ASSESSING AFFORDABILITY OF FRUITS AND VEGETABLES IN THE BRAZOS VALLEY-TEXAS

A Dissertation

by

JUSTUS LOTADE-MANJE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2011

Major Subject: Agricultural Economics

Assessing Affordability of Fruits and Vegetables in the Brazos Valley-Texas

ASSESSING AFFORDABILITY OF FRUITS AND VEGETABLES IN THE BRAZOS VALLEY-TEXAS

A Dissertation

by

JUSTUS LOTADE-MANJE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee:	Richard A. Dunn
Committee Members:	Rodolfo M. Nayga, Jr.
	Alan Love
	Joseph Sharkey
	Ximing Wu
Head of Department:	John P. Nichols

December 2011

Major Subject: Agricultural Economics

ABSTRACT

Assessing Affordability of Fruits and Vegetables in the Brazos Valley-Texas.

(December 2011)

Justus Lotade-Manje, D.U.E.S. I&II.; M.S., Université Abdou Moumouni de Niamey, Niger; M.S., Colorado State University-Fort Collins, Colorado

Chair of Advisory Committee: Dr. Richard A. Dunn

The burden of obesity-related illness, which disproportionately affects low income households and historically disadvantaged racial and ethnic groups, is a leading public health issue in the United States. In addition, previous research has documented differences in eating behavior and dietary intake between racial and ethnic groups, as well as between urban and rural residents. The coexistence of diet-related disparities and diet-related health conditions has therefore become a major focus of research and policy. Researchers have hypothesized that differences in eating behavior originate from differing levels of access to and affordability of healthy food options, such as fresh fruits and vegetables. Therefore, this dissertation examines the affordability of fresh produce in the Brazos Valley of Texas.

This study uses information on produce prices collected by taking a census of food stores in a large regional area through the method ground-truthing. These are combined with responses to a contemporaneous health assessment survey. Key innovations include the construction of price indices based on economic theory, testing

the robustness of results to different methods of price imputation, and employing spatial econometric techniques.

In the first part of the analysis, I evaluate the socioeconomic and geographical factors associated with the affordability of fresh fruits and vegetables. The results based on Ordinary Least Squares (OLS) regression show that except housing values (as median value of owner-occupied units) and store type, most factors do not have significant effects on the prices for these food items. In addition, the sizes and signs of the coefficients vary greatly across items. We found that consumers who pay higher premiums for fresh produce reside in rural areas and high proportion of minorities neighborhoods. We then assess how our results are influenced by different imputation methods to account for missing prices. The results reveal that the impacts of the factors used are similar regardless of the imputation methods. Finally, we investigate the presence of spatial relationships between prices at particular stores and competing stores in the neighborhoods. The spatial estimation results based on Maximum Likelihood (ML) indicate a weak spatial correlation between the prices at stores located near each others in the neighborhoods. Stores selling vegetables display a certain level of spatial autocorrelation between the prices at a particular store and its neighboring competitors. Stores selling fruits do not present such relations in the prices.

DEDICATION

To my dad, mom, brothers and sisters, my committee members, my friend Mourdoumngar, and all those who supported me spiritually, morally, financially and materially through the years.

ACKNOWLEDGEMENTS

It is an honor for me to take this opportunity to thank all my committee members for all their support and encouragement. I would like to address a special thanks to Dr. Nayga for his guidance, kind words and willingness to provide me with necessary resources to get to this point. I dearly appreciate the financial and also material support he provided me. I am thankful for having Dr. Dunn as my Chair, and, for that, I thank Dr. Nayga for his wisdom. Dr. Dunn has been a mentor and fundamental guide who never stopped communicating and helping. Dr. Dunn opened my eyes on academic approaches and was always there to graciously help. I present Dr. Sharkey all my gratitude for the data, encouragement and financial support. It is important to mention that without all the support from Dr. Sharkey, I would not have a dissertation on this topic. Also I wish to thanks Dr. Love for his help, comments and encouragement. I would like to thank Dr. Wu for his time and his insight for improving my dissertation. Special thanks to Dr. Maria Loureiro for all her help throughout the years.

Also, I will not end without saying a special thanks to Dr. Leatham for his support, Dr. Siebert for his guidance and Vicki Heard for all the assistance. In addition, I offer special recognition for the assistance and coaching of the Thesis Office on the 6th floor, The University Writing Center and GIS offices located on the second floor of Evans library at Texas A&M University.

I would also like to thank all my colleagues with whom I had to spend nights studying for exams and also who shared with me some class materials.

Aussi, un grand remerciement a ma mere, mes frères et soeurs pour leur encouragements, leur patience, and leur amour pour durant toutes ces annees que j'ai passé sans loin d'eux.

Finally, I acknowledge that all remaining errors are my own.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Many researchers have hypothesized that differences in eating behavior originate from differing levels of access to healthy food options (Andreyeva et al. 2008; Inagami et al. 2006; Morland et al. 2002a; Morland et al. 2002b; Rose and Richard 2004). For example, consumption of fruit and vegetables is recommended through the Dietary Guidelines for Americans (U.S. Department of Health and Human Services and U.S. Department of Agriculture 2005), but these foods are often not easily accessible by racial and ethnic minority groups in large urban centers or populations in rural areas (Dubowitz et al. 2008; Liese et al. 2007; Morland and Filomena 2007; Morton and Blanchard 2007; Powell et al. 2007b; Sharkey and Horel 2008; Shaw 2006; McClelland et al. 1998; Zenk et al. 2005; Zenk et al. 2006). Along with reduced access, fresh fruit and vegetables may also be less affordable to rural populations and racial/ethnic minority groups (Morland and Filomena 2007; Liese et al. 2007; Ard et al. 2007; Block and Kouba 2006; Ball, Timperio, and Crawford 2009). Therefore, the aim of this dissertation is to examine the affordability of fresh fruits and vegetables in the Brazos Valley region of Texas.

This study uses information on produce prices collected by taking a census of food stores in a large regional area through the method ground-truthing. These are combined with responses to a contemporaneous health assessment survey. Key

---

This dissertation follows the style of *American Journal of Agricultural Economics*.

innovations include the construction of price indices based on economic theory, testing the robustness of results to different methods of price imputation, and employing spatial econometric techniques.

This study is divided into three parts that comprise Chapters II-IV of the dissertation:

Chapter II: Previous research has documented differences in affordability of healthy food items according to the demographic and socio-economic profiles of neighborhoods. Therefore, the first part of the dissertation investigates whether stores located in rural areas, in neighborhoods with lower socio-economic status, and in neighborhoods with higher proportions of African-American and Hispanic residents charge more for fresh fruit and vegetables. The results show that proportion of minority residents is positively associated with the cost of purchasing fresh produce. Convenience stores tend to sell these items at higher prices compared to supermarkets and grocery stores. In addition, rural consumers (those who live outside Brazos County) also tend to pay higher prices.

Chapter III: This chapter examines the robustness of estimation results to the choice of imputation method when price data is missing. The price indices presented in Chapter II are constructed using three different imputation methods: zero, mean and regression imputation. The regression analysis of Chapter II is repeated using each of these methods and the resulting coefficient estimates are then compared. I find that there is no meaningful effect of the imputation method on the conclusions of the study.

Chapter IV: This chapter examines the spatial relationships between fruits and vegetables prices among stores in proximity to each other. The main objective was to determine whether store prices are correlated based on geographic location once confounding neighbor characteristics were controlled. The results show the presence of a weak spatial relationship between stores selling vegetables, but not fruits. Stores that sell vegetables at relatively low prices tend to be located near stores that sell vegetables at relatively high prices. This suggests that stores differentiate themselves based on characteristics beside price.

These three chapters provide better understanding of the factors determining the affordability of healthy food items in a highly diverse—demographically and socio-economically—rural area of the southern United States. Given the many public health challenges facing these populations, policy-makers can use these results to benefit historically disadvantaged populations.

CHAPTER II

DETERMINATION OF FACTORS AFFECTING AFFORDABILITY OF HEALTHY FOOD IN THE RURAL COUNTIES OF BRAZOS VALLEY

2.1 Introduction

The burden of obesity-related illness, which disproportionately affects low income households and historically disadvantaged racial and ethnic groups, is a leading public health issue in the United States (Mokdad et al. 2003; Ogden et al. 2006). In addition, previous research has documented differences in eating behavior and dietary intake between racial and ethnic groups (Dubowitz et al. 2008). The coexistence of diet-related disparities and diet-related health conditions has therefore become a major focus of research and policy (Satia et al. 2009).

Many researchers have hypothesized that differences in eating behavior originate from differing levels of access to healthy food options (Andreyeva et al. 2008; Inagami et al. 2006; Morland et al. 2002a; Morland et al. 2002b; Rose and Richard 2004). For example, consumption of fruit and vegetables is recommended through the Dietary Guidelines for Americans (U.S. Department of Health and Human Services and U.S. Department of Agriculture 2005), but these foods are often not easily accessible by racial and ethnic minority groups in large urban centers or populations in rural areas (Dubowitz et al. 2008; Liese et al. 2007; Morland and Filomena 2007; Morton and Blanchard 2007; Powell et al. 2007b; Sharkey and Horel 2008; Shaw 2006; McClelland et al. 1998; Zenk et al. 2005; Zenk et al. 2006). Along with reduced access, fresh fruit and vegetables may also be less affordable to rural populations and racial/ethnic

minority groups (Morland and Filomena 2007; Liese et al. 2007; Ard et al. 2007; Block and Kouba 2006; Ball, Timperio, and Crawford 2009). Therefore, the aim of this paper is to examine whether stores located in rural areas, in neighborhoods with lower socio-economic status, and in neighborhoods with higher proportions of African-American and Hispanic residents charge more for fresh fruit and vegetables.

Our approach is novel in several respects. Unlike previous work, our information on the price of fresh fruits and vegetables comes from data collected by taking a census of food stores in a large regional area through the method ground-truthing. Moreover, this region in central Texas is home to a socio-economically and demographic diverse population spread over six rural counties and one medium-sized urban county. Finally, we handle missing prices through an imputation strategy more firmly grounded in the economic theory surrounding the decision of a profit-maximizing store owner to stock a particular item for sale.

2.2 Data and Methods

2.2.1 Data

The seven contiguous counties of the Brazos Valley are situated between the Dallas and Houston metropolitan areas. The region is home to 300,000 residents, of which 51.4% reside in one of six predominately rural counties. The seventh, Brazos County, includes the medium-sized urban center of Bryan-College Station. Socioeconomic characteristics were extracted from the 2000 decennial census Summary Files 3 (SF-3) at the level of the census block group (CBG) since the CBG is the smallest unit of census geography for which the detailed “long-form” social and economic data

from the census are tabulated (17, 23). SF-3 data were merged for the six rural counties. The rural areas included 101 CBG and the urban county included 93 CBG.

Table 2.1 presents selected socio-economic information for the counties in the Brazos Valley from the US Census Bureau. Median household income ranges from just

Table 2.1. By County Demographic Repartition and Store Per Type.

	Brazos	Burleson	Grimes	Leon	Madison	Robertson	Washington
Population	170,954	16,598	25,603	16,462	13,379	15,819	32,034
Median income ($)	33,186	31,174	33,327	29,443	28,963	29,983	35,852
Bachelor's degree (%)	37	13.2	10.3	12.1	11.5	12.7	19.0
Black (%)	10.7	14.3	18.2	10.1	21.8	22.9	17.8
Hispanic (%)	20.8	16.5	18.2	10.9	18.9	16.8	11.6
Store types							
Supermarket	11	2	2	1	2	3	2
Grocery	3	3	2	4	0	2	0
Convenience	114	19	25	25	12	18	32

Source: US Census Bureau and BVFEP.

over $34,000 in Madison County to nearly $45,000 in Washington County. Robertson County has the largest proportion of Blacks at 22.9%, while neighboring Leon County is only 10.1% Black. Leon County also has the lowest percentage of Hispanic residents. In contrast, Hispanics account for 20.8% of the population in Brazos County, the largest

county in the region. As expected, it also has the most education population with 37% holding at least a Bachelor's Degree. Thus, the Brazos Valley region allows us to study the effects of urbanicity, education levels, income levels and demographic make-up on affordability within a compact, contiguous area.

Information about prices comes from the Brazos Valley Food Environment Project (BVFEP). As part of the BVFEP, trained observers enumerated all food stores and food service places by driving all Interstates, US Highways, Texas State Highways, Texas Farm-to-Market Roads and other major thoroughfares to locate all stores that could sell food items (Sharkey and Horel 2008). As previously published, the BVFEP used ground-truth methods in a two-stage approach to determine the location of all food stores and the availability of fresh produce. Identification and surveying occurred between September 2006 and July 2007. Food stores were classified into several categories: supercenter, supermarket, grocery store, convenience store, dollar store, mass merchandiser, and pharmacy. The BVFEP indentified 2 supercenters, 22 supermarkets, 14 grocery stores and 254 convenience stores across the seven counties (their geographic distribution is reported in Table 2.1). Investigators then entered all food stores with an extensive list of food items on a tally sheet in order to catalogue which items were sold and at what price (Bustillos et al. 2009). Based on input from local residents and nutrition professionals, ten types of fresh fruit (apples, avocado, bananas, berries, grapes, mango, melons, oranges, peaches and pears) and eleven types of fresh vegetables (broccoli and cauliflower, carrots, corn, green beans, leafy greens, lettuce, okra, onions, potatoes, squash and tomatoes) were included in this catalogue. The following

information was recorded: whether each type (e.g. apples) was available for purchase; the number of varieties of each type of fruit or vegetable available for purchase; and the lowest-priced variety of each type of fruit or vegetable. Because in-store prices were posted in several forms—per item, per ounces, per pound—all prices were later transformed or recalculated into a uniform price per pound. To do so when prices were posted per item, surveyors weighed the items using a sensor scale. The price of food items that were either not sold or not displayed were recorded as missing. Contrary to earlier studies (Latham and Moffat 2007; Cummins and Macintyre 2002), surveyors did not purposefully interact with store managers or employees during the data collection process. Although there are many fruit and vegetable varieties that were not included in the BVFEP, the 9 fruits used here accounted for 80% of consumption and expenditure according to the Fresh Look Marketing data. Lemons, limes and tangerines are the most commonly consumed fruits not included. The ten vegetables account for 72% of all fresh whole vegetable expenditure and 75% of consumption. The most common varieties not included are celery, cucumber, mushrooms and peppers. We omit okra and mango from the subsequent analysis given their limited availability and low consumption shares.

It is worth noting that the ground-truthing methods utilized here differ from those typically employed in the literature. The direct observation approach more closely approximates the behavior of actual shoppers. When consumers want to buy a food product, they do not call stores to ask for the price of the items; they walk or drive to the stores. In addition, when the consumers go to the stores, they rarely ask for prices. They directly go to the section where the produces are displayed and buy the items needed or a

substitute if price is high. A full census of stores is another superlative aspect of our data. Previously, many researches use secondary data either available online or via third parties. These data tend to be targeted at only a subset of stores (Cassady et al. 2007 Morland and Filomena 2007; Morris et al. 1992;), locations (Block and Kouba 2006; Hendrickson et al. 2006; Jetter and Cassady 2006; Morris et al. 1992) and types of food (Liese et al. 2007). These data are therefore less able to accurately describe the actual food environment.

2.2.2 Price Imputation

In previous studies, missing prices have typically been imputed by taking the mean price over the stores that do sell the item (Block and Kouba 2006; Lee et al. 2002). This imputation strategy may be misguided since stores that do not sell a particular item are likely not comparable to the average store that does. Instead, one could view the decision of a store-owner not to stock an item for sale as the result of profit-maximizing behavior. The price that consumers are willing to pay for the missing type is below the cost that a store owner faces to offer the type for sale. Nevertheless, there is still some reservation price that would lead the store owner to stock the item. Thus, the proper price for missing types is this unobserved *shadow price* and intuitively, it should be higher than the mean observed price. Ignoring the underlying reasons for missing price information may not be benign to the purpose at hand. If economically disadvantaged neighborhoods suffered from both low availability and affordability of fresh produce, then using the mean price calculated from stores in economically advantaged

neighborhoods would understate the true relationship between economic status and affordability.

Alternatively, one may take a statistical perspective and assume that over some period of time, all stores eventually stock all types of fruit. Since inventory and prices are only observed once for each store, it is possible that unobserved fruit types were recently sold or will be sold again soon. The goal is then to reasonably estimate these unobserved prices given the observable price data. Both the economic and statistical perspectives suggest that one method to overcome this issue is to employ a price imputation that takes advantage of the observed prices in each store. For example, a store that sells apples and bananas above the mean price found in others stores would likely charge an above average price for berries, as well. Therefore, in the current study the price of each fruit or vegetable item was first estimated as a linear function of the store type, the county in which the store was located and the prices of the most common fruit or vegetable types—apples, oranges and bananas for fruit and onions, potatoes and tomatoes for vegetables. The coefficient estimates from these regressions, which are available in a supplemental appendix, were then used to impute values for the missing prices of other types (Bradley 2003).

2.2.3 Price Indices

The actual and imputed prices (the former when available, the latter when missing) were then used to calculate two types of price indices for both fresh fruit and fresh vegetables: a *high variety* and a *basic* index. The high variety index includes the full set of items, while the basic index includes only the most common items (apples,

bananas and oranges for fruit and carrots, lettuce, onion, potatoes and tomatoes for vegetables). Such a distinction may be relevant for policy-makers, for example, if the intention of a particular program is to provide assistance to increase the amount of fruit and vegetable consumption with little importance attached to the variety of types consumed. Each index is calculated as the weighted mean price per pound multiplied by the recommended number of pounds consumed per week for a representative family of two adults and two children from the most recent USDA Thrifty Food Plan: 24.5 pounds of fruit and 31.5 pounds of vegetables. The Thrifty Food Plan expects that fruit and vegetable consumption will also come from a mix of sources (e.g. fresh whole, frozen, canned, dried, etc.), but the choice of multiplicative factors affects the magnitude of the subsequent coefficient estimates, but not their statistical significance. If only half of total fruit consumption should come from fresh whole items, then the appropriate adjustment is to either halve the coefficient estimate or reinterpret it as biweekly expenditure. The weights are equal to the consumption shares calculated for the Dallas metropolitan area from Fresh Look Marketing, Inc. (Chicago, Illinois) and represent all supermarkets (sales of at least $2 million) with about 70% of all commodity volume (ACV) in the Dallas market (Timothy Richards, personal communication).

The price indices calculated for each store were linked to socio-economic information for the CBG in which the store was located from the 2000 decennial Census (U.S. Census Bureau. 2009 [Available from: http://www.census.gov]). To explore the role of economic status and demographic composition on the affordability of fresh produce, the following variables were utilized: the median value of owner-occupied

housing, the median household income, the percent of the population under 200% of the poverty line, the proportion of residents with a high school diploma, the proportion of residents who do not own an automobile, the proportion of residents above age 65, the proportion of the population that self-reports Black race, the proportion of the population that self-reports Hispanic ethnicity and an indicator for being located in Brazos County, i.e. urban.

It is difficult *a priori* to sign the coefficient on our wealth and income measures. As normal goods, the demand for fresh fruits and vegetables should be increasing in monetary resources, which should also lead to higher prices. However, if higher demand for grocery items leads to the opening of supermarkets, economies of scale may actually lower prices. This leads us to hypothesize that conditional on store-type affordability is decreasing in income and wealth.

Although education is positively correlated with income, we expect that the coefficient on education and affordability are positively related. Educated shoppers may be better able to compare the full menu of prices at different stores or understand various discounts. As more mobile consumers, stores may respond by offering lower prices.

Since individuals without transportation are less likely to comparison shop, we expect that the relationship between transportation availability and affordability is negative. Because of mobility issues, it is possible that older residents are less able to shop at stores outside their immediate community, suggesting the age profile of the neighborhood is negatively association with affordability. For those who are able to choose among various firms, however, retirement likely provides additional time to

comparison shop. Additional price information for consumers would tend to lower the prices charged by firms and thus the overall effect is ambiguous a priori.

We expect that convenience stores will charge a higher price for fresh fruits and vegetables, but have no prior on whether supermarkets charge less than grocery stores. We also do not have a strong prior on how race/ethnicity is related to affordability, though previous work in urban areas tends to find that neighborhoods with higher proportions of non-White residents pay more for fresh fruits and vegetables (Morland et al. 2002b; Morland and Filomena 2007; Jetter and Cassady 2006), though the work of Block and Kouba (2006) is an important exception..

2.2.4 Statistical Analysis

A linear relationship between these explanatory variables and each price index was then estimated using ordinary least squares regression models. Therefore, each observation was a store, with the store-level price index being the dependent variable and the characteristics of the CBG in which the store is located being the explanatory variables (Model 1). Since the affordability of fresh produce could vary both within and between store types, a second set of regressions were estimated that include control variables for store type along with the CBG characteristics (Model 2). By comparing coefficient estimates across models, it is possible to determine whether a characteristic is correlated with affordability through the location choices of different store types. In each regression, the median value of owner-occupied housing and the median household income were taken in their natural logarithm; hence the coefficient estimates are interpreted as the effect of a 100% increase in variable.

It is important to note that the number of observations available for the regression analysis was relatively small. Even with imputation of missing prices, only 34 stores posted enough price information to calculate a high variety price index for fruit and 36 stores posted enough price information to calculate a high variety price index for vegetables. This further highlights the importance both of using an appropriate price imputation strategy and selecting a parsimonious set of explanatory variables. Moreover, the typical issues associated with small sample sizes—large standard errors and low powered hypothesis tests—would be exacerbated by collinearity between explanatory variables. Table 2.2 provides the correlation matrix of our explanatory variables and only four of these are above 0.5 in absolute value and none are greater than 0.7. Nevertheless, we calculate variance inflation factors (VIF) to assess collinearity (O'Brien 2007). Since multiple stores can be located in a single CBG, robust standard errors clustered at the CBG-level were calculated (Moulton 1990).

2.3. Results

Among the 39 stores that sold at least 3 fruit items (2 supercenters, 22 supermarkets, 11 grocery and 4 convenience), apples, oranges, avocado and banana were the most commonly found (Table 2.3 provides both the number of stores with price data

Table 2.2: Matrix of Key Variables Used in the Regression Analysis.

	Income	House	HS	Age	Auto	Hispanic	Black	Pov
Median household income	1.000							
Median value of owner-occupied housing	0.058	1.000						
High school degree (%)	0.046	0.302	1.000					
Age>65 (%)	-0.121	-0.269	-0.816	1.000				
Without automobile (%)	-0.110	-0.557	-0.130	0.068	1.000			
Hispanic (%)	-0.057	-0.454	0.253	-0.329	0.306	1.000		
Black (%)	-0.092	-0.588	-0.093	0.018	0.679	0.214	1.000	
Below 200% of poverty (%)	-0.082	-0.400	0.138	-0.203	0.483	0.504	0.437	1.000

Income: Median household income
HS: High school degree (%)
Age: Age>65 (%)
Auto: Without automobile (%)
Pov: Below 200% of poverty (%)

by item along with the proportion of stores doing so). On a per weight basis, bananas were the cheapest fruit type, while berries were the most expensive. Among the 49 stores that sold at least 4 vegetable items (2 supercenters, 22 supermarkets, 14 grocery stores and 11 convenience stores), carrots, lettuce, onions, potatoes and tomatoes were the most common. Potatoes were the least expensive vegetable type, while green beans were the most expensive.

Although there were 39 stores that sold at least 3 types of fruit, only 35 stores posted the requisite information—the prices of apples, oranges and bananas—to calculate basic and high variety price indices (2 supercenters, 22 supermarkets, 9 grocery and 2 convenience). Similarly, of the 49 stores that sold at least four types of vegetables, 37 posted the requisite price information to calculate a basic vegetable price index (2 supercenters, 22 supermarkets, 10 grocery and 3 convenience) and 36 posted the requisite information to calculate a high variety index (2 supercenters, 22 supermarkets, 10 grocery and 2 convenience). The mean cost of meeting the USDA recommended level of fruit consumption from a high variety basket of fruit types was just under $27 per week (Table 2.4). In contrast, relying on only the three most common fruits lowered the weekly expense to just under $17 per week, a reduction of 37%. The effect of moving from a high variety to a low variety basket was much less when considering

Table 2.3. Price Availability, Mean Price and Consumption Shares of Fresh Fruit and Vegetables Types.

	Stores with non-missing price[1]	Mean price[2] ($)	Minimum price ($)	Maximum price ($)	Consumption Share[4] (%)
Fruit					
apples	36 (100)	1.12 ±0.30	0.61	1.99	11.7
avocado	35 (97.2)	2.19 ±1.00	0.59	5.47	9.1
bananas	35 (97.2)	0.47 ±0.11	0.29	0.69	33.8
berries	26 (72.2)	2.79 ±0.58	1.50	3.99	6.2
grapes	31 (86.1)	1.72 ±0.48	0.89	2.79	8.9
melon	30 (83.3)	0.81 ±0.24	0.33	1.12	18.2
oranges	36 (100)	0.98 ±0.37	0.33	1.89	7.7
peaches	24 (66.7)	1.58 ±0.30	1.27	2.29	2.6
pears	23 (63.9)	1.30 ±0.45	0.35	1.79	1.8
All fruit types[3]	18 (50)				100
Vegetables					
carrots	36 (73.5)	1.04 ±0.44	0.49	2.00	7.8
corn	25 (51.0)	0.83 ±0.36	0.45	1.82	5.7
cruciferous	29 (59.2)	0.95 ±0.33	0.32	1.59	3.8
green beans	22 (44.9)	1.55 ±0.52	0.99	2.79	2.1
greens	29 (57.1)	0.98 ±0.24	0.70	2.01	1.1
lettuce	39 (79.6)	0.75 ±0.30	0.49	1.98	7.7
onions	39 (77.6)	0.97 ±0.36	0.39	1.99	16.4
potatoes	38 (75.5)	0.60 ±0.36	0.30	2.39	33.8
tomatoes	38 (77.6)	1.33 ±0.43	0.69	2.39	4.8
squash	30 (61.2)	1.23 ±0.39	0.50	1.88	17.0
All veg. types[3]	21 (42.8)				100

[1] For fruit, sample is all stores selling at least three types: n=36. For vegetables, sample is all stores selling at least four types: n=49. Proportion of stores with non-missing price in parentheses.
[2] Means are reported ± SD.
[3] *All types* summarizes the number of stores selling all types of fruits or vegetables, respectively.
[4] Consumption share of each type from Fresh Look Marketing, Inc for Dallas market.

Table 2.4. Summary Statistics of Produce Availability, Store Types and CBG Characteristics.

	n	Mean	Minimum	Maximum
Price indices, $				
High variety fruit	34	26.98 ± 3.97	19.07	37.89
Basic fruit	35	16.87 ± 2.62	10.16	23.63
High variety vegetable	36	28.72 ± 6.75	17.19	48.19
Basic vegetable	37	27.58 ± 7.54	15.88	53.24
Store types, %				
Proportion supermarkets/supercenters	38	63.2 ±48.9	0	1
Proportion grocery stores	38	26.3 ±44.6	0	1
Proportion convenience stores	38	10.5 ±31.1	0	1
CBG characteristics				
Median owner-occupied housing value, $	38	70,739 ± 39,191	0	187,500
Median family income, $	38	32,379 ± 15,124	13,292	88,172
Proportion below 200% of poverty, %	38	45.5 ±18.5	2.0	78.1
Proportion with HS diploma, %	38	73.4 ±16.6	40.8	100
Proportion age 65 or older, %	38	13.9 ±6.8	1.6	55.5
Proportion without a vehicle	38	9.6 ±7.7	0	38.0
Proportion African American, %	38	18.8 ±17.7	0.2	75.5
Proportion Hispanic, %	38	16.9 ±11.8	4.9	52.6

Notes: Price indices are the cost of purchasing the recommended weekly servings of fruits and vegetables according to the USDA Thrifty Food Plan for a representative household of 2 adults and 2 children from fresh, whole items: 24.5 pounds (11.1kg) of fruit and 31.5 pounds (14.3kg) of vegetables). Not all price indices can be calculated for all stores because of variation in which prices for individual goods are available. Store types and CBG characteristics calculated over the set of stores for which a basic fruit index or a basic vegetable index or both could be calculated. Means are reported ± SD. n is the number of stores, CBG is census block group. Vegetable consumption: a 3.9% decline from $28.72 to $27.58 per week.

As summarized on Table 2.3, the set of stores (35) that had the requisite information to calculate a basic fruit price index is not a subset of the stores (37) that had the requisite information to calculate a basic vegetable price index. Thus, the remaining summary statistics were calculated over the 38 stores for which we were able to calculate a basic fruit index, a basic vegetable index or both. Among these, there are 24 supermarkets or supercenters (65%), 10 grocery stores (27%) and 4 convenience stores (8%). It is evident that the CBG in which stores were located are exceptionally diverse. The median value of owner-occupied housing ranged from $0, indicating that the CBG was entirely comprised of commercial and rental units, up to $187,500. Because one supermarket was located in a CBG with no reported owner-occupied housing, this observations was dropped. The median household income was highly variable, with a standard deviation just under half the mean (COV=0.46). Moreover, the highest median income was more than 6.6 times larger than the lowest. This was also reflected in the poverty rate, which ranged from 2.0% to 78.1%. There were also several minority-majority CBG in which more than 50% of the population was either African American or Hispanic.

Regression analysis of the store-level price indices on the characteristics associated with the CBG in which the store was located consistently revealed that stores in the urban county charged less for fresh produce (Table 2.5 and 2.6). The limited number of observations leads to rather imprecise coefficient estimates, but the sign pattern can still be instructive. The coefficient on the urban indicator was negative in

each regression, though it was never statistically significant. For example, the mean weekly expenditure required to satisfy the USDA recommended level of vegetable

Table 2.5. Association between Fruit Price Indices and CBG Characteristics, Geographic Location, and Store Type: Regression Results from Pooled Sample of Urban and Rural Food.

Price Index:	Fresh Fruit			
	High variety[2]		Basic[3]	
	Model 1	Model 2	Model 1	Model 2
CBG characteristics				
Median value of owner-occupied	3.223	6.777*	9.894**	10.00**
	(3.568)	(3.420)	(4.623)	(4.390)
Median household income[4]	1.616	3.01	-0.293	-3.297
	(4.267)	(3.670)	(6.617)	(5.626)
Proportion below 200% of poverty,	0.0956	0.127	-0.166	-0.243
	(0.123)	(0.099)	(0.236)	(0.203)
Proportion with HS diploma, %	-0.180*	-0.0518	-0.118	-0.134
	(0.094)	(0.131)	(0.181)	(0.171)
Proportion age 65 or older, %	-0.0135	-0.0573	-0.316	-0.305
	(0.157)	(0.193)	(0.305)	(0.288)
Proportion without a vehicle	-0.145	-0.09	0.117	0.306
	(0.172)	(0.247)	(0.433)	(0.454)
Proportion African American	-0.00835	0.0498	0.0881	0.0348
	(0.106)	(0.112)	(0.153)	(0.168)
Proportion Hispanic	-0.0906	-0.0619	0.0388	0.0492
	(0.104)	(0.084)	(0.181)	(0.191)
Geographic Location				
Urban	-3.826	-4.945	-7.257	-7.397
	(2.200)	(3.404)	(5.113)	(4.458)
Store Types[5]				
Grocery store		1.412		-4.265
		(1.786)		(3.139)
Convenience store		9.909***		2.141
		(2.612)		(3.360)
N	33	33	35	35
R^2	0.242	0.485	0.248	0.332

[1]Coefficient estimates from linear regression. Robust standard errors clustered at CBG level in parentheses below coefficient estimates. $^{*}P<0.1$ $^{**}P<0.05$ ***P<0.01

[2]High variety indices include all produce types listed in Table 2.1.

[3]Basic fruit index only includes apples, bananas, and oranges. Basic vegetable index only includes carrots, lettuce, onion, potatoes and tomatoes.

[4] Taken in natural logarithm.

[5]Referent category is supercenter/supermarkets.

Table 2.6. Association between Vegetable Price Indices and CBG Characteristics, Geographic Location, and Store Type: Regression Results from Pooled Sample of Urban and Rural Food.

Price Index:	Fresh Vegetables			
	High variety[2]		Basic[3]	
	Model 1	Model 2	Model 1	Model 2
CBG characteristics				
Median value of owner-occupied	-0.529	0.676	10.58*	11.28**
	(1.713)	(1.881)	(5.420)	(5.110)
Median household income[4]	-2.823	-1.31	0.597	-3.025
	(3.131)	(2.658)	(7.229)	(6.140)
Proportion below 200% of poverty,	-0.064	-0.00589	-0.198	-0.281
	(0.093)	(0.093)	(0.268)	(0.232)
Proportion with HS diploma, %	-0.00823	0.0376	-0.146	-0.15
	(0.063)	(0.072)	(0.208)	(0.191)
Proportion age 65 or older, %	-0.104	-0.0754	-0.334	-0.309
	(0.128)	(0.127)	(0.317)	(0.280)
Proportion without a vehicle	-0.0822	-0.15	0.124	0.431
	(0.102)	(0.131)	(0.466)	(0.498)
Proportion African American	0.00534	0.044	0.121	0.0416
	(0.058)	(0.066)	(0.159)	(0.166)
Proportion Hispanic	0.00698	0.00577	0.0696	0.0691
	(0.091)	(0.084)	(0.166)	(0.162)
Geographic Location				
Urban	-1.467	-2.038	-6.479	-7.353
	(1.834)	(1.686)	(5.169)	(4.543)
Store Types[5]				
Grocery store		2.732		-5.891
		(1.730)		(3.633)
Convenience store		3.400*		4.091
		(1.973)		(3.209)
N	34	34	36	36
R^2	0.117	0.263	0.226	0.367

[1]Coefficient estimates from linear regression. Robust standard errors clustered at CBG level in parentheses below coefficient estimates. $^{*}P<0.1$ $^{**}P<0.05$ ***P<0.01
[2]High variety indices include all produce types listed in Table 2.1.
[3]Basic fruit index only includes apples, bananas, and oranges. Basic vegetable index only includes carrots, lettuce, onion, potatoes and tomatoes.
[4] Taken in natural logarithm.
[5]Referent category is supercenter/supermarkets.
HS: High School; CBG: census block group.

consumption using the basic set of items was over $6.00 higher when shopping at a store located in one of the rural counties, while fruit consumption was over $7.00 higher.

We also found that the value of owner-occupied housing was associated with more expensive produce at local stores. The coefficient estimate on median value of owner-occupied housing was positive in all but one regression and statistically significant in five. As expected, the coefficient on our measure of education was negative in each regression and was statistically significant in one. The coefficients on income and poverty rate did not show a consistent pattern, which is particularly problematic given the imprecision of our estimates.

The coefficient on the proportion of residents who are African American was positive in seven regressions, but was never significant. Except for the high variety fruit basket, the coefficient on the proportion of Hispanic residents also positive and never approached significance (the smallest P-value is 0.340). It is also worth noting that coefficient estimates did not respond strongly to the inclusion of controls for the type of store when the price indices for fresh vegetables were the dependent variables, e.g. the size and significance of coefficients were similar. The effect when the fruit price indices were the dependent variables was stronger, but the estimates were nonetheless qualitatively similar.

2.4. Discussion

The first goal of this article was to describe the relationship between the prices charged by stores for fresh produce and the characteristics of the surrounding communities. We demonstrated that individuals who shop at food stores located in the rural counties of the Brazos Valley region must pay significantly more to attain the USDA recommended level of fresh fruit and vegetable consumption through fresh whole items than residents who shop in the urban area. Moreover, the difference in cost between urban and rural stores was not explained by differences in the type of stores that locate in these areas. This result is consistent with previous work in the literature and further illustrates the challenges that rural households face with respect to making healthy lifestyle decisions. Of course, individuals who reside in rural areas but work in the urban area are able to shop at urban retailers with little additional transportation or time cost. Nevertheless, numerous at-risk groups such as the older adults, the unemployed, those without access to transportation and parents who cannot afford child-care either do not have this option or experience costs that make it prohibitively expensive. Therefore, in future work we plan to study how the affordability of healthy food items affects the decision of where rural residents shop.

We also found that higher housing values are positively associated with the cost of fresh fruits and vegetables. Since median household income was also included in the set of regressors, two explanations are possible. First, housing values reflect the economic value of a location, which is capitalized through rent and property taxes. Thus, median housing values capture some of the operating costs of a store owner and higher

operating costs naturally imply higher prices. Second, housing is an asset, and often the largest one for a household. It is possible that two areas with identical median household incomes nevertheless exhibit large differences in asset wealth. Wealth differences could *cause* differences in demand for healthy food items, thereby raising price. Alternatively, the attitudes and preferences that lead individuals to accumulate asset wealth, e.g. greater patience or greater appreciation of the long-run consequences of current decisions, may be associated with the attitudes and preferences that encourage healthy eating behavior. This suggests that failing to distinguish the supply-side effects of land values from the demand side effects of income and wealth may misstate the role of community socio-economic status on affordability since the two are positively correlated.

Given the lack of precision in coefficient estimates, it is impossible to judge the association between medium household income and the cost of fresh produce. There was suggestive evidence, however, that stores located in areas with higher proportions of African-Americans charge more for fresh fruits and vegetables. The evidence was much weaker with respect to the proportion of Hispanic residents.

In addition to providing a fuller description of affordability differences between urban and rural areas, the current paper also makes several methodological improvements over previous work. The analysis confronts the common problem of missing prices using a price imputation strategy that is more firmly grounded in economic and statistical theory than has previously been employed. The inclusion of a control for local housing values along with local income information is also a positive step in separating potential supply-side influences on availability from demand-side

explanations. In addition, we used data from the Brazos Valley Food Environment Project, which identified all traditional and non-traditional food stores using ground-truthed methods and conducting a comprehensive assessment of the availability and price of fresh fruit and vegetables. By including supercenters, supermarkets, grocery stores, convenience stores, dollar stores, mass merchandisers, and pharmacies, this study provided a more complete picture of availability and price of fresh fruit and vegetables

The analysis suffers from several limitations. First, the number of food stores in the Brazos Valley region is relatively small and future work should consider canvassing a larger area to increase the precision of estimates. Doing so could not only increase the number of observations in an analysis similar to the one undertaken here, but also allow for separate regressions for urban and rural areas. Of course, the cost associated with completing a census of food stores should not be underestimated. Second, we are unable to translate differences in local affordability into differences in eating behavior. Third, it must be acknowledged that fresh whole items from food stores are not the only source of fruit and vegetables, though the majority of fruit and vegetable consumption is in the form of fresh whole items. In food stores, fruit and vegetables can also be purchased in frozen, canned, dried and juiced forms. Additionally, fruit and vegetable consumption may also occur in restaurant settings. Since the nutritional value of consumption likely varies by the form consumed, the affordability of these different options, both in absolute terms and relative to each other, is also worthy of future study. Fourth, sales shares were not available for the stores in our sample, and thus we were unable to weight observations in the regression analysis. Future data collection effects should do so in

order to account for differences in the relative importance that each store plays in the actual purchasing decisions of households. Finally, while our imputation strategy allows us to calculate a hypothetical measure of affordability for stores that do not sell all items, these stores may typically exhibit limited availability Researchers and policy makers should keep both aspects—availability and affordability—in mind when considering improvements in the food environment.

Despite these limitations, this study extends prior work by examining the affordability of fresh fruit and vegetables from traditional and non-traditional food stores in a large rural area; and how access to an affordable supply of fresh fruit and vegetables differs by neighborhood and geographic inequalities. The approach and findings of this study are relevant and have important research and policy implications for understanding access and availability of affordable, healthy foods. Access to a good variety of affordable healthy foods, such as fruit and vegetables, can play a pivotal role in the nutritional health of rural families. Many of these families live in socioeconomically-deprived neighborhoods; many have a low household income, are unemployed, older, or lack access to a vehicle. In order for rural families to be food secure and have access to fruit and vegetables, food resources need to be available and affordable in local stores.

CHAPTER III

COMPARISON OF THREE IMPUTATION METHODS FOR MISSING ITEMS: IMPLICATION OF THE PRICE INDICES FOR FRUIT AND VEGETABLES

3.1 Introduction

The increasing burden of obesity-related illness has become a leading public health issue in the United States. Many researchers hypothesize that the local food environment—the availability and affordability of both healthy and unhealthy eating options—plays an important role in obesity outcomes. The food options available to households can constrain eating choices and the characteristics of households including the type of neighborhood can affect the set of goods and services that firms offer for sell and the prices they charge. In order to explain differences in obesity outcomes among different subpopulations (e.g. wealthy versus poor; urban versus rural; White versus non-White), a number of recent papers have investigated how the affordability of healthy food items varies with neighborhood characteristics (Ball, Timperio, and Crawford 2009; Jetter and Cassady 2006; Kaufman et al. 1999; Lotade-Manje et al. 2009; Pollar et al. 2002). These investigations require the use of information about the prices of food items, but in practice this information is often incomplete. While the appropriateness of different imputation strategies has been considered in medical and biological science, it has received scant attention in food policy.

Therefore, this paper uses store-level price information to compare how different imputation methods affect the results from an analysis of the association between neighborhood socio-economic conditions and the affordability of healthy food options.

We focus on the prices of fresh fruits and vegetables, since both have been identified as important components of a healthy diet. In addition, previous work in this area has used a variety of imputation strategies without considering robustness to the method chosen (Block and Kouba 2006; Dunn et al. 2011; Hendrickson et al. 2006; Latham and Moffat 2007; Lee et al. 2002).

3.2 Missing Data and Imputation

3.2.1 Missing Data

Missing data can arise because of non-response to survey questions, errors in processing responses or constructing the dataset or the impossibility of collecting certain information (Blend and Marwala 2008). Moreover, missingness may be either temporary or permanent (Armknecht and Maitland-Smith 1999). The data are temporarily missing when the observations are not available for a limited period—for example due to seasonality or the availability may occur in the near future due to human interventions or technical manipulations. The data are permanently unavailable when there is no possibility to retrieve them—for example when the market does not carry the product anymore, the respondents of surveys cannot be reached or are unwilling to answer to certain items, or in general the sources of information cease to exist.

For the purposes of statistical analysis, the most important characteristic of a missing datum is whether the data generating process that led to missingness was *ignorable* or *non-ignorable*. Ignorable processes are associated with data that are either *missing completely at random* (MCAR) or *missing at random* (MAR). The former occurs whenever the probability of missingness is the same for all the observations, and

it is difficult to distinguish complete data with incomplete data (Heitjan 1997). In this case, the data are missing simply by "accident" and cannot be related to any event associated with the characteristics of the data. For example, during a census, a surveyor may sneeze and omit the price of a particular good sitting on a store shelf. As long as higher price goods did not increase the likelihood of a sneeze, omission is completely by accident. Even though there is incomplete data set, results with available data on hand are still consistent and provide valid inferences[1]. The latter arises when the reason for missingness is random conditional on observable characteristics, i.e. the behavior of observationally equivalent units is random with respect to the completeness of data. For example, if males were more likely than females to withhold income information from a survey, but neither sex withheld based on income level, then the data generating process would be ignorable conditional on gender. Since the data are MAR, one could calculate unbiased means of income for males and income for females.

A non-ignorable process is associated with data that is *not missing at random* (NMAR). This arises when a datum is more likely to be missing because of its value. Suppose that in our hypothetical survey, wealthier males were more likely than poorer males to report income, while the opposite was true for females. Because the data are MNAR, the mean income of males calculated from this survey would be biased downward and the mean income of females would be biased upward. It is thus intuitive that the assumption of data that are MAR conditional on a set of observable

[1] Carpenter and Kenward (2005) at www. Missingdata.org.uk or http://missingdata.lshtm.ac.uk/jargon_web/node4.html

characteristics is the minimum necessary condition for most imputation strategies to be valid.

In our data, missingness arises either when stores do not offer particular items for sale or the price for an available item is not displayed, which may result from either a human error or a managerial decision. The decision not the sell a product can be viewed as an outcome of a profit-maximizing store-owner: the price that consumers are willing to pay for the missing type is below the cost that a store owner faces to offer the type for sell. Nevertheless, there is still some reservation price that would lead the store owner to stock the item. In this context, the proper price for missing items is this unobserved *shadow price*. It is also clear that in this framework, the data generating mechanism for prices is non-ignorable, and thus unconditioned prices are not MAR.

Alternatively, one may take a statistical perspective and assume that over some period of time, all stores eventually stock all types of produce. Since inventory and prices are only observed once for each store, it is possible that unobserved types were recently sold or will be sold again soon. An argument for an ignorable mechanism conditional on store type or location may be more palatable under this scenario, but there is no way to test whether the economic framework or the statistical is closer to the truth.

3.2.2 Imputation Methods

Different types of missing values and the techniques used to deal with the missing observations are well described by Nordholt (1998) and Hawthorne and Elliott (2005). In this study, we investigate the effect of three methods of handling missing

prices of fruit and vegetables. Following pages provide more detailed information regarding the conception and implementation of the methods.

Zero substitution—replacing missing values with zero—as used by Hendrickson et al. (2006), is obviously the least tenable since prices are not zero and implementation would bias the mean downward (Gan et al. 2006). Its impact on regression analysis is less clear, however.

Mean imputation, which is one of the most used techniques in the literatures (Raymond 1986), replaces missing observations with the mean value over observed prices. If an item is not stocked because the reservation price of store owners is high relative to prevailing demand, the mean observed price would tend to underestimate the true value, suggesting that mean imputation may be inappropriate if missingness arises from profit-maximizing behavior. Separating the analysis by store-type or location to calculate subgroup means would simply result in biased imputation with each subgroup, though the size of the bias may be less than using a grand mean over all observations.

An additional problem with mean imputation relates to the spike it creates at the mean of the price distribution. This spike tends to reduce the correlations between price and other variables, as well as the standard errors of estimates. Given that the mean of a variable X_i is $\sum X_i/n$ (i.e. summation of the values over the number of observation) and the variance $S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$, adding a mean value to the variables does not change the general mean of the variable, but the variance declines because the numerator remains unchanged while the denominator increases. Moreover, in a multivariate

regression setting, an attenuated covariance between price and one explanatory would bias the coefficient estimates of all characteristics. When each of the covariances is attenuated, the direction of this bias for any estimate cannot be signed *a priori* since each of the attenuated covariances interacts with the covariance matrix of explanatory variables.

Regression imputation (RI) involves estimating a relationship between observed prices and characteristics, then using the estimated relationship to predict prices. One can view mean imputation as a restricted form of RI when only a constant term is estimated. Similarly, calculation of subgroup means based on observable characteristics like county of location or store-type is equivalent to regression imputation with a constant and a collection of dummy variables. More generally, RI permits inclusion of continuous explanatory variables. Valid imputation requires that conditional on the included explanatory variables missing data are MAR. This makes the selection of explanatory variables a vital component of the imputation process, a problem taken up subsequently. RI also suffers from the same form of attenuation found with mean imputation since observationally identical observations receive the same imputed value. One can partially overcome this through stochastic substitution (RISS) or multiple imputation (MI), whereby the imputed value is the regression prediction plus a randomly selected residual from the estimation.

3.3 Previous Investigations of Imputation Method Performance

Numerous studies in the medical and physical sciences have compared the performance of different imputation strategies. In most applications, zero imputation

performs poorly relative to mean imputation. For example, Fraser et al. (2009) compare the results of a food questionnaire survey that has missing items and full data (the surveyors did a follow up to fill in the missing values) and find that the correlation between the data with zero imputation in the original (full) data is at least 0.9. But the authors also added that the true use of zero imputation may depend on the importance of the missing variables. In this sense, imputing zero for food not eaten may make more sense compared to imputing a zero for price of the food. In other scenarios, however, zero imputation is clearly suboptimal. Sehgal et al. (2008) tested the performance of different imputation methods including zero imputation. The test results show that among the techniques used, zero imputation has the highest error rate across the missing values.

Although mean imputation is often preferable to zero imputation, it is not without its own performance issues. For instance, when evaluating the performance of several methods values in gene microarray data, Troyanskaya et al. (2001) find that mean imputation performs better than zero imputation, but is less accurate than other methods and performs poorly in non-time series data. Using the mean imputation technique, Armknecht and Moulton (1995) show that the method can result in different outcomes if the quality of the products is accounted for. The authors mentioned that the quality should be considered whenever the variety of the replacing product is incomparable.

When it comes to the simple regression method, Musil et al. (2002) found that it performs better than mean imputation; the latter was actually the least accurate method to be used among the methods used (listwise deletion, mean substitution, simple

regression, regression with an error term, and the expectation maximization [EM] algorithm). This result is confirmed by Shrive et al. (2006), who found that the regression method outperform the mean imputation among other imputation techniques. Raymond and Roberts (1987) also encourage the use of regression imputations. The authors constructed missing values by dropping existing values and proceeded to compare different methods of imputations. The results of the comparisons were that regression imputation is a method of choice when facing 10%-40% percentage of missing items. Using simulation techniques to compare the results full data and the results with imputed values, the authors found that regression methods provide the most accurate regression estimates. Olinsky, Chen and Harlow (2003) compared the efficacy of mean and regression imputation techniques in structural equation modeling using two sample sizes with seven levels of incomplete data. They also found that regression imputation tended to outperform mean imputation.

Although regression imputation tends to be preferred to zero and mean imputation, Tanguma (2000) notes that under certain conditions mean substitution has its advantages relative to regression imputation. I therefore conclude that at the very least, researchers who face missing data should check whether their results are robust to the choice of method and if they are not, should attempt to understand why.

3.4 Data

Information on the prices of fresh fruits and vegetables come from the Brazos Valley Food Environment Project (BVFEP). The Brazos Valley area is comprised of seven contiguous counties situated between the Austin, Dallas and Houston metropolitan

areas and is home to over 300,000 residents. The region is geographically diverse with 51.4% of the population residing in one of six predominately rural counties (Burleson, Grimes, Leon, Madison, Robertson and Washington). The seventh, Brazos County, includes the medium-sized urban center of Bryan-College Station.

As part of the BVFEP, trained observers enumerated all food stores and food service places by driving all Interstates, US Highways, Texas State Highways, Texas Farm-to-Market Roads and other major thoroughfares to locate all stores that could sell food items (Musil et al. 2002). Food stores were classified into several categories: supercenter, supermarket, grocery store, convenience store, dollar store, mass merchandiser, and pharmacy. Investigators then entered all food stores with an extensive list of food items in order to catalogue which items were sold and at what price (Sharkey and Horel 2008).

Ten types of fresh fruit (apples, avocado, bananas, berries, grapes, mango, melons, oranges, peaches and pears) and eleven types of fresh vegetables (broccoli and cauliflower, carrots, corn, green beans, leafy greens, lettuce, okra, onions, potatoes, squash and tomatoes) were included in this catalogue. The following information was recorded: whether each type (e.g. apples) was available for purchase; the number of varieties of each type of fruit or vegetable available for purchase; and the lowest-priced variety of each type of fruit or vegetable. Because in-store prices were posted in several forms—per item, per ounces, per pound—all prices were later transformed or recalculated into a uniform price per pound. To do so, when prices were posted per item,

surveyors weighed the items using a sensor scale. The price of food items that were either not sold or not displayed were recorded as missing.

Socioeconomic characteristics of communities were extracted from the 2000 decennial census Summary Files 3 (SF-3) at the CBG levels, since the CBG is the smallest unit of census geography for which the detailed "long-form" social and economic data from the census are tabulated. The socio-economic variables examined in this study are: the median income of the families living in the county; the median value of houses in a given Census Block Group (CBG); the proportion of residents that live below the 200% poverty level; the proportion of residents that are at least of 65 years old; the proportion of residents who have at least graduated from high school; the proportion of the residents that do not have access to transportation (vehicle); and the proportion of the population that is Black of Hispanic.

3.5 Method

This section describes each of the methods used to impute missing prices for the current dataset; the calculation of price indices over the individual produce items; and the methods used to compare the performance of the various indices.

3.5.1 Zero Imputation

Observed values are assigned when displayed, while missing prices are replaced by zero.

3.5.2 Mean Imputation

Several mean imputations are considered. First, the missing price of each item is replaced with the mean price of the item over stores that stocked it. Second, the missing

price of each item is replaced with the mean price of the item over stores of the same type (supermarket/supercenter, grocery store and convenience store) that stocked it. Third, the missing price of each item is replaced with the mean price of the item over stores in the same county that stocked it.

3.5.3 Regression Imputation

This imputation technique predicts the values of the missing prices based on the available prices of assumed related products using a linear regression. The process is as follow: the missing price of each fruit item is predicted based on an estimated relationship with county, store type and the prices of common fruit types (apples, avocados, bananas, grapes and oranges). There are some stores that sell four of the five common types. In this case, the missing price of the common type is imputed from an estimated relationship with county, store type and the prices of remaining common types.

Similarly, the missing price of each vegetable item is predicted based on an estimated relationship with county, store type and the prices of common vegetable types (carrots, lettuce, onions, potatoes and tomatoes). For stores that sell only three of the five common types, the missing price of the common type is first imputed from an estimated relationship with county, store type and the prices of remaining common types. Imputation was not attempted over stores that sell less than three of the common types.

3.5.4 Calculation of Price Indices

Two types of price indices are calculated for fruits and for vegetables. The first index is based on a *fixed basket* method. In this case, each index is calculated as the

weighted average price per pound multiplied by the recommended number of pounds consumed per week for a representative family of two adults and two children from the most recent USDA Thrifty Food Plan: 24.5 pounds of fruit and 31.5 pounds of vegetables[2]. The weights are equal to the consumption shares calculated for the Dallas metropolitan area from Fresh Look Marketing, Inc. (Chicago, Illinois) and represent all supermarkets (sells of at least $2 million) with about 70% of all commodity volume (ACV) in the Dallas market[3] using the consumption shares for each type.

The general expression for the fixed price index is:

$$P(p) = n\sum_i c_i p_i \tag{1}$$

where *P(p)* represents the price index for a designated food type under the fixed method assumption, c_i is consumption share pertaining to the specific food item i, p_i is the price per unit vector for the food type i, and n is the designated number of servings.

The second index (*economic)* is based on the solution to a consumer's expenditure minimization problem. Unlike the fixed-basket, an economic index allows for substitution behavior as the relative price of goods change. Assuming that fresh fruits and fresh vegetables are weakly separable with subgroup utilities that are Cobb-Douglas, Dunn el al. (2011) demonstrate that the magnitude of this substitution bias can be large, estimating a lower bound of at least 8%. We adopt their economic price index here, which takes the form:

---

[2] The choice of multiplicative factors affects the magnitude of the subsequent coefficient estimates, but not their statistical significance. If only half of total fruit consumption should come from fresh whole items, then the appropriate adjustment is to either halve the coefficient estimate or reinterpret it as biweekly expenditure.

[3] We thank Timothy Richards for providing this information.

$$e(p,u^*) = n\left[\sum_i \left(\frac{\alpha_i}{p_i}\right)\right]^{-1} \tag{2}$$

where p represents the per unit price vector for a designated food type, u represents a desired attainable level of utility; α_i is proportion of expenditure on a certain food type pertaining to the specific food item i, p_i is the price per unit vector for the food type i, and n is the designated number of servings.

3.5.5 Comparison of Imputation Methods

After imputing the missing prices, we compute both the fixed-basket and economic indices defined above for fruits and vegetables. We then compare the mean values of these indices to determine the robustness our affordability measure to different imputation strategies. We also use these price indices as dependent variables in OLS regressions to assess the robustness of the relationship between the affordability of fruit and vegetables and socio-economic characteristics. In these regressions, both median household income and median housing value are taken in their natural log. In addition, we define categorical variables for types of store and county location.

3.6 Results

Table 3.1 provides descriptive statistics of availability and price for the fruit and vegetables items in our study. On average 14% of fruit items are missing compared to 19% for vegetables. Of the stores selling at least three fruits, the prices of apples, bananas and oranges were available at all of them. Of the stores selling at least three vegetables, the prices of onions and potatoes were never missing. Bananas exhibited the lowest average price among fruit items, while potatoes tended to be the cheapest

vegetable on a per pound basis. On average, the most expensive fruits are berries followed by avocado—both above $2 per pound. It is also worth noting that these prices varied widely: the cheapest avocados were available at $1 per pound, while the most expensive were $5.50. The price of berries started at $1.50 per pound in some stores, reaching as high as $4 per pound at others. For the vegetables, greens are the far the most expensive but cost only $1.50 on average; some stores in the areas sell as high as $2.79 per pound or lower at the price of $0.99 per pound.

Descriptive statistics for the various prices are summarized in Table 3.2 and several noteworthy patterns emerge. As expected, the mean with zero imputation is always lower than when using mean or regression imputation. For six of the eight indices, the mean value using regression imputation is larger than when using mean imputation. This is consistent with the economic argument presented earlier that stores chose not to offer items for sale when the reservation price of the store-owner is above what consumer demand will support. Finally, for fruits, the high variety index always larger than the basic index, whereas the reverse is observed in the case of the vegetable prices.

Table 3.1: Percentage of Stores with Non-Missingness and Description Statistic of Price Per Pound of Fresh Produce Items.

	Non-missing price (%)	Mean	Std. Dev.	Min	Max
Fruits					
apples	100	$1.12	0.30	$0.61	$1.99
avocado	97.1	$2.19	1.02	$0.59	$5.47
bananas	100	$0.47	0.11	$0.29	$0.69
berries	74.3	$2.79	0.58	$1.50	$3.99
grapes	88.6	$1.72	0.49	$0.89	$2.79
mango	74.3	$1.08	0.40	$0.36	$2.19
melon	82.9	$0.82	0.25	$0.33	$1.12
oranges	100	$0.97	0.37	$0.33	$1.89
peaches	74.3	$1.58	0.30	$1.27	$2.29
pears	71.4	$1.30	0.45	$0.35	$1.79
Vegetables					
carrots	94.6	$1.10	0.45	$0.49	$2.00
corn	70.3	$0.83	0.36	$0.45	$1.82
cruciferous	78.4	$0.93	0.32	$0.32	$1.49
green beans	56.8	$0.97	0.24	$0.70	$2.01
greens	75.7	$1.50	0.47	$0.99	$2.79
lettuce	97.3	$0.73	0.29	$0.49	$1.98
okra	29.7	$1.89	1.02	$0.50	$3.49
onions	100	$0.95	0.37	$0.39	$1.99
potatoes	100	$0.62	0.38	$0.30	$2.39
squash	83.8	$1.31	0.44	$0.69	$2.39
tomatoes	100%	$1.22	0.40	$0.50	$1.88

Notes: Fruit availability is calculated over 33 stores with observable prices for apples, oranges and bananas. Vegetable availability is calculated over 33 stores with observable prices for onions, potatoes and tomatoes. Std. Dev. is standard deviation.

Table 3.2. Description Statistic of the Prices and the Socioeconomic Factors.

Approach	Variables	Explanation	Mean	Std. Dev.	Minimum	Maximum
Fixed Approach	Price indices	Imputation Method				
	High variety					
	Fruit	Regression	$1.114	0.164	$0.778	$1.547
		Mean	$1.110	0.161	$0.778	$1.547
		Zero	$1.002	0.21	$0.553	$1.547
	Vegetable	Regression	$0.910	0.219	$0.557	$1.493
		Mean	$0.905	0.213	$0.587	$1.493
		Zero	$0.873	0.228	$0.506	$1.493
	Basic variety					
	Fruit	Regression	$0.730	0.196	$0.415	$1.421
		Mean	$0.736	0.201	$0.415	$1.421
		Zero	$0.721	0.200	$0.415	$1.421
	Vegetable	Regression	$1.015	0.369	$0.500	$1.899
		Mean	$0.999	0.341	$0.500	$1.899
		Zero	$0.923	0.305	$0.500	$1.899
Economic Approach	*High variety*					
	Fruit	Regression	$1.007	0.137	$0.712	$1.292
		Mean	$1.015	0.135	$0.712	$1.289
		Zero	$0.963	0.128	$0.712	$1.133
	Vegetable	Regression	$0.837	0.188	$0.531	$1.254
		Mean	$0.836	0.186	$0.549	$1.254
		Zero	$0.869	0.204	$0.591	$1.254
	Basic variety					
	Fruit	Regression	$0.739	0.19	$0.451	$1.424
		Mean	$0.746	0.197	$0.451	$1.424
		Zero	$0.737	0.192	$0.451	$1.424
	Vegetable	Regression	$0.918	0.276	$0.446	$1.533
		Mean	$0.902	0.21	$0.509	$1.334
		Zero	$0.869	0.204	$0.591	$1.254
Explanatory	Income		$30,303	1.547	$6,248	$119,014
	Median house value		$67,643	1.500	$23,500	$188,716
	Poverty level (%)		44.521	18.473	1.081	83.505
	Age 65 or older (%)		12.138	4.669	7.700	20.00
	Education (%)		76.019	5.517	67.300	81.300
	No Transportation (%)		9.162	7.168	0	38.037
	Hispanic (%)		17.842	15.752	1.557	80.651
	African-American (%)		16.486	16.009	0.186	75.534
	Convenience stores		0.809	0.394	0.00	1.00
	Supermarket		0.079	0.271	0.00	1.00

Std. Dev. is standard deviation.

Table 3.2 also reports descriptive statistics for the CBGs in which the stores are located. The result shows also that on average 76% of the residents have at least earned high school diploma. The average median income among these CBGs is just over $30,300, while the average median value of owner-occupied housing is about $67,600 for mortgage. The average poverty rate is higher than both the state and national average with 44% living below the poverty level. Only 9% of the residents do not have own transportation. Among the CBGs in this sample, on average, 34% of the population identified as either Hispanics or African-Americans. In addition at least 88% of the food stores in the region serving consumers are convenience stores.

We proceed by testing the equality of the means for the different price indices under alternative scenarios. We report the results of the t-test for the mean prices that are not statistically significant different on Table 3.3. This presentation is due to the fact that only few of the prices are not significantly different. The high variety prices for fruit and vegetables based on the fixed and economic assumption are not significantly different when the regression and mean imputations methods are applied. This equality is also the case for the basic variety in addition to the equality between zero and regression imputations.

To examine how estimations respond to the choice of imputation method, Tables 3.4-3.7 report coefficient estimates from regressions where a price index is the dependent variable and socio-economic characteristics of the CBG are the independent variables. For simplicity of assessment and comparison of the results, we present the

fixed methods on the left hand side of the tables and the economic method on the right hand side of the tables.

The results provided on Table 3.4 shows that the median value of housing is positive associated with the high variety fruit index, regardless of how it is computed or missing prices are imputed. The association is larger by at least 25% using the fixed-basket index rather than the economic one. The differences across imputation type are much smaller, however, with the smallest associations found when using regression imputation. The density of the African-American residents has significantly positive association with all but one index (fixed-basket with regression imputation). Stores classified as convenience stores exhibit significantly higher prices of fruit prices when the imputation methods do not involve replacing missing values with zero. In contrast, the coefficient on supermarkets is only significant when using zero imputation. The proportions of residents who are at least of 65 years of age or older and also those with high school diplomat are both negatively associated with the high variety fruit index, but only significant when using the economic index.

Table 3.3. T-tests for Equality of Means of the Dependent Variables (High Variety Prices).

Null Hypothesis	Alternative Hypothesis	Test Statistic	P-Values	Decision
Fruit_Fixed Method:				
-High Variety				
$Price_{\_Mean} = Price_{\_Regression}$	$Price_{\_Mean} \neq Price_{\_Regression}$	0.987	0.331	Fail to reject the null hypothesis
-Basic Variety				
$Price_{\_Regression} = Price_{\_Zero}$	$Price_{\_Regression} \neq Price_{\_Zero}$	-1.03	0.331	Fail to reject the null hypothesis
$Price_{\_Regression} = Price_{\_Mean}$	$Price_{\_Regression} \neq Price_{\_Mean}$	-1.000	0.324	Fail to reject the null hypothesis
Fruit_Economic-Method:				
-High Variety				
$Price_{\_Mean} = Price_{\_Regression}$	$Price_{\_Mean} \neq Price_{\_Regression}$	0.188	0.331	Fail to reject the null hypothesis
-Basic Variety				
$Price_{\_Regression} = Price_{\_Zero}$	$Price_{\_Regression} \neq Price_{\_Zero}$	1.000	0.324	Fail to reject the null hypothesis

Table 3.3 continued.

Null Hypothesis	Alternative Hypothesis	Test Statistic	P-Values	Decision
Vegetables_Fixed Method:				
-High Variety				
$Price_{\_Mean} = Price_{\_Regression}$	$Price_{\_Mean} \neq Price_{\_Regression}$	1.292	0.206	Fail to reject the null hypothesis
-Basic Variety				
$Price_{\_Regression} = Price_{\_Mean}$	$Price_{\_Regression} \neq Price_{\_Mean}$	1.625	0.111	Fail to reject the null hypothesis
$Price_{\_Regression} = Price_{\_Zero}$	$Price_{\_Regression} \neq Price_{\_Zero}$	1.625	0.110	Fail to reject the null hypothesis
Vegetables_Economic Method:				
-High Variety				
$Price_{\_Regression} = Price_{\_Mean}$	$Price_{\_Regression} \neq Price_{\_Mean}$	0.635	0.530	Fail to reject the null hypothesis
-Basic Variety				
$Price_{\_Regression} = Price_{\_Zero}$	$Price_{\_Regression} \neq Price_{\_Zero}$	-0.657	0.506	Fail to reject the null hypothesis
$Price_{\_Regression} = Price_{\_Mean}$	$Price_{\_Regression} \neq Price_{\_Mean}$	1.186	0.241	Fail to reject the null hypothesis

Table 3.4. Regression Results for High Variety Fruit Price Index by Imputation Method and Computation Method.

		Fixed Method			Economic Method	
High variety fruit	Variables	Zero Imputation	Mean Imputation	Regression Imputation	Mean Imputation	Regression Imputation
		-0.114	-0.038	-0.057	-0.056	0.148
Fixed	Income	(0.081)	(0.08)	(0.08)	(0.065)	(0.15)
	Housing	0.481***	0.485***	0.459***	0.359***	0.332**
		(0.132)	(0.125)	(0.13)	(0.11)	(0.122)
	Poverty	0.001	0.0005	0.001	-0.002	0.002
		(0.002)	(0.001)	(0.002)	(0.002)	(0.005)
	Education	-0.025	-0.028	-0.028	-0.058***	-0.052**
		(0.031)	(0.026)	(0.026)	(0.016)	(0.015)
	Transport	-0.0001	0.002	0.002	0.002	0.003
		(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
	65 yrs old	-0.012	-0.016	-0.015	-0.064**	-0.056**
		(0.044)	(0.039)	(0.039)	(0.023)	(0.021)
	African-American	0.004*	0.005*	0.004	0.005**	0.004**
		(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
	Hispanics	0.003	0.003	0.003	0.004	0.002
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	Supermarkets	0.231**	0.009	0.011	0.004	0.006
		(0.086)	(0.049)	(0.052)	(0.059)	(0.070)
	Convenience stores	0.141	0.365**	0.409***	0.248	0.260*
		(0.124)	(0.142)	(0.142)	(0.149)	(0.150)
	Burleson-Madison	-0.011	0.144	0.120	0.195*	0.190*
		(0.145)	(0.125)	(0.130)	(0.098)	(0.109)
	Grimes-Washington	0.127	0.137	0.135	-0.11	-0.043
		(0.187)	(0.162)	(0.166)	(0.133)	(0.120)
		0.327	0.318	0.400	0.625***	0.636**
	Leon-Robertson	(0.343)	(0.315)	(0.315)	(0.195)	(0.192)
		-1.563	-1.853	-1.490	2.600	0.074
	Intercept	(3.114)	(2.309)	(2.407)	(1.905)	(2.054)
	RMSE	0.1416	0.1266	0.1296	0.110	0.118
	R-squared	0.734	0.611	0.641	0.605	0.561

* 10 percent significance level, ** 5 percent significance level, ***1 percent significance level

Table 3.5 reports analogous results for the high variety vegetable price index. Only the median value of owner-occupied housing exhibits a statistically significant association with the cost of purchasing vegetables and the coefficient estimates are similar across both computation and imputation method. Age, ethnicity, transportation, store types and Burleson-Madison county indicator all have positive relationship with the prices of vegetables. The level of poverty has negative relationship. However these relationships are not statistically at standard levels.

The results using the basic index over stores that at least three varieties of fruit (Table 3.6) differ from those using the high variety index over stores that sell at least five different varieties (Table 3.4). Whereas the median value of housing is large and statistically significant in the former, the coefficient estimates are small and statistically insignificant in latter. The proportion of African-American residents is also not significantly associated with the basic fruit index. There are some commonalities between the two sets of results, however. For example, both the proportion of residents above age 65 and the proportion of residents who have completed high school are negatively associated with the cost of purchasing fruit. In addition, there is little difference in coefficient estimates across computation or imputation method.

Table 3.5. Regression Results for High Variety Vegetable Price Index by Imputation Method and Computation Method.

		Fixed Method			Economic Method	
High variety veg	Variables	Zero Imputation	Mean Imputation	Regression Imputation	Mean Imputation	Regression Imputation
	Income	0.047	0.059	0.063	-0.101	-0.108
Fixed		(0.15)	(0.284)	(0.145)	(0.099)	(0.243)
	Housing	0.485*	0.429*	0.408	0.438*	0.448*
		(0.238)	(0.237)	(0.246)	(0.222)	(0.229)
	Poverty	-0.002	-0.002	-0.002	-0.002	-0.005
		(0.004)	(0.004)	(0.004)	(0.004)	(0.008)
	Education	0.03	0.024	0.022	-0.004	-0.0001
		(0.054)	(0.053)	(0.056)	(0.029)	(0.027)
	Transport	0.017	0.016	0.014	0.007	0.009
		(0.015)	(0.014)	(0.015)	(0.009)	(0.009)
	65 yrs old	0.065	0.060	0.057	0.018	0.021
		(0.073)	(0.072)	(0.076)	(0.040)	(0.035)
	African-American	0.002	0.002	0.002	0.005	0.005
		(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
	Hispanics	0.004	0.003	0.003	0.005	0.006
		(0.005)	(0.005)	(0.005)	(0.004)	(0.006)
	Supermarkets	0.141	0.100	0.098	0.085	0.065
		(0.148)	(0.145)	(0.147)	(0.120)	(0.100)
	Convenience stores	0.011	0.073	0.036	0.052	0.117
		(0.160)	(0.168)	(0.180)	(0.137)	(0.140)
	Burleson-Madison	0.097	0.062	0.055	0.113	0.125
		(0.150)	(0.148)	(0.154)	(0.125)	(0.135)
	Grimes-Washington	0.439	0.399	0.402	0.165	-0.035
		(0.348)	(0.342)	(0.370)	(0.185)	(0.336)
	Leon-Robertson	-0.245	-0.221	-0.204	-0.024	0.217
		(0.549)	(0.540)	(0.565)	(0.335)	(0.160)
	Intercept	-8.439	-7.304	-6.925	-3.215	-3.432
		(6.023)	(5.793)	(6.104)	(3.322)	(3.170)
	RMSE	0.2111	0.2087	0.2207	0.1684	0.173
	R-squared	0.474	0.415	0.384	0.462	0.431

* 10 percent significance level, ** 5 percent significance level, ***1 percent significance level

Table 3.6. Regression Results for Basic Fruit Price Index by Imputation Method and Computation Method.

		Fixed Method			Economic Method	
Basic variety fruits	Variables	Zero Imputation	Mean Imputation	Regression Imputation	Mean Imputation	Regression Imputation
Fixed (regular)	Income	-0.08	-0.068	-0.073	-0.116	-0.134
		(0.168)	(0.174)	(0.171)	(0.139)	(0.144)
	Housing	0.014	0.062	0.042	-0.037	-0.045
		(0.138)	(0.136)	(0.132)	(0.136)	(0.133)
	Poverty	-0.007	-0.00004	-0.0003	-0.001	-0.005
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
	Education	-0.056*	-0.060*	-0.059*	-0.059**	-0.045**
		(0.029)	(0.032)	(0.031)	(0.027)	(0.014)
	Transport	-0.006	-0.010	-0.009	-0.010	-0.006
		(0.010)	(0.011)	(0.010)	(0.010)	(0.008)
	65 yrs old	-0.065*	-0.067	-0.067	-0.067*	-0.050**
		(0.037)	(0.042)	(0.040)	(0.035)	(0.019)
	African-American	0.00002	0.001	0.001	0.001	0.0004
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	Hispanics	-0.001	-0.0007	-0.001	0.0006	0.001
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
	Supermarkets	-0.148	-0.222*	-0.191	-0.179	-0.168
		(0.127)	(0.121)	(0.116)	(0.112)	(0.095)
	Convenience stores	0.206	0.167	0.183	0.115	0.169
		(0.257)	(0.231)	(0.240)	(0.212)	(0.198)
	Burleson-Madison	-0.168	-0.078	-0.115	-0.085	-0.117
		(0.137)	(0.150)	(0.141)	(0.151)	(0.141)
	Grimes-Washington	-0.340	-0.387	-0.375	-0.425*	0.209
		(0.280)	(0.294)	(0.287)	(0.250)	(0.200)
	Leon-Robertson	0.314	0.312	0.313	0.285	-0.293
		(0.244)	(0.293)	(0.270)	(0.249)	(0.159)
	Intercept	6.662	6.394	6.505	7.952*	7.000
		(4.582)	(4.862)	(4.726)	(4.046)	(3.118)
	RMSE	0.1989	0.1941	0.1899	0.1795	0.1707
	R-squared	0.397	0.429	0.423	0.492	0.479

* 10 percent significance level, ** 5 percent significance level, ***1 percent significance level

Table 3.7. Regression Results for Basic Vegetable Price Index by Imputation Method and Computation Method.

		Fixed Method			Economic Method	
Basic variety vegetable	Variables	Zero Imputation	Mean Imputation	Regression Imputation	Mean Imputation	Regression Imputation
Fixed (regular)	Income	-0.132	0.072	0.117	-0.03	-0.042
		(0.164)	(0.141)	(0.149)	(0.084)	(0.193)
	Housing	0.332	0.352**	0.368**	0.317**	0.348**
		(0.198)	(0.163)	(0.160)	(0.120)	(0.131)
	Poverty	-0.004	-0.005	-0.006*	-0.002	-0.005
		(0.003)	(0 .003)	(0.003)	(0.002)	(0.005)
	Education	0.035	0.031	0.031	-0.0007	0.015
		(0.060)	(0.059)	(0.057)	(0.025)	(0.030)
	Transport	0.006	0.003	0.005	-0.002	0.003
		(0.014)	(0.014)	(0.014)	(0.007)	(0.008)
	65 yrs old	0.078	0.066	0.064	0.015	0.035
		(0.079)	(0.077)	(0.075)	(0.033)	(0.041)
	African-American	0.004	0.007	0.007	0.006*	0.006*
		(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
	Hispanics	0.005	0.007*	0.008**	0.005*	0.007*
		(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
	Supermarkets	0.173	0.048	0.055	0.040	0.074
		(0.112)	(0.112)	(0.118)	(0.076)	(0.072)
	Convenience stores	0.444***	0.579***	0.665***	0.338***	0.425***
		(0.152)	(0.147)	(0.161)	(0.092)	(0.093)
	Burleson-Madison	-0.158	-0.083	-0.075	0.013	-0.039
		(0.150)	(0.137)	(0.137)	(0.102)	(0.119)
	Grimes-Washington	0.147	0.122	0.132	0.064	-0.119
		(0.368)	(0.354)	(0.343)	(0.149)	(0.311)
	Leon-Robertson	-0.532	-0.280	-0.243	0.011	0.143
		(0.550)	(0.538)	(0.523)	(0.255)	(0.178)
	Intercept	-5.151	-7.013	-7.681	-2.661	-4.280
		(5.775)	(5.957)	(5.843)	(2.590)	(2.800)
	RMSE	0.2633	0.2415	0.2473	0.1559	0.1867
	R-squared	0.452	0.636	0.676	0.581	0.62

* 10 percent significance level, ** 5 percent significance level, ***1 percent significance level

The relationship between the affordability of vegetables and the socio-economic characteristics of the community in which the store is located also varies depending upon whether affordability is measured using a high variety (Table 3.5) or basic (Table 3.7) price index. The value of housing has a large, positive association with the basic price index, though there is relatively little variation across computation or imputation method. The proportion of residents with at least a high school education and the proportion of residents at least 65 years old are both negatively associated with the basic index, but again, differences across computation and imputation method are slight. However, the coefficient estimates on the convenience store indicator are sensitive to these choices. Zero imputation yields the smallest coefficient estimate, while regression imputation produces the largest. Using a fixed basket index also tends to produce larger coefficient estimates than when using an economic index.

3.7 Discussion and Conclusions

In this study, we have examined the robustness of an empirical analysis of the affordability of fruits and vegetables are to different methods of computing price indices, different methods for imputing missing prices, and different definitions of affordability. Our results suggest that the definition of affordability, specifically the decision to include or omit less common types of items, has the far greater effect on our conclusions than either the computation (fixed versus economic index) or imputation method (zero, mean or regression). Both the mean cost of purchasing fruits and vegetables and the coefficient estimates from our regression analysis vary greatly across high variety and basic indices.

This is not to say, however, that imputation or computation methods are unimportant. Zero imputation always produces the smallest average affordability measure and intuition suggests this must be biased downward from the true cost of purchasing fruits and vegetables. It is also notable that mean imputation tends to produce smaller average costs than regression imputation using the prices of common items, which is consistent with the argument that stores that decide not to sell a particular item do so because their reservation price is higher than the price consumers are willing to pay.

Furthermore, we found a handful of cases where the economic index yielded smaller (closer to zero) coefficient estimates than the fixed-basket index. Since a fixed-basket index does not allow for substitution across items in response to differences in relative price, it would tend to exhibit greater variation and potentially inflated coefficient estimates.

Although many of our coefficient estimates were similar across computation and imputation method, this may simply be due to lack of precision because the number of available stores in our dataset is relatively small. Future work that utilized a larger sample of stores, e.g. scanner data, would be a useful extension. It would also allow for a sample that was more representative of the United States than is currently available just using stores in the Brazos Valley.

CHAPTER IV

SPATIAL PRICE COMPETITION IN THE HEALTHY FOOD MARKET IN THE BRAZOS VALLEY REGION OF TEXAS

4.1 Introduction

The consequences of poor dietary habits in terms of increased disease risk and medical costs are well-documented (Brown et al. 1995; You and Nayga 2005; Joshipura et al. 1999; Bazzano et al. 2002; Liu et al. 2000; Hung et al. 2004). The affordability and accessibility of healthful food options have been identified as barriers to consuming a healthy diet (Caraher et al. 1998; Flournoy 2006). Moreover, disparities in affordability and accessibility across different racial, ethnic and socio-economic groups have been proposed as contributing to the observed disparities in nutrition-related illness (Jetter and Cassady 2004; Young et al. 2008).

In much of the existing literature, affordability and accessibility are implicitly conceived as existing in two orthogonal dimensions. But, if food stores act as profit-maximizing firms, then economic theory suggests that the pricing strategy of one firm depends upon the presence and pricing strategies of other firms. For example, firms engaged in oligopolistic competition will compete for customers by lowering their prices. As the degree of competition increases (greater accessibility), the cost of purchasing items should decrease (greater affordability).

In this paper, the prices of fruits and vegetables (F&V) collected in the Brazos Valley (BV) as part of a census of the local food environment are used to investigate the spatial correlation in prices between stores. Specifically, the costs of purchasing F&V

are modeled using spatially autoregressive regression analysis (Anselin 1988). Regression results imply that fruit prices are positively spatially autocorrelated, which is consistent with competition effects. Vegetable prices are also positively spatially autocorrelated, but the relationship is weaker than for fruit.

The results of this study hold important implications for evaluating policies intended to improve the food environment, particularly for disadvantaged groups. For example, numerous studies have found that inner-city urban and poor rural areas exhibit both low access and high costs for purchasing healthy food options. Many studies have also found that neighborhoods with higher proportions of Black residents tend to have limited access to healthy food and must pay a higher price (Graddy 1997; Chung and Meyers 1999; Hayes 2000; Powell et al. 2007a). Interventions that increase the availability of fresh fruits and vegetables in these types of areas, but neglected competition effects on pricing would understate the true benefits of such programs. In addition, previous work that has examined the affordability of fruits and vegetables across neighborhood characteristics has ignored the spatial component of price competition (Miller and Coble 2007; Ball, Timperio, and Crawford 2009; Block and Kouba 2006). This is problematic since failure to account for spatial lags can result in inconsistent parameter estimates and/or improper interpretation of coefficients (Anselin 1988).

4.2 Literature Review

While previous theoretical and empirical research has examined the effect of competition on pricing strategies (Donkin et al. 2000; Hotelling 1929; Greenhut and

Ohta 1973; Lundberg and Lundberg 2008), very little attention has been directed toward fruit and vegetable prices specifically. Furthermore, existing studies have tended to use competition measures such as market concentration, market density or distance to the nearest competitor. Compared to many researches on housing markets (Bourassa et al. 2007; Can 1992; Jeanty et al. 2010; Pace et al. 2009), however, the explicit use of spatial econometric techniques is missing. For example, some authors in the food sector use spatial factors as dependency criteria while others (Claycombe 1991) prefer to choose characteristics of stores, and consider evenly distributed consumers around the stores.

Many authors who focus on competition between stores assess the relationship between location choice and market power. Built upon Hotelling's model (1929) of a profit-maximizing firm, Greenhut and Ohta (1973), Stern (1972) and Tirole (1988) among others examine spatial competition as a two part process: location choice followed by the selection of a price strategy. All of these authors use a location theory where distance and (market) concentration are considered in price competition. The key components of their models include the cost of entry and exit, the geography of the market (linear versus circular), the distribution of consumers and potential competitors, transportation costs and the expected time horizon (one-shot versus repeated interaction). Assumptions regarding these market characteristics can influence the implied relationship between competition and pricing (Ohta 1980 1981). For example, Pal (1998), uses the Cournot method of competition to assess interaction between firms. He finds that when the market is of linear shape, stores cluster to increase their market share, while firms locate at equal distance if the market is circular.

Claycombe and Mahan (1993) find a strong association between commuting variables and concentration of retail stores on the price of beef. These authors find that in general, the longer the commuting distance to work, the lesser the concentration of stores and lower the price of beef. Further explanation given to this finding is that of price competition between neighboring stores with perfect information regarding competitors' pricing strategies. In addition, limited number of competitors allows stores to take into consideration the ability of consumers to have better information regarding prices and lower their search costs.

Kalnins (2003) assesses how the price of hamburgers at one restaurant affects the prices at peers in the surrounding area. He finds are a price increase is associated with price changes at restaurants of the same chain but there is no direct evidence of cross-chain competition.

Hess and Gerstner (1991) consider price matching between supermarkets and grocery stores and conclude that such policies inhibit, rather than foster competition. Price matching leads to price coordination and thus higher prices. These authors use simple linear regressions by including the relative matching percentages.

Fik (1988) uses the price of a market basket for chain stores to assess spatial competition. Prices and socioeconomic variables are collected by census tract and treated as time series. Although the model specification does not employ spatial lag dependence, spatial indices are included in the analysis. Fik (1988) found a significant spatial competition between the stores.

4.3 Methods

4.3.1 Study Site

The data used in this paper were collected in the seven counties (Brazos, Burleson, Grimes, Leon, Madison, Robertson and Washington) of the Brazos Valley region of central Texas (BV). Brazos County, which includes the adjointing cities of Bryan and College Station, is the most populous in the region with nearly 200,000 residents accounting for nearly two-thirds of the total population. In addition, almost half of all food stores are located in Brazos County. It has the highest percentage of households living in poverty (Table 4.1) and 31.5% of its residents are ethnic/racial minorities. Madison County is the least populous county with the lowest median household income and follows Brazos County with the second highest poverty rate. In contrast, residents of Washington County exhibit the highest socio-economic status.

4.4 Data

Trained surveyors identified then visited all food stores in the BV. Surveyors recorded the latitude and longitude at each location then entered the establishment to collect information on item availability and cost. The fresh fruits included in the survey instrument were apples, avocado, bananas, berries, grapes, melons, oranges, peaches and pears. Fresh vegetables were broccoli or cauliflower, carrots, corn, green beans, leafy greens, lettuce, onions, potatoes, squash and tomatoes.

Table 4.1. By County Demographic Repartition and Store Types.

Store type	Brazos	Burleson	Grimes	Leon	Madison	Robertson	Washington
Population	170,954	16,598	25,603	16,462	13,379	15,819	32,034
Median income ($)	33,187	31,175	33,328	29,443	28,964	29,984	35,852
Education (%): Bachelor or higher	37	13.2	10.3	12.1	11.5	12.7	19.0
Percentage Black	10.7	14.3	18.2	10.1	21.8	22.9	17.8
Percentage Hispanic	20.8	16.5	18.2	10.9	18.9	16.8	11.6
Supermarket	11	2	2	1	2	3	2
Grocery	3	3	2	4	0	2	0

Source: US Census Bureau and BVFEP.

Table 4.2. Description of the Variables Used in the Regression Models.

Variable	Description	Type
Dependents		
hFPI	High variety fruit price index	> 0
bFPI	Basic variety fruit price index	> 0
hVPI	High variety vegetable price index	> 0
bVPI	Basic variety vegetable price index	> 0
Independents		
Income	Log household median income	> 0
Nberstores	Number of stores per census tract	≥ 0
HScl	Percentage of residents with at least high school diploma	≥ 0
Blacks	Percentage of residents who are African Americans	≥ 0
Hispanics	Percentage of residents who are Hispanics	≥ 0
Grocery	Store type grocery stores	0, 1

Prices are standardized into a price per pound ($/lb). For example, apples are recorded by price per bag or price per apple, so they have to be transformed into price per pound using their respective weights. Surveyors also identified the store as a supermarket, grocery or convenience stores based upon observation store characteristics such as size, but this information was not recorded.

Prices of individual items were aggregated into a consumption weighted price index using consumption shares from FreshLook Marketing for the Dallas metropolitan area. Reported in Table 4.2, stores that sold at least 5 fruit items, a high variety index over all fruit types was calculated. For stores selling 3 to 4 fruit items, a basic variety index over apples, oranges and bananas was calculated. Similarly, for stores that sold at

least 5 vegetables items, a high variety index over all vegetable types was calculated. For stores selling 3 to 4 vegetables items, a low variety index over carrots, lettuce, onions, potatoes and tomatoes was calculated (Dunn et al. 2009). The summary of the prices (Table 4.3) shows that the average high variety prices for fruits is 1.01/lb and for basic variety is 0.74/lb; the average high variety price for vegetables is 0.84/lb and 0.92/lb for basic variety prices.

Table 4.3. Descriptive Statistics of Factors Used in the Study.

Variables	Means	Std. Dev.
Price Indices[a]		
High Fruits	1.01	0.137
Basic Fruits	0.74	0.189
High Vegetables	0.84	0.188
Socio-economic Indicators		
Median Income	30,292	1.546
Population	1464	672.252
Education	76.02	5.517
Age 65 and older	12.14	4.669
Poverty below 200%	42.02	15.96
African-Americans	16.49	16.009
Hispanics	17.84	15.752
Number of stores	3.5	1.942
Grocery stores	0.05	0.210

[a] At the store and CBG level of computation. Std. Dev. is standard deviation.

The summary statistics for the CBGs with a store for which a price index can be calculated are also reported in Table 4.2.

The latitude and longitude were used to assign stores to Census Block Groups (CBG). Socio-economic characteristics of CBGs taken from the 2000 Census were then linked to stores.

4.5 Empirical Model

To analyze price competition between stores, spatial econometric modeling was applied to assess spatial F&V price dependence. The key components of a spatial price model account for three sources of spatial dependence: spatial autoregression, spatial autocorrelation and heterogeneity (Anselin 1988). The general model developed in Cliff-Ord (1973 1981) and Ord (1975) is formulated as:

$$Y = \rho WY + \beta X + \varepsilon \qquad (3)$$

$$\varepsilon = \lambda W\varepsilon + \mu \qquad (4)$$

$$\mu \sim N(0,\Omega),\ \varepsilon \sim N(0, \sigma^2)$$

where Y is a vector ($n \times 1$) of observations for the dependent variable, i.e. the prices of F&V, and X is a matrix ($n \times k$) of exogenous variables, such as neighborhood socioeconomic and store characteristics. The parameter ρ is the spatial dependence parameter that introduces spatial lags, i.e. spatial autoregression (Wall 2004). The parameter λ introduces spatial autocorrelation into the error structure through the residual ε, while the residual μ is assumed spatially independent. W is the spatial weight matrix ($n \times n$) and β is a vector ($k \times 1$) of parameters associated with the explanatory variables (Anselin 1988).

Previous empirical research has implicitly assumed that $\rho = \lambda = 0$, so that (1) and (2) reduce to:

$$Y = \beta X + \varepsilon \tag{5}$$

$$\varepsilon \sim N(0, \sigma^2)$$

However, stores do not price their items independently of other stores, and as noted by Tobler (1979), "everything is related to everything else, but near things are more related than distant things." Therefore, the assumption that $\rho = 0$ is relaxed, while maintaining the assumption that $\lambda = 0$. This results in a mixed regressive spatial autoregressive model (Anselin, 1988) is:

$$Y = \rho WY + \beta X + \varepsilon \tag{6}$$

$$= (I - \rho W)^{-1}\beta X + (I - \rho W)^{-1}\varepsilon \tag{7}$$

with $\varepsilon \sim N(0, \sigma^2)$,

Estimation of equation 4 directly is problematic since the dependent variable is explained by itself, generating endogeneity. Simple estimation of OLS will not be consistent due to the correlation between Y and ε (Anselin and Bera 1998). Estimation of the reduced form equation 5 avoids this issue along with other problems that are caused by a correlation between errors and regressors (Viton 2010).

As in all spatial econometric analyses, incorporating spatial dependence through the spatial weight matrix, W, is the key specification choice (Hui et al. 2007; Getis 2009). Spatial interactions are dictated by W and the choice different W's will lead to different regression results (Leenders 2002). Further, W is an $n \times n$ matrix so

computational cost of estimation is increasing in the size of the dataset as estimation of equation 5 is typically by Maximum Likelihood methods.

Researchers in the social sciences generally construct W using either distance or neighboring elements. In the former, the elements of W are defined as $w_{ij} = 1/d_{i,j}$ where $d_{i,j}$ is the distance between locations i and j. By this definition, as the distance between locations increases, the influence of one on the other decreases. In the latter, the elements of W are defined as $w_{i,j} = 1$ if i and j are neighbors, e.g. in localities that share a common border or are within a specified distance of each other, and 0 otherwise. Under both defintions, W is symmetric. Alternatively, the elements of W can be defined using the length of shared borders. This is computed by taking the ratio of the common border by the total length (perimeter) of the particular unit being considered. Doing so allows W to be asymmetric. Typically, W is standardized so that each row sums to unity. In all cases, the diagonal of the matrix is zero because the stores cannot be their own neighbors.

In the subsequent analysis, the elements of the weight matrix are defined as $w_{i,j}=1$ if $d_{i,j}<1$. That is, the prices of stores within one mile of each other are allowed to exert influence on each other. The weight matrix is left un-standardized.

4.5.1 Model Specification

The (reduced-form of the) following spatial autoregressive specification is estimated for each of the four price indices defined previously: high variety fruit price (hFPI), basic variety fruit price (bFPI), high variety vegetables price (hVPI), and basic variety vegetables price (bVPI). The model specification is presented as follow:

$$\text{Prices} = \beta_0 + \rho * W * \text{Prices} + \beta_1 * \text{Log of median Income} + \beta_2 * \text{Number of stores} + \beta_3 * \text{Percentage Blacks} + \beta_4 * \text{Percentage Hispanics} + \beta_5 * \text{Grocery} + \varepsilon \quad (8)$$

Where *Prices* is a vector for each price index and appears on both sides of the outcome equation, β_0 is the intercept of the regression, ρ is the parameter representing the spatial dependence intrinsic to our data collection and measuring the spillovers of a particular location in the neighboring areas, and β_i represents coefficients associated with each of the explanatory variables. It is important to mention that the weight matrix *W* is not standardized for the intensities of neighborhood relationships are not equally distributed; standardizing might change existing and intended economic relationships between the stores (Hao 2008). In addition, *W*Prices* is not specified in the model but generated by the inherent program codes (Pisati 2001), so the reduced form version (equation 6) is actually estimated.

Log of median Income and Log of housing value are respectively the logarithm of the household median income per county and logarithm of average housing values per CBG. Poverty represents percentage of the population living below 200% poverty line per CBG; Age is the percentage of persons who are 65 years of age or older. Grocery and Convenience are indicator variables for store type and number of stores represents the total number of stores observed in each Census tract. The number of competitors per area was directly included in the specification because it reinforced the fact that prices might vary by density of the stores associated with their geographical locations. Therefore, the greater the number of competitors the greater the interaction between stores, and the lower the prices.

4.6 Data Analysis

The extent of spatial autoregression among prices is examined using Moran's *I* and Geary's *C* whose respectively formulae are:

$$I = \frac{n}{S_0} \frac{\sum_i^n \sum_j^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{9}$$

where $S_0 = \sum_i^n \sum_j^n w_{ij}$

$$C = \frac{n-1}{2S_0} \frac{\sum_i^n \sum_j^n (x_i - x_j)^2}{\sum_i^n (x_i - \bar{x})^2} \tag{10}$$

where *i* and *j* denote stores and x_i and x_j are observations (prices) for *i* and *j*; w_{ij} is the itth, jth element of the spatial weight matrix *W*; and n is the total number of locations. For positive spatial autocorrelation, the expected value of *I* tends toward 1 and *C* toward 0. For negative spatial autocorrelation, the expected value of I tends toward -1 and *C* is greater than 1. The first method is the main focus in this study due to its popularity.

In addition, the ML estimates are assessed using Wald, Likelihood Ratio (LR) and Lagrange Multiplier (LM) tests (Anselin 1988) to examine the appropriateness of allowing for spatial autoregression. A robust version of the latter test can be utilized when preliminary results are statistically significant. The software used for the analysis was the statistical package STATA (v. 11.0, Stata Corp, College Station, TX).

The importance of the effects of all the factors used is based on the *t test* statistics but the primary focus is to determine if the coefficient associated with the lag dependent

variable ρ, was statistically significant (based on the p-values at 5% significance level). This allowed for an understanding of the fruits and vegetables price dependence on nearby store prices. The rest of the factors used were based on their significant impacts on the price indices.

We are working with the hypothesis of the existence of spatial relationships between store prices of fruits and vegetables. The following section contains the results of the regression models and related figures. The focus is about assessing the presence of spatial dependencies in price indices for fruits and vegetables in the Brazos Valley.

It is important to mention that the term spatial relationship(s) is used to cover both spatial lagged dependency that is the average of the nearby price indices and spatial autocorrelation in the error.

4.7 Results

It can be observed (Table 4.4) that the proportion of Black residents is negatively correlated with standard measures of socio-economic status. However, the relationships are positive between Hispanics and convenience stores.

Table 4.4. Correlation Between Factors Used in the Study.

	Income	Grocery	Convenience	Blacks	Hispanics	Education	Age	House	poverty	Number of stores
Income	1									
Grocery	-0.469	1								
Convenience	-0.118	-0.193	1							
Blacks	-0.0754	0.039	0.399	1						
Hispanics	0.083	-0.101	0.603	0.300	1					
Education	0.257	-0.316	0.128	-0.154	0.169	1				
Age	-0.331	0.412	-0.243	-0.068	-0.309	-0.771	1			
House	0.249	-0.284	-0.473	-0.667	-0.441	0.417	-0.229	1		
Poverty	0.084	-0.097	0.191	0.384	0.421	0.094	-0.293	-0.310	1	
Number of stores	0.262	-0.151	-0.016	0.075	0.096	-0.015	0.154	-0.095	-0.134	1

Refer to the description of the variables on page 61

A priori assessment conducted using a graph of residuals versus the predicted values shows that the residuals are not randomly distributed. The specific pattern and clustering of the residuals plotted in the figure contribute to the argument that stores are not randomly distributed with respect to price. Diagnostic tests were performed to assess the existence of spatial relationships in the prices.

The results, based on Moran's *I*, are presented in Figures 4.1– 4.4. The south-west and north-east quadrants indicate positive spatial relationships, and the south-east and north-west indicate negative spatial relationships between the prices. It is important to mention that the directional location and density of the numbers explain de relations in the quadrants, the numbers represent solely store (no economic value).

4.7.1 Fruit Price Indices

The Moran scatterplots for the high variety price indices (Figure 4.1) show a negative relationship between the price *z* (on the horizontal axis) and *Wz* (on the vertical axis), weighted average price index for neighboring stores. Moran's *I* score for this relationship of -0.029, an indication of weak spatial dependence. In addition, the negative slope implies clustering of stores with opposing price schemes. For example, plots in the upper-left quadrant indicate the presence of stores with low price indices surrounded by stores with high price indices.

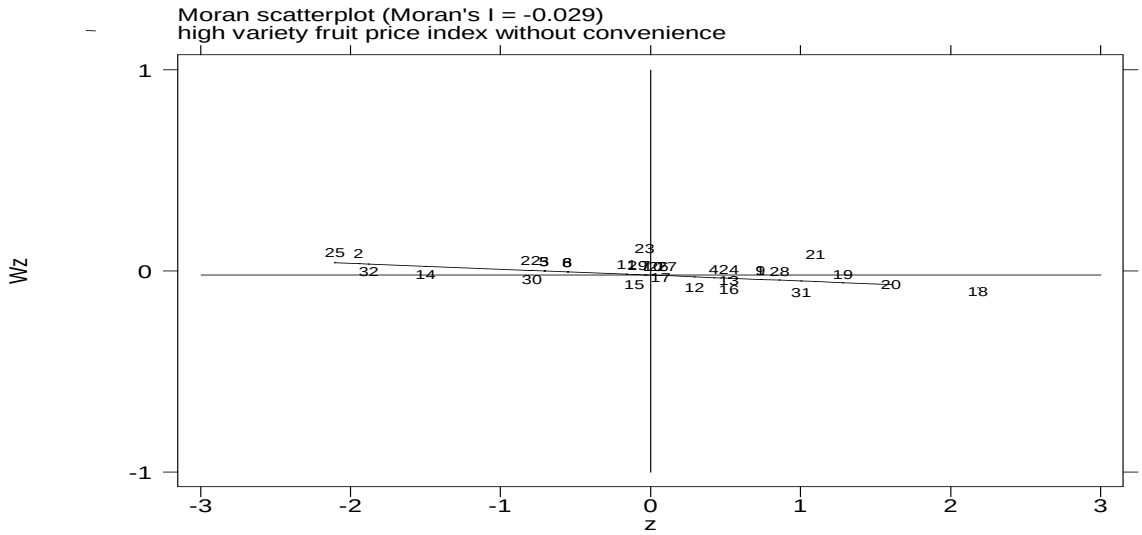

Figure 4.1. Moran scatterplot for high variety fruit price index.

Similar to the high variety index, Figure 4.2 reveals a negative spatial relationship between stores pricing of the low variety fruit price index. This relationship is weaker (-0.015), however. Based on the direction, slope, and value of Moran's *I*, the scatterplots for the fruit price indices demonstrate that stores with lower prices tend to be surrounded by those with higher prices, and vice versa.

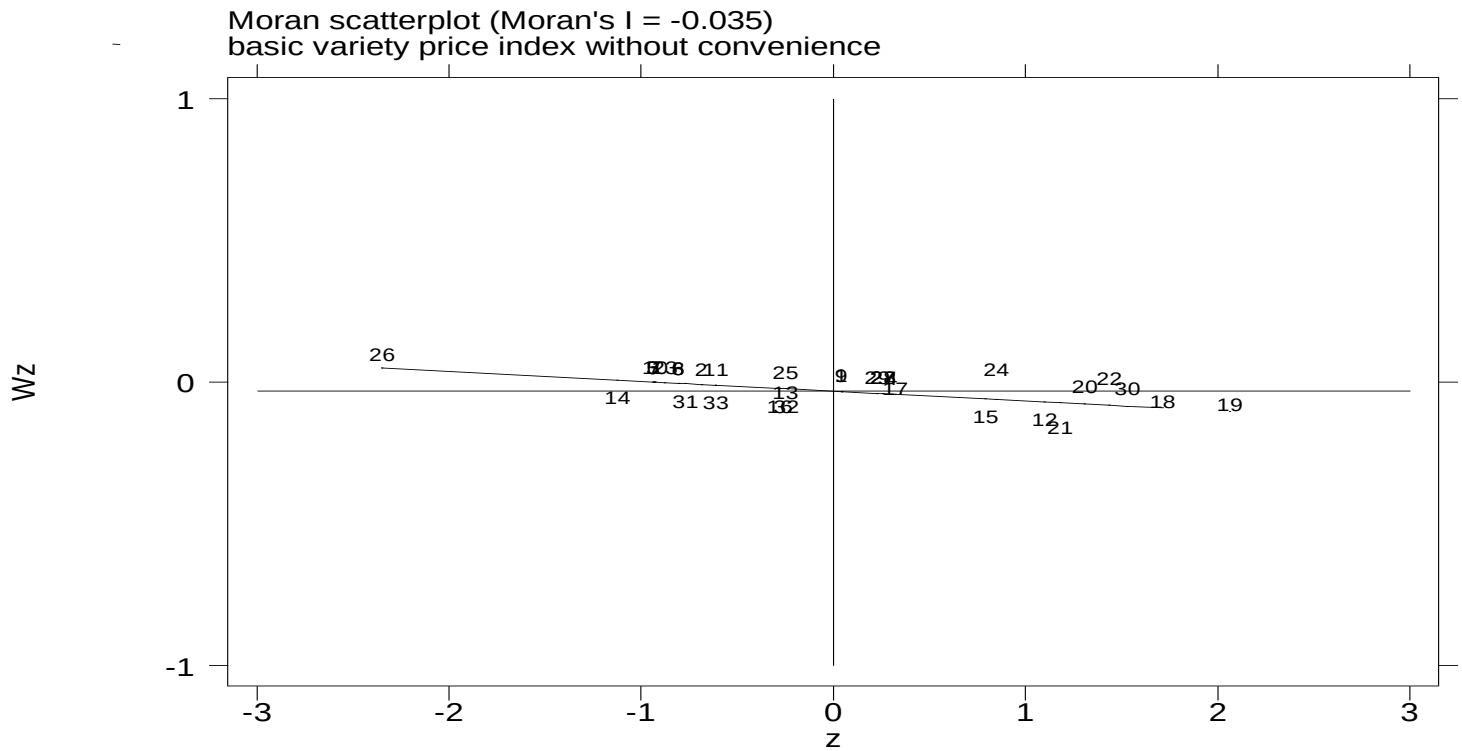

Figure 4.2. Moran scatterplot for basic variety fruit price index.

4.7.2 Vegetable Price Indices

This section is similar to the fruit price indices. The results for the high and low vegetable price indices (Figures 4.3 and 4.4) are similar. For both, Moran's *I* is negative, -0.029 for the former and -0.027 for the latter (se *z* values). Given the magnitudes of the Moran's I scores, the results indicate only a week presence of spatial autocorrelation. One interpretation is that store pricing lacks a spatial dimension. Alternatively, difference is local characteristics which have not been accounted for obscure the true spatial pricing relationship. Therefore, we also conduct spatial diagnostic tests from OLS regression results (Tables 4.7 and 4.8).

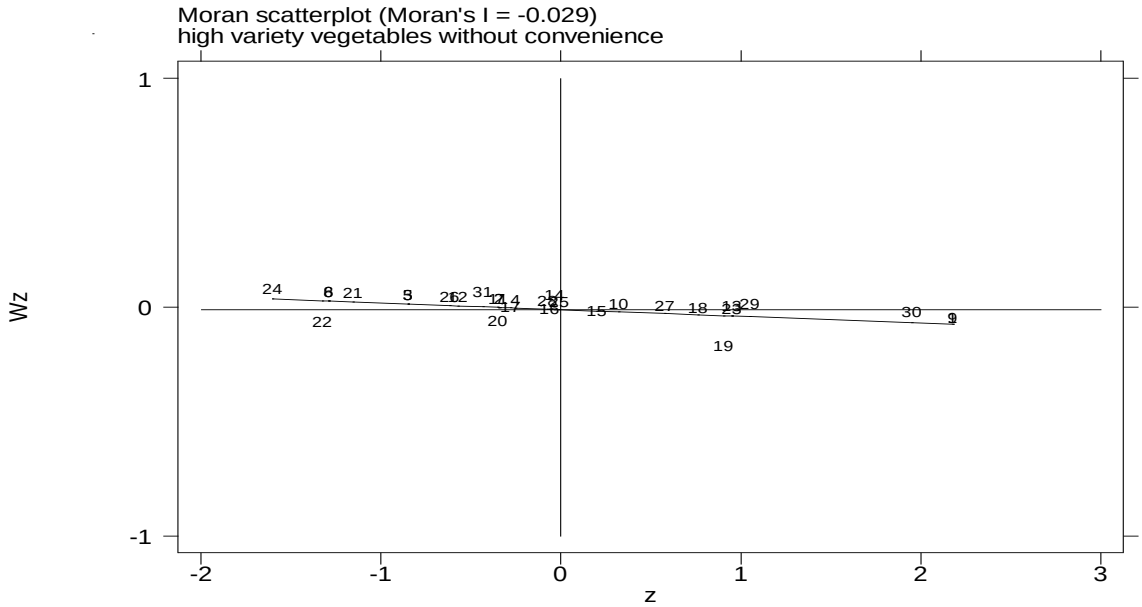

Figure 4.3. Moran scatterplot for high variety vegetable price index.

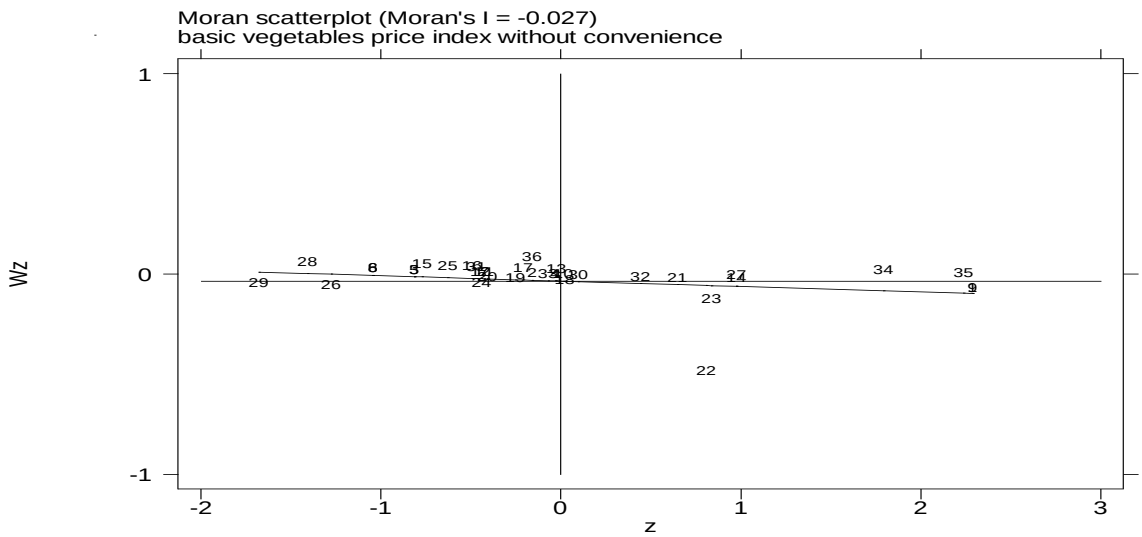

Figure 4.4. Moran scatterplot for basic variety vegetable price index.

4.8 Estimation of the Spatial Regression Models

Three different types of regressions, OLS, spatial lagged and spatial error, were conducted. Diagnostic tests for spatial dependency were performed and results are summarized in the tables on pages 76 and 77. Six tests—specifically, Moran's *I*, Lagrange Multiplier (LM), Robust Lagrange Multiplier, Wald (W), and Likelihood Ration (LR)—were performed for each price index under the null hypothesis that there is no spatial autocorrelation (dependence/relationship) between a price index at a particular store and neighboring stores. The first three tested for spatial error dependence in the OLS regressions, and the first and the last two were performed for the spatial lagged and spatial error based on the ML regressions.

Results for the high variety fruit price index are reported in Table 4.5. Assessing the residuals from OLS regression (column 1), Moran's *I* is positive and highly

statistically significant ($p < 0.05$) indicating the presence of spatial autocorrelation. From the second to the third column, however, neither ρ nor λ are significant indicating that one cannot reject the null hypothesis of no spatial relationships between a price index for a given store and the prices of competitors located in the surrounding areas. Looking to the LM, LR and W tests, these similarly indicate that the modeling of spatial relationships is not statistically significant (even at 10% level). Thus, except for Moran's *I*, all the tests (reported and not reported in column 1) based on OLS and ML fail to indicate the presence of spatial dependence parameters.

Table 4.5. Impacts of Nearby High Variety Fruit Price Stores.

	High variety fruits price index		
Variables	OLS (I)	Lag (II)	Error(III)
Income	-0.138 (0.093)	2.346 (0.869)**	2.335 (0.903)**
Number stores	0.003 (0.013)	-0.101(0.093)	-0.104 (0.098)
Blacks	-0.0004 (0.002)	-0.010 (0.013)	-0.0005 (0.013)
Hispanics	-0.0005 (0.003)	-0.0001(0.001)	-0.0001 (0.001)
Grocery	-0.029 (0.070)	-0.00015(0.002)	-0.0002 (0.002)
Constant	2.448 (0.967)**	-0.031(0.062)	-0.029 (0.064)
Test Results			
Moran's *I*	3.368***		
rho		-0.009(0.011)	
lambda			-0.004 (.007)
Wald		0.764	0.367
Likelihood Ratio		0.755	0.537
Lagrange Multiplier		0.592	0.319
Robust Lagrange multiplier		0.341	0.068

** denote significance at 5% level,
*** denote significance at 1% level.

Table 4.6 presents the results for the basic variety fruit price index. Again examining the OLS regression coefficients using Moran's *I* reveals spatial correlation. ML regressions results further reveal spatial correlation in the error, though no evidence of lagged dependency. Unlike the lag parameter ρ, the spatial relationship parameter λ is highly statistically significant ($p < 0.01$). In addition the Wald and LR of the spatial error specification support the inclusion of a spatial error component. In contrast, none of the tests are significant in the spatial lagged estimation.

Table 4.6. Impacts of Nearby Basic Variety Fruit Price Stores.

	Basic variety fruits price index		
Variables	OLS (I)	Lag(II)	Error(III)
Income	0.049 (0.095)	0.069 (0.875)	0.365 (0.138)***
Number stores	-0.001 (0.014)	-0.101(0.093)	0.105 (0.091)
Blacks	-0.0002 (0.002)	-0.007(0.012)	0.032(0.010)***
Hispanics	-0.002 (0.003)	0.0002 (0.001)	-0.004(0.001)**
Grocery	-0.176 (0.068)**	-0.002 (0.002)	-0.001(0.003)
Constant	0.377 (0.952)	0.174 (0.060)**	0.399 (0.033)***
Test Results			
Moran's *I*	3.412***		
rho		-0.020 (0.013)	
lambda			-0.566 (0.043)***
Wald		2.345	174.469***
Likelihood Ratio		2.268	43.620***
Lagrange Multiplier		1.669	0.312
Robust Lagrange multiplier		1.389	0.032

** denote significance at 5% level,
*** denote significance at 1% level.

In Table 4.7, Moran's I indicates significant ($p < 0.01$) presence of spatial correlations for the high variety vegetable price index. ML regression results indicate that both the

spatial relationship parameters ρ and λ are statistically ($p < 0.05$ and $p < 0.01$, respectively) different from zero. This significance indicates the presence of spatial relationships in the dependent variable and in the error term. The Wald test supports the inclusion of a spatial lag in the explanatory variables, while the Wald and LR support the inclusion of a spatial error.

Table 4.7. Impacts of Nearby High Variety Vegetable Price Stores.

	High variety vegetable price index		
Variables	OLS (I)	Lag(II)	Error(III)
Income	-0.0676 (0.126)	1.406 (1.102)	-1.620 (0.016)***
Number stores	-0.030 (0.018)	-0.101(0.093)	0.057(0.120)
Blacks	0.0009 (0.002)	-0.041(0.016)***	-0.006 (0.002)***
Hispanics	0.0008 (0.003)	0.002 (0.002)	-0.004 (0.0004)***
Grocery	0.165 (0.098)**	0.002 (0.003)	0.007(0.0003)***
Constant	1.489 (1.261)	-0.146 (0.082)*	-0.316 (0.004)***
Test Results			
Moran's *I*	4.516***		
rho		-0.043 (0.020)**	
lambda			-1.000 (0.003)***
Wald		4.859**	86000***
Likelihood Ratio		4.511	226.37***
Lagrange Multiplier		2.604	0.627
Robust Lagrange multiplier		2.209	0.232

* denote significance at 10% level,
** denote significance at 5% level,
*** denote significance at 1% level.

As with the high variety vegetable price indices, Moran's I is positive and significant for the basic variety (Table 4.8) while ρ and λ are negative and statistically significant. Comparing the two spatial models (columns II and III), all the tests are

significant for spatial lag model results whereas Wald and Likelihood ratio are significant for the spatial error model.

Table 4.8. Impacts of Nearby Basic Variety Vegetable Price Stores.

	Basic variety vegetable price index		
Variables	OLS (I)	Lag(II)	Error(III)
Income	-0.086 (0.134)	1.844 (1.192)	-0.157(0.134)
Number stores	-0.037 (0.019)*	-0.101(0.093)	-0.040 (0.116)
Blacks	0.001(0.002)	-0.040 (0.016)**	-0.049 (0.017)***
Hispanics	0.0005 (0.003)	0.001(0.002)	0.0004 (0.002)
Grocery	0.203 (0.092)**	0.002 (0.003)	-0.009 (0.006)
Constant	1.663 (1.338)	-0.227(0.080)***	-0.244 (0.044)***
Test Results			
Moran's *I*	2.884***		
rho		-0.018 (0.008)**	
lambda			-0.567(0.101)***
Wald		4.908**	31.201***
Likelihood Ratio		4.601*	29.992*
Lagrange Multiplier		3.850*	0.503
Robust Lagrange multiplier		3.354*	0.006

* denote significance at 10% level,
** denote significance at 5% level,
*** denote significance at 1% level.

Regarding socioeconomic factors included in the estimations, contrary to the expectations, income has positive impacts ($p < 0.01$) on both basic food price indices. However, the impact is negative ($p < 0.01$) for high variety vegetable price indices. This indicates for an increase in income by a factor of 1, basic prices will increase whereas high variety prices will go down. As expected, the number of stores has negative impacts ($p < 0.01$ for high variety, and $p < 0.05$ and 0.01 for basic variety) the vegetable price

indices. This is contrary to the fruits where the number of stores has positive impacts on the basic price indices. The effects of the residents classified as minorities are dissimilar. Blacks have negative impacts for basic fruit and high variety vegetable prices when the estimation is based on spatial error. Hispanics has positive effects ($p < 0.01$) for these price indices.

Overall, the estimations for the vegetable price indices have more significant parameters compared to the fruits. This might be due to the longevity of each food item.

4.9 Discussion and Conclusions

This paper applied spatial econometric techniques to investigate the relationship between fruit and vegetable (F&V) prices at stores located nearby that could be potential competitors. Based on data collected in the Brazos Valley region of Texas, we conducted an assessment of the presence of spatial price dependence and determined the spatial effects in terms of both lagged dependency and error structure. These spatial relationships covered the prices of high and basic variety vegetable consumption baskets and the basic variety fruit consumption basket.

These results are contrary to our expectation that low-price stores would tend to drive down prices at potential competitors, i.e. positive spatial dependence. Several explanations are possible. First, fruit, particularly items in the high variety basket such as berries and melon, may be highly seasonal with pricing subject to market forces largely outside the control of local outlets.

Second, the strong negative relationship among vegetable prices may be the result of contrasting pricing strategies among larger outlets. National chains can have a

broader pricing policy compared to local stores so that the latter might be more flexible in their pricing schemes. In addition, national chain stores with pricing strategies such as Every Day Low Price (EDLP) might be less flexible than those with strategies categorized as promotions (Promo). Stores adopting EDLP strategy have more stable prices compared to the latter (Lal and Rao 1997). When the two types of stores are located in the same neighborhoods, prices at promo can be lower than EDLP during the promotions generating negative spatial dependence. This type of relationship, well detailed by Lal and Rao (1997), is characteristic of the competition between Albertson's (EDLP) and Kroger or Safety (Promo).

Although this study provides additional knowledge about fruit and vegetable market pricing behavior, there are limitations worth mentioning. First, the small geographic region considered here might hinder the generalization of the findings to other locations. Another limitation is the fact that this study did not involve time variation during the data collection. As fruits and vegetables are in general seasonal products, the prices vary over time and space. Including time and space in the study might provide ample understanding of the pricing strategies and factors affecting the prices.

Finally, store may be competing along multiple dimensions, e.g. quality, variety, etc. It may be more profitable for a high price-high quality store to open near a low price-low quality store than to open near another high price-high quality store. This type of market segmentation would also generate negative spatial dependence.

Despite these limitations, our findings are that there exist spatial relationships between stores in term of vegetable prices. The implication for policy makers is understanding how to improve the food environment through competition between stores. For example, if stores with high prices compete with those charging high prices for one market segment, and stores with low prices compete with those charging low prices for another segment, the end result will be lower prices for fruits and vegetables in the Brazos Valley conditional on the market segment.

CHAPTER V

CONCLUSIONS

The burden of obesity-related illness, which disproportionately affects low income households and historically disadvantaged racial and ethnic groups, is a leading public health issue in the United States. In addition, previous research has documented differences in eating behavior and dietary intake between racial and ethnic groups, as well as between urban and rural residents.

This study has extended prior work by examining the affordability of fresh fruit and vegetables from traditional and non-traditional food stores in a large rural area; how access to an affordable supply of fresh fruit and vegetables differs by neighborhood and geographic inequalities; whether results are robust to the choice of empirical methods, such as the definition of the market basket and price imputation technique; and how the pricing strategies of stores interact in a spatial economic framework.

First, the determinants of affordability of fruits and vegetables are assessed in Chapter II. We demonstrated that individuals who shop at food stores located in the rural counties of the Brazos Valley region must pay significantly more to attain the USDA recommended level of fresh fruit and vegetable consumption through fresh whole items than residents who shop in the urban area. In addition, stores located in neighborhoods with higher proportions of minority residents also charged more for fresh produce. These results are consistent with previous works in the literature and further illustrates the challenges that historically disadvantaged rural households face with respect to making healthy lifestyle decisions.

In Chapter III, we have examined the robustness of an empirical analysis of the affordability of fruits and vegetables are to different methods of computing price indices, different methods for imputing missing prices, and different definitions of affordability. Our results suggest that the definition of affordability, specifically the decision to include or omit less common types of items, has the far greater effect on our conclusions than either the computation (fixed versus economic index) or imputation method (zero, mean or regression). Both the mean cost of purchasing fruits and vegetables and the coefficient estimates from our regression analysis vary greatly across high variety and basic indices. This is not to say, however, that imputation or computation methods are unimportant. Mean imputation tends to produce smaller average costs than regression imputation using the prices of common items, which is consistent with the argument that stores that decide not to sell a particular item do so because their reservation price is higher than the price consumers are willing to pay.

Finally, Chapter IV applied spatial econometric techniques to investigate the relationship between fruit and vegetable (F&V) prices at stores located nearby that could be potential competitors. Based on data collected in the Brazos Valley region of Texas, we conducted an assessment of the presence of spatial price dependence and determined the spatial effects in terms of both lagged dependency and error structure. These spatial relationships covered the prices of high and basic variety vegetable consumption baskets and the basic variety fruit consumption basket. These results are contrary to our expectation that low-price stores would tend to drive down prices at potential competitors, i.e. positive spatial dependence.

The approach and findings of this study are relevant and have important research and policy implications for understanding access and availability of affordable, healthy foods. Access to a good variety of affordable healthy foods, such as fruit and vegetables, can play a pivotal role in the nutritional health of rural families. Many of these families live in socio-economically deprived neighborhoods; many have a low household income, are unemployed, older, or lack access to a vehicle. In order for rural families to be food secure and have access to fruit and vegetables, food resources need to be available and affordable in local stores.

REFERENCES

Andreyeva, T., D.M. Blumenthal, M.B. Schwartz, M.W. Long, and K.D. Brownell. 2008. "Availability and Prices of Foods across Stores and Neighborhoods: The Case of New Haven, Connecticut." *Health Affairs* 27:1381–1388.

Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. and A. Bera. 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." In A. Ullah and D.E. Gilles, eds. *Handbook of Applied Economic Statistics.* New York NY: Marcel Dekker, pp. 237–289.

Ard, J.D., S. Fitzpatrick, R.A. Desmond, B.S. Sutton,M. Pisu, A.B. Allison,F. Franklin, and M. Baskin. 2007. "The Impact of Cost on the Availability of Fruits and Vegetables in the Homes of Schoolchildren in Birmingham, Alabama." *American Journal of Public Health* 97:367–72.

Armknecht, P.A. and F. Maitland-Smith. 1999. "The Price Imputation and Other Techniques for Dealing with Missing Observations, Seasonality and Quality Change in Price Indices" Working Paper. Vol., pp. 1-28, 1999. Accessed on July 03, 2009. http://ssrn.com/abstract=880605.

Armknecht, P.A. and B.R. Moulton. 1995. "Quality Adjustment in Price Indices: Methods for Imputing Price and Quality Change." Paper Prepared for the Second

International Conference on Price Indices. Accessed on July 19, 2009. >http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/8e98d9c3d6e9363eca25727500004401/$FILE/1995%202nd%20Meeting%20%20Quality%20adjustment%20in%20price%20indices%20Methods%20for%20imputing%20price%20and%20quality%20change.pdf<

Ball, K., A. Timperio, and D. Crawford. 2009. "Neighbourhood Socioeconomic Inequalities in Food Access and Affordability." *Health & Place*.15:578–85.

Bazzano, L.A., J. He, L.G. Odgden, C.M. Loria, S. Vupputuri, L. Myers, and P.K. Whelton. 2002. "Fruit and Vegetable Intake and Risk of Cardiovascular Disease in US Adults: The First National Health and Nutrition Examination Survey Epidemiologic Follow-up Study." *American Journal of Clinical Nutrition* 76:93–99.

Blend, D. and T. Marwala. 2008. "Comparison of Data Imputation Techniques and their Impact" Accessed on June 28, 2009. >http://arxiv.org/ftp/arxiv/papers/0812/0812.1539.pdf<

Block, D. and J. Kouba. 2006. "A Comparison of the Availability and Affordability of a Market Basket in Two Communities in the Chicago Area." *Public Health Nutrition* 9:837–845.

Bourassa, S.C., C. Eva, and H. Martin. 2007. "Spatial Dependence, Housing Submarkets, and House Price Prediction." Journal of Real Estate Financial Economics 35: 143–160.

Bradley, R. 2003. "Price Index Estimation Using Price Imputation for Unsold Items." In R.C. Feenstra & Matthew D. Shapiro, eds. *Scanner Data and Price Indexes*, NBER

Books, National Bureau of Economic Research, Inc., Studies in Income and Wealth 64, Chicago IL: University of Chicago Press, pp. 349–379.

Brown, L.M., C.A. Swanson, G. Gridley, G.M. Swanson, J.B. Schoenberg, R.S. Greenberg, D.T. Silverman, L.M. Pottern, R.B. Hayes, A.G. Schwartz, J.M. Liff, J.F. Fraumeni, and R.N. Hoover Jr. 1995. "Adenocarcinoma of the Esophagus: Role of Obesity and Diet." *Journal of the National Cancer Institute* 87 (2): 104–109.

Bustillos, B.D., J.R. Sharkey, J. Anding, and A. McIntosh. 2009. "Availability of Healthier Food Alternatives in Traditional, Convenience, and Non-traditional Types of Food Stores in Two Rural Texas Counties." *Journal of American Dietetic Association* 109:883–889.

Can, A. 1992. "Specification and Estimation of Hedonic Housing Price Models." *Regional Science and Urban Economics* 22: 452–474.

Caraher, M., P. Dixon and T. Lang. 1998. "Access to Healthy Foods: Part I. Barriers to Accessing Healthy Foods: Differentials by Gender, Social Class, Income and Mode of Transport." *Health Education Journal* 57:191–201.

Carpenter, J.R. and M.K. Kenward. 2005. "Multiple Bias Modeling for Observational Data." *Journal of the Royal Statistical Societies Series* 168: 267–308.

Cassady, D., K.M. Jetter., and J. Culp. 2007. "Is Price a Barrier to Eating More Fruits and Vegetables for Low-Income Families?" *Journal of American Dietetic Association.* 107:1909–1915.

Chung, C., and S.L. Meyers. 1999. "Do the Poor Pay More for Food? An Analysis of Grocery Store Availability and Food Price Disparities." *The Journal of Consumer Affairs*. 33 (2): 276–296.

Claycombe, R. J. 1991. "Spatial Retail Markets." *International Journal of Industrial Organization* 9: 303–313.

Claycombe, R. J. and T.E. Mahan. 1993. "Spatial Aspects of Retail Market Structure:Beef Pricing Revisited." *International Journal of Industrial Organization* 11: 283–291.

Cliff, A.D., and J.K. Ord. 1973. *Spatial Autocorrelation*. London: Pion Press.

Cliff, A.D., and J.K. Ord. 1981. *Spatial Process*. London: Pion Press.

Cummins, S., and S. Macintyre. 2002. "A Systematic Study of an Urban Foodscape: The Price and Availability of Food in Greater Glasgow." *Urban Studies* 39 (11): 2115–2130.

Donkin, A.J., A.E. Dowler, S.J. Stevenson, and S.A. Turner. 2000. "Mapping Access to Food in a Deprived Area: The Development of Price and Availability Indices. *Public Health Nutrition* 3(1): 31–38.

Dubowitz, T., M. Heron, C.E. Bird, N. Lurie, B.K. Finch, R. Basurto-Davilá, L. Hale, amd J.J. Escarce. 2008. "Neighborhood Socioeconomic Status and Fruit and Vegetable Intake Among Whites, Blacks, and Mexican Americans in the United States." *American Journal of Clinical Nutrition* 87:1883–91.

Dunn, R. A., J. R. Sharkey, J. Lotade-Manje, Y. Bouhlal, and R.M. Nayga. 2011. "Socio-economic Status, Racial Composition, and the Affordability of Fresh Fruits

and Vegetables in the Neighborhoods of a Large Rural Region in Texas." *Nutrition Journal* 10:6

Dunn, R. A., R. Nayga, and J. R. Sharkey. 2009. "Calculating a Store-level Price Index for Fresh Fruits and Vegetables: Comparing the Thrifty Food Basket with an index that Accounts for Substitution." Working paper. Department of Agricultural Economic, Texas A&M University, College Station TX.

Fik, T.J. 1988. "Spatial Competition and Price Reporting in Retail Food Markets." *Economic Geography* 64 (1): 29–44.

Flournoy, R. 2006. "Healthy Food, Strong Communities: Fresh Fruits and Veggies are Good for More than Just Your Health." Shelterforce Online, Issue no. 147 available at http://www.nhi.org/online/issues/147/healthyfoods.html.

Fraser, G.E., R. Yan, T.L. Butler, K. Jaceldo-Siegl, W.L. Beeson, and J. Chan. 2009. "Missing Data in a Long Food Frequency Questionnaire: Are Imputed Zeroes Correct?" *Epidemiology* 20 (2): 289–294.

Fraumeni, J.F., R.N. Hoover Jr. 1995. "Adenocarcinoma of the Esophagus: Role of Obesity and Diet." *Journal of the National Cancer Institute* 87 (2):104–109.

Gan, X., A. WC. Liew, and H. Yan. 2006. "Microarray Missing Data Imputation Based on a Set Theoretic Framework and Biological Knowledge" *Nucleic Acids Research* 34(5): 1608–1619

Getis, A. 2009. "Spatial Weights Matrices." Geographical Analysis 41: 404–410.

Graddy, K. 1997. “Do Fast-food Chains Price Discriminate on the Race and Income Characteristics of an Area?” *Journal of Business & Economic Statistics* 15 (4): 391–401.

Greenhut, M.L., and H. Ohta. 1973. “Spatial Configurations and Competitive Equilibrium.” *Weltwirtschaftl iches Archiv* 109: 87–104.

Hawthorne, G., and P. Elliott. 2005. “Imputing Cross-sectional Missing Data: Comparison of Common Techniques.” *Australian and New Zealand Journal of Psychiatry* 39:583–590

Hao, Q. 2008. “Review on Spatial Econometric Analysis.” International Seminar on Future Information Technology and Management Engineering pp. 644–647. >http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4746576&isnumber=4746420<

Hayes, L.R. 2000. “Do the Poor Pay More? An Empirical Investigation of Price Dispersion in Food Retailing.” Working paper, Department of Economics, Rutgers University, New Brunswick NJ.

Heitjan, D.F. 1997. “Annotation: What Can Be Done about Missing Data? Approaches to Imputation.” *American Journal of Public Health* 87:548–550.

Hendrickson ,D., C. Smith, and N. Eikenberry. 2006. “Fruit and vegetable access in four low-income food deserts communities in Minnesota.” *Agriculture and Human Values.* 23: 371–383.

Hess, J. D. and E. Gerstner. 1991. “Price-matching Policies: An Empirical Case.” *Managerial and Decision Economics* 12: 305–315.

Hotelling, H. 1929. “Stability in Competition.” *Economic Journal* 39 (149): 41–57.

Hui E. C. M, C. K. Chau, L. Pun, and M.Y. Law. 2007. “Measuring the Neighboring and Environmental Effects on Residential Property Value: Using Spatial Weighting Matrix.” *Building and Environment* 42:2333–2343

Hung, H. C, K. J. Joshipura, R. Jiang, Hu FB, D. Hunter, Smith-Warner SA, Colditz GA, B. Rosner, D. Spiegelman, and W. C. Willett. 2004. “Fruit and Vegetable Intake and Risk of Major Chronic Disease.” *Journal of the National Cancer Institute* 96: 1577–1584.

Inagami, S., D.A. Cohen, B.K. Finch, and S.M. Asch. 2006. “You Are Where You Shop: Grocery Store Locations, Weight, and Neighborhoods.” *American Journal of Preventive Medicine* 31:10–7.

Jeanty, P. W., M. Partridge, and E. Irwin. 2010. “Estimation of a Spatial Simultaneous Equation Model of Population Migration and Housing Price Dynamics.” *Regional Science and Urban Economics, doi: 10.1016/j.regsciurbeco.2010.01.002*.

Jetter, K.M. and D.L. Cassady. 2006. “The Availability and Cost of Healthier Food Alternatives” *American Journal of Preventive Medicine* 30 (1): 38–44.

—. 2004. Explaining Disparities in the Cost of Healthier Food. Paper presented at the 2004 AAEA Annual Meeting, Denver, 1–4 August 2004.

Joshipura, K.J., A. Ascherio, J.E. Manson, M.J. Stampfer, E.B. Rimm, F.E. Speizer, C.H. Hennekens, D. Spiegelman, and W.C. Willett. 1999. “Fruit and Vegetable Intake in Relation to Risk of Ischemic Stroke.” *The Journal of the American Medical Association* 282:1233–1239.

Kalnins, A. 2003. "Hamburger Prices and Spatial Econometrics." *Journal of Economics and Management Strategy* 12(4): 591–616.

Kaufman, P.R., J.M. MacDonald, S.M. Lutz, and D.M. Smallwood. 1999. "Do the Poor Pay More for Food? Item Selection and Price Differences Affect Low-Income Household Food Costs" *Food and Rural Economic Division,* Washington DC: Economic Research Service, U.S. Department of Agriculture (Agricultural Economic Report 759).

Lal, R. and R. Rao. 1997. "Supermarket Competition: The Case of Every Day Low Pricing." *Marketing Science* 16(1): 60–80.

Latham, J., and T. Moffat. 2007. "Determinants of Variation in Food Cost and Availability in Two Socioeconomically Contrasting Neighborhoods of Hamilton, Ontario, Canada." *Health & Place.* 13: 273–287.

Lee, A.J., A.M. Darcy, D. Leonard, A.D. Groos, C.O. Stubbs, S.K. Lowson, S.M. Dunn, T. Coyne, and M.D. Riley. 2002. "Food Availability, Cost Disparity and Improvement in Relation to Accessibility and Remoteness in Queensland." *Australian and New Zealand Journal of Public Health* 26:266–272.

Leenders, R.T.AJ. 2002. "Modeling Social Influence Through Network Autocorrelation: Constructing the Weight Matrix." *Social Network* 24: 21–47.

Liese, A.D., K.E. Weis, and D. Pluto. 2007. "Food Store Types, Availability and Cost of Foods in a Rural Environment." *Journal of the American Dietetic Association*.107:1916–1923.

Liu, S., J.E. Manson, I.M. Lee, S.R. Cole, C.H. Hennekens, W.C. Willett, and J.E. Buring. 2000. "Fruits and Vegetables and Risk of Cardiovascular Disease: The Women's Health Study." *The American Journal of Clinical Nutrition* 72 (4): 922–928.

Lotade-Manje, J., R. Dunn, R. Nayga, and J. Sharkey. 2009. "Determinants of Affordability of Healthy Food in the Rural Counties: Case of the Brazos Valley Area." Working paper, Department of Agricultural Economics, Texas A&M University, College Station TX.

Lundberg, J. and S. Lundberg. 2008. "Using Spatial Econometrics to Test for Price Competition between Food Stores Based on Swedish Data." Working paper, Center for Regional Science at Umea University (CERUM), Umeå Sweden.

McClelland, J.W., W. Demark-Wahnefried, R.D. Mustian, A.T. Cowan, and M.K. Campbell. 1998. "Fruit and Vegetable Consumption of rural African Americans: Baseline Survey Results of the Black Churches United for Better Health 5 A Day Project." *Nutrition and Cancer* 30:148–157.

Miller, J.C. and H.K. Coble. 2007. "Cheap Food Policy: Fact or Rhetoric?" *Food Policy* 32: 98–111.

Mokdad, A.H., E.S. Ford, B.A. Bowman, W.H. Dietz, F. Vinicor, V.S. Bales, and J.S. Marks. 2003. "Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors," *Journal of the American Medical Association* 289:76–79.

Morland, K. and S. Filomena. 2007. "Disparities in the Availability of Fruits and Vegetables between Racially Segregated Urban Neighborhoods." *Public Health Nutrition* 10:1481–1489.

Morland, K., S. Wing and A.D. Roux. 2002a. "The Contextual Effect of the Local Food Environment on Residents' Diets: The Atherosclerosis Risk in Communities Study." *American Journal of Public Health* 92:1761–1767.

Morland, K., S. Wing, A.D. Roux, and C. Poole. 2002b. "Neighborhood Characteristics Associated with the Location of Food Stores and Food Service Places." *American Journal of Preventive Medicine* 22:23–29.

Morris, P.M., L. Neuhauser, and C. Campbell. 1992. "Food Security in Rural America: A Study of Availability and Costs of Food." Journal of Nutrition Education 24:52S–8S.

Morton, L.W. and T.C. Blanchard. 2007. "Starved for Access: Life in Rural America's Food Deserts." *Rural Realities* 1:1–10.

Moulton, B.R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *The Review of Economics and Statistics* 72:334–38.

Musil, C., C. Warner, P. Yobas and S. Jones. 2002. "A Comparison of Imputation Techniques for Handling Missing Data." *Western Journal of Nursing Research* 24(7):815–29.

Nordholt, E.S. 1998. "Imputation: Methods, Simulation Experiments and Practical Examples." *International Statistical Review / Revue Internationale de Statistique*. 66 (2): 157–180.

O'Brien, R.M. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality and Quantity* 41:673–690.

Ogden C, Caroll M, Curtin L, McDowell M, Tabak C, Flegal K. 2006. "Prevalence of Overweight and Obesity in the United States, 1999-2004." *Journal of the American Medical Association* 295:1549–1555.

Ohta, H. 1980. "Spatial Competition, Concentration and Welfare." *Regional Science and Urban Economics* 10: 3–16.

—. 1981. "The Price Effects of Spatial Competition." *Review of Economic Studies* 48 (2): 317–325.

Olinsky, A., S. Chen, and L. Harlow. 2003. "The Comparative Efficacy of Imputation Methods for Missing Data in Structural Equation Modeling." *European Journal of Operational Research* 151 (1): 53–79.

Ord, J.K. 1975. "Estimation Methods for Models of Spatial Interaction. Journal of American Statistical Association 70 (349): 120–126.

Pace, K. R., J. LeSage, and S. Zhu. 2009. "Impact of Cliff and Ord on the Housing and Real Estate Literature." *Geographical Analysis* 41: 418–424.

Pal, D. 1998. "Does Cournot Competition Yield Spatial Agglomeration?" *Economics Letters* 60: 49–53.

Pisati, M. 2001. "sg162: Tools for Spatial Data Analysis." *Stata Technical Bulletin* 60: 21–37.

Pollar, J., S.F.L. Kirk and J.E. Cade. 2002. "Factors Affecting Food Choice in Relation to Fruit and Vegetable Intake: A Review" *Nutrition Research Reviews* 15: 373–387.

Powell, L.M., M.C. Auld, F.J. Chaloupka, P.M. O'Malley, and L.D. Johnston. 2007a. "Association between Access to Food Stores and Adolescent Body Mass Index." *American Journal of Preventive Medicine*.33:S301–S7.

Powell, L. M., S. Slater, D. Mirtcheva, Y. Bao, and F.J. Chaloupka. 2007b. "Food Store Availability and Neighborhood Characteristics in the United States." *Preventive Medicine* 44: 189–195.

Raymond, M. and D. Roberts. 1987. "A Comparison of Methods for Treating Incomplete Data in Selection Research." *Educational and Psychological Measurement* 47: 13–26.

Raymond, M.R. 1986. "Missing Data in Evaluation Research." *Evaluation and the Health Profession* 9: 395–420.

Rose, D. and R. Richards. 2004. "Food Store Access and Household Fruit and Vegetable Use among Participants in the US Food Stamp Program." *Public Health Nutrition* 7:1081–1088.

Satia, J.A. 2009. "Diet-Related Disparities: Understanding the Problem and Accelerating Solutions." *Journal of American Dietetic Association* 109:610–615.

Sehgal, M.S.B., I. Gondal, L.S. Dooley, and R. Coppel. 2008. "Ameliorative Missing Value Imputation for Robust Biological Knowledge Inference." *Journal of Biomedical Informatics* 41: 499–514.

Sharkey, J. and S. Horel. 2008. "Neighborhood Socioeconomic Deprivation and Minority Composition Are Associated with Better Potential Spatial Access to the Food Environment in a Large Rural Area." *Journal of Nutrition* 138:620–627.

Shaw, H.J. 2006. "Food Deserts: Towards the Development of a Classification." *Geografiska Annaler* 88 B (2):231–247.

Shrive, F.M., H. Stuart, H. Quan, and W.A. Ghali. 2006. "Dealing with Missing Data in a Multi-question Depression Scale: A Comparison of Imputation Methods." BMC Medical Research Methodology 6:57.

Stern, H. N. 1972. "The Optimal Size of Market Areas." *Journal of Economic Theory* 4: 154–173.

Tanguma, J. 2000. "A Review of the Literature on Missing Data." Paper presented at Mid-South Education Research Association annual meeting, Bowling Green KY, 15–17 November.

Tirole, J. 1988. The Theory of Industrial Organization. Cambridge MA: The MIT Press.

Tobler, W. 1979. "Cellular Geographiy." In S. Gale and G. Olson, eds. *Philosophy in Geography,* Dordrecht The Netherlands: Reidel, pp. 379–386.

Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R.B. Altman. 2001. "Missing Value Estimation Methods for DNA Microarrays." Bioinformatics 17 (6): 520–525.

U.S. Census Bureau. Accessed May 21, 2009. >http://www.census.gov<

U.S. Department of Health and Human Services and U.S. Department of Agriculture. Dietary Guidelines for Americans 2005. 2005 [cited 2005 January 14]; Accessed May 21, 2009. >http://www.health.gov/dietaryguidelines/dga2005/document/pdf/DGA2005.pdf<

Viton, P.A. 2010. "Notes on Spatial Econometric Models." City and Regional Planning 870.03.

Wall, M.M. 2004. "A Close Look at the Spatial Structure Implied by the CAR and SAR Models." *Journal of Statistical Planning and Inference* 121 (2): 311–324.

You, W. and R.M. Nayga. Jr. 2005. "Household Fast Food Expenditures and Children's Television Viewing: Can They Really Significantly Influence Children's Dietary Quality?" *Journal of Agricultural and Resource Economics* 30: 302–314.

Young, C.M., B.C. Batch, and L.P. Svetkey. 2008. "Effect of Socioeconomic Status on Food Availability and Cost of the Dietary Approaches to Stop Hypertension (DASH) Dietary Pattern." *Journal of Clinical Hypertension* (Greenwich) 10:603–11.

Zenk, S.N, A.J. Schulz, B.A. Israel, S.A. James, S. Bao, and M.L. Wilson. 2005. "Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit." *American Journal of Public Health* 95:660–667.

—. 2006. "Fruit and Vegetable Access Differs by Community Racial Composition and Socioeconomic Position in Detroit, Michigan." *Ethnicity and Disease* 16:275–80.

VITA

Name: Justus Lotade-Manje

Address: Department of Agricultural Economics, Room 304F Blocker Building
Texas A&M University
College Station, TX 77843-2124

Email Address: jlotade@yahoo.com or lmanje@gmail.com

Education: D.U.E.S. I&II. Université Abdou Moumouni de Niamey, 1992
M.S., Université Abdou Moumouni de Niamey, Niger, 1997
M.S., Agricultural Economics, Colorado State University, 2002
Ph.D., Agricultural Economics, Texas A&M University, 2011