advantage of cost saving from sharing suppliers, information, customers, etc.. Bitler and Haider (2011) and Bonanno (2012) explain that transportation infrastructure and distance from distribution centers influence input costs, and therefore where retailers locate and the prices that they charge. In addition to poor transportation infrastructure, on many occasions, low-income areas have insufficient budgets for other public goods, like police and fire protection, which can make locating in these areas more risky and costly for firms. There may also be a shortage of qualified labor in areas where there are barriers to education. However, labor and property costs are also lower in these low-income areas, so it is possible that some of these costs may offset.

In a perfectly competitive market, an increase in the number of firms offering the same product should lead to a decrease in price. However, in a market facing monopolistic competition, firms will use advertising to differentiate products and services to maintain some degree of pricing power. They will remain imperfect substitutes, so firms will not have complete control over prices, but more than in the case of perfect competition. Hatzenbuehler, Gillespie and O'Neal (2012) find that in higher-income areas, there are multiple large supermarkets; however, more competition by these superstores results only in more locations, not lower prices. Many news sources and consumer reports show an increase in competition for supermarkets from non-primary-food retailers, like dollar stores and Walgreens (Jacobson, 2018). Although the same products are being sold, this has not resulted in lower prices at either type of retail location. It appears that an increase in competition has not led to lower prices. Conversely, Bonanno and Li (2015) find that, lack of competition in low access areas leads to higher prices due to monopolistic position and/or cost inefficiencies. Again, Caspi et al. (2017) compares similar findings inside and outside areas designated food deserts and concludes that prices are higher at isolated stores than at non-isolated stores in both areas. So, it appears that while competition does not lower prices because of factors of monopolistic competition, it does prevent them from rising because of factors of oligopoly or even true monopoly.

Income and prices are the two biggest influences on the demand side of the food retail market. Economic theory tells us that we would expect to see fewer stores in areas with lower income, controlling for population (Bonanno, 2012). Government programs such as the Supplemental Nutrition Assistance Program (SNAP) and Women, Infants and Children (WIC) Supplemental Nutrition Program provide subsidies that increase purchasing power for food. Bitler and Haider (2011) and Bonanno (2012) explain that this helps to increase demand for food in low-income areas. Kozlova (2016) finds that when individuals with low incomes have more money to spend on food, they do not prefer healthy food over unhealthy food. However, Weatherspoon et al. (2012a & 2012b) finds that both price and income elasticities of demand for fruits and vegetables, for consumers using SNAP benefits in designated food desert areas in Detroit, are larger than the national average, in other words, they do prefer healthy food over unhealthy food. In Weatherspoon's studies, for both fruits and vegetables, income plays a much bigger role than price. So, it is clear that an increase in income, or an increase in food subsidies, leads to an increase in demand for food; however, it is less clear whether there is a greater increase in demand for healthy food or if the increase in demand for healthy and unhealthy food is equal.

Although income may be the primary barrier in purchasing healthy food it is not the only one. Constraints on time and/or transportation can also have an impact (Bonanno & Li, 2015) (Hillier et al., 2015). Douangchai (2011) found mobility to be a key factor when access to food was studied at the census tract level. Raja, Ma and Yadav (2008) and Dutko (2012) find that access to vehicles allows some to overcome food access problems by eliminating transportation disparities. Transportation to food retail locations is part of the cost of purchasing food. For individuals with low incomes, overcoming this boundary becomes increasingly more difficult, and more expensive, the fewer and further apart food retail locations become.

Even if there are no physical barriers, heterogeneity of preferences may still cause consumers to choose unhealthy foods over healthy foods. Bitler and Haider (2011) explain that this may be due to discount rates, or the inability of individuals of lower socioeconomic status to invest in the future. They suggest that this includes investing in their future health by increasing spending on healthy food. Alcott, Diamond and Dube (2017) take the position that low-income populations across the U.S., which disproportionately include racial/ethnic minorities, just prefer not to eat as much healthy food as a result of several factors including lack of information. However, in Milwaukee County, Warsaw (2018a & 2018b) uses the hedonic pricing model and heterogeneous consumer preferences for housing characteristics, including food access, to show that African-American households have a much higher willingness to pay for access to food controlling for income. Although there is evidence of heterogeneity of preferences, there is also growing evidence that healthy food is a normal good and that demand increases as income increases, and at an even greater rate among low-income and minority consumers.

Geographic Information Systems and the Spatial Distribution of Food Access

The way that we measure access to food is constantly evolving. GIS technology has been a useful tool in the development of this analysis, with some limitations. Moore and Diez-Roux (2006) measured food access by using GIS to count the number of food retailers located within the geographical boundaries of the census tracts. Census tracts are much larger than CBGs, and census blocks are smaller than CBGs but do not provide median household income data. This makes the CBGs the smallest area for which this analysis can be completed with publicly available data. In this paper, CBG and the more relatable term "neighborhood" will be used interchangeably where consistent with clarity. Although, socially constructed neighborhoods may not coincide with the geographical boundaries of CBGs. Raja, Ma and Yadav (2008) use both counts within the geographical boundaries of CBGs and travel distances to retail locations from CBG centroids. Their choice to use travel times was informed by the work of Helling and Sawicki (2003) who incorporated opportunity cost of travel time into their food access analysis. Caldwell (2011) uses GIS to map "provisional food markets", specifically community supported agriculture and food buying clubs. Caldwell uses data at the census tract level and finds that the majority of community supported agriculture in the city of Milwaukee is located in predominately white areas. However, although their future is uncertain, *Growing Power*[1], the last urban farm in Milwaukee, has served a low-income predominantly black area since 1993 (Satterfield, 2018). Conversely, most food buying clubs locate in moderately to extremely poor census tracts in the central city. Caldwell's study is complimentary to this as his focus is on different types of locations where food may be accessed.

GIS has proven to be a useful tool; however, the limitations of GIS based spatial analysis must also be acknowledged. Caldwell (2011) notes that either an address, or latitude and longitude coordinates, are needed to locate food retailers on a map. This means that nontraditional retailers such as produce stands and mobile markets are excluded. He also finds that the identification codes[2] used to classify retailers by type are commonly inaccurate. This leads to the omission and misidentification of retail locations offering healthy food options and results in inaccurate analyses. Pettygrove and Ghose (2016) express concern about additional deficiencies in traditional spatial analysis of food access. These include individual's social behaviors, the varied ways that people obtain food, and broader political and economic issues that contribute to inequalities of resources and power. Although they do not explicitly state the use of GIS, Leonard et al. (2014) address the question of social influence when they apply Durlauf's (2004) theory of neighborhood effects to food choice by combining measurements of proximity to food sources and dietary intake of spatially-based social networks. Their results support the hypothesis that there is a relationship between nutrition intake and social ties in the predominantly black Fair Park neighborhood of Dallas. However, it is not clear if this is due to peer-to-peer interactions or social norms; more research is also needed in this area.

III. Methods & Data

Methods

This study will replicate an empirical study by Raja, Ma & Yadav (2008) with data from Milwaukee County, Wisconsin with additional methods. Publicly available data for CBGs and food retailers are matched, and preliminary analysis done in a geographic information system (GIS), with adjustments to the categorization of food retailers. The main two analyses are of the inequalities of spatial distribution of food retailers and of their correlates in neighborhood racial

[1] Growing Power is now known as Will's Roadside Stand.
[2] Caldwell uses NAICS and Raja, Ma & Yadav use SIC codes, both have the same accuracy problems.

composition. The first uses Gini coefficients to measure the inequality across CBGs in counts of within-CBG food retailers. Lorenz curves also illustrate these inequalities. The second analysis uses a reduced form Poisson regression model to test if there is a significant difference in access to specific types of food retailers in CBGs with more people of color versus predominantly white CBGs. The model combines the effects of demand and supply on the food retail market. Additional models -- negative binomial regression, zero-inflated Poisson regression and semiparametric Poisson regression -- are also used to analyze the same data. These models were not used in the study being replicated; however, may be a better fit for the data being used.

GIS design and preliminary analysis

Retailer data were matched to other data with the Census Geocoder. It is a web-based tool that for each retailer's address provided its geographic coordinates – latitude and longitude – and a geographic code matching those in the American Community Survey, the source of other data. In addition to placing food retailers of different types within their CBGs, origin-destination cost matrix analysis in ArcGIS Network Analyst was used to estimate travel times from each CBG's centroid to each food retailer. From those travel times, calculated separately for driving, bicycling and walking, a count of retailers within a five-minute travel time was totaled for each CBG and travel mode. Travel speeds were adapted from state and federal government guidelines, matched for driving to categories of functional classification codes in geo-referenced street and road data, and adjusted with the global turn delay evaluator of ArcGIS for intersections and turns.

Inequalities of spatial distributions of food retailers

This analysis employed counts of the number of each type of food retailer located within the geographical boundaries of each CBG. Using the *ineqdec0* module of Stata to estimate a Gini coefficient for the inequality of the distribution of each of several types of food retailers across CBGs.[3] If a Gini coefficient were equal to 0, the retailers would be perfectly equally distributed and if equal to 1, the retailers would be concentrated in just one CBG. Intermediate values reflect differing degrees of spatial inequality. Patterns of these inequalities are illustrated with Lorenz curves constructed in Stata with its *glcurve* module.[4] The Gini coefficient estimates the ratio of

[3] http://www.lisdatacenter.org/wp-content/uploads/2011/03/C3-3-7-2-self-teaching-stata.pdf
[4] https://www.statalist.org/forums/forum/general-stata-discussion/general/288264-multiple-lorenz-curve-and-graph-line-45

the area between a Lorenz curve and the diagonal 45-degree line of perfect equality to the entire area under the line of perfect equality. Hence, the further a Lorenz curve is below the line of perfect equality, the larger the Gini coefficient and the more concentrated the spatial distribution of retailers.

Correlates of food access across neighborhoods, focusing on racial composition

The relationships of multiple neighborhood (CBG) characteristics to each type of food retailer count were analyzed using Poisson regressions, employing the *Poisson* module of Stata.[5] The Poisson regression model is a reduced form equation combining demand and supply factors as independent variables with a retailer count as a dependent variable that could represent an equilibrium of demand and supply. A separate estimate was calculated for each type of food retailer and its separate counts within the neighborhood and within a five-minute travel time by each travel mode. The independent variables used are median household income, population, land area, and indicator variables for the racial composition of the neighborhood.

A composite of Poisson regression estimates is the focus for greater accuracy in estimating the variation of retailer counts with neighborhood racial composition. The key slope coefficients of the Poisson equations themselves are log differences in expected retailer counts between neighborhoods with different categorical characteristics (racial composition in this model). If E_j is the exposure, the expected counts for the j^{th} observation, C_j, for a particular type of food retailer, will be

$$C_j = e^{\ln(E_j)+\beta_0+\beta_1 x_{1j}+\cdots\beta_i x_{ij}+\cdots+\beta_k x_{kj}}$$

With exposure assumed to be constant at 1 across neighborhoods, this coincides with the incidence rate. Where x_i is a predominant neighborhood racial composition dummy variable, the corresponding incidence rate ratio (IRR) divides the expected retailer count of a neighborhood with that characteristic by the expected retailer count of a neighborhood without it:

$$e^{\beta_i} = \frac{e^{\ln(E)+\beta_1 x_1+\cdots+\ \beta_i(x_i+1)+\cdots+\beta_k x_k}}{e^{\ln(E)+\beta_1 x_1+\cdots+\ \beta_i x_i+\cdots+\beta_k x_k}}$$

Both can be used -- directly for log differences, and indirectly for IRRs (subtracting 1) -- to measure proportional (percentage) differences between neighborhoods with different characteristics. Specifically, expected retailer counts (including those within five minutes by

5 https://www.stata.com/manuals13/rpoisson.pdf

three travel modes) in predominantly black and racially-mixed neighborhoods are compared with those in predominantly white neighborhoods. The inherent inaccuracy of using log differences for this purpose increases, however, as the proportional difference increases and is quite large at the observed levels of disparities between neighborhoods with different predominant racial groups. Hence, following Raja, Ma and Yadav (2008) IRRs are used, either interpreting them as ratios, or subtracting 1 to interpret the proportional differences. Using IRRs also changes how the null hypothesis is stated. It could be stated, as is common, with $H_0\colon \beta_i = 0$. However, because the point estimate is a ratio, no difference of expected retailer counts between neighborhoods of different characteristics is an IRR of one, not zero, so the null hypothesis is stated as $H_0\colon IRR = 1$.

Analysis using additional models

The accuracy of the Poisson regression itself is dependent on the following assumption: the distribution of counts must follow a Poisson distribution with mean of the observed random variable(s) equal to its variance. To test if this assumption has been violated, a test for overdispersion is appropriate. The goodness-of-fit Chi-square test can be completed in Stata with its *estat gof* module.[6] When overdispersion is detected the negative binomial model will be more accurate. This model has been formulated with overdispersion as an end itself, or as a consequence of incorporating unobserved individual heterogeneity (Greene, 1994). The negative binomial regression is employed in Stata using module *n breg*[7]. The Poisson distribution may be generalized by including a gamma noise variable which has a mean of 1 and a scale parameter of ν. The Poisson-gamma mixture (negative binomial) distribution that results is

$$\Pr(Y = y_i | u_i, a) = \frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1 + au_i}\right)^{a^{-1}} \left(\frac{au_i}{1 + au_i}\right)^{y_i}$$

where

$$u_i = \exp(ln(E_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}) \;\; and \;\; a = \frac{1}{v}$$

Half of the Poisson regressions tested failed the goodness-of-fit Chi-square test. There is evidence of overdispersion; so, use of the negative binomial regression model is appropriate.

Excess zeros in a data set can also masquerade as overdispersion; however, this is not a case of overdispersion, or greater variability, but a greater number of data points equal to zero

---

[6] https://www.stata.com/help13.cgi?logistic+estat+gof
[7] https://stats.idre.ucla.edu/stata/dae/negative-binomial-regression/

than expected. The zero-inflated Poisson regression model has been modified to handle these excess zeros. The appropriateness of this model can only be evaluated by testing the modification against the base. The modified zero-inflated Poisson regression is run simultaneously with the base by employing the *zip* module in Stata.[8] Suppose that for each observation, there are two possible cases. Suppose that if case 1 occurs, the count is zero. However, if case 2 occurs, counts (including zeros) are generated according to a Poisson model. Suppose that case1 occurs with probability π and case 2 occurs with probability 1 - π. Therefore, the probability distribution of the ZIP random variable y_i can be written

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1-\pi_i)\exp(-u_i) & if\ j = 0 \\ (1-\pi_i)\dfrac{u_i^{y_i}\exp(-u_i)}{y_i!} & if\ j > 0 \end{cases}$$

Where the logistic link function π_i is given by

$$\pi_i = \frac{\lambda_i}{1+\lambda_i}$$

where

$$\lambda_i = \exp(in(E_i) + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \cdots + \gamma_m z_{mi})$$

The logistic component includes an exposure and a set of *m* regressor variables (the *z*'s). Note that the *z*'s and the *x*'s may or may not include terms in common.

The Poisson component can include an exposure and a set of *k* regressor variables (the *x*'s). The expression relating these quantities is also

$$u_i = \exp(ln(E_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})$$

The zero-inflated Poisson model is most appropriate when zero is a type of default data point. For example, when counting offspring in a sample that contains individuals that have not yet reached reproductive maturity. In this case, should a zero from an organism that is not yet reproductively mature be counted the same as a zero from a mature organism? For the purpose of this study, a similar question could be asked about zoning. Should a zero that occurs in an CBG with no commercial zoning be counted the same as a zero in a CBG where retailers are able to locate but choose not to? This question has not been definitively answered so results from this regression are included. For the ease of comparison, the results of both the negative binomial

[8] https://stats.idre.ucla.edu/stata/dae/zero-inflated-poisson-regression/

regression and the zero-inflated Poisson regression are also expressed as IRRs. Therefore, the null hypotheses remain the same.

Finally, a semiparametric Poisson regression is proposed in modeling spatially clustered count data. Semiparametric estimation methods are generally more advantageous over the traditional approaches when the linear model fit is poor. In cases where there is a good linear fit, the proposed method is inferior to the traditional methods, but can still be advantageous when there are several covariates involved since the back-fitting algorithm yields computational simplicity in the estimation process.

The mean of the count data Y is affected by explanatory variables (Xi's) and the heterogeneous Poisson model is $Y_i \sim P_0(u + X_i'\beta)\ or\ logE(Y_i) = u + X_i'\beta$. The expected mean of the response variable in this model is heterogeneous and depends on the explanatory variables. However, in phenomenon that exhibit spatial dependence, the cluster where the observation belongs can further contribute homogeneous effects on the response variable. A Poisson regression that can account for cluster effects is $Y_{ij} \sim P_0\left(u + X_{ij}'\beta\right)$ where Y_{ij} refers to the j^{th} observation in the i^{th} cluster and u_i is the cluster-specific intercept, a random component.

It is hypothesized that because of clustering, the effects of X_{ij} vary across the clusters. The model becomes $Y_i^c \sim P_0\left(u_c + X_{ij}^c\beta_j\right)$ where Y_i^c is the i^{th}observation in the c^{th}cluster, and is highly vulnerable to overparametrization. To resolve this issue, transforming the model into an additive combination of parametric and nonparametric specifications, i.e. $logE(Y_i^c|u_c) = u_c + f(X_{ij}^c)$. The cluster-specific intercept u_k is formulated parametrically through the random effects, while the covariates are specified in a nonparametric way (Barrios & Vera, 2011). The semiparametric model is then estimated iteratively through the back-fitting algorithm. This model can be extended to more than one covariate as

$$logE(Y_i^c|u_c) = u_c + f_1(X_{i1}^c) + \cdots + f_{ip}\left(X_{ip}^c\right)$$

The above is an additive model. By introducing a distribution and link function into the additive model, a generalized additive model (GAM) is created.

$$logE(Y_i^c|u_c) = u_c + f_1(X_{i1}^c) + \sum_j f_j\left(X_j^c\right)$$

The GAM alleviates the curse of dimensionality and becomes easy to fit computationally. In this study, the intercept and race indicator are modeled with parameters and all other variables are modelled in additive nonparametric form, which results in a semiparametric GAM as shown above. There is not module in Stata (until Stata 15) for the semiparametric Poisson regression model. To complete this analysis, the *'mgcv'* package in the open source statistical software R Studio is available. The implementation of GAM in this package is not based on back-fitting, but on the Lanczos algorithm, a way of efficiently calculating truncated matrix decompositions and is restricted to splines with no mixing of local polynomials. This makes it possible to simultaneously fit the model and optimize the smoothing parameters (Breheny, 2018). There is currently no code available in R to calculate the IRR of semiparametric regressions; so, an evaluation of the coefficients is necessary. Therefore, the null hypotheses for the semi-parametric Poisson regressions is stated as $H_0: \beta = 0$.

Data sources

The reduced form model includes variables that are from the supply and demand sides of the market. The demand side factors are median household income, population, land area and the racial composition of neighborhoods (CBGs). Data on these factors were obtained from the 2012-2016 American Community Survey (ACS) 5-year estimates of the U.S. Census Bureau. Of the 858 CBGs in Milwaukee County, 846 were included; the other 12 CBGs lacked data for median household income and were omitted. The racial composition indicators were created by calculating the percentages of white and black residents in the population for each CBG. One with 60% or more white residents is indicated as predominantly white, one with 60% or more black residents as predominantly black, and all others as racially mixed as done by Raja, Ma & Yadav (2008). Income and area are scaled in large units to make results easier to interpret.

The supply outcomes of food retailer locations were obtained from the Reference USA database. Among other data, it provided each retailer's name and address, its longitude and latitude, up to six Standard Industrial Classification (SIC) codes and the number of employees. The database was searched using the same SIC codes used by Raja, Ma & Yadav (2008), who searched only the primary code. Since the share of food purchases from retailers with other primary activities has grown substantially in recent years, the search was expanded to include all SIC codes applied to a retailer not just the primary code as seen in Table 1 in the appendix. According to Progressive Grocer Staff (2018), for example, supercenters now account for more than 25% of retail grocery sales by value. Wal-Mart supercenters coded primarily as department stores, but secondarily as grocery stores, would be omitted if only primary codes were used.

Similarly, many gas stations and liquor stores function substantially as convenience stores, but that function would be missed if selecting only on the primary code. In these cases and several others, recognizing retailers as food sellers based on any SIC code matching the list is more appropriate than using only the primary code.

The SIC codes and number of employees were used to classify each food retailer as a: supermarket (grocery store with over 50 employees), grocery store (with under 50 employees), convenience store, variety store, specialty meat or produce market, or restaurant. The food retail categories in this study's descriptive summary and regression analysis differ slightly from those of Raja, Ma & Yadav (2008). There are so few specialty meat and produce markets in Milwaukee County that regressions on these categories were inevitably inconclusive and are omitted. Stand-alone candy stores and bakeries were even scarcer in Milwaukee County and omitted also from descriptive summaries. Other specialty stores were problematic, including vitamins and supplements rather than food itself, and were omitted. Using non-primary SIC codes, ethnic and kosher food retailers were included in the small grocery store category. A variety store category was added to include dollar stores and pharmacies coded primarily or non-primarily as variety stores. As mentioned in the literature review, these types of stores capture enough of the food retail market to warrant their inclusion. The data does not include community-supported agriculture/gardens, emergency food locations or institutional food venues like school cafeterias. The categories of primary importance in this study are: Supermarkets, Grocery stores, Convenience stores (which include variety stores), and Restaurants.

After matching the retailer data with geographic coordinates, CBGs and their demand-side factors in ACS data, geo-referenced street and road centerline data from the Milwaukee County Land Information Office (MCLIO) were added to the GIS. These were used to link retailers with CBGs in a network dataset for estimation of travel times between them, in most cases using default settings of ArcGIS Network Analyst, such as the global turn delay evaluator. The Wisconsin Department of Transportation (2009) documents its criteria for functional classification and related speed limits of streets and roads. Its specific codes for those classifications and the speed limits applied specifically to Milwaukee County streets and roads, however, have not yet been located. Hence, the functional classification interpretations of codes in the MCLIO database were inferred from State of Michigan and federal documentation to guide coding into the ArcGIS network dataset of travel speeds from Wisconsin DOT speed limits.

IV. Results

Results are presented as a comparison with results from Raja, Ma and Yadav (2008), in Erie County, New York. Milwaukee County and Erie County are similar in size and racial composition. They are both predominantly white but with a significant number of both predominantly black and racially-mixed neighborhoods. Erie County has 912 CBGs of which 897 were included in their study, with a racial composition distribution of: 16% (141) predominantly black CBGs, 9% (79) mixed-race CBGs and 75% (677) predominantly white CBGs. and Milwaukee County has 858 CBGs of which 846 were included in the study, with a racial composition distribution of: 26% (218) predominantly black CBGs, 16% (134) mixed-race CBGs and 58% (494) predominantly white CBGs. The main metropolitan area in Erie County is Buffalo, New York and the main metropolitan area in Milwaukee County is Milwaukee, Wisconsin. According to reports, Milwaukee is more racially segregated than Buffalo and, as expected, there is also more disparity in food access in Milwaukee County than in Erie County.

Gini coefficients & Lorenz curves

In Milwaukee County, there are 1,763 restaurants (see Table 2) with the Gini coefficient closest to 0 at 0.72 (see Table 3). Compared to Erie County, they have 1,685 restaurants with a Gini coefficient of 0.18, or fewer restaurants that are much more evenly distributed. The Gini coefficient for convenience stores is 0.80 and an exact match in both counties (even though Milwaukee County has 87 more locations). When convenience and variety stores, which were omitted from the Erie County analysis, are combined the distribution became more even at 0.76. This makes sense because combining these categories also increased the number of retail locations. Grocery stores have a Gini coefficient of 0.83 a near match to Erie County (even though Milwaukee County has 66 more locations). Supermarkets are the most unequally distributed with a Gini coefficient of 0.95 compared to 0.89 in Erie County. This means that in Milwaukee County 95% of CBG lack a supermarket. Lorenz curves (see Figure 1) give a visual illustration of the Gini coefficients explained above. Based on these results, the null hypothesis of Gini coefficient = 0 is rejected.

Poisson, Negative binomial, & Zero-inflated Poisson regression using Incidence rate ratios

Incidence rate ratio (IRR) point estimates in Table 4 compare predominantly black and racially-mixed CBGs with predominantly white CBGs. It also compares the IRRs calculated from 3 of the models – Poisson, Negative Binomial, and Zero-inflated Poisson regressions. The null hypothesis is $H_0: IRR = 1$ and "statistically significant" means rejecting the null with an IRR significantly different from 1. Following convention, the test is formally two-tailed, with an alternative hypothesis of IRR $\neq 1$, rather than specifically IRR> 1. We are careful to distinguish in discussion cases with IRR > 1 from those with IRR < 1. Convenience and variety stores are combined when doing the Poisson regressions (see Table 4). The results are almost identical with smaller standard errors when combining the two categories. The maximum acceptable significance level has been set at 5% and is indicated in Table 4 with * for a p-value of 5% to 1% and ** for a p-value of less than 1%. Results that are not marked are not statistically significant at the maximum level.

It is also important to note the sampling distributions of IRR point estimates from a Poisson distribution follow *not* normal distributions, but chi-square distributions, since they are ratios[9]. Following accepted practice in economic journal articles, rather than only reporting p-values with the coefficient point estimates, (robust) standard errors are included in the results table so that confidence intervals can be more easily and accurately constructed by the reader if desired, along with ranges of p-values marked with * as discussed above.

Supermarkets

The IRR point estimates of supermarkets follow the same pattern as in Erie County, but their standard errors are substantial, and they are not directly comparable statistically. When retailers are measured within CBG, and within a 5-minute drive, 5-minute bike ride, and 5-minute walk of the CBG centroids only the findings for predominantly black CBGs are statistically significant. Predominantly black CBGs have considerably <u>less</u> access to large supermarkets than predominantly white CBGs, all else equal. As expected, having access to a vehicle closes this gap; however, it does not eliminate it entirely. The IRR point estimates for the median household income variable produced puzzling results. Despite lack of significance at the 5% level for two of the four proximity measurements, the IRRs are significantly negative at the

[9] Raja, Ma & Yadav (2008) reported their (apparent) p-values as "chi-square values"

<1% significance level for both the driving and walking modes of transportation. We have presumed that large supermarkets are a normal good, and this is supported by the work of Warsaw, which makes these results unexpected.

As expected, results from the negative binomial regression are extremely close to the Poisson regression both in direction and significance. Results from the zero-inflated Poisson (ZIP) regression varied significantly relative to the number of data points that are equal to zero. The only statistically significant results are at the 5-minute drive distance and are a close match to those of the original regression.

Grocery Stores

The IRR point estimates of grocery stores do not follow the same pattern as in Erie County; however, the results for grocery stores are considerably higher than supermarkets which does match. When measuring the number of grocery stores within CBGs only the results for predominantly black CBGs are statistically significant; however, when measuring all three travel modes, all results are statistically significant at at-least the 5% level. Both predominantly black and racially mixed CBGs have access to considerably <u>more</u> grocery stores than predominantly white CBGs, all else equal. In some cases, they have access to more than twice as many retailers. Another pattern that emerges with grocery stores is that an increase of $10,000 in median household income results in a decrease of between .87 - .77 times the number of locations, which is statistically significant at a < 1% level in all four regressions. An increase in income that results in a decrease of supply could indicate that grocery stores are considered an inferior good. The results from the negative binomial and ZIP regressions follow the same pattern, with <u>more</u> access to grocery stores in predominantly black and racially mixed neighborhoods than predominantly white with statistically significant results < 1% for both the 5-minute driving and biking travel modes.

Fruit and meat markets

The results for fruit and meat are inconclusive. In replicating Raja, Ma and Yadav (2008) specialty retailers had their own categories. However, the number of locations for these types of retailers are so small that the only statistically significant results are for the 5-minute drive time and are heavily skewed to the locations where the few retailers are located. For example, there are only 22 meat markets in all of Milwaukee County and none of them are located within a 5-

minute walk of a racially mixed CBG centroid. Therefore, these retailers were included in the grocery store category and the separate specialty categories were eliminated.

Convenience & Variety Stores

As mentioned above, secondary SIC codes were used when creating the counts for retail locations by type. In the case of convenience stores, the decision to include retailers coded primarily as variety stores and secondarily as either grocery or convenience stores was also made. This is justified based on the reasoning that variety stores offer the same types of food items that convenience stores offer. Raja, Ma and Yadav (2008) excluded variety stores; however, based on the literature review, food shopping at variety stores, such as dollar stores, is common and on the rise.

There are significantly more convenience and variety stores in Milwaukee County (297 of which 58 are variety stores) than in Erie County (152); however, relatively fewer are located in predominantly black and racially mixed CBGs. For within-CBG food access, none of the finding are statistically significant at the 5% level and this is also true in Erie County. At the 5-minute drive and 5-minute bike ride times, results are statistically significant. Both predominantly black and racially mixed CBGs have <u>more</u> access to convenience and variety stores than predominantly white CBGs, all else equal. The negative binomial regression produced nearly identical results. The ZIP regression had more variation but followed the same pattern in both direction and significance. A change in median household income has the same effect on convenience stores as with grocery stores. A $10,000 increase to median household income results in between .91 and .73 times the number of convenience and variety stores locating in that area with statistical significance of < 1% in all four regressions.

Restaurants

There are also more restaurants in Milwaukee County (1,763) than in Erie County (1,685), and again relatively fewer are located in predominantly black and racially-mixed CBGs. Within the geographical boundaries of the CBG and at all three travel distances, predominantly black CBGs have <u>less</u> access to restaurants than predominantly white CBGs at statistically significant levels of at-least 5%, all else equal. Except for the 5-minute drive time, the same is true for racially mixed CBGs. This is true using all 3 statistical models. An increase of median household income leads to a decrease of .92 - .80 times the number of restaurants with statistical significance in all four regressions. There also seems to be a pattern with land area in the restaurant category. An increase in land area leads to an increase in restaurants of over 50% with statistical significance at every measurement *except* the 5-minute drive time. However, it is unclear exactly what this point estimate means. If CBGs are defined in a way that results in fairly similar populations, then land area may be inversely related with population density. How this relates to median income, not to mention other variables not considered like zoning laws, has also not been measured. So, this point estimate is not particularly meaningful to this study at this time.

Semiparametric Poisson regression

The code to calculate the IRR for a semiparametric Poisson regression is not currently available. Based on this limitation, the coefficients estimated by the semiparametric and traditional Poisson regressions will be compared (see Table 5) and the null hypotheses $H_0: \beta = 0$ will be tested. Coefficients for all retailer types, both within the geographical boundaries of CBGs and all 3 travel modes, in both predominantly black and racially mixed neighborhoods are very similar for the semiparametric and traditional Poisson regressions. The R code used to calculate the semiparametric regression does not provide the coefficient point estimate for the control variables that are modelled nonparametrically. Rather, a graph showing the conditional expection, all else equal, on the Y axis and the observed value of the variable on the X axis (see Figures 2-7). Therefore, only the statistical significance, not the numerical value, of the independent variables of the land area, median household income, and total population will be compared.

For supermarkets, the coefficients for predominantly black neighborhoods are negative and statistically significant at <5%, for all travel modes and within the geographical boundaries of CBGs. These results also indicate <u>fewer</u> supermarkets in predominantly black neighborhoods. For racially-mixed CBGs, the results for the travel mode of walking were significant but the coefficient was very small (-0.03). Median household income coefficient estimates are only significant when driving and walking.

For grocery stores, the coefficients for predominantly black and racially-mixed neighborhoods are positive and statistically significant for all travel modes and within the geographical boundaries of CBGs for predominantly black and racially-mixed neighborhoods. These results also indicate <u>more</u> grocery stores in predominantly black and racially-mixed neighborhoods than predominantly white neighborhoods. The median household income coefficient estimates are significant in all regressions and total population and land area coefficient estimates are significant for all three travel modes.

For convenience and variety stores, the coefficients for predominantly black and racially-mixed neighborhoods are positive and statistically significant for the travel modes of driving and biking. This is the same for both the semiparametric and traditional Poisson regressions. Coefficient estimates for land area and income are statistically significant for all regressions in this category.

For restaurants, the coefficients for predominantly black and racially-mixed neighborhoods are negative and statistically significant for all travel modes and within the geographical boundaries of CBGs. These results also indicate <u>fewer</u> restaurants in predominantly black neighborhoods and racially-mixed neighborhoods than in predominantly white neighborhoods. Coefficient estimates for land area, total population and median household income are statistically significant for all regressions in this category.

V. Discussion

The consistency of results across all four regression models provides a thorough robustness check. It is with great confidence that the conclusion is reached that based on these results, we must reject both H_0: IRR = 1 and $H_0: \beta = 0$ for all food retail categories.

However, several findings need additional clarity in order to fully understand them and their implications. There needs to be further data collected on the cost and variety of foods offered at retail locations throughout Milwaukee County, especially at small grocery, convenience and variety stores. Also, the data that is available on access to vehicles available in the American Community Survey needs to be added as a variable in the regression. Finally, other conditions that are linked to both income and race need further consideration. The variables that are used in the reduced form models were a good place to start, but additional research is still needed in several areas.

More data needs to be collected on what retailers are offering and at what prices. In both Milwaukee County and Erie County, predominantly black CBGs have considerably less access to large supermarkets and considerably more access to smaller grocery stores. This is only a problem if smaller grocery stores are *not* able to provide healthy food at reasonable prices like we assume that supermarkets do. Raja, Ma and Yadav suggest that different neighborhoods “specialize” in different types of food retailers. Stepping away from the framework that is typically used -- of large supermarkets located relatively far apart -- a walkable neighborhood where small retailers meet specialized demands for healthy and ethnically appropriate foods may be a more viable and successful option, both in terms of social justice and sustainability, in urban neighborhoods. However, there is not enough data on food retailers in Milwaukee County to determine if retailers in the small grocery store category should be classified as healthy or unhealthy food retailers.

The currently available research shows the following: SIC codes are inconsistent and at times inaccurate. The most recent price and availability study for Milwaukee (Jonson et al., 1996) was completed over 20 years ago and needs to be updated. Caldwell's (2011) study introduces the wholesale buyer's club as an important resource for healthy food in low-income areas. Gorski et al. (2018) finds that people who shop at large stores that sell more than grocery items tend to be more obese. Therefore, the presence/absence of a superstore/large supermarket cannot be used as a proxy for the health of a neighborhood. Further research needs to be done on both what is being sold and what is being bought at retailers regardless of their size or SIC codes.

This study has focused on race; however, the importance of income cannot be understated. An individual's income is linked to their ability to own and maintain a vehicle. Table 6 shows that with access to a vehicle shoppers can access an average of over 3 supermarkets and over 21 grocery stores, plus other specialty markets, in just 5 minutes. The more people in Milwaukee County who have access to a vehicle, the less that disparity in access is an issue. The publicly available data in the American Community Survey shows that on average about 40% of the total population has access to a vehicle to travel to work (see Table 7). It can be inferred that if individuals have a vehicle to travel to work, they can also travel to shop. However, simple summary statistics show deep disparities in vehicle access as well. A range of 6% to over 70% of the total population of a given CBG may have access to a vehicle. Adding this variable to a more sophisticated regression analysis may produce more nuanced results.

In addition to disparities in transportation, the disparity in median household income also needs attention. Glasmeier (2017) finds that in 2015 the living wage in Wisconsin for a household with 1 adult and 2 children is approximately $59,000. Only 12 predominantly black and 24 racially-mixed CBGs have median household income of at least $50,000 (see Table 8). This means that only 5.5% of predominantly black CBGs and 17.9% of racially-mixed CBGs have median household income near a living wage in Milwaukee County. At the same time, over

64% of predominantly white CBGs have median household income of at least $50,000. The income disparities between predominantly white CBGs and CBGs of color requires further investigation.

Finally, the differences in income and racial compositions in neighborhoods are likely to cause endogeneity in our model. Both race and income are linked to other conditions in neighborhoods that influence retailers' decisions on where to locate. These can include disparities in local infrastructure and/or municipal services, like police and fire protection. The models used in this study only control for difference in income and racial composition of neighborhoods and not the effects of those differences.

IV. Conclusion

The more quantitative empirical analysis that is completed on the topic of food access the better able we will be to determine what is typical. The results of the analysis in Milwaukee County, WI follow the same pattern as Erie County, NY; however, only 2 studies cannot be considered conclusive. Further research is needed in Milwaukee County to update the food pricing and availability data; without this data a conclusion on the food environment in this area cannot be reached. If further research confirms higher costs and less variety of healthy foods at smaller grocery stores and convenience and variety stores, then these results have troubling implications for public health in the context of a metropolitan area with high levels of racial segregation like Milwaukee. However, if smaller food retail locations can provide healthy and ethnically appropriate foods with shorter travel times at reasonable prices, then the greater number of locations would indicate the opposite of a food desert.

V. References

(2017). A look inside food deserts. *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention. Retrieved from www.cdc.gov/features/fooddeserts/ index.html.

Alcott, H., Diamond, R. & Dube, J-P. (2017). The geography of poverty and nutrition: Food deserts and food choices across the United States (Working Paper No. 3095779*). NBER*. doi:10.3386/w24094

Barrios, E.B. & Vera, E. (2011). Semiparametric Poisson regression model for clustered data. Retrieved from https://pdfs.semanticscholar.org/4a19/ac61b3e584de14145770197158 7af1beee3e.pdf

Bayatpour, A. J. (2016, August 17). Unwanted label: How did Milwaukee become known as "America's most segregated major city?" Retrieved from https://fox6now.com/2016/08/ 17/unwanted-label-how-did-milwaukee-become-known-as-americas-most-segregated-major-city/

Bitler, M., & Haider, S. J. (2011). An economic view of food deserts in the United States. *Journal of Policy Analysis and Management, 30(1),* 153-176. Retrieved from: http://onlinelibrary. wiley.com/journal/10.1002/%281SSN%291520-6688/issues

Bonanno, A. (2012). Food deserts: Demand, supply, and economic theory. *Choices, 27(3),* 1-4

Bonanno, A. and Li, J. (2015). Food insecurity and food access in US metropolitan areas. *Applied Economic Perspectives and Policy, 37(2),* 177-204

Breheny, P. (2018). Multiple regression and additive models. Retrieved from https://web.as. uky.edu/statistics/users/pbreheny/621/F12/notes/11-29.pdf

Caldwell, M.A. (2011). Finding hidden wells on food desert islands: Racial segregation, poverty concentration and provisional food options. *Department of Sociology, University of Wisconsin-Milwaukee*

Caspi, C.E., Pellerie, J.E., Harnack, L., Erickson, D.J., Laska, M.N. (2016). Differences in healthy food supply and stocking practices between small grocery stores, gas-marts, pharmacies and dollar stores. *Public Health Nutr., 19(3):* 540-547. doi:10.1017/S1368980015002724

Douangchai, V.L. (2011). Capturing Atlanta's food environment: A community level assessment of three disparate areas. *Georgia Institute of Technology.*

Durlauf, S. (2004). Neighborhood effects. Chapter 50 in *Handbook of Regional and Urban Economics*, vol. 4, pp. 2173-2242, eds. J. Henderson & J. Thisse. Amsterdam, Netherlands: Elsevier. Retrieved from https://www.ssc.wisc.edu/~sdurlauf/includes /pdf/Neighborhood%20effects.pdf.

Dutko, P. (2012). Food deserts suffer persistent socioeconomic disadvantage. *Choices, 27(3)*

Gateway to health communication & social marketing practice. (2017, September 15). Retrieved from: https://www.cdc.gov/healthcommunication/toolstemplates/entertainmented/tips/ FoodDesert.html

Glasmeier, A. K. (2017). Living wage calculation for Wisconsin. Retrieved from http://living wage.mit.edu/states/55

Gorski Findling, M. T., Wolfson, J. A., Rimm, E. B., Bleich, S. N. (2018). Differences in the neighborhood retail food environment and obesity among U.S. children and adolescents by SNAP participation. *Obesity Journal*, *26(6)*, 1063-1071. Retrieved from: https://doi.org/10.1002/oby.22184.

Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. (Working Paper No. EC-94-10). Retrieved from https://ssrn. com/abstract=1293115

Hatzenbuehler, P.L., Gillespie, J.M., and O'Neal, C.E. (2012). Does healthy food cost more in poor neighborhoods? An analysis of retail food cost and spatial competition. *Agricultural and Resource Economics Review 41(1),* 43-56.

Helling, Amy, and David S. Sawicki. (2003). Race and residential accessibility to shopping and services. *Housing Policy Debate, 14(2),* 69-101.

Hillier, A., Smith, T., Cannuscio, C. C., Karpyn, A., Glanz, K. (2015). A discrete choice approach to modeling food store access. *Environment and Planning B: Planning and Design, 42,* doi: 10,1068/b39136

Jacobsen, J. (2018). Investing in upgrades, expansions: Dollar, limited-assortment retailers poised for growth. *Beverage Industry, 109(2),* 30-31.

Johnson, K. et. al. (1996). Comparative study of food pricing and availability in Milwaukee. *UWM Center for Urban Initiatives and Research*

Kozlova, O. (2016). Is healthy food a luxury for the low-income households in the U.S.? (Working Paper). *Duke University*. Retrieved from: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=RESPhDConf2017&paper_id=163

Leonard, T., McKillop, C., Carson, J. A., & Shuval, K. (2014). Neighborhood effects on food consumption. *Journal of Behavioral and Experimental Economics, 51*, 99-113. doi: 10.1016/j.socec.2014.04.002

Moore, Latetia, & Diez Roux, A. V. (2006). Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health 96 (2):* 325-31.

Pettygrove, M. and Ghose, R. (2016). Mapping urban geographies of food and dietary health: A synthesized framework. *Geography Compass, 10/6:* 268-281. doi:10.1111/gec3.12268

Ploeg, M. V., & Rhone, A. (2017, December 5). Food Atlas Research Atlas: Documentation. Retrieved from https://www.ers.usda.gov/data-products/food-access-research-atlas /documentation/

Progressive Grocer Staff. (2018, April 15). Optimistic Outlook Isn't Stopping Grocers in Race to Finish Line. 85th Annual Report of the Grocery Industry. Retrieved from https://progressivegrocer.com/optimistic-outlook-isnt-stopping-grocers-race-finish-line-annual-report

Protect the harvest. (2015). A weapon in the fight against food deserts: tax credits for grocery stores. Retrieved from http://protecttheharvest.com/2015/12/21/a-weapon-in-the-fight-against-food-deserts-tax-credits-for-grocery-stores/

Raja, S., Ma, C. & Yadav, P. (2008). Beyond food deserts: Measuring and mapping racial disparities in neighborhood food environments. *Journal of Planning Education and Research, 27,* 469-482. doi: 10.1177/0739456X08317461

Satterfield, S. (2018). Behind the rise and fall of growing power. *Civil Eats.* Retrieved from: http://civileats.com/2012/03/13/behind-the-rise-and-fall-of-growing-power/

Verschay, R.J. (2009). A fresh fruit and vegetable outlet audit of Clark County, Wisconsin. *The Graduate School University of Wisconsin-Stout.*

Warsaw, P. (2018a). The implicit price of food access in an urban area: Evidence from Milwaukee property markets. *Center for Integrated Agricultural Systems, University of Wisconsin-Madison.*

Warsaw, P. (2018b). Estimating consumer preferences for food access in a residential sorting model of Milwaukee, WI. *Center for Integrated Agricultural Systems, University of Wisconsin-Madison.*

Weatherspoon, D. D., Oehmke, J. F., Coleman, M. A., Dembele, A., & Weatherspoon, L. J. (2012a). Will long term food desert consumers purchase fresh fruits and vegetables?. *Choices*, *27*(3).

Weatherspoon, D.D., Dembélé, A., Weatherspoon, L., Coleman, M., & Oehmke, J. (2012b). Price and expenditure elasticities for vegetables in an urban food desert. 2012 AAEA/EAAE Food Environment Symposium, Boston, MA. *Agricultural and Applied Economics Association*. Retrieved from: http://ideas.repec.org/p/ags/aaeafe/123392.html

Wisconsin Department of Transportation (2009). Chapter 13 Traffic Regulations, Section 5 Speed Limits. *Traffic Engineering, Operations & Safety Manual*. Retrieved from https://wisconsindot.gov/dtsdManuals/traffic-ops/manuals-and-standards/teops/13-05.pdf

VI. Appendix: Figures and Tables

Table 1

Food destination classification, Milwaukee County, Wisconsin

Food Destination	*SIC Codes*	*Definition*	*Example*
Supermarket	5411: 01, 02, 04 to 08	Grocery stores and food markets with more than 50 employees. Note: Destinations primarily coded as natural food, and specialty/ethnic markets (SIC codes 5499: 09 & 16) are included.	Pick N Save, Wal-Mart Supercenter
Grocery	5411: 01, 02, 04 to 08	All other grocers and food markets. Note: Destinations primarily coded as natural food, and specialty/ethnic markets (SIC codes 5499: 09 & 16) are included.	Viet Hoa Supermarket, Palmer Food Market LLC, Lloyd Food, All African Market
Convenience	5411: 03	Convenience stores. Note: Destinations primarily coded as liquor stores (SIC code 5921:02) and gas stations (SIC code 5541:01) with non-primary grocery or convenience sic codes are included.	Phil's Gas & Grocery Inc., Open Pantry
Variety	5311: 01	Variety stores. Note: Destinations primarily coded as pharmacies (SIC code 5912: 05) with non-primary grocery or variety SIC codes are included.	Dollar store, Walgreens
Meat and Fish	5421	Seafood, butchers, poultry	Amana Halal Meat LLC
Fruit and Vegetable	5431	Farmer's markets, vegetable markets	Dave's Fruit Stand
Restaurant	5812	Restaurants, cafes, and delis	McDonald's, Riverwalk Café

Note: SIC = Standard Industrial Classification

Table 2

Available food destinations, Milwaukee County, Wisconsin

Food Destination	County Total	% of Total	Number per 10,000 Population
Restaurants	1763	74.48%	18.45
Convenience & Variety stores	297	12.55%	3.11
Convenience stores	*239*	*10.10%*	*2.5*
Variety stores	*58*	*2.45%*	*0.61*
Grocery stores	212	8.96%	2.22
Supermarkets	59	2.49%	0.62
Meat & Fish stores	22	0.93%	0.23
Fruit & Vegetable markets	14	0.59%	0.15
Total	2367	100.00%	24.78
Total Population	955,306		

Table 3

Gini coefficients of food destinations, Milwaukee County, Wisconsin

Type of retailer	Gini Coefficient	
	0	Perfectly equal distribution
Restaurants	0.72	
Convenience & Variety stores	0.76	
Convenience stores	0.80	
Grocery stores	0.83	
Variety stores	0.94	
Supermarkets	0.95	
	1	Perfectly unequal distribution

Figure 1

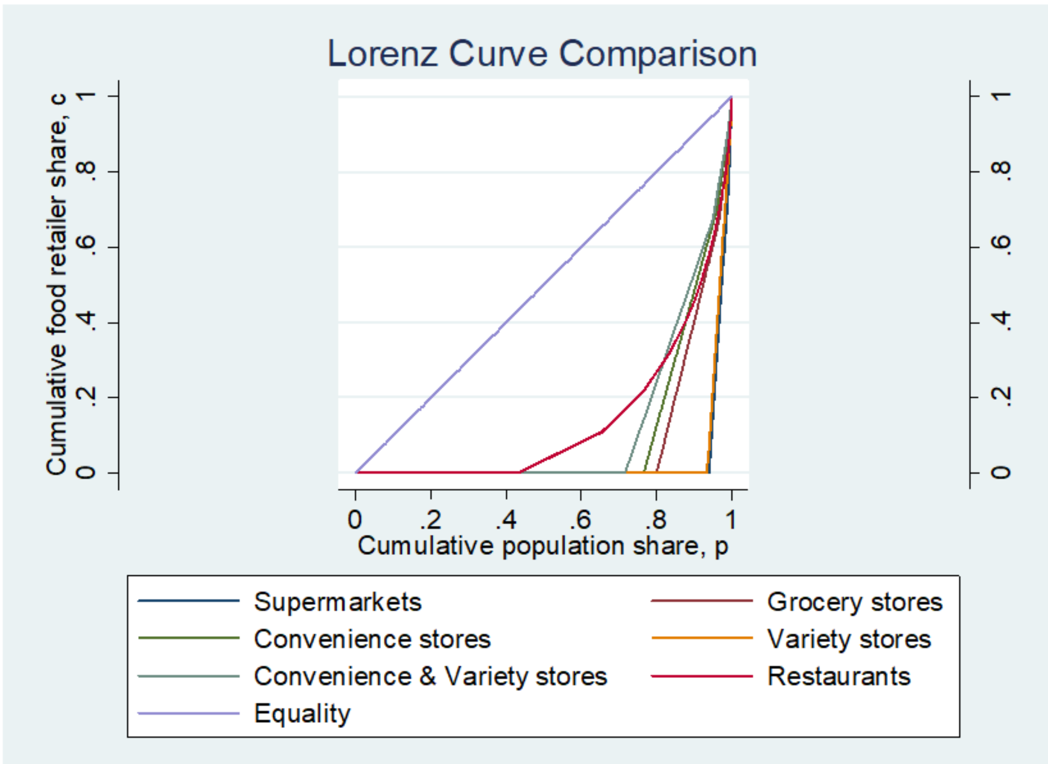
Lorenz Curve Comparison
Cumulative food retailer share, c
0
.2
.4
.6
.8
1
Cumulative population share, p
Supermarkets
Grocery stores
Convenience stores
Variety stores
Convenience & Variety stores
Restaurants
Equality

Table 4

Incidence Rate Ratio Point Estimates

Relative to Predominantly White CBGs

		Poisson Regression			Negative Binomial Regression		Zero Inflated Poisson Regression		
		Predominantly	Racially	Goodness-of-fit	Predominantly	Racially	Predominantly	Racially	
Retailer	**Travel mode**	**Black CBGs**	**Mixed CBGs**	**Result**	**Black CBGs**	**Mixed CBGs**	**Black CBGs**	**Mixed CBGs**	**# data = 0**
Supermarket		0.25*	0.95	pass	0.25*	0.97	1.16	0.98	
		(0.1638)	(0.3650)		(0.1425)	(0.3865)	(0.2618)	(0.1242)	794
	Drive	0.73**	1.05		0.73**	1.05	0.77**	1.10*	
		(0.0344)	(0.0493)	pass	(0.0371)	(0.0537)	(0.0346)	(0.0507)	31
	Bike	0.54**	0.93		0.34**	0.90	0.87	0.82**	
		(0.0805)	(0.1508)	pass	(0.0735)	(0.1678)	(0.0744)	(0.0617)	628
	Walk	0.13*	0.81		0.12*	0.75	0.74	1.0	
		(0.1104)	(0.4405)	pass	(0.1055)	(0.4988)	(0.2074)	(0.1928)	828
Grocery Store		1.61*	1.54		1.65*	1.48	1.12	1.02	
		(0.3228)	(0.3258)	pass	(0.3420)	(0.3325)	(1.023)	(0.1029)	678
	Drive	1.92**	1.65**		2.0**	1.76**	1.92**	1.66**	
		(0.1173)	(0.1109)	fail	(0.1437)	(0.1393)	(0.1175)	(0.1118)	22
	Bike	1.99**	2.01**		2.06**	2.05**	1.56**	1.65**	
		(0.1970)	(0.0758)	fail	(0.1906)	(0.2023)	(0.1206)	(0.1377)	313
	Walk	2.09**	1.76*		2.08**	1.77*	1.12	1.18	
		(0.5411)	(0.5000)	pass	(0.5274)	(0.4718)	(0.1263)	(0.1380)	725
Convenience Store		0.83	0.96		0.82	0.96	0.96	1.11	
		(0.1315)	(0.1881)	pass	(0.1413)	(0.1724)	(0.660)	(1.227)	608
	Drive	1.31**	1.28**		1.28**	1.25**	1.32**	1.28**	
		(0.491)	(0.0509)	fail	(0.0539)	(0.0582)	(0.0494)	(0.0513)	4
	Bike	1.13*	1.33**		1.12	1.31**	1.14*	1.37**	
		(0.0706)	(0.0971)	fail	(0.0743)	(0.0928)	(0.0656)	(0.0930)	155
	Walk	0.79	1.70*		0.80	1.71*	0.88	1.10	
		(0.2121)	(0.4637)	pass	(0.1990)	(0.4110)	(0.1058)	(0.1288)	722
Restaurant		0.32**	0.48**		0.33**	0.49**	0.51**	0.56**	
		(0.0571)	(0.0733)	fail	(0.0476)	(0.0783)	(0.0746)	(0.0730)	367
	Drive	0.65**	1.07		0.67**	1.06	0.65**	1.07	
		(0.0456)	(0.0681)	fail	(0.0407)	(0.0727)	(0.0456)	(0.0681)	0
	Bike	0.39**	0.70**		0.41**	0.71**	0.41**	0.74**	
		(0.0476)	(0.0697)	fail	(0.0357)	(0.0714)	(0.0503)	(0.0750)	55
	Walk	0.21**	0.53**		0.21**	0.51**	0.38**	0.59**	
		(0.450)	(0.1012)	fail	(0.0464)	(0.5918)	(0.0587)	(0.0930)	576
						Robust Standard Errors in Parentheses			
						**P < 0.00, *P < 0.05 for IRR = 1			

Table 5					
Parametric Coefficient Point Estimates For the Independent Variable of the Racial Composition of CBGs					
Retailer Type	**Travel Mode**	**Predominantly Black CBG**		**Racially-Mixed CBG**	
		Poisson Regression	Semi-Parametric Poisson Regression	Poisson Regression	Semi-Parametric Poisson Regression
Supermarket		-1.40* (0.6673)	-1.23* (0.5687)	-0.05 (0.3836)	0.17 (0.5180)
	Drive	-0.31** (0.0470)	-0.34** (0.0537)	0.05 (0.0467)	-0.00 (0.0536)
	Bike	-1.06** (0.2318)	-1.11** (0.2103)	-0.77 (0.1628)	-0.17 (0.1726)
	Walk	-2.03* (0.8409)	-1.90* (0.8318)	-.21 (0.5424)	-0.03* (0.5799)
Grocery Store		0.47* (0.2009)	0.45* (0.1948)	0.37 (0.2241)	0.34 (0.2101)
	Drive	0.65** (0.0612)	0.65** (0.0225)	0.50** (0.0671)	0.44** (0.0244)
	Bike	0.70** (0.0922)	0.74** (0.0653)	0.69** (0.0992)	0.57** (0.0691)
	Walk	0.74** (0.2584)	0.74** (0.2384)	0.56* (0.2844)	0.54* (0.2479)
Convenience Store		-0.19 (0.1590)	-0.12 (0.1702)	-0.04 (0.1954)	0.05 (0.1748)
	Drive	0.27** (0.0374)	0.31** (0.0199)	0.25** (0.0398)	0.25** (0.0216)
	Bike	0.13* (0.0622)	0.16** (0.0599)	0.29** (0.0728)	0.25** (0.0626)
	Walk	-0.24 (0.2683)	-0.24 (0.2370)	0.53* (0.2721)	0.40 (0.2261)
Restaurant		-1.15** (0.1801)	-1.04** (0.0880)	-0.73** (0.1529)	-0.58** (0.0835)
	Drive	-0.43** (0.7000)	-0.48** (0.0090)	0.07 (0.0634)	-0.03** (0.0088)
	Bike	-0.93** (0.1205)	-0.96** (0.0304)	-0.35** (0.0994)	-0.48** (0.0288)
	Walk	-1.54** (0.2107)	-1.64** (0.1159)	-0.64** (0.1919)	-0.85** (0.0986)
			Standard Errors in Parentheses **P < 0.00, * P < 0.05 for β = 0		

Figure 2
Land area
when Supermarkets are accessible
by a 5-minute drive

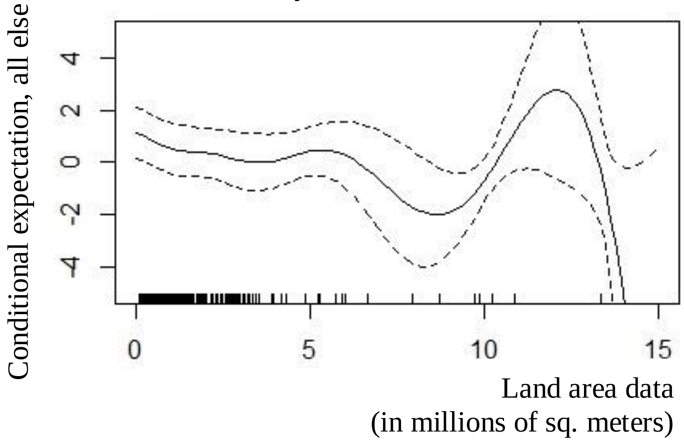

Figure 3
Total population
when Supermarkets are accessible
by a 5-minute drive

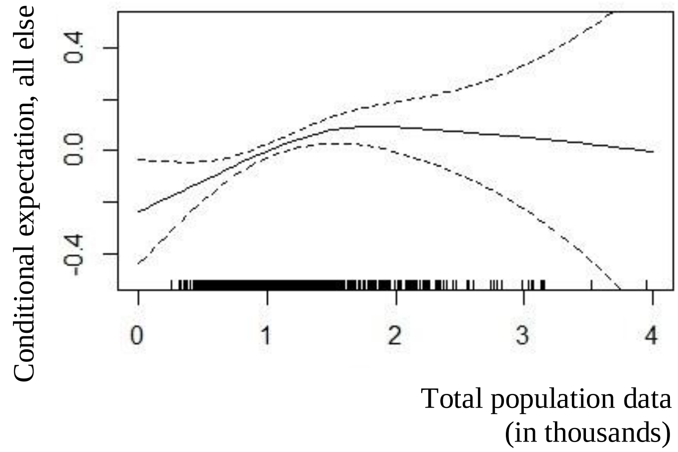

Figure 4
Median household income
when Supermarkets are accessible
by a 5-minute drive

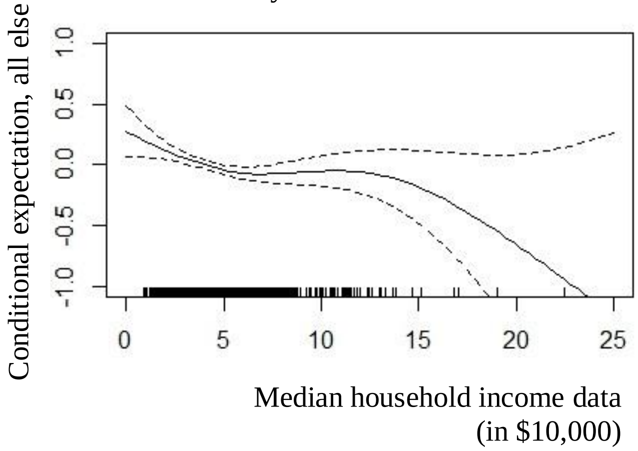

Figure 5
Land area
when Grocery stores
within the geographical boundaries of CBGs

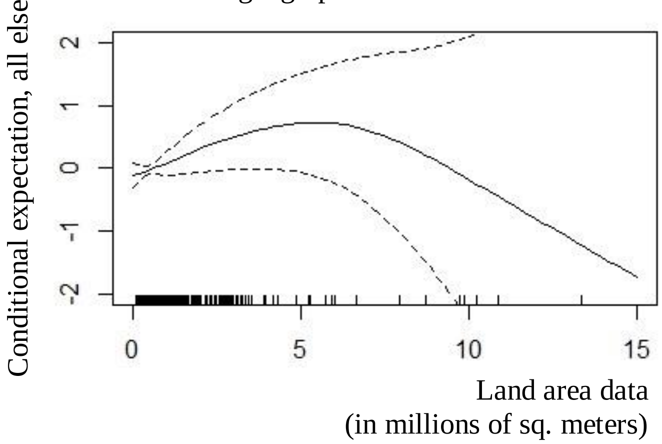

Figure 6
Total population
when Grocery stores
within the geographical boundaries of CBGs

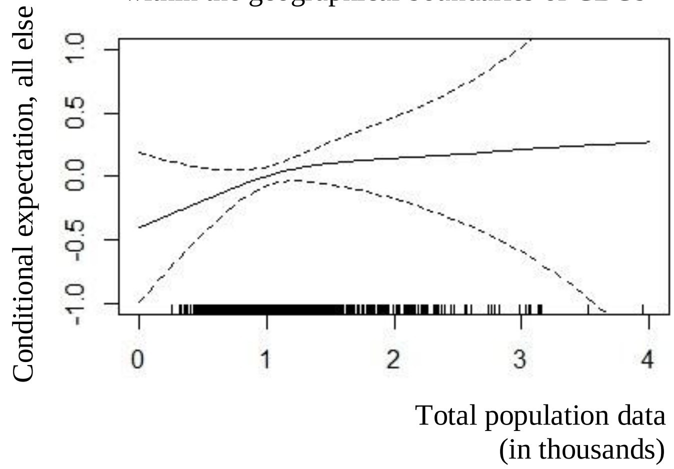

Figure 7
Median household income
when Grocery stores
within the geographical boundaries of CBGs

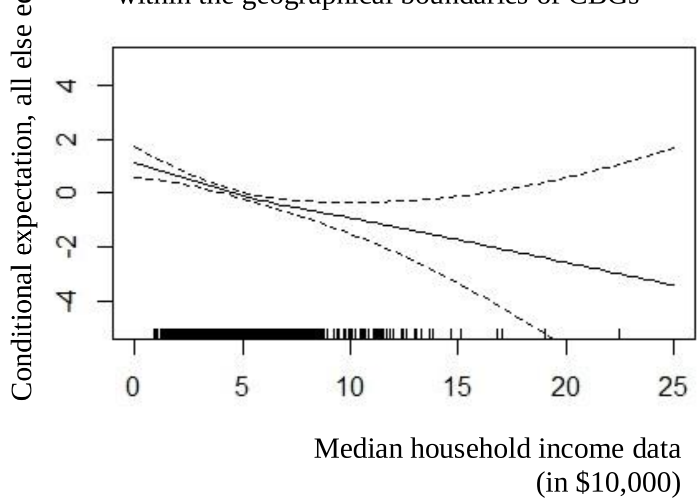

Table 6				
Average Number of Retail Locations in Milwaukee County Census Block Group:				
	Within 5-minute travel time of CBG centroid by:			
Type of Retail Location	*geographical boundaries*	*Walking*	*Biking*	*Driving*
Supermarkets	0.07	0.02	0.35	3.84
Grocery Stores	0.25	0.19	2.59	21.6
Convenience & Variety Stores	0.34	0.18	2.81	24.88
Meat & Fish Markets	0.02	0.01	0.21	2.13
Fruit & Vegetable Markets	0.01	0.01	0.15	1.4
Restaurants	2.05	1.07	13.83	137.88

Table 7		
Means of Transportation to Work		
Car, Truck, or Van		
as a percent of total CBG population		
Mean	39.0%	
Median	40.5%	
Minimum	6.4%	
Maximum	76.0%	
Standard Error	0.0043	
Count	857	CBGs

Table 8

Number of Census Block Groups in Milwaukee County

Divided by Median Household Income

Median Household Income	*Predominantly White*	*Predominantly Black*	*Racially-Mixed*
$0 - $19,867	6	39	21
	1.21%	17.89%	15.67%
$19,868 - $38,162	75	135	68
	15.18%	61.93%	50.75%
$38,163 - $64,417	241	41	38
	48.79%	18.81%	28.36%
$64,418 - $108,039	145	3	7
	29.35%	1.37%	5.22%
$108,040 +	27	0	0
	5.47%	0%	0%
Total	494	218	134
	100%	100%	100%