8-2012

Food Environment and Childhood Obesity

Yucong Jiao
University of Arkansas, Fayetteville

FOOD ENVIRONMENT AND CHILDHOOD OBESITY

FOOD ENVIRONMENT AND CHILDHOOD OBESITY

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Agricultural Economics

By

Yucong Jiao
University of East China Normal University
Bachelor of Science in Economics, 2010

August 2012
University of Arkansas

ABSTRACT

This thesis examines whether children's food environment, especially food stores that have fresh produce, affects obesity prevalence among elementary school children in the state of Arkansas. Misclassified food outlet types in the Dun and Bradstreet commercial data set were first corrected and then food environment measures were computed and aggregated to geographic regions corresponding to school attendance areas. After applying classical panel estimation, it was found that the fixed effects model fit the data best. Results indicate that an additional supermarket within a one-mile radial of the census neighborhood block center will bring down childhood obesity prevalence by 0.58 percent, whereas associations between densities of supermarkets within farther buffers and children's overweight status were not found. In addition, distance from neighborhood block center to closest supermarkets did not seem to play a role in determining children's BMI, nor did presence of dollar, convenience and drug stores. Finally, fixed effects models incorporating spatial lags and spatial errors were estimated. Results showed no significant spatial effects.

This thesis is approved for recommendation
to the Graduate Council.

Thesis Director:

____
Dr. Michael Thomsen

Thesis Committee:

____
Dr. Rodolfo M. Nayga, Jr.

____
Dr. Bruce L. Dixon

____
Dr. Michael Thomsen

THESIS DUPLICATION RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed ______________________________________

Yucong Jiao

Refused ______________________________________

ACKNOWLEDGMENTS

Special thanks are due to my advisor Dr. Thomsen who has patiently guided me through the thesis writing process and spent time discussing each part of the thesis with me. It would be impossible to finish this thesis without his generous help. Also, special thanks go out to Dr. Nayga and Dr. Dixon for their valuable suggestions and constant help on improving the thesis. I would also like to thank Diana Danforth, Jim Smartt, Pete Alviola, Zhongyi Wang, Jebaraj Asirvatham and other members of the childhood obesity project for making this wonderful research opportunity happen.

I am most grateful to all the faculty, staff and students in the Department of Agricultural Economics and Agribusiness, University of Arkansas, Fayetteville, for they have created a stimulating and welcoming academic and social environment to help me focus on the thesis as well as enjoy the learning process.

Finally, I appreciate the financial support that helped make this research possible. This was provided by the Arkansas Agricultural Experiment Station and by the Agriculture and Food Research Initiative Competitive Grant no. 2011-68001-30014 from the USDA National Institute of Food and Agriculture.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

During the last two decades, the United States has experienced a dramatic increase of childhood obesity. The prevalence of obesity among children aged 6 to 11 years increased from 6.5% in 1980 to 19.6% in 2008, and the prevalence of obesity among adolescents aged 12 to 19 years increased from 5.0% to 18.1% (Ogden and Carroll 2010).

Childhood obesity can have many adverse impacts on health outcomes. A consensus has been reached that childhood obesity has a positive correlation with cardiovascular disease risk (Freedman et al. 2007; Ingelsson et al. 2007; Baker, Olsen, and Sorensen 2007). It has also been shown that obese children are at approximately a 3-fold higher risk for hypertension than normal weight children (Sorof and Daniels 2002). In addition, children who are overweight or obese are more likely to remain so as adults (Freedman et al. 2005).

These negative health effects have increasingly attracted public attention and motivated research into better understanding the causes of childhood obesity. Although many factors could contribute to weight gain, it is undeniable that what one eats should play an important role. This is the focus of this thesis. Specifically, I address the following research question: Does the food environment contribute to childhood obesity? One hypothesis is that children who live in neighborhoods that have low access to large grocery stores might have higher BMI than children from communities that have ample access to fresh fruits and vegetables. The topic has received heated discussion and it is important in directing policy. Policy makers in several major cities have started to employ zoning laws to regulate the availability of fast food restaurants (Sturm and Cohen 2009; Abdollah 2007) due to obesity risk concerns. To date, the empirical findings seeking to assess the relationship between food environment and obesity have not shown consistent results and so offer no clear direction guiding policy.

This thesis examines children's food environment –the availability of food stores and restaurants of various types within school neighborhoods and the impact of this environment on obesity prevalence among elementary school children in the state of Arkansas. Drawing from a dataset of geo-coded food establishments from Dun and Bradstreet (D&B), I first define different types of stores and restaurants according to my research question. Then using a geographical information system (Arc-GIS) and statistical software (SAS and R), I was able to construct two types of school-level measures, both based on the census blocks of residents surrounding elementary schools. One is the distance of the census block center to closest food outlets of different types. The other is the density of food establishments within given radial distances of the census block centers. Combining school level obesity prevalence from the Arkansas Center for Health Improvement (ACHI) and school-level measures from the Arkansas Department of Education (ADE) with food outlets measures, I ended up with a six year balanced panel of 234 schools. A pooled OLS model, random effects model and fixed effects model were then estimated and compared. In addition, motivated by recent research findings that people living in nearby neighborhoods might share similar health outcome due to similar food environment (Chen et al. 2010), a spatial panel model was also employed.

This thesis contributes to the obesity literature in the following ways. First, I corrected misclassified food outlet types in D&B. Unlike previous studies which applied assignment directly from D&B, I divided food stores into four major types: large grocery stores which provide fresh fruits and vegetables; convenience store or small grocery stores; dollar stores; and drug stores. Restaurants were also assigned to four types based on their format and menu content: full-service restaurants; fast food restaurants; sandwich places; and pizza places. This data preparation process involved manually checking establishments through Internet search engines and by telephone calls to certain food stores.

Second, the unique way that food environment measures were computed offers more accurate proxy than previous studies. In the literature, most of the studies computed density measures within a neighborhood by counting number of food establishments. The approach I use here differs in that it accounts for the possibility that residents living on the border of a neighborhood have access to stores in nearby neighborhoods. Specifically, two types of measures were constructed. Their first is distance to closest food outlets. The second is density of food outlets within one mile, two miles and five miles. Both of the measures were first computed on the census block level (a census block is relatively small and treated as a proxy of residential address since the average block has a population of 15 people). Afterwards, the block-level measures and were then aggregated up to a school neighborhood average weighted by block population. Thus, the measure conforms to a school boundary but reflects establishments outside of the school boundary which are accessible to residents within the boundary. It is assumed that the impact of food stores over childhood obesity can either come in the form of density within certain radial distances around one's home or the distance to the closest food stores. The reason why I selected school neighborhood instead of commonly used census tract as a food consumption area is data driven. The prevalence of obesity and characteristics of schoolchildren are reported at the school level. Some studies in the literature have also used certain buffer zones around school as a proxy for children's food consumption area, and like mentioned before, this measure might not accurately represent food purchase opportunities for people living on the border of the buffer.

Third, the model employed in the thesis considers spatial dependence across school neighborhoods and unobserved school effects. It is assumed that the dataset of school BMI and food environment measures are spatial in nature because of agglomeration along socioeconomic status and shared food environments. This hypothesis is motivated by recent research findings which suggested that people with similar BMI tend to cluster together (Christakis and Fowler 2007; Mobley et al. 2004; Eid et al. 2008; Chen and Florax 2010) and that the location of food retailers follows certain geographical patterns in which the low income minority neighborhoods seem to be in a disadvantaged status (Donkin et al. 1999; Morland et al. 2002; Moore and Roux 2006; Frank et al. 2006; Zenk et al. 2005). To account for these spatial effects, which could potentially exist in the dependent variable or in the explanatory variables or both, a spatial fixed effects model was employed using the open-source statistical software R. This approach provides some evidence as to whether children in nearby school neighborhoods share more similar BMI outcomes as opposed to faraway neighborhoods after controlling for the schools' unobserved specific effects.

Finally, dollar stores have been defined as a separate type of food store in this study. Dollar stores in my data as shown later in chapter 3 take up about 9 percent of total food stores in Arkansas and represent a major feature in the food environment. In addition, it is suspected that dollar stores might potentially be a major source of calories for lower income rural residents. Thus dollar stores are included separately and their presence, especially in rural areas, is expected to be associated with higher childhood obesity rates.

The aim of this thesis is to offer some policy guidance as to whether government, in order to curb the trend of childhood obesity, should encourage new openings of supermarkets in communities that have low access to fresh produce. In addition, I am curious to learn whether other types of food outlets play a role in determining children's weight so that more comprehensive policy guidance can be provided about food landscapes. The rest of the thesis is organized as follows. Chapter 2 covers a review of existing literature. Chapter 3 introduces the data and describes how food environment measures were computed. Chapter 4 presents the theoretical model. Chapter 5 shows results and Chapter 6 discusses conclusion, some limitations of the thesis and policy implications.

CHAPTER 2: EXISTING LITERATURE

There is a large empirical literature studying the impact of food outlets on obesity. Previous research has generally focused on the proximity of supermarkets vs. convenience stores and fast food outlets vs. full-service restaurants based on the premise that supermarkets and full-service restaurants typically offer healthier foods than convenience stores and fast food outlets (Sallis et al. 1986). Thus, usually a negative relationship between access to supermarkets and the obesity rate is expected, whereas access to fast foods is assumed to be associated with a higher prevalence of obesity. However, prior findings have offered confounding results. Many studies have indeed indicated such relationship between proximity to food outlets and BMI (Currie et al. 2009; Alviola et al. 2011; Brennan and Carpenter 2009; Powell et al. 2007; Maddock 2004; Mehta and Chang 2008; Morland, Diez Roux, and Wing 2006; Rundle et el. 2009; Chen et al. 2010), whereas a few papers have found no such association (Harris et al. 2011; Strum and Datar 2005; Lee 2012; Powell 2009). In addition, several studies reviewed have shown mixed or unexpected results (Lopez 2007; Wang et al. 2007; Zick et al. 2009). Below is a review of literature in terms of food environment measures, data sources and methods.

Food Environment Measures Review

Food Environment Studies Designed for Children and Adolescents

In the obesity literature, a number of studies have focused on childhood and adolescent obesity (Currie et al. 2009; Alviola et al. 2011; Brennan and Carpenter 2009; Harris et al. 2011; Powell et al. 2007; Strum and Datar 2005; Lee 2011). Among those studies, the majority identified buffer zones around school or used school zip-code as an adequate neighborhood for children's food consumption environment. For example, Currie et al. (2009) in their California public school study, observed obesity rate for 9th graders over several years and found that a fast food establishment within a tenth of a mile around a school was associated with approximately 5.2 percent increase in obesity rates. Alviola et al. (2011) have found similar results using a sample of Arkansas public school children. Specifically, they find that the number of fast food restaurants within a quarter mile of a middle or high school is positively related to childhood obesity rates. Brennan and Carpenter (2009) also used a one half mile buffer around schools as food exposure area in their California Healthy Kids Survey and came up with the conclusion that students with fast food outlets around their schools within one half mile consume less fresh fruits and vegetables whereas they consume more soda and are more likely to be overweight. Harris et al. (2011) have computed food establishments within 2 kilometers of high schools in Maine and found that there were no significant relationships between the BMI and density of food stores around schools. Powell et al. (2007) drawing repeated cross sections of individual-level data on adolescents from Monitoring the Future (MTF) surveys which provided food store availability at the school zip-code level, found that the increased presence of chain supermarkets was significantly associated with lower adolescent BMI and that greater availability of convenience stores was associated with higher BMI. In Powell (2009)'s paper, she used county level measures to perform the analysis. With improvements in geographic information systems, many

researchers have started to characterize a child's market for food in a more refined manner. For example, Strum and Datar (2005) count fast food and food stores using children's home zip code. Combining food price as explanatory variable, they found that lower real prices for vegetables and fruits could predict a significantly lower gain in BMI between kindergarten and third grade. However, after controlling for individual characteristics no significant effect was found for dairy or fast-food prices nor was a significant effect found for outlet density. Another study by Lee (2012) defined children's food purchase area as individual's census tract. She constructed three measures of local food availability: the number of stores or restaurants types per 1000 population; the number per square mile and the shares of each establishment type out of all food outlets.

After applying a multi-level modeling procedure, she came to the conclusion that food outlets do not independently explain weight gain over time in this sample of elementary school-aged children.

Food Environment Studies Designed for Adults

Research interested in adult health outcome usually focus on home neighborhood instead of school (Maddock 2004; Mehta and Chang 2008; Lopez 2007; Morland, Diez Roux, and Wing 2006; Wang et al. 2007; Zick et al. 2009; Rundle et el. 2008; Chen et al. 2010) So far there has been no consensus reached as how to capture an individual's neighborhood for food purchase and therefore geographic scales for food environment measures vary widely across studies. For example, using samples from the Behavioral Risk Factor Surveillance System (BRFSS), Maddock (2004) examined state-level aggregated means for square miles per fast food restaurants and population per fast food restaurant and used them as measures for food environment. He revealed a correlation between both measures with state-level obesity prevalence. In another study, Mehta and Chang (2008) drew samples from BRFSS but used five-year Individual-level data. Numbers of fast food establishment per 10,000 individuals were calculated on the county level. Their results indicated a significant association between fast food density and increased BMI. In another BRFSS study, Lopez (2007) incorporated zip-code level variables into the analysis and unexpectedly found that the presence of supermarkets or large grocery stores is positively associated with obesity risk while the presence of fast food restaurants is not. Morland, Diez Roux, and Wing (2006) considered the census tract as an individual's neighborhood for food access in their Atherosclerosis Risk in Communities (ARIC) study. They find that the availability of supermarkets was associated with a lower prevalence of obesity and overweight and that the availability of convenience stores was associated with a higher prevalence. Wang et al. (2007) defined neighborhood by a combination of <u>census tract and block groups.</u> Store proximity and count of stores per square mile were both used in the multi-level modeling procedure. They found that higher density of small grocery stores, like convenience stores, is associated with higher BMI. Their result about supermarkets, however, is a bit intriguing. They expected that living closer to chain supermarkets would decrease people's risk of being obese. On the contrary, they found women living closer to chain supermarkets are actually at higher risk of being obese. Another study, also on <u>census block groups</u>, by Zick et al. (2009) have incorporated city walkability measures and showed that for individuals living in non-low income neighborhoods, having one or more convenience stores, full-service restaurants, or fast food restaurants is associated with reduced obesity risk, and the presence of at least one healthy grocery option in low income neighborhoods is also associated with a reduction in obesity risk. Taking the neighborhood measures at a more detailed level, Rundle et al. (2008), in their New York City study, defined an individual's market for food as a <u>half-mile buffer</u> around the individual's home address. Their study showed that density of food outlets that have fresh produce was inversely correlated with BMI while density of less "healthy" stores such as convenience stores was not significantly associated with BMI, after controlling for city's walkability. Chen et al. (2010) in their Marion County, Indiana study treated the characteristics of one's nei ghbors as his or her food community. They found that increased access to chain grocers in low-income communities decreased the average BMI by approximately 0.3 BMI points.

Data Sources Review

Data sources of food outlets, which play a crucial role in determining the accuracy of results, also vary across studies. Among the literature examined, Harris et al. (2011) conducted mail surveys to determine height, weight, and calorie-dense food consumption for 552 students at 11 Maine high schools. The food stores within 2 km of each school were then visited to record the type. Such surveys are the most valid way to obtain food landscape data (Galvez et al. 2007), but are impractical when there is a need to apply findings to a large scale or when the period of study spans multiple years. A few researchers collected data from government or public directories such as state department of health lists , company yellow page or website listings (Morland, Diez Roux, and Wing 2006; Wang et al. 2007; Chen et al. 2010; Maddock 2004), which have been criticized for the lack of detailed information on the scale of operations and slow update schedules (Lee 2012). Some analyses draw data from government business patterns like Zip Code Business Patterns (Strum and Datar 2005) and County Business Patterns (Lopez, 2007). The limitation of these data sources is that they usually offer a pattern too general for food landscape measures and are hard to match to smaller neighborhoods. A majority of studies in the literature obtained food outlet data from commercial databases, such as D&B (Currie et al. 2009; Alviola et al. 2011; Powell et al. 2007; Zick et al. 2009; Rundle et al. 2008; Lee 2012), InfoUSA (Black et al. 2010; Li et al. 2009; Babey et al. 2008), Microsoft Streets and Trips (Brennan and Carpenter 2009), and The Economic Census (Mehta and Chang 2008). There are, to my knowledge, no papers in the field to have compared accuracy of all these commercial data sets. Several studies however have been conducted specifically to examine the validity of two most commonly used commercial data sets – D&B and InfoUSA. D&B is a public company that licenses information on businesses and corporations for use in credit decisions. InfoUSA is a provider of end-to-end marketing solutions for small-and medium-size businesses. It offers business mailing lists to researchers. Both commercial data sets used 4 digit, 6 digit and eight-digit Standard Industry Classification (SIC) codes to identify different types of food outlets. In a ground-truthing study by Liese et al. (2010), 77.7 percent of all food outlets from D &B address list were identified as " located and open" with the rest either "closed", "not found" or "PO address". Furthermore, about two thirds as many additional stores as were listed in D&B were found in Liese's field trips. The accuracy looks better for InfoUSA, for about 89.2 percent were "located and open" and about one half as the number of listed additional stores was "found but not listed". Another comparison study of the two data source by Powell et al. (2011) has also suggested that overall agreement between outlets listed and on the ground is higher in InfoUSA than D&B but Powell et al. 's study found a lower rate (D&B 49% and InfoUSA 64%)than Liese's study. Both studies have shown that despite the massive application of large commercial data such as D&B and InfoUSA, a validity issue might exist in these data sets. Since the majority of empirical studies has employed and will continue to rely on these commercial data sets, a systematical way to match and clean up the secondary data is required.

Methods Review

The previous literature has employed different empirical models to explain the impact of food outlets on health outcomes, among which multilevel regressions are common due to the different levels of BMI and neighborhood measures.

Many of the examined papers used cross-sectional analysis. For example, Brennan and Carpenter (2009), Zick et al. (2009) estimated a cross-sectional ordinary least squares model (OLS) for BMI and a logistic regression for dichotomous overweight and obesity outcomes. Similarly, Harrison et al. (2011) employed cross-sectional logistic models for their overweight/obese indicator in the Maine high school student study. Maddock (2004) presented a cross-sectional multilevel hierarchal analysis on a state-wide basis. Following the same strategy, Rundle et al. (2008) also performed a cross-sectional multi-level estimation but also included a city walkability index. Morland, Diez Roux, and Wing (2006) performed binomial regressions with a random intercept for each census tract to estimate prevalence ratios (PRs) of risk to be overweight associated with the presence of different types of food stores. Remaining studies in the literature reviewed above performed panel estimation. For instance, Strum and Datar (2005) used multi-level models with school random effects to explain the BMI change between first/third and kindergarten. Lee (2012) estimated cross-classified random-effects models (CCREM), which is a special case of multi-level modeling for data that are not purely hierarchical or nested. The benefits of this approach were described as follows:

> "This framework allows for a systematic analysis of how characteristics conceptualized and measured at various levels of non-nested structures (i.e., children in the same school may live in different neighborhoods, and vice versa) affect child weight gain over time."and " These models rely on an assumption that the random effects across different levels and the random effects across different groupings in the same level are uncorrelated.(Lee 2012 p. 1197)."

The county level analysis conducted by Mehta and Chang (2008) also employed a two-level hierarchic panel model. The first level analysis was focused on the individual health outcome with a random effects intercept, and second level regression examined logistic models with census region dummy variables to account for unmeasured regional characteristics. Lopez (2007) estimated similar models to Mehta and Chang, but was modeling the second level regression using store measures from ZCTAs. In another paper that involved individuals nested within neighborhood, Wang et al. (2007) used SAS's MIXED procedure to conduct iterative maximum likelihood estimation. Currie et al. (2009) estimated a fixed affects model while controlling for individual and neighborhood characteristics. In Alviola's 2011 paper, an instrumental variable approach was employed with fast-food restaurant proximity being instrumented by proportion of the population within the 15 to 24 year-old age group and nearness of the school to a major highway. Powell et al. (2007) used an OLS model with dummy variables for the years in their seven year MTF survey study. Powell (2009) adopted panel methods for her national longitudinal study on county level. Chen et al. (2010) estimated a spatial lag model, which considered the influence of socioeconomics and food purchasing characteristics of one's neighbors on one's own health outcomes. They reported spatial OLS estimates as well as regular OLS results.

It is not surprising to see the mixed findings in literature given the various data sources and levels of aggregation involved in different studies. Therefore, how to accurately measure the food environment becomes my first objective in chapter 3. In this review of literature, it is noticed that not many studies have taken unobserved individual effect into consideration while this heterogeneity can easily exist and cause biased or inefficient estimates. This issue can be handled in panel data methods by observing repeated observations from the same unit over time, and thus is discussed in chapter 4. In addition, the clustering of BMI and food environments indicates the need to consider spatial effects and thereby a spatial fixed effects model is also presented in the method section.

CHAPTER 3: DATA AND SUMMARY STATISTICS

My thesis analyzes school-level data on obesity prevalence for students in Arkansas public elementary schools which have second grades. These data were provided by the Arkansas Center for Health Improvement (ACHI). The data for food outlets were obtained from D&B and included geographic coordinates in addition to business addresses. Data preparation involved the use of SAS and ArcGIS software to match food store locations to the 2010 census blocks, assign those blocks to elementary schools, and compute measures of the neighborhood food environment confronting the student body at each school.

School BMI and Control Variables

Obesity is defined as proportion of children with BMI scores greater than or equal to the 95th percentile as defined by gender specific growth chart of Center for Disease Control and Prevention (CDC). A notable distinction of the ACHI data is that the BMI was calculated based on measured height and weight as opposed to self-reported height and weight data common in many earlier studies. Time varying control variables were collected from the Arkansas Department of Education (ADE). These include the school-level percentage of free and reduced lunch participation, the school-level percentage of students enrolled in each grade, and school level proportions for race and ethnicity. Other control variables of socio-demographic and economic characteristics were measured at the school district level and reflect the 2000 census.

The 2000 census block file contains a school district identifier which could then be matched with the ADE school district codes. Since those control variables were based on 2000 census file, they do not have any variation from 2005 to 2010, the period considered in my study.

Table 1 in page 29 contains the summary statistics of BMI data and control variables which were obtained from ACHI, ADE and the 2000 census. As shown, the sample mean of the school-level obesity rate is about 20 percent which is slightly higher than the national level childhood obesity rate. On average, the Arkansas elementary schools have a 63 percent participation rate in free and reduced lunch. Since free and reduced lunch participation depends on income eligibility, this can be viewed as an indicator of low family income. In addition, it is worth noting that in Arkansas school districts, for people aged 25 or above, on average about 59 percent people have a high school degree or equivalent while only about 17 percent achieved a bachelor or higher level of education.

Assignment of Food Stores to Types

Data on Arkansas food outlets are from D & B. Our purpose for food stores is to separate the ones that offer healthy options from the ones that do not. Although the D&B data source has been widely employed in the previous literature, these data have been criticized for validity problems. In this thesis, I further found that the classification of stores by SIC codes are not always reliable if the objective is to separate stores based on whether they provide healthy options. For example, Wal-Mart Supercenters were often assigned to the department store SIC code, along with companies such as Sears and JCPenny. Furthermore, a large number of establishments classified as grocery stores by SIC code were probably too small to carry a broad range of healthy food options. In addition, many specialty stores and pharmacies appear in the grocery store listing in D & B, which would have falsely increased supermarket density if not classified otherwise. As for the restaurant data, a similar problem exists. D&B-provided SIC codes sometimes distinguish between fast food and full service restaurants but often do not. Many are simply listed under the general code for "eating places". Thus, taking steps to assure classification of food stores into accurate types of food establishment was my first objective.

Food Store Selection According to Primary SIC Codes

After studying carefully each of the primary SIC codes (four digit primary SIC) codes, I decided to focus on potential food stores from among the following codes: 5171(petrol bulk station terminals), 5172 (petroleum products), 5182(wine distilled beverages), 5191(farm supplies), 5194(tobacco products), 5199(nondurable goods), 5311(department stores), 5331(variety stores), 5399(miscellaneous merchandise stores), 5411(grocery stores), 5421(meat and fish markets), 5431(fruit and vegetable markets), 5441(candy, nuts, confectioneries), 5451(dairy products), 5461(retail bakeries), 5499(miscellaneous food stores), 5541(gasoline service stations) and 5912(drug proprietary stores).

Establishments from among these codes that represented headquarters locations (not retail locations) were then deleted. Starting with 35,061 records, I initially assigned all food stores into the following twelve types and then later, based on the research interest, focused on four major types.

- Supermarkets or large grocery stores with fresh produce departments
- Traditional Wal-Mart, Target, and similar discount retailers (no produce departments)
- Dollar stores
- Gas stations or convenience stores
- Specialty vegetables and fruits
- Specialty meat, seafood, dairy, eggs, poultry, and/or cheese
- Specialty candy, ice cream, nuts, popcorn, or pretzels
- Specialty bakery, cookies, donuts, cakes, pastries
- Specialty ethnic stores
- Drug stores
- Stores deemed irrelevant to the study

- Stores of undetermined type

Manual Checks of Stores With Fresh Produce

Since misclassifying large grocery stores would have a significant impact on my results, close attention was paid to those establishments. I first looked at eight-digit SIC codes to help define these store types. A large grocery store is mostly likely to be found in the three following categories: Supermarkets (eight-digit SIC codes in 53110000, 53119901, 53999906, 54110100, 54110101, 54110103 and 54110102), grocery stores (eight-digit SIC codes in 54110000, 54119904 and 54119905) and health stores (eight-digit SIC codes in 54990000, 54990100 and 54990102). Establishments in each of these three categories were then analyzed for accuracy. Establishments were screened, sometimes with the help of Internet search engines, to guarantee irrelevant department stores, misclassified convenience stores, or other stores were excluded in the supermarket category. Grocery stores were further divided into three categories according to sales information. Kaufman (1999) used the $500,000 annual sales as a cut –off value to distinguish large grocery stores from small grocery stores. Following his study, this thesis also adopted the $500,000 value as a cut-off threshold. Stores with annual sales more than $500,000 or missing were checked randomly, again often with the help of Internet search engines and were confirmed to be mostly supermarkets and large grocery stores. If annual sales were under $500,000, one of my colleagues selected certain records randomly within each 100,000 increment and called hundreds of stores over the phone (Wang 2010); health stores were briefly checked through the Internet and phone, and only two records were considered to be a supermarket or large grocery store.

Wal-Mart stores were found either to be a traditional Wal-Mart, a large discount department store, or a supercenter, which contains a full line of grocery items, including a full fresh produce department. Since Wal-Mart stores are a major player in Arkansas, further attention was needed to assure Wal-Mart stores were assigned correctly. All Wal-Mart establishments were pulled out and compared to the listings on the Wal-Mart official website. Indeed some traditional Wal-Marts were identified and their types were not recognizable from D & B records. In addition, a few non-existing D&B records were found and were excluded.

<u>Assignment by ultimate Duns number</u>

An ultimate Duns number refers to the corporate parent or as stated in the D&B data dictionary “the DUNS Number of the top-most domestic (U.S./Canada) member in the corporate family”(Dun & Bradstreet 2009) . A frequency table was first run on the prevalence of ultimate Duns numbers within the dataset. These were assigned to 102 major companies, such as “EZ Mart”, “Freds”, “Kroger”, ”Kmart”. There were however, some left-over data that actually belong to the chain company but somehow did not have an ultimate Duns number.

Assignment by Eight-digit SIC Codes

After cleaning up most of the major food store companies, I then turned to assignment by eight-digit SIC codes. Again a frequency table about eight-digit SIC codes prevalence was run and 49 eight-digit SIC codes were assigned to the several food store types based on SIC code description.

Assignment by Keywords

The unassigned establishments were pulled into a separate file and a keyword search was run on the company name and trade style[1]. For example, if a keyword "stop" was found in either the company name or trade style, it was a good indicator that the establishment was a convenience store.

Assignment by Eight-digit SIC Codes after Keyword

In this final step, all the stores left with missing type were assigned according to eight-digit SIC codes.

---

1 Trade style refers to the typical "doing business as" name the company uses. For example, "McDonalds" would be listed as the trade style for a McDonald's restaurant. The formal company name may be for a holding company or other entity that owns the franchise and may not, on its own, indicate that the establishment in question is a McDonald's restaurant.

Thus far, all the stores have been assigned to a relatively accurate type. Table 2 in page 29 describes different methods used for assignment. Among 35,061 records, 760 records (2.2%) were assigned by phone calls to the establishment. 4,446 (12.7%) were assigned after verifying business type through Internet search engines, 6,917(19.7%) by ultimate duns number, 15,661 (44.7%) by eight-digit SIC code, 2,874(8.2%) by keyword and 4,403 (12.6%) by eight-digit SIC code after keyword searches. From the frequency table we can conclude that only about 59% (45%+14%) data got assigned by D&B's original SIC codes. The rest 41% were either manually checked through the Internet or telephone, or were assigned by company name, trade style, keywords, or by parent company ultimate Duns number.

Table 3 in page 30 shows the frequency of main food store types. In the state of Arkansas, convenient stores are most frequent and account for 39.5% of all stores. Stores with fresh produce have the second largest number of establishments at around 11%. Drug stores rank the third with a share of 10%. Dollar stores rank a near fourth and account for 8.94% percent of store types.

School Neighborhood Boundaries

In this study, school service areas were used to develop school-level measures of the food environment. Unfortunately, school boundary shape files are not available statewide and so I used the following procedure in ArcGIS to approximate service areas. Raw data for this procedure consist of geo-coordinates for elementary schools and the shape file for 2010 census blocks. First, my focus is on elementary schools and so I include only those schools containing a second grade class. Second, using the ArcGIS network analyst tool, routes from each block centroid to closest school based on actual driving distance were drawn and a file was created containing each block ID linked with the closest elementary school. Finally, I used the dissolve tool in ArcGIS to disaggregate each block boundary and was left with only the school service area boundaries. The concept is that each block can be viewed as a residential address and that children in each block would go to the closest school. This assumption can be violated if a student goes to a farther school in search of better education quality or for some other reason. However, this would be a problem even if I had actual as opposed to approximate school service area shape files. Since the design is not strictly intended for attendance purposes but rather to approximate the neighborhood food environment confronting students that likely attend the school, it can be deemed as an adequate measure for the purpose of capturing food environment features.

To illustrate the size of my designed neighborhood, I made a comparison between counts of school service areas and census tracts. In the year 2010, 504 elementary school boundaries were created in the state of Arkansas whereas 425 census tracts were found in the same period.

(See figure 1)So one would expect the size of food environment measures in this study to be slightly smaller than the conventional census-tract measures employed in literature.

School-level Food Store Measures

Two types of food store measures on school level-density measure and distance measure were constructed using SAS and ArcGIS. First, I drew buffers of 0.5, 1, 2, 5 and 10 miles around each block centroid, and then counted the number of grocery, convenience, dollar and drug stores. Second, distance from each block to the closet food store of a given type was calculated. Finally, combined with the school boundary files, I was able to average the block level measures to the school level using block populations as weights. School boundaries in 2010 with their associated density measure for large grocery stores within one mile buffer are shown in figure1. In addition, I added cites which have a population of more than 20,000 along with major rivers inside a ten mile buffer around Arkansas to validate that people in bordering neighborhoods within Arkansas do generally shop within the state.

Table 4 in page 31contains summary statistics for school-level food store measures focused on large grocery stores, dollar stores, convenience stores and drug stores. As shown, on average students living in school neighborhoods need to travel 4.2 miles to the closet large grocery store. This neighborhood average distance can be as close as 0.2 miles or as far as 23.8

miles which points out rather uneven access to large grocery stores across the state. The distance to closet convenience store is much nearer, about 2.3 miles, indicating easier access to less healthful food choices. It is noticed that the average distance taken to the closest dollar store is 4.1 miles, about the same as large grocery stores. As for food store density, on average, children in a school neighborhood have access to 0.6 large grocery stores, 1.8 convenience stores and 0.5 dollar stores within a one mile radial distance around the census block of their residence. The access increases to 1.1 grocery stores, 3.6 convenient stores and 0.9 dollar stores within a one to two mile buffer (the donut area consisting of the complement of the one mile buffer with respect to the two mile buffer). Within the two to five mile buffer, the access becomes 4 grocery stores, 2.9 dollar stores and 14 convenience stores. These density measures are used to cover the food store landscape within five miles of each census block center.

Table 1. Descriptive Statistics for BMI and Control Variables (N=2304)

Variable (Percent of Students)	Level	Mean	Std. Dev.	Min	Max
Obese students	school	20.00	4.95	4.59	44.21
African American	school	23.14	28.57	0.00	100.00
Hispanic	school	8.64	13.35	0.00	79.02
Free and reduced lunch	school	62.82	20.53	4.30	100.00
Pre-kindergarten and kindergarten	school	25.29	11.87	0.00	100.00
First and second grade	school	27.17	12.76	0.00	100.00
Third and fourth grade	school	24.04	11.04	0.00	100.00
High school degree or equivalent	school-district	58.90	5.12	47.35	71.98
Bachelors and advanced degree	school-district	16.56	8.36	4.28	38.69
Rural places	school-district	51.03	38.34	0.42	100.00

Table2. Food Store Assignment Methods (N=35061)

Assignment Method	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Keyword	2,874	8.20	2,874	8.20
Eight-digit SIC codes after Keyword	4,403	12.56	7,277	20.76
Phone call	760	2.17	8,037	22.92
Internet follow-up	4,446	12.68	12,483	35.60
Eight-digit SIC codes	15,661	44.67	28,144	80.27
Ultimate duns number	6,917	19.73	35,061	100.00

Table 3. Food Store Types (N=35061)

Store Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Large grocery stores[a]	3,864	11.02	3,864	11.02
Discount retailers[b]	217	0.62	4,081	11.64
Dollar stores	3,134	8.94	7,215	20.58
Convenience store or Gas station	13,849	39.50	21,064	60.08
Specialty vegetables and fruits	484	1.38	21,548	61.46
Specialty meat, seafood, dairy, eggs, poultry, and/or cheese	1,535	4.38	23,083	65.84
Specialty candy, ice cream, nuts, popcorn, pretzels	810	2.31	23,893	68.15
Specialty bakery, cookies, donuts, cakes, pastries	1,848	5.27	25,741	73.42
Specialty ethnic stores	1,034	2.95	26,775	76.37
Drug stores	4,122	11.76	30,897	88.12
Stores deemed irrelevant to the study	3,550	10.13	34,447	98.25
Stores of undetermined type	614	1.75	35,061	100.00

[a]Large grocery stores are stores with fresh produce department such as Wal-Mart, HARPS ect.

[b]Discount retailers include discount stores without fresh produce department such as traditional Target, traditional Wal-Mart etc.

Table 4. Descriptive Statistics for School-level Food Store Measures (N=2304)

Variables	Mean	Std Dev	Min	Max
Distance to nearest large grocery stores	4.21	4.44	0.27	23.79
Distance to nearest dollar stores	4.06	4.26	0.31	25.04
Distance to nearest convenience stores	2.27	2.61	0.19	20.30
Distance to nearest drug stores	3.81	3.79	0.25	20.72
Density of large grocery stores within one mile buffer	0.61	0.79	0.00	5.25
Density of dollar stores within one mile buffer	0.51	0.60	0.00	3.61
Density of convenience stores within one mile radial distance	1.78	2.11	0.00	13.70
Density of drug stores within one mile radial distance	0.77	1.05	0.00	5.83
Density of large grocery stores within one to two mile radial distance	1.14	1.38	0.00	7.35
Density of dollar stores within one to two mile radial distance	0.89	1.03	0.00	5.68
Density of convenience stores within one to two mile radial distance	3.64	4.44	0.00	25.65
Density of drug stores within one to two mile radial distance	1.41	1.99	0.00	11.82
Density of large grocery stores within two to five mile radial distance	3.99	5.14	0.00	26.01
Density of dollar stores within two to five mile radial distance	2.86	3.14	0.00	15.13
Density of convenience stores within two to five mile radial distance	14.10	17.59	0.00	94.85
Density of drug stores within two to five mile radial distance	4.87	7.21	0.00	36.93

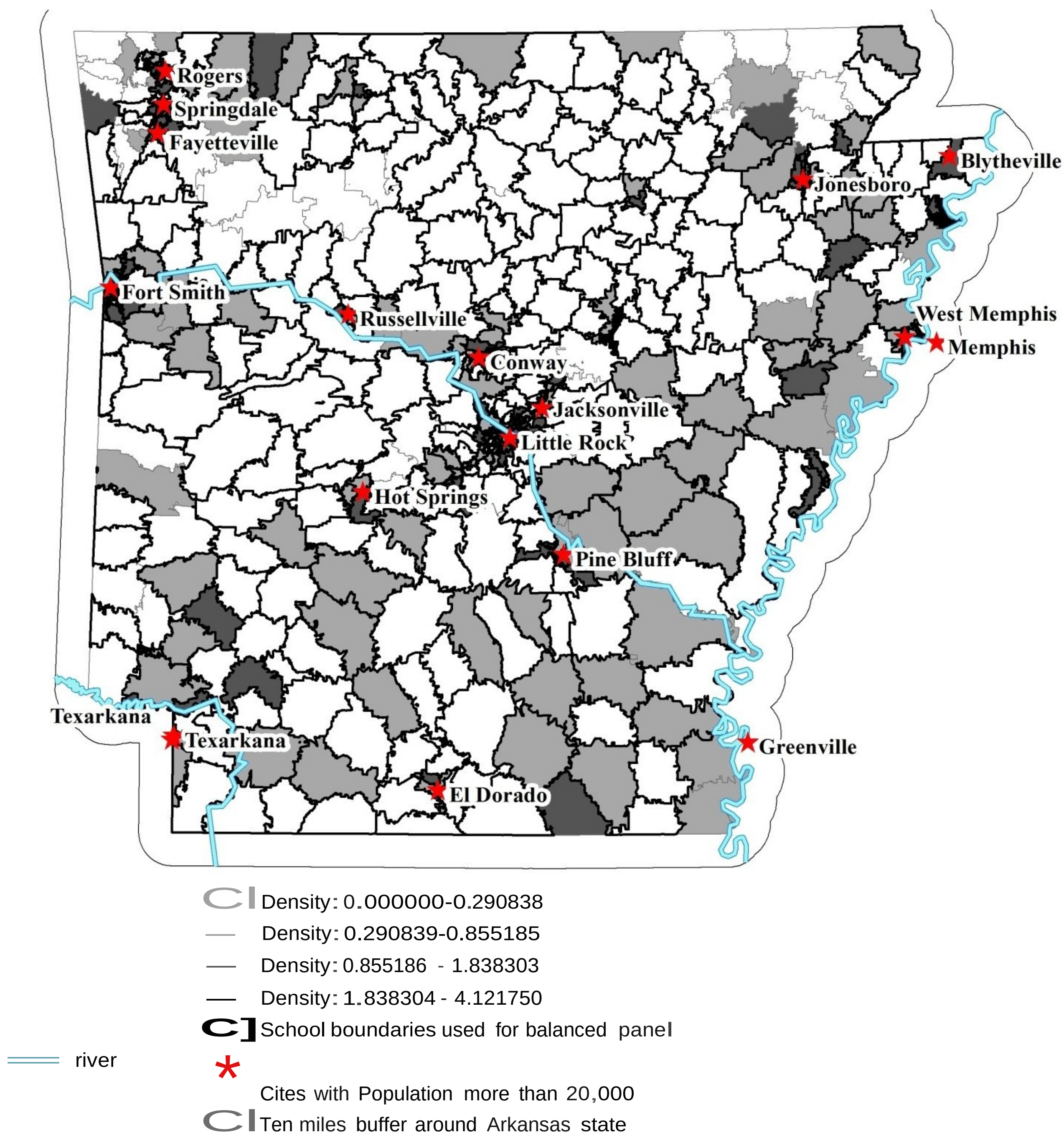

Figure 1. Map of2010 School level Large Grocery Store Density Measure

Note: The Density Measure is the count of large grocery stores within one mile radial distance around block center then aggregated to school level weighted by block population.

CHAPTER 4: MODEL SPECIFICATION

In this thesis, I evaluated how children's food environment measures after taking control of ethnicity, income and education, relate to obesity prevalence. A generalized panel model in matrix notation can be expressed as:

(1)

where is an vector of prevalence of school-level obesity rate with representing the number of cross-section units and equal to the number of time periods, is an matrix of food environment measure variables, is an matrix of control variables which could help explain , a vector of time- invariant unobserved school specific effect, and randomly distributed error with zero mean and constant variance. If is zero which indicates that unobserved school specific effect does not exist, the ordinary least square (OLS) estimator will be a best linear unbiased estimator (BLUE). If however is not zero, the OLS estimator will no longer be BLUE and other methods should be employed. To be more specific, if unobserved individual heterogeneity does exist and has not been accounted for, the model will contain omitted variable bias and the actual effect will depend on whether is correlated with the regressors. When is uncorrelated with any column of , equation (1) becomes a random effects model in which failing to consider the unobserved individual effect will result in consistent but inefficient estimates. That is to say the coefficients are still correct whereas their standard errors become bigger which might lead to reject of the null hypothesis when it actually should not. To solve this problem, one needs to resort to some feasible generalized least squares (GLS) estimators. On the other hand, when is correlated with any column of , equation (1) is treated as a fixed effects model and failure to account for will yield inconsistent results. In other words, under this circumstance the estimated coefficients of equation (1) will be wrong, as well as their standard errors and statistical inference. Thereby a mean-differencing method is employed (usually referred to as “within model”). Although the assumptions of ordinary least squares (OLS), random effects (RE) and fixed effects (FE) models can be clearly specified in theory, it is generally hard to determine whether the unobserved individual effect is correlated with regressors or not. Therefore a Hausman test is performed to determine whether RE model or FE model fits the data better. The null hypothesis of Hausman test is that unobserved heterogeneity is uncorrelated with regressors, under which both the RE and FE estimators are consistent. If this assumption is not true, only FE estimator will be consistent. A significantly large test statistic which indicates the big difference between the two sets of coefficients. Therefore, a large test statistic suggests the reject of null hypothesis and that FE model is the more appropriate specification.

In this thesis, I propose two specifications as baseline models which differ in terms of food environment measures. The first model is based on neighborhood average density of food stores and the second on average distance to the nearest food store of a given type.

In the density based measure model, consists of the neighborhood average density of large grocery stores within one mile radial distance around block center, the average density of large grocery stores within a one to two mile buffer and the average density within a two to five mile buffer. The decision of the radial distance of buffer zones employed requires some general knowledge of the scale of Arkansas and what has been used in literature. Most studies on individual levels would use a one-mile buffer to approximate the food consumption market and that those studies generally have a more compact scale of focused areas. It is noticed that according to the Census Bureau, the state of Arkansas has a population density of 56.43 inhabitants per square mile, ranking 34th among all the states (U.S. Census Bureau 2010). This low population density indicates the study area in this thesis is relatively spread out and therefore a larger radial distance measure might be needed. In addition, it is found that the average distance needed to travel to closest large grocery stores in the state of Arkansas is about four miles instead of 0.9 mile in literature (Handy and Clifton 2001) or 1 mile (Chen et al. 2010), and that within one mile buffer, about 32.29 percent people in a school neighborhood on average do not have access to food stores, whereas in literature this number is about 10 percent (Chen et al.

2010). The percent of people that have zero access to large grocery stores drops to 12.33 when the buffer is extended to five miles. So, after carefully studying the layout pattern of large grocery stores, a one mile buffer, a one to two mile buffer and a two to five mile buffer were included in the baseline model to capture the overall large grocery geography.

In the distance based measure model, represents the log form of neighborhood average distance travelled from block center to closest big grocery store. Note that in this model, a non-linear relationship between distance to closest big grocery store and obesity prevalence is suspected based on the assumption that the further a grocery store is located from a block center the slower its rate of impact on BMI of children in that block might become. The following example helps explain this hypothesis: a child's health outcome is considered to be affected by what his or her parents purchased for food. If a parent has to travel, say two miles to the closest supermarket to purchase fresh fruits and vegetables, he or she might choose to go there more or less often than if the store were one mile closer or further. However, if a parent has to travel ten miles to the closet big grocery store, relocating that store one mile nearer or further will probably have less of an effect on the parent's food shopping behaviors.

In the next step, I expand both baseline models to include other food environment measures. To account for collinearity problems, measures of other types of food stores -

convenience stores, dollar stores and drug stores, were designed in ratio form. The ratio is computed as distance or density of the outlet in question divided by distance or density of large grocery stores. Thus in the expanded density model further includes variables of ratio density measures of dollar, convenience and drug stores within a one mile buffer. The density of large grocery stores, which is treated as the denominator in the ratio measures, can be zero within a one mile radial distance, and that would result in an undefined measure. Therefore, I assigned density of large grocery equal to 0.003 in cases where measured density was zero. The 0.003 replacement number is determined by a frequency table which shows 0.003 to be the second smallest value, after zero of the average density of large grocery stores within a one mile buffer. Likewise, in the expanded distance model, further represents log form of ratio measures of distance relative to the nearest supermarket for the rest of the three store types.

As outlined before, bordered school neighborhoods might share similar characteristics of ethnicity, income level, food environment and physical environment such as work-out facilities, sidewalks and trials. Since these characteristics could have explanatory power over health outcomes, it is possible that children's obesity rate is clustered through these channels. As demonstrated by previous literature, BMI is also found to be clustered across space. The spatial correlation can generally be specified in three formats-spatial lag, spatial error, or both, based on where the spatial component comes in. If a spatial lag term is ignored, it is equivalent to committing an omitted variable misspecification error and thus introduced biased results. If a spatial error term is neglected, the estimates will still be unbiased but inefficient. (Luc Anselin and Daniel Arribas-Bel, 2011) In this thesis, I used open-source software R 2.15.0 splm package which offers a spatial panel data model estimation algorithm to account for the spatial effect. The model is specified as:

,

where is an vector of school level obesity prevalence**,** is a matrix of food environment measure variables, is a matrix of exogenous control variables, an identity matrix of dimension T, is the spatial weights matrix of values between 0 and 1, and the corresponding spatial-lag parameter. The disturbance error vector consists of two parts:

,

where is a vector of ones, an identity matrix of dimension N, is a vector of time-invariant school specific effects, and a vector of spatially autocorrelated errors that follow a spatial autoregressive process of the form

,

with ( ) as the spatial autoregressive parameter (the spatial error term), the spatial weight matrix of dimension N, and randomly distributed error term.

As in the previously presented classical panel model, the school specific effect can be treated as zero, fixed or random. Correspondingly, a spatial pooled OLS, FE or RE model can be specified to account for differences in . In this thesis, only the results of spatial FE models were reported due to the strong preference of the FE model based on Hausman tests from the previous classical panel model. To insure this preference still holds in spatial setting, a spatial RE and pooled OLS have all been estimated and spatial Hausman test suggested the same as the Hausman test from classic panel model.

There are various ways to construct a spatial weight matrix including contiguity, inverse distance and N nearest neighbors. Following the literature, a spatial weight matrix based on queen contiguity was constructed in this thesis. The queen contiguity defines neighbors as polygons sharing common borders and corners. So for example, say school A is on border of school B but not of school C, then on the weight matrix, the element corresponding to school A and B will be assigned a value of 1 whereas the element corresponding to A and C will have a value of 0. The diagonal elements of W are all 0s since one is not considered as a neighbor of oneself. Then the weight matrix is row standardized so that the sum of each row equals 1. It is worth noting that the weight matrix in this model was based on boundaries in one cross- time period and repeated over the six years. However in real data, due to school mergers, construction and exit, the school attendance areas do have variations across years. If the spatial weight matrix is misspecified, it can lead to biased estimates. This becomes one limitation of this thesis and will be discussed later in the last chapter. In the results, I report estimates using spatial weight matrix developed from the 2010 school boundaries.

The splm package from R offers two major functions to estimate spatial panel models- spml and spgm. The spml approach implements maximum likelihood (ML) estimation and spgm employs generalized methods of moments (GMM). In this thesis, I applied ML estimation to the spatial panel data. The spml function has an argument “model” which controls the specification. For example, “model” takes up the value “within” for fixed effects, “random” for random effects, and “pooling” for no effects. The spatial structure is controlled by arguments lag and spatial error, which can be specified by the user. General panel models were also estimated using R with the package plm.

CHAPTER 5: RESULTS

Benchmark Results

Table 5 in page 52-53 shows the baseline estimates of the effect of density measures along with other control variables on school obesity prevalence. The first three specifications consist of a pooled OLS, random effects model and fixed effects model. Estimates of the impact of grocery store density vary a lot across three specifications in terms of sign, magnitude and statistical significance. The pooled OLS model shows that a higher density of large grocery stores within two to five mile radial distance is statistically associated with lower obesity prevalence. Adding one large grocery store within two to five mile buffer zone around children's block center will bring down the school-level obesity prevalence by 0.07 percent. The densities within one mile and one to two mile buffers are also negatively correlated with obesity but are not statistically significant. In the random effects model, densities of large grocery stores within one mile and two to five mile buffers are associated with lower obesity rate while density within one to two miles are associated with higher obesity rate. But none of these estimates are significant. The fixed effects model, on the other hand, provides quite different results. The density of large grocery store within one mile radial distance in fixed effects model is significantly (at the 10 percent level) and negatively associated with obesity prevalence while

densities within one to two or two to five mile buffer show a positive but insignificant association. Increasing the density by one store within one mile radial distance would reduce the obesity prevalence by 0.58 percent, which is about three times larger than in pooled OLS model. The from Hausman test is proximately 31 and so I reject the null hypothesis that the unobserved school effect is not correlated with regressors, which suggests that fixed effects estimates are preferred. Comparison between pooled OLS and the fixed effects model suggests that once the unobserved school specific effect is captured, the impact of large grocery stores on childhood obesity seem to come from density measures within one mile radial distance instead of the other two father measures. In addition, this impact becomes much larger when school heterogeneity is accounted for.

The last two specifications consist of a spatial lag and a spatial error fixed effects model, arguing the form of spatial spillover is either in the dependent variable or error term. As results show, the spatial components (both the spatial lag and error estimates) are extremely small and statistically insignificant. The estimates of other explanatory variables are robust as the magnitudes and signs stay the same as those found in the classical fixed effects estimates. However it is worth emphasizing that the statistical significance of large grocery store density within one mile increased in both the spatial lag and error model when compared to the general fixed effects model. This is due to smaller standard errors in the models that account for space. The coefficients of the spatial lag model cannot be directly interpreted as a marginal effect. The marginal impact of covariates on dependent variable consists of two parts: direct impact and indirect impact. The direct impact on school-level obesity prevalence is due to a one unit change in the explanatory variables, whereas the indirect impact is caused by change of neighboring school's obesity prevalence which again is determined by the same covariates. However, for the model estimated in this thesis, the indirect effect is so small, as suggested by spatial lag estimator, that it can almost be ignored. Therefore, we can interpret the marginal effect directly from coefficients in the spatial lag model here, which is basically, the same as that found in the classical fixed effects model.

The estimates for control variables generally display consistent results with the literature. For example, in the fixed effects model, the percentage of African American students, percentage of Hispanic students and percentage of third and fourth graders amongst all measured students show a significant and positive relationship with obesity prevalence. On the other hand, percentage of free and reduced lunch participation, which is a proxy for income level, shows no significant association with BMI. This is unexpected as earlier work generally shows a relationship between income and obesity. The percentage of pre-kindergarten and kindergarten among total measured students is also associated with lower obesity prevalence. The percentage measures for different grades are included in the analysis as age controls. This hypothesis is supported by results in table 5 as the estimate on percentage of pre-kindergarten and kindergarten students is significant and negative while the estimate on percentage of 3rd and 4th graders is significant and positive. The education controls are time invariant and therefore were excluded in the fixed effects model, whereas in OLS, both the percentage of people with a high school degree, and the percentage of people with a bachelors or advanced degree show a statistically significant and inverse relationship to BMI, which is again in line with literature. Finally, the percentage of rural blocks, which is again time invariant, is associated with higher obesity rate in OLS model. It is noted that estimates of control variables in both the spatial lag and error models have slightly lower standard errors.

Table 6 in page 54-55 shows results from both the classical panel and spatial panel methods based on distance measures of large grocery stores. Log transformation of distance to nearest large grocery store from school neighborhood census block center was employed. This non-linear relationship between distance and obesity prevalence is assumed because it is believed that the impact of closest large grocery store over obesity should increase at a decreasing rate as the distance increases. The OLS model supported this assumption by rendering a significant positive estimate for log form of distance to nearest large grocers. However the coefficient cannot be interpreted directly as marginal effect. In fact, as mentioned before, the marginal effect of distance to nearest large grocery store is not fixed, but varying as it takes on different values. If we take two arbitrary values of distance to nearest large grocery store as D1 and D2, the expected mean difference in obesity prevalence is, holding the other control variables constant, $0.479\times(\log(D2) - \log(D1)) = 0.479\times\log(D2/D1)$. Therefore, as long as the percent increase in distance measure is fixed, we will see the same difference in obesity rate regardless of what underlying distance is. For example, if there is 10% increase in distance to closest large grocery store, the childhood obesity prevalence will go up by $0.479\times\log1.1$ percent, which is approximately 0.02 percent. The logged distance measure is insignificant in both the random effects and fixed effects models. The estimate in the fixed effects model is of much smaller magnitude. As for the spatial models, it is noticed that both the spatial lag and error term show no statistical significance indicating the non-existence of spatial effects. In addition, the spatial coefficients are too small to have any influential explanatory power on obesity rates. Estimates of other predictors generally agree with the density model. Finally, as seen in the earlier estimates presented in table 5, the spatial model shrinks the standard errors and thus increases the statistical significance of some of the control variables.

Expanded Model Results

To gain a more comprehensive view of the food environment's impact on childhood obesity, I expanded the baseline models to include ratio measures of other store types. Table 7 in page 56-57 provides results based on density measures from OLS, random effects, fixed effects, spatial lag and spatial error models. The OLS model shows statistically significant positive relationship between density of drug stores relative to density of large grocery store within one mile and childhood obesity prevalence. A one point increase in this ratio will bring up childhood obesity prevalence by 0.008 percent. This effect is rather small which indicates the impact of drug store on obesity rate is not major. The random effects model also shows similar relationship between the drug store ratio measure and obesity rate but with an even smaller magnitude at 0.004. The fixed effects estimates including those from spatial models all show a positive but insignificant relationship between the drug store density ratio and obesity rate. For other types of store ratio measures, none of the five specifications provides any significant results on any measure other than the average density of grocery stores within one mile of residence. It should also be mentioned that when compared with baseline density model, the expanded model provides a slightly bigger estimate for the density of large grocery stores within this one-mile buffer. In addition, like in previous estimations, both the spatial lag and error estimates are extremely small and insignificant. Therefore, it is concluded that spatial effects almost do not exist in these data regardless of specification.

Table 8 in page 58-59 shows the results from expanded models based on the distance measures. In the OLS model, the log ratio of distance to closest convenience store relative to large grocery store is statistically significant and associated with higher obesity rates whereas distance to closest drug store is associated with a lower obesity rate. As explained in the baseline model, estimates in the log form of regressors cannot be directly interpreted as marginal effect but one can calculate the effect caused by a certain percentage change in the regressors. So after some simple calculations, it is found that a 10% increase in the distance to closest convenience store relative to large grocery store will increase the childhood obesity prevalence by 0.02 percent, whereas a 10% rise in the distance to closest drug store relative to large grocery store will decrease the childhood obesity prevalence by 0.018 percent. The mixed results are hard to explain in that I would expect similar adverse impact of drug store and convenience stores on children's obesity outcomes considering neither convenience stores nor drug stores offer fresh produce. One possible reason would be that convenience stores usually cluster in urban areas, and thus the results in which presence of convenience stores seem to reduce the obesity rate is in fact due to urbanity. To better understanding the role of stores with no fresh produce and easy access, a sensitivity analysis will be performed that combines convenience stores and drug stores into one category. As for dollar stores, none of the five models show any significant estimates indicating that the distance to dollar stores generally would not affect childhood obesity. Although dollar stores are not influencing children's health status on the whole, it is suspected they might exert some impact on childhood obesity in rural areas due to the fact that dollar stores usually differentiate themselves as low-price retailers and target people with low income. Therefore a subsample analysis of rural areas and dollar stores is performed.

Before turning to these follow-up estimations, it is worth pointing out that since the splm package for spatial estimation in R doesn't have the option of using robust standard errors to account for general hetroskedasticity, for table 5, 6, 7 and 8 non-robust standard errors were reported for both classical and spatial models for comparison reasons. I have estimated the normal OLS, random effects and fixed effects models with robust standard errors in R package plm which accounts for hetorskedasticity, and the major results on food store measures generally do not change except that, in the expanded model, the log ratio measures of convenience store density and distance become significant at the 10% level. Another issue which might deserve our attention is that while R did give us the option of adding spatial lag and error terms in one model, when I estimated the density model using this method, no inference information could be provided and a warning message was produced indicating the quasi-variance of the predicator is negative and that may be due to the strong correlation of the spatial lag and error term. Once I remove one term of the spatial model, the models converged without incident. Since the data

does not seem to show any spatial pattern this may not be a major issue in this thesis. Finally, the reported fixed effects estimates were based on "one-way" and did not consider the time-effect. This is because the "two-way" fixed effects were run first but showed no significant time effects.

Results of Dollar Stores in Rural Areas

Following the previous discussion, a subsample of rural schools was selected to further decide whether dollar stores exert any influence on rural children's health outcomes. Here "being a rural school" was defined as whether the percentage of rural residents in a school service area account for 90% of the population or more. The 90% cut off is rather arbitrary, and therefore, subsamples using thresholds ranging from from 60% to 100% were performed to test robustness to this definition. When the 90% cut off was used, a balanced panel which consists of 130 schools over six years was constructed. Table 9 in page 60-61 presents the panel estimation based on both density and distance measures of dollar stores. Since the previous spatial estimation reflects no spatial effect, only the classical panel estimation was reported. The OLS results based on density measures indicate density of dollar stores within a one mile buffer around the neighborhood block center is associated with higher obesity rates. A one-store increase in the density of dollar stores within one mile radial distance will bring childhood obesity prevalence up by 2.24 percent. The 2.24 is a very big impact and it shows strong statistical significance. On the contrary, the density of stores within one to two mile buffer indicates a negative relationship between density and childhood obesity prevalence. The marginal impact is -4.05, greater than the dollar store density within one mile. The results from OLS are hard to explain and it might be subject to misspecification due to the neglect of unobserved school heterogeneity. Indeed when random effects and fixed effects estimates are examined, the significant effect on dollar store density measures within a one mile or one to two mile buffer disappeared.

A LM Breusch Pagan test was performed. The null hypothesis in the LM test is that variance across entities is zero which means no significant random effects and the OLS estimator is preferred. The test yields a of 320, which strongly rejects the null hypothesis. Thereby random effects model is preferred over OLS in this analysis. In addition, the Hausman test yields a of 13.01 which shows no statistical significance on 10% level to reject its null hypothesis and thus suggests the random effects model is preferred over the fixed effects model. As for the distance based models, none of the three models show any statistical significance for the distance to closet dollar store.

Estimations of rural schools being defined as rural residence accounts for 60%, 70% and 80% of population were performed and yielded consistent results with 90% cut off value. This analysis indicates dollar stores do not have an impact over rural children's BMI once school specific effects are taken into account regardless of the definition of "rural school".

Sensitivity Analysis

To test the robustness of the models, I explored various specifications. First, as discussed before, I added convenience stores and drug stores together and performed the expanded model analysis. The new expanded density model shows that the combined measure has no statistical explanatory power over childhood obesity prevalence. Second, as mentioned before, one drawback of estimating spatial model in R is that the software does not allow for variations of spatial weight matrix and thus the school boundaries have to be fixed over all time periods. I previously used school boundary of 2010 which is the latest year in the dataset. However the results could change if school boundaries of previous years were used. Therefore I re-estimated baseline models using 2007 boundaries. 2007 is the middle year of my sample. Similar results were shown in the spatial models and again no spatial effect was detected. Finally, due to previous findings in the literature which indicates that the existence of restaurants might exert some influence on childhood obesity, ratio measures of fast foods were added to the expanded models to prevent the mistake of omitting variables. The results are reported in table 10 in page 62. First it is noticed that the density of fast food within a one mile radial distance of neighborhood block centers is negatively associated with obesity rates, which is the opposite of what we would expect. However this coefficient is not significant across the three panel specifications. Second, when compared with the previous expanded model (see table 7), a slightly increase in the estimate for large grocery stores within a one mile buffer is found. This indicates when the impact of fast food restaurants is accounted for, the influence of large grocery stores on childhood obesity prevalence increased about 0.12 percent.

Table 5a. Pooled OLS, Random Effect and Fixed Effect Models: Baseline Density Results(T=6,N=384)

Variables	OLS			REM			FEM(within)		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
(Intercept)	25.366	***	(1.546)	26.722	***	(2.594)			
Free and reduced lunch	0.052	***	(0.007)	0.036	***	(0.009)	0.009		(0.014)
African American	0.047	***	(0.005)	0.054	***	(0.008)	0.060	**	(0.029)
Hispanic	0.089	***	(0.009)	0.095	***	(0.014)	0.113	***	(0.036)
Rural places	0.018	***	(0.004)	0.022	***	(0.007)			
High school degree or equivalent	-0.116	***	(0.021)	-0.134	***	(0.037)			
Bachelors and advanced degree	-0.146	***	(0.019)	-0.183	***	(0.031)			
Pre-kindergarten and kindergarten	-0.057	***	(0.007)	-0.042	***	(0.007)	-0.034	***	(0.008)
First and second grade	-0.014	**	(0.007)	-0.007		(0.007)	-0.003		(0.008)
Third and fourth grade	0.009		(0.008)	0.014	*	(0.008)	0.017	**	(0.008)
NearGrocery_1mile	-0.199		(0.152)	-0.340		(0.210)	-0.579	*	(0.306)
NearGrocery_1to2mile	-0.090		(0.104)	0.059		(0.149)	0.317		(0.227)
NearGrocery_2to5mile	-0.073	***	(0.028)	-0.044		(0.042)	0.016		(0.071)
Adj. R-Squared	0.333			0.144			0.016		
Hausman test(chisq)				31.493	***				
F-test (F)	5.582	***							

Note: NearGrocery_1mile represents the density of large grocery store within one mile buffer around children's neighborhood block center.

NearGrocery_1to2mile represents the density of large grocery store within the donut area of two mile buffer minus one mile buffer.

NearGrocery_2to5mile represents density of large grocery store within the donut area of five mile buffer minus two mile buffer.

Table 5b. Spatial Fixed Effect Models: Baseline Density Results(T=6,N=384)

Variables	Spatial Lag FEM			Spatial Error FEM		
	Coef.		SE	Coef.		SE
Free and reduced lunch	0.009		(0.012)	0.009		(0.012)
African American	0.060	**	(0.027)	0.060	**	(0.027)
Hispanic	0.113	***	(0.033)	0.113	***	(0.033)
Pre-kindergarten and kindergarten	-0.034	***	(0.008)	-0.034	***	(0.008)
First and second grade	-0.003		(0.007)	-0.003		(0.007)
Third and fourth grade	0.017	**	(0.008)	0.017	**	(0.008)
NearGrocery_1mile	-0.579	**	(0.278)	-0.579	**	(0.278)
NearGrocery_1to2mile	0.317		(0.207)	0.317		(0.207)
NearGrocery_2to5mile	0.016		(0.065)	0.016		(0.065)
Spatially lagged dependent variable	-0.0009		(0.030)			
Spatially lagged error				-0.0006		(0.030)

Table 6a. Pooled OLS, Random Effect and Fixed Effect Models: Baseline Distance Results(T=6,N=384)

Variables	**OLS**			**REM**			**FEM(within)**		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
(Intercept)	24.759	***	(1.546)	26.061	***	(2.598)			
Free and reduced lunch	0.050	***	(0.007)	0.035	***	(0.009)	0.008		(0.014)
African American	0.045	***	(0.005)	0.053	***	(0.008)	0.060	**	(0.029)
Hispanic	0.085	***	(0.008)	0.093	***	(0.014)	0.116	***	(0.036)
Rural places	0.015	***	(0.005)	0.019	**	(0.008)			
High school degree or equivalent	-0.111	***	(0.021)	-0.128	***	(0.037)			
Bachelors and advanced degree	-0.168	***	(0.017)	-0.190	***	(0.029)			
Pre-kindergarten and kindergarten	-0.057	***	(0.007)	-0.042	***	(0.007)	-0.034	***	(0.008)
First and second grade	-0.012	*	(0.007)	-0.006		(0.007)	-0.003		(0.008)
Third and fourth grade	0.009		(0.008)	0.014	*	(0.008)	0.016	*	(0.008)
Log(DistGrocery)	0.479	***	(0.152)	0.336		(0.208)	0.073		(0.296)
Adjusted R-squared	0.333			0.144			0.014		
Hausman test (chisq)				22.629	***				
F test (F)	5.578	***							

Note: log(DistGrocery) represents the log form of distance to nearest big grocery store from children's neighborhood block center

Table 6b. Spatial Fixed Effect Models: Baseline Distance Results(T=6, N=384)

Variables	**Spatial Lag FEM**			**Saptial Error FEM**		
	Coef.		**SE**	**Coef.**		**SE**
Free and reduced lunch	0.008		(0.012)	0.008		(0.012)
African American	0.060	**	(0.027)	0.060	**	(0.027)
Hispanic	0.116	***	(0.033)	0.116	***	(0.033)
Pre-kindergarten and kindergarten	-0.034	***	(0.008)	-0.034	***	(0.008)
First and second grade	-0.003		(0.007)	-0.003		(0.007)
Third and fourth grade	0.016	**	(0.008)	0.016	**	(0.008)
Log(DistGrocery)	0.073		(0.270)	0.073		(0.270)
Spatially lagged dependent variable	-0.0004		(0.030)			
Spatially lagged error				0.0003		(0.030)

Table 7a. Pooled OLS, Random Effect and Fixed Effect Models: Expanded Density Results(T=6,N=384)

Variables	**OLS**			**REM**			**FEM(within)**		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
(Intercept)	23.689	***	(1.601)	25.753	***	(2.634)			
Free and reduced lunch	0.054	***	(0.007)	0.037	***	(0.009)	0.009		(0.014)
African American	0.047	***	(0.005)	0.054	***	(0.008)	0.060	**	(0.029)
Hispanic	0.088	***	(0.009)	0.095	***	(0.014)	0.114	***	(0.036)
Rural places	0.016	***	(0.004)	0.021	***	(0.007)			
High school degree or equivalent	-0.092	***	(0.022)	-0.120	***	(0.037)			
Bachelors and advanced degree	-0.139	***	(0.019)	-0.179	***	(0.031)			
Pre-kindergarten and kindergarten	-0.057	***	(0.007)	-0.042	***	(0.007)	-0.034	***	(0.008)
First and second grade	-0.014	**	(0.007)	-0.007		(0.007)	-0.003		(0.008)
Third and fourth grade	0.009		(0.008)	0.014	*	(0.008)	0.017	**	(0.008)
NearGrocery_1mile	-0.164		(0.155)	-0.333		(0.213)	-0.583	*	(0.309)
NearGrocery_1to2mile	-0.093		(0.104)	0.055		(0.149)	0.315		(0.227)
NearGrocery_2to5mile	-0.074	***	(0.028)	-0.043		(0.042)	0.020		(0.071)
NearDollar_1mile/NearGrocery_1mile	0.000		(0.002)	0.002		(0.003)	0.003		(0.004)
NearConv_1mile/NearGrocery_1mile	-0.001		(0.001)	-0.001		(0.001)	-0.002		(0.001)
NearDrug_1mile/NearGrocery_1mile	0.008	***	(0.002)	0.004	*	(0.002)	0.002		(0.003)
Adjusted R-squared:	0.3372			0.147			0.017		
Hausman test (Chisq)				22.639	***				
F test (F)	5.506	***							
RESET (F)	18.689	***							

Note: NearDollar_1mile/NearGrocery_1mile represents the density of dollar stores relative to large grocery stores within one mile buffer.

NearConv_1mile/NearGrocery_1mile represents the density of convenience stores relative to large grocery stores within one mile buffer.

NearDrug_1mile/NearGrocery_1mile represents the density of drug stores relative to large grocery stores within one mile buffer.

Table 7b. Spatial Fixed Effect Models: Expanded Density Results(T=6,N=384)

Variables	**Spatial Lag FEM**			**Spatial Error FEM**		
	Coef.		**SE**	**Coef.**		**SE**
Free and reduced lunch	0.009		(0.012)	0.009		(0.012)
African American	0.060	**	(0.027)	0.060	**	(0.027)
Hispanic	0.114	***	(0.033)	0.114		(0.033)
Pre-kindergarten and kindergarten	-0.034	***	(0.008)	-0.034	***	(0.008)
First and second grade	-0.003		(0.007)	-0.003		(0.007)
Third and fourth grade	0.017	**	(0.008)	0.017	**	(0.008)
NearGrocery_1mile	-0.583	**	(0.281)	-0.584	**	(0.281)
NearGrocery_1to2mile	0.315		(0.207)	0.314		(0.207)
NearGrocery_2to5mile	0.020		(0.065)	0.021		(0.065)
NearDollar_1mile/NearGrocery_1mile	0.003		(0.004)	0.003		(0.004)
NearConv_1mile/NearGrocery_1mile	-0.002		(0.001)	-0.002		(0.001)
NearDrug_1mile/NearGrocery_1mile	0.002		(0.002)	0.002		(0.002)
Spatially lagged dependent variable	-0.002		(0.030)			
Spatially lagged error				-0.003		(0.030)

Table 8a. Pooled OLS, Random Effect and Fixed Effect Models: Expanded Distance Results(T=6,N=384)

Variables	OLS			REM			FEM(within)		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
(Intercept)	24.054	***	(1.583)	25.287	***	(2.627)			
Free and reduced lunch	0.051	***	(0.007)	0.037	***	(0.009)	0.010		(0.014)
African American	0.046	***	(0.005)	0.054	***	(0.008)	0.061	**	(0.029)
Hispanic	0.088	***	(0.008)	0.097	***	(0.014)	0.121	***	(0.036)
Rural places	0.015	***	(0.005)	0.016	*	(0.008)			
High school degree or equivalent	-0.098	***	(0.022)	-0.113	***	(0.037)			
Bachelors and advanced degree	-0.162	***	(0.017)	-0.185	***	(0.029)			
Pre-kindergarten and kindergarten	-0.058	***	(0.007)	-0.043	***	(0.007)	-0.035	***	(0.008)
First and second grade	-0.012	*	(0.007)	-0.007		(0.007)	-0.003		(0.008)
Third and fourth grade	0.008		(0.008)	0.013	*	(0.008)	0.016	*	(0.009)
Log(DistGrocery)	0.479	***	(0.185)	0.516	*	(0.284)	0.595		(0.548)
Log(DistDollar/DistGrocery)	-0.203		(0.209)	0.093		(0.261)	0.376		(0.343)
Log(DistConv/DistGrocery)	0.496	**	(0.218)	0.479	*	(0.251)	0.443		(0.307)
Log(DistDrug/DistGrocery)	-0.435	**	(0.201)	-0.364		(0.259)	-0.261		(0.372)
Adjusted R-squared:	0.335			0.145			0.016		
Hausman test (chisq)				25.470	***				
F test (F)	5.550	***							

Note: Log(DistDollar/DistGrocery) represents the log form of distance to nearest dollar store relative to distance to nearest large grocery store from block center.

Log(DistConv/DistGrocery) represents the log form of distance to nearest convenience store relative to distance to nearest large grocery store.

Log(DistDrug/DistGrocery) represents the log form of distance to nearest drug store relative to distance to nearest large grocery store.

Table 8b. Spatial Fixed Effect Models: Expanded Distance Results(T=6,N=384)

Variables	Spatial Lag FEM			Spatial Error FEM		
	Coef.		SE	Coef.		SE
Free and reduced lunch	0.010		(0.012)	0.010		(0.012)
African American	0.061	**	(0.027)	0.061	**	(0.027)
Hispanic	0.121	***	(0.033)	0.121	***	(0.033)
Pre-kindergarten and kindergarten	-0.035	***	(0.008)	-0.035	***	(0.008)
First and second grade	-0.003		(0.007)	-0.003		(0.007)
Third and fourth grade	0.016	**	(0.008)	0.016	**	(0.008)
Log(DistGrocery)	0.596		(0.499)	0.596		(0.499)
Log(DistDollar/DistGrocery)	0.376		(0.312)	0.376		(0.312)
Log(DistConv/DistGrocery)	0.443		(0.280)	0.443		(0.280)
Log(DistDrug/DistGrocery)	-0.261		(0.339)	-0.260		(0.339)
Spatially lagged dependent variable	0.001		(0.030)			
Spatially lagged error				0.002		(0.030)

Table 9a. Balanced Panel Estimations for Dollar Store in Rural Areas: Density Results(T=6, N=130)

Variables	**Density**								
	OLS			**REM**			**FEM**		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
(Intercept)	52.438	***	(14.512)	50.183	**	(23.532)			
Free and reduced lunch	-0.020		(0.016)	-0.004		(0.020)	0.019		(0.028)
African American	0.091	***	(0.014)	0.091	***	(0.023)	0.090		(0.122)
Hispanic	0.057	*	(0.032)	0.055		(0.053)	0.080		(0.163)
Rural places	-0.182		(0.142)	-0.178		(0.233)			
High school degree or equivalent	-0.107	**	(0.043)	-0.107		(0.076)			
Bachelors and advanced degree	-0.569	***	(0.070)	-0.536	***	(0.111)			
Pre-kindergarten and kindergarten	-0.033	*	(0.017)	-0.028	*	(0.016)	-0.032	*	(0.017)
First and second grade	-0.002		(0.020)	0.002		(0.018)	0.000		(0.019)
Third and fourth grade	0.001		(0.019)	0.021		(0.022)	0.026		(0.023)
NearDollar_1mile	2.244	***	(0.843)	0.930		(1.365)	-0.754		(2.177)
NearDollar_1to2mile	-4.048	**	(1.595)	-3.144		(2.439)	-1.904		(4.008)
NearDollar_2to5mile	0.188		(0.469)	0.053		(0.781)	-0.633		(1.979)
Adjusted R-squared	0.238			0.100					
Hausman test (chisq)				13.009					
LM bp test (chisq)	320.71	***							

Note: Robust standard errors are in parentheses

Note: NearDollar_1mile represents the density of large grocery store within one mile buffer around children's neighborhood block center.

NearDollar_1to2mile represents the density of large grocery store within the donut area of two mile buffer minus one mile buffer.

NearDollar_2to5mile represents density of large grocery store within the donut area of five mile buffer minus two mile buffer.

Table 9b. Balanced Panel Estimations for Dollar Store in Rural Areas:Distance Results (T=6, N=130)

Variables	**Distance**								
	OLS			**REM**			**FEM**		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
(Intercept)	56.116	***	(13.110)	52.476	***	(20.155)			
Free and reduced lunch	-0.023		(0.016)	-0.005		(0.020)	0.020		(0.021)
African American	0.097	***	(0.013)	0.093	***	(0.023)	0.082		(0.065)
Hispanic	0.058	**	(0.032)	0.053		(0.052)	0.064		(0.098)
Rural places	-0.118	***	(0.042)	-0.202		(0.199)			
High school degree or equivalent	-0.207		(0.128)	-0.112		(0.077)			
Bachelors and advanced degree	-0.588	***	(0.070)	-0.544	***	(0.112)			
Pre-kindergarten and kindergarten	-0.035	**	(0.017)	-0.030	*	(0.016)	-0.032	*	(0.017)
First and second grade	-0.004		(0.020)	0.002		(0.018)	0.000		(0.019)
Third and fourth grade	-0.001		(0.019)	0.022		(0.022)	0.026		(0.023)
Log(DistDollar)	-0.043		(0.294)	0.2147		(0.444)	0.613		(0.588)
Adjusted R-squared	0.2333			0.0982					
Hausman test (chisq)				13.938	*				
LM bp test (chisq)	325.3	***							

Note: Robust standard errors are in parentheses

Note: Log (DistDollar) represents the log form of distance to nearest dollar from children's neighborhood block center.

Table 10. Panel Models: Expanded Density Results with Fast Food Restaurants(T=6,N=384)

Variables	**OLS**			**REM**			**FEM(within)**		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
(Intercept)	23.683	***	(1.601)	25.707	***	(2.637)			
Free and reduced lunch	0.054	***	(0.007)	0.037	***	(0.009)	0.009		(0.014)
African American	0.046	***	(0.005)	0.054	***	(0.008)	0.061	**	(0.029)
Hispanic	0.088	***	(0.009)	0.096	***	(0.014)	0.115	***	(0.036)
Rural places	0.016	***	(0.004)	0.020	***	(0.007)			
High school degree or equivalent	-0.092	***	(0.022)	-0.119	***	(0.037)			
Bachelors and advanced degree	-0.138	***	(0.019)	-0.178	***	(0.031)			
Pre-kindergarten and kindergarten	-0.057	***	(0.007)	-0.042	***	(0.007)	-0.034	***	(0.008)
First and second grade	-0.014	**	(0.007)	-0.007		(0.007)	-0.003		(0.008)
Third and fourth grade	0.009		(0.008)	0.014	*	(0.008)	0.017	**	(0.008)
NearGrocery_1mile	-0.169		(0.155)	-0.344		(0.214)	-0.595	*	(0.309)
NearGrocery_1to2mile	-0.095		(0.104)	0.052		(0.149)	0.311		(0.227)
NearGrocery_2to5mile	-0.074	***	(0.028)	-0.043		(0.042)	0.021		(0.071)
NearDollar_1mile/NearGrocery_1mile	0.001		(0.003)	0.003		(0.003)	0.004		(0.004)
NearConv_1mile/NearGrocery_1mile	-0.001		(0.001)	-0.001		(0.001)	-0.002		(0.001)
NearDrug_1mile/NearGrocery_1mile	0.008	***	(0.002)	0.005	*	(0.002)	0.004		(0.003)
NearFastFood_1mile/NearGrocery_1mile	-0.001		(0.002)	-0.002		(0.002)	-0.003		(0.002)
Adjusted R-squared:	0.337			0.147			0.017		
Hausman test (chisq)				22.629	***				

Note: NearFastFood_1mile/NearGrocery_1mile represents the density of fast food restaurants relative to large grocery stores within one mile buffer.

CHAPTER 6: DISCUSSIONS AND CONCLUSIONS

This thesis set out to examine whether access to different types of food stores would have any impact over childhood obesity. OLS, random effects and fixed effects model methods were then applied, and according to the Hausman test, the fixed effects model which accounts for unobserved school specific effect fits the data best. As spatial dependence was suspected in the variables of overweight prevalence and other characteristics such as race, education and income, spatial panel methods were employed and compared with classical panel estimations. Finally, analysis focused on dollar stores in rural areas were conducted in an effort to answer whether dollar stores contribute to childhood obesity in rural areas.

One major finding of this thesis is that density of large grocery stores within a one mile radial distance from neighborhood block centers do exert an impact on childhood obesity prevalence, however this influence varies across specifications. To be more specific, a one store increase in the density will bring down childhood obesity prevalence by 0.58 percent. On the other hand, associations between densities of large grocery stores within farther buffers and children's overweight status were not found in my analyses. This might indicate that parents make shopping decisions primarily based on the density of large grocery stores close to home, and once the closest store that sells fresh produce is farther than one mile, they do not care about the density any more. One possible explanation for the different impact between densities of large grocery stores within one mile and farther buffers might be the different cost of walking and driving. Taken that an average adult can walk for one mile in about 18 minutes, stores within a one mile radial distance from home can be easily visited by walking, whereas farther stores might force the parents to drive. Once the parents are driving, the cost of going to a farther store is greatly reduced. Assuming the average driving speed is 40 mph, it only takes approximately an additional 1.5 minutes for the parents to shop at a store that is one-mile farther from home. Therefore, I suspect the results shown in this analysis can be attributed to means of shopping transportation. However, my explanation could be violated if most people in a neighborhood which have at least one large grocery store within a one mile radial distance, own cars and do not choose to walk even if the store is very close. It is also interesting to find that distance from neighborhood block centers to closest large store does not seem to affect children's BMI. This is to say parents shopping patterns are generally not affected by the closest large grocery store available but the density of shops within one mile buffer zone which offer fresh produce. In other words, parents' shopping preference might play a more important role than access issues when the store is really close to residential address.

Presence of dollar stores and convenience stores although suspected as contributors to childhood obesity, showed no statistical associations with children's obesity prevalence. In addition, subsample analysis of rural school service areas backs up the finding of no impact of dollar stores on rural children's weight outcomes.

The spatial panel estimation suggests neither the assumed spatial lagged term nor spatial autocorrelation in error term were present in my data. Although the spatial method increased significance of estimates of interest, the spatial effects were extremely small and strongly statistically insignificant.

Finally, this study is similar to previous literature in finding that children in African American and Hispanic neighborhoods are more likely to be overweight. It is also discovered that younger children in pre-kindergarten and kindergarten are less likely to be overweight than their older counterparts. One result that is not in line with the literature is that once school specific effects are considered, income, which is approximated by participation of free and reduced lunch, shows no impact over BMI. It is worth noting that in the analysis of dollar stores in rural areas, the fixed effects model, which takes unobserved school hetrogeniety into consideration, is no longer preferred by the Hausman test. This might hint that the school unobserved effects are largely due to the local income level, and once controlled for, income alone will not play a role.

There are also several limitations of this thesis which hopefully could be addressed by later studies. For example, this analysis focused on food environment at the school level and measures of access to large grocery store is based on block level then aggregated to each school. Although I believe this is a finer measure than direct count of stores within school service areas, it still has neglected much individual information. Actually this might be the reason why spatial effect did not show up in the analysis. The transmission of behavior that can be linked to health outcome through geography might be within school boundary instead of between schools. Therefore, future studies might be needed to employ a more refined level-individual level data in order to check the spatial pattern of the childhood obesity problem. Another limitation of this thesis is that because of software limitations, varying spatial weight matrices across years could not be incorporated into the spatial panel frame, whereas in reality, the school service areas do vary over the years. Failing to account for this will result in biased estimates and wrong statistical inference. So, further studies using spatial panel models in R should make an effort to adjust the splm package to allow for time-varying spatial weight matrices and unbalanced panel estimation. In addition, if the study is to be based on individual level data, no physical boundary is available to construct spatial weight matrix based on contiguity, and thus other method of drawing the weight matrix might be required and should be justified. Finally, in the literature, prices of fast food, prices of food at home and physical facilities are sometimes incorporated into food environment analysis since they all potentially have explanatory power in deciding children's weight status. The results in the literature are rather mixed. For example, Powell (2009) has found that price of fast food is significantly negatively correlated with childhood obesity, whereas price of food at home and exposure to physical facilities do not receive any statistical significance in explaining adolescents' BMI. If the food prices and physical facilities do play a role in determining children's health outcome, omitting these measures will result in model misspecifications and thus wrong estimates (if omitted variables are correlated with regressors) and inference. Therefore, future studies need to include food prices and physical facilities data to obtain a more comprehensive analysis in order to detect true determinants of childhood obesity.

Back to the research question raised at the beginning: Does food environment play a role in determining childhood obesity? The analysis presented in this thesis reveals one possible answer: Yes, food environment does play a role but perhaps not a strong and essential role. If that is indeed the case, it is worth thinking twice for the policy makers to decide targeting food desert issue as a strategy to prevent childhood obesity. Of course, the results from this thesis are far from being conclusive. Given the importance of the childhood obesity issues, there is need for more studies to be conducted in the future to offer a well-founded policy direction based on extensive evidence.

REFERENCES

Abdollah, T. 2007. "A strict order for fast food." *Los Angeles Times,* September. http://articles.latimes.com/2007/sep/10/local/me-fastfood10/2 (assessed May 5, 2012)

Alviola, P., R. M. Nayga, Jr., M. Thomsen, D. Danforth, and J. Smartt. 2011. "The Effect of Fast-Food Restaurants on School-Level Obesity." Working Paper, Dept. of Agricultural Economics and Agribusiness, University of Arkansas, Fayetteville.

Anselin, L., and D. Arribas-Bel. 2011. "Spatial Fixed Effects and Spatial Dependence." Working Paper, GeoDa Center for Geospatial Analysis and Computation, Arizona State University, Tempe.

Babey, S. H., A. L. Diamant, T. A. Hastert, and S. Harvey. 2008. "Designed for Disease: The Link Between Local Food Environments and Obesity and Diabetes." Recent Work, UCLA Center for Health Policy Research, UC Los Angeles.

Baker, J. L., L. W. Olsen, and T .A. Sorensen. 2007. "Childhood Body-Mass Index and the Risk of Coronary Heart Disease in Adulthood." *The New England Journal of Medicine* 357(23): 2330-2337.

Black, J. L., J. Macinko, L. B. Dixon, and G. E. Fryer, Jr. 2009. "Neighborhoods and Obesity in New York City." *Health & Place* 16(3)489-499.

Brennan, D., and C. Carpenter. 2009. "Proximity of Fast-Food Restaurants to Schools and Adolescent Obesity." *American Journal of Public Health* 99(3):505-510.

Chen, S. E., and R. J. G. M. Florax. 2010. "Zoning for Health: The Obesity Epidemic and Opportunities for Local Policy Intervention." *The Journal of Nutrition* 140:1181–1184.

Chen, S., R. J. G. M. Florax, S. Snyder, and C. C. Miller. 2010. "Obesity and Access to Chain Grocers." *Economic Geography* 86(4):431-452.

Christakis, N. A., and J. H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *The New England Journal Medicine* 357:370-9.

Currie, J., S. D. Vigna, E. Moretti, and V. Pathania. 2009. "The Effect of Fast Food Restaurants on Obesity and Weight Gain." Working Paper, National Bureau of Economic Research.

Donkin, A. J.M., E.A. Dowler, S. J. Stevenson, and S. A. Turner. 1999. "Mapping access to food at a local level." *British Food Journal* 101(7): 554-564.

Dun & Bradstreet. 1999. *Samples and Descriptions: Prospecting Records Layout*. https://www.dnb.com/products/samples/dmilayout.htm (accessed June 12, 2011).

Eid, J., H. G. Overman, D. Puga, and M. A. Turner. 2008. “Fat city: Questioning the Relationship between Urban Sprawl and Obesity.” *Journal of Urban Economics* 63:385-404.

Frank, L., K. Glanz, M. McCarron, J. Sallis, B. Sealens, and J. Chapman. 2006. “The Spatial Distribution of Food Outlet Type and Quality around Schools in Differing Built Environment and Demographic Contexts.” *Berkeley Planning Journal* 19(1): 79-95.

Freedman, D. S., L. K. Khan, M. K. Serdula, W. H. Dietz, S. R. Srinivasan, and G.S. Berenson. 2005. “The Relation of Childhood BMI to Adult Adiposity: The Bogalusa Heart Study.” *Journal of the American Academy of Pediatrics* 115(1):22-27.

Freedman, D., Z. Mei, S. R. Srinivasan, G.S. Berenson, and W. H. Dietz. 2007. “Cardiovascular Risk Factors and Excess Adiposity among Overweight Children and Adolescents: The Bogalusa Heart Study.” *The Journal of Pediatrics* 150(1):12-17.

Galvez, M. P., K. Morland, C. Raines, J. Kobil, J. Siskind, J. Godbold, and B. Brenner. 2007. “Race and Food Store Availability in an Inner-City Neighborhood.” *Public Health Nutrition* 11(6):624–631.

Handy, S. L., and Clifton, K. J. 2001. “Local Shopping as a Strategy for Reducing Automobile Travel.” *Transportation* 28:317–46.

Harris, D. E., J. W. Blum, M. Bampton, L. M. O’Brien, C. M. Beaudoin, M. Polacsek, and K. A. O’Rourke. 2011. “Location of Food Stores near Schools does not Predict the Weight Status of Maine High School Students.” *Journal of Nutrition Education and Behavior* 43(4):274-278.

Ingelsson, E., L. M. Sullivan, C. S. Fox, J. M. Murabito, E. J. Benjamin, J. F. Polak, J. B. Meigs, M. J. Keys, C. J. O’Donnell, T. J. Wang, R. B. D’Agostino, P. A. Wolf, and R. S. Vasan. 2007. “Burden and Prognostic Importance of Subclinical Cardiovascular Disease in Overweight and Obese Individuals.” *Journal of American Heart Associations* 116:375-384.

Kaufman, P. 1999. “Rural Poor Have Less Access to Supermarkets, Large Grocery Stores.” *Rural Development Perspectives* 13 (3): 19-26.

Lee. H. 2012. “The Role of Local Food Availability in Explaining Obesity Risk among Young School-aged Children.” *Social Science & Medicine* 74:1193-1203.

Li, F., P. Harmer, B. J. Cardinal, M. Bosworth, and D. Johnson-Shelton. 2009. "Obesity and the Built Environment: Does the Density of Neighborhood Fast-Food Outlets Matter?" *American Journal of Health Promotion* 23(3):203-209.

Liese, A. D., N. Colabianchi, A. P. Lamichhane, T. L. Barnes, J. D. Hibbert, D. E. Porter, M. D. Nichols, and A. B. Lawson. 2010. "Validation of 3 Food Outlet Databases: Completeness and Geospatial Accuracy in Rural and Urban Food Environments." *American Journal of Epidemiology* 172(11):1324-1333.

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

Maddock, J. 2004. "The Relationship between Obesity and the Prevalence of Fast Food Restaurants: State-Level Analysis." *American Journal of Health Promotion* 19(2)137-143.

Mehta, N. K., and V. W. Chang. 2008. "Weight Status and Restaurant Availability: A Multilevel Analysis." *American Journal of Preventive Medicine* 34(2):127–133.

Mobley, L. R, E. A. Finkelstein, O. A. Khavjou, and J. C. Will. 2004. "Spatial Analysis of Body Mass Index and Smoking Behavior among WISEWOMAN Participants." *Journal of Women's Health* 13(5): 519-528.

Moore, L. V., and A. V. Diez Roux. 2006. "Associations of Neighborhood Characteristics With the Location and Type of Food Stores." *American Journal of Public Health* 96(2)325-331.

Morland, K., A. V. Diez Roux, and S. Wing. 2006. "Supermarkets, Other Food Stores, and Obesity: The Atherosclerosis Risk in Communities Study." *American Journal of Preventive Medicine* 30(4):333–339.

Morland, K., S. Wing, A. Diez Roux, and C.s Poole. 2002. "Neighborhood Characteristics Associated with the Location of Food Stores and Food Service Places." *Journal of Preventive Medicine* 22(1):23-29.

Ogden, C., and M. Carroll. 2010. " Prevalence of Obesity Among Children and Adolescents: United States, Trends 1963–1965 Through 2007–2008." Division of Health and Nutrition Examination Surveys, National Center for Health Statistics.

Powell, L. M. 2009. "Fast Food Costs and Adolescent Body Mass Index: Evidence from Panel Data." *Journal of Health Economics 28*: 963-970.

Powell, L. M., E. Han, S. N. Zank, T. Khan, C. M. Quinn, K. P. Gibbs, O. Pugach, D. C. Barker, E. A. REsnick, J. Myllyluoma, and F. J. Chaloupka. 2011. "Field Validation of Secondary

Commercial Data Sources on the Retail Food Outlet Environment in the U.S." *Health & Place* 17(5):1122-1131.

Powell, L. M., M. C. Auld, F. J. Chaloupka, P. M. O'Malley, L. D. Johnston. 2007. "Associations between Access to Food Stores and Adolescent Body Mass Index." *American Journal of Preventive Medicine* 33(4S): S-301-S-307.

Rundle, A., K. M. Neckerman, L. Freeman, G. S. Lovasi, M. Purciel, J. Quinn, C. Richards, N. Sircar, and C. Weiss. 2009. "Neighborhood Food Environment and Walkability Predict Obesity in New York City." *Environmental Health Perspectives* 117(3):442–447.

Sallis, J. F., P. R. Nader, J. W. Rupp, C. J. Atkins, and W.C. Wilson. 1986. "San Diego Surveyed for Heart-Healthy Foods and Exercise Facilities." *Public Health Report* 101(2):216-219.

Sorof, J., and S. Daniels. 2002. "Obesity Hypertension in Children: A Problem of Epidemic Proportions." *Journal of the American Hearts Associations* 40:441-447.

Sturm R., and A. Datar. 2005. "Body Mass Index in Elementary School Children, Metropolitan Area Food Prices and Food Outlet Density." *Public Health* 119:1059–1068.

Sturm, R., and D. A. Cohen. 2009. "Zoning for Health? The Year-Old Ban on New Fast-Food Restaurants in South LA." *Health Affairs* 28(6): 1088-1097.

U.S. Census Bureau. 2010. *Resident Population Data - 2010 Census.* http://2010.census.gov/2010census/data/apportionment-dens-text.php# (Accessed June 3, 2012).

Wang, M. C., S. Kim, A. A., Gonzalez, K. E., MacLeod, and M. A. Winkleby. "Socioeconomic and Food-Related Physical Characteristics of the Neighborhood Environment are Associated with Body Mass Index." 2007. *Journal of Epidemiology Community Health* 61:491–498.

Wang, Z. 2010. "Call Records of Large Grocery Stores in Arkansas." Unpunished, Dept. of Agricultural Economics and Agribusiness, University of Arkansas, Fayetteville.

Zenk, A. N., A. J. Schulz, T. Hollis-Neely, R. T. Campbell, N. Holmes, G. Watkins, R. Nwankwo, and A.Odoms-Young. 2005. "Fruit and Vegetable Intake in African Americans: Income and Store Characteristics." *American Journal of Preventive Medicine* 9(1):1-9.

Zick, C. D., K. R. Smith, J. X. Fan, B. B. Brown, I. Yamada, and L. Kowaleski-Jones. 2009. "Running to the Store? "The Relationship between Neighborhood Environments and the Risk of Obesity." *Social Science and Medicine* 69(10):1493–1500.