STUDIES CONCERNING SPENDING BEHAVIOR OF HOUSEHOLDS BY STORE TYPE AND INCOME LEVEL AND EFFECTS OF CHANGES IN IMMIGRATION POLICY ON STATE-LEVEL WIC PARTICIPATION RATES

A Dissertation

by

KEEHYUN LEE

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Oral Capps, Jr
Committee Members,	Gary Williams
	Ariun Ishdorj
	Jenna Anding
Head of Department,	Mark L. Waller

August 2021

Major Subject: Agricultural Economics

ABSTRACT

This research is composed of three essays. The first essay examines how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type. The second related essay goes a step further investigating household food and beverage expenditures not only by store type but also by income level. The outlets considered in this study are grocery stores, convenience stores, discount stores, club stores, drug stores, and dollar stores. The third essay evaluates the impact of policy regime change implemented by the Trump administration on state-level WIC (Special Supplemental Nutrition Program for Women, Infants, and Children) program participation.

We employ a dynamic correlated random effect Tobit model in both essays. The source of data for this analysis is the Nielsen Homescan Panel over the period between 2011 and 2015. A differentiated feature of our empirical analysis relates to transforming the dependent variables which include zero observations using the inverse hyperbolic sine function. In the second essay, we form sub-samples by three categories of income levels (low, mid, and high-income level).

The results suggest that habitual spending behavior is undoubtedly a key factor in affecting nominal food and beverage expenditures across all store formats. This finding also holds across the three respective income sub-samples. Household income is not a statistically significant factor. However, household size, age, urbanization, education, race and ethnicity, region, time-invariant socio-demographic variables, indeed are drivers of household food and beverage expenditures at the six store outlets across the income categories.

For the third essay, we employ a Triple Difference estimator to investigate impact of immigration policy changes implemented by the Trump Administration on state-level WIC participation rates. We use Current Population Survey Annual Social and Economic Supplement data (CPS-ASEC) provided by IPUMS (Integrated Public Use Microdata Series) in this analysis. We find that state-level WIC participation rates of Hispanic non-citizens are significantly lower after the immigration policy change implemented by the Trump administration. But this finding only holds for Hispanic non-citizens, not non-Hispanic non-citizens.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Oral Capps, Jr. and my committee members, Dr. Gary W. Williams, Dr. Ariun Ishdorj, Dr. Jenna Anding, and Dr. Senarath Dharmasena, for their guidance and support throughout the course of this research. I wish to express my appreciation to Loren Burns for editing dissertation.

Finally, and most importantly, huge thanks to my wife, Bora Hwang and two sons, Uijin and Hajin Lee for their patience, encouragement, and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Oral Capps, Jr., Gary Williams, Ariun Ishdorj, and Senarath Dharmasena of the Department of Agricultural Economics and Professor Jenna Anding of the Department of Nutrition.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

This study was supported by a graduate research assistantship from AFCERC (Agribusiness, Food, and Consumer Economics Research Center) and a graduate teaching assistantship from the Department of Agricultural Economics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

This research is composed of three essays. The first essay examines how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type. The second related essay goes a step further investigating household food and beverage expenditures not only by store type but also by income level. The outlets considered in this study are grocery stores, convenience stores, discount stores, club stores, drug stores, and dollar stores. The third essay evaluates the impact of policy regime change implemented by the Trump administration on state-level WIC (Special Supplemental Nutrition Program for Women, Infants, and Children) program participation.

A number of choices is evident beyond traditional supermarkets or grocery stores owing to the increasingly diverse U.S. retail food landscape. Despite the plethora of previous studies that largely focus on factors affecting store choice, one area of research that has received relatively little attention is how the magnitude of household food and beverage expenditures is impacted by the type of store outlets. In this light, the purpose of the first and the second essay is to examine how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type. The list of socio-demographic factors includes: (1) household income; (2) household size; (3) age; (4) urbanization; (5) education; (6) race and ethnicity; and (7) region. Characteristics of the retail environment relate to the number of club stores, the number of convenience stores, the number of grocery stores and supercenters and the number of drug stores within the zip code area of the of the household. Whether traditional or non-traditional, store outlets differ in prices, product assortment, advertising strategies, and location (Volpe, Kuhns, and Jaenicke, 2017). The outlets considered in this study are grocery, convenience, discount, club, drug, and dollar store types.

As mentioned previously, prior works mainly highlighted store choice. To differentiate our study from the extant literature, we explore the factors which directly affect household food expenditure by store outlet. Indeed, Volpe, Jaenicke, and Chenarides (2018) estimated the impacts of expenditure share by store format, but in our study, we quantify the magnitude of the impact of household socio-demographics, the retail food environment, and spending habits on food and beverage expenditures by diverse store types. Hence, by analyzing factors that impact household food expenditure across the aforementioned six store types, this study contributes to the economic literature. Another contribution is that our study also considers habitual persistence or spending habits, a dynamic property of household expenditure on food and beverages. However, in the previously mentioned studies, habitual behavior was not included in the set of explanatory variables.

To further differentiate our study from previous studies, we employ a dynamic correlated random effect Tobit model to incorporate habitual purchasing behavior. The source of data for this analysis is the Nielsen Homescan Panel over the period between 2011 and 2015. Specifically, we use a balanced panel of 28,109 households who participated in the survey for all five years from 2011 to 2015. The total number of observations available for analysis is 140,545. The panel structure allows us to incorporate dynamic modeling by including lagged dependent variables as explanatory variables to account for spending habits.

Another advantage of the use of this model is that we are in a position to handle corner solution problems. The dependent variables reflect household purchasing history according to store type and indeed have zero values; hence the dependent variables are left censored. A differentiated feature of our empirical analysis relates to transforming the dependent variables which include zero observations using the inverse hyperbolic sine (arcsinh) function (Bellemare and Wichman 2020). A notable problem with taking the logarithm of any variable is that it does not allow retaining zero-valued observations because the ln (0) is undefined. As pointed out by Bellemare and Wichman (2019), "applied econometricians are typically loath to drop those observations for which the logarithm is undefined." Consequently, researchers often have resorted to ad hoc means of accounting for this situation when taking the natural logarithm of a variable, such as adding 1 to the variable prior to its transformation (MaCurdy and Pencavel, 1986). In recent years, the inverse hyperbolic sine (or arcsinh) transformation has grown in popularity among applied econometricians due to the fact that it is similar to the behavior of the logarithm function, it allows retaining zero-valued observations without any arbitrariness, and it often results in normal distributions (Burbidge et al. 1988; Yen and Jones 1997; MacKinnon and Magee 1990; Pence 2006; Van den Heuvel et al. 2011; Bellemere, Barrett, and Just 2013; Brown et al. 2015; Bellemere and Wichman 2020).

The third essay deals with the impact of changes in immigration policy implemented by the Trump administration on state-level WIC participation rates. Within five days of taking office, President Trump issued a series of executive orders that promised major changes to the U.S. immigration system. These executive orders demonstrated the Trump administration's focus to make changes in border security and interior enforcement. Concerning border security, the construction of barriers along the southern border and zero-tolerance to all individuals crossing the border illegally were the predominant changes taken. In another executive order, a new interior enforcement regime was mentioned, expanding the classes of non-citizens who are priorities for removal and directing agencies to execute U.S. immigration laws against "all removable aliens." With this regime change, the Trump administration abandoned the prosecutorial discretion guideline under the Obama administration, wherein non-citizens prioritized for removal were only those who had criminal convictions, who recently crossed the border illegally, or who had been ordered removed. With changes in immigration enforcement, the Trump administration made policies that were disadvantageous to non-citizens who hold a legal visa or permanent residence status and to unauthorized immigrants. For example, aliens who applied for adjustment of status or extension of stay who receive public benefits, such as Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance to Needy Families (TANF), Medicaid, can be denied their application by United States Citizenship and Immigration Service (USCIS) due to Inadmissibility on Pubic Charge Grounds final rule. Also, those aliens are inadmissible to the United States and ineligible to become a lawful permanent resident (Green Card).

Scholars have studied how immigration policy change affects the fear of deportation of non-citizens. Hispanic families have deportation fear due to their immigration status, affecting their food security status, school enrollment, and access to social benefits (Berk and Schur, 2001; Jefferies, 2014; Sullivan and Enriquez, 2015; Becerra, 2016). Unlike other social programs such as TANF, SNAP, and Medicaid, the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and the National School Lunch Program (NSLP) have no eligibility restrictions based on applicants' immigration status or legal status. Immigrants face fewer barriers to WIC and NSLP programs relative to other social programs (Vericker et al., 2010). Vargas and Pirog (2016) reported decreases in the participation rate in WIC attributed to increases of deportation fear.

Because of recent changes in immigration policy, fear of deportation or losing legal immigration status of non-citizens has been growing (Hing, 2018; Torres et al., 2018; Tummala-Narra, 2019; Alif et al., 2019, Fleming, 2019). On the basis of the UCLA Luskin Los Angeles County Quality of Life Index, more than one-third of Los Angeles County residents were concerned about deportation of their immigrant's friends and family members, and almost half of the county residents believed that a new federal health law proposed under the Trump administration may make them hard to access health care programs. The fear of deportation clearly affects decision-making of non-citizens as to whether to participate in social benefits and food assistance programs during the Trump administration (Bleich and Fleischhacker, 2019; Callaghan et al., 2019, Laird et al., 2019). Non-citizens avoid revealing their status information to the government authority because revealing this information may increase their risk to be deported.

This effect may be larger in the Hispanic community because almost half of immigrants are from Mexico, the Caribbean, and South America, and more than 30 percent of the population of undocumented immigrants is of Hispanic ethnicity. Watson (2014) and Alsan and Yang (2019) reported decreases in Hispanic participation of social benefit programs after implementing specific immigration policies. Specifically, Alsan and Yang (2019) detected direct and indirect effects of immigration policy changes of Hispanics. The direct effect is the effect from immigration policy change within the non-citizen Hispanics population. But immigration policy changes also may affect citizens Hispanic households (indirect effect) because of concern about their non-citizen Hispanic neighbors. Callaghan et al. (2019) also highlight that participation in health care programs in Texas by Hispanics has decreased due to immigration policies enacted by the Trump administration. Another study deals with decreases in SNAP participation after implementation of immigration policy changes (Laird et al., 2019).

The third essay deals with investigating how fear of deportation from immigration policy changes affects participation in food assistance programs of non-citizens. We raise two research questions. First, does immigration policy change affect noncitizens' public benefit participation rate.? Despite many articles from the popular press which addressed negative impacts of immigration policy regime change during the Trump administration on non-citizen households' public benefit participation, these claims have not yet been substantiated. In fact, the decreasing pattern in public benefits participation rate may be caused by other factors, such as changes in income, employment status, or immigration policy. Therefore, a systematic analysis done via regression analysis concerning public benefit participation incorporating relevant factors is needed to identify and assess the impact of immigration policy change. Second, how does immigration policy change affect Hispanics' public benefits participation? The impact of immigration policy change may vary by race and ethnicity of non-citizens. Hispanics occupy a large portion of the non-citizen population in United States. As well, more than 30 percent of undocumented immigrants are Hispanics.

To identify the fear effect associated with public benefit participation, we focus exclusively on changes in WIC program participation rates for a couple of reasons. First, regardless of the specific immigration status of non-citizen, all non-citizens who meet income and categorical requirements are eligible to participate in the WIC program. But, to be approved for other social benefits (SNAP, Medicaid, TANF, and SSI), non-citizen applicants have to be 'qualified-aliens'. Second, any change in the WIC participation rate after immigration policy changes can be considered as a fear effect. Non-citizens who have benefited from participation in the WIC program are not targeted by any immigration policy change after the Trump Administration. The only policy revised by the Trump administration that related to use of social benefits is the Public

Charge rule implemented by USCIS (United States Citizenship and Immigration Services). But the WIC program is not considered concerning the revised Public Charge rule. So, although actual policy changes in the Trump administration are not related to the WIC program, decreases in the WIC participation rate by non-citizens after immigration policy changes may reflect the effect of fear of deportation from non-citizens.

To address the previously mentioned research questions, we estimate the change in immigration policy pre- versus post- Trump administration in state-level WIC participation rates by citizenship and ethnicity. We use the Triple Difference (Difference-in-Difference-in-Difference) methodology to compare the program participation rate for non-citizen Hispanic households to the participation rate for non-citizen non-Hispanic and citizen non-Hispanic households before versus after the Trump administration. We use data from the CPS-ASEC (Current Population Survey- Annual Social and Economic Supplement), publicly accessible at the IPUMS (Integrated Public Use Microdata Series) website, to estimate differences of WIC participation rates of non-citizens between the second Obama administration and the Trump administration (2013-2018). The CPS-ASEC data provide repeated cross-sections surveyed every March by different panelists in each year from 2013 to 2018. The CPS-ASEC data also provide socio-demographic information, including income and citizenship, region up to the county level, social benefit participation in various programs (e.g., SNAP, Medicaid, TANF, SSI, and WIC), and employment status of survey participants.

We hypothesize that if eligible non-citizens express fear of deportation from immigration policy changes, those individuals forfeit participating in food assistance programs. Moreover, we investigate whether fear of deportation affects specific ethnic groups. As previously mentioned, Hispanics may be more prone to fear of deportation than other ethnic groups. Furthermore, we identify how fear of deportation affects non-citizens by different immigration status. Reactions of non-citizens to immigration policy changes may vary by their legal status (legal and illegal immigrants) because government authorities such as ICE (U.S. Immigration and Customs Enforcement) have the authority to remove undocumented immigrants. We hypothesize that non-citizens who possess immigration status that does not guarantee stable residence in the United States may have deeper fears of deportation from recent immigration policy changes.

This study extends the existing literature by estimating causal effects of immigration policy regime change on WIC participation of non-citizens. Because immigration policy has been changed under the Trump administration, identifying recent trends in the WIC participation rate of non-citizens is important. Second, our research design investigating the fear of deportation on WIC participation by immigration status and ethnicity provides a clearer understanding of WIC participation by non-citizens.

CHAPTER II

HABITUAL BEHAVIOR OF HOUSEHOLD FOOD EXPENDITURE BY STORE TYPE AND INCOME LEVEL[1]

II.1 Introduction

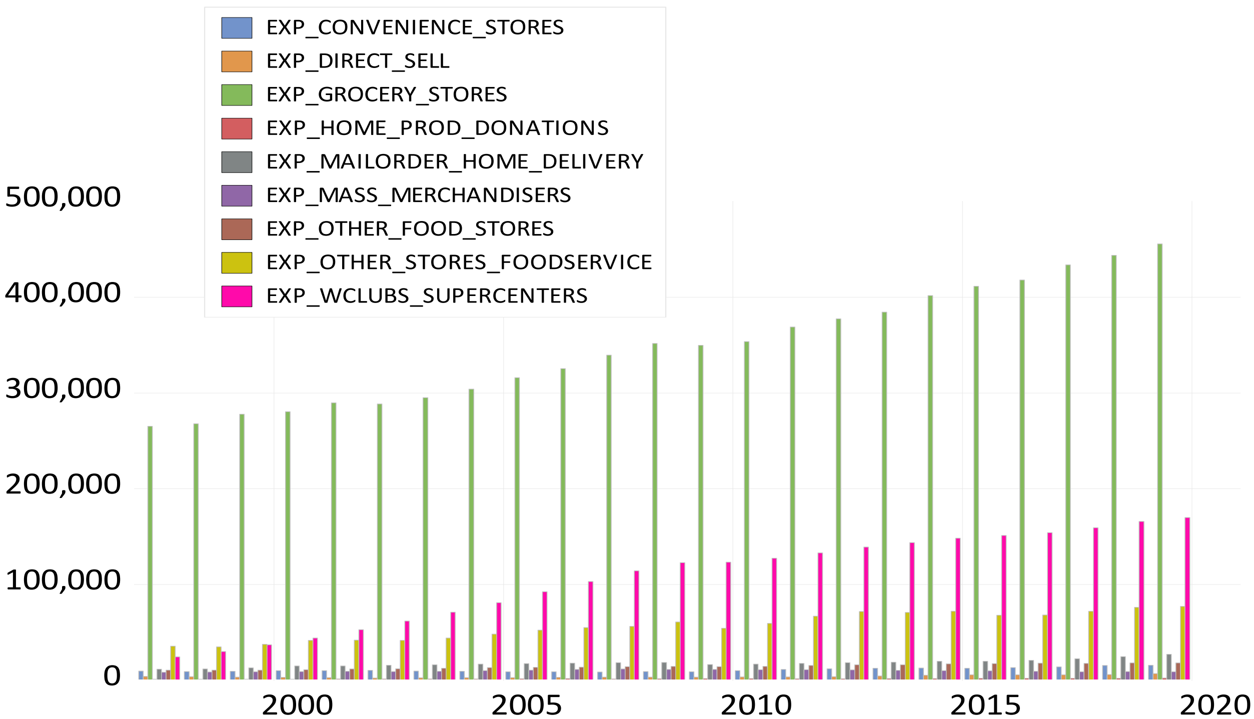
Without question, the food retail environment has changed over the past few decades (Capps and Griffin, 1998; Goldman and Hino, 2005; and Volpe, Kuhns, and Jaenicke, 2017). Over the past 25 years, a number of nontraditional store formats—including supercenters (such as Wal-Mart), dollar stores, and club stores—have gained market share and prominence in the retail food landscape. As exhibited in Figure II-1, the Economic Research Service (ERS) breaks down nominal food expenditures into nine categories: (1) convenience stores; (2) grocery stores; (3) mail order/home delivery; (4) mass merchandisers; (5) warehouse clubs/supercenters; (6) direct sales; (7) other food stores; (8) other stores foodservice; and (9) donations. In particular, over the period 1997 to 2019, nominal expenditures from convenience store were $10.93 billion on average; currently $15.75 billion; from grocery stores $$347.83 billion on average; currently $455.73 billion; from mail order/home delivery $17.80 billion on average; currently $27.00 billion; from

---

[1] Researcher's own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC, and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

mass merchandisers $9.92 billion on average; currently $8.78 billion; from warehouse clubs/supercenters $106.53 billion on average; currently $169.90 billion.

Figure II-1. Breakdown of Nominal Food at Home Expenditures, 1997 to 2019 (Millions of Dollars)



Source: Economic Research Service, USDA.

As shown in Figure II-2, shares of nominal food at home expenditures over the period 1997 to 2019 were as follows: (1) convenience stores, 1.96% on average, ranging from 1.51% to 2.58%; currently 2.02%; (2) grocery stores, 61.91% on average, ranging from 57.97% to 71.63%; currently 58.31%; (3) mail order/home delivery, 3.15% on average, ranging from 2.80% to 3.61%; currently 3.45%; (4) mass merchandisers, 1.82% on average, ranging 1.12% to 2.27%; currently 1.12%; and (5) warehouse/supercenters, 17.70% on average, ranging from 6.61% to 21.78%; currently 21.74%. Accounting for about 80% of at-home food expenditures, the major outlets unequivocally are grocery stores and warehouse clubs/supercenters. That said, other longstanding outlets such as convenience stores, discount stores, and dollar stores have expanded their food offerings to better attract grocery shoppers (Volpe, Kuhns, and Jaenicke, 2017).

Figure II-2. Share of Nominal Food at Home Expenditures, 1997 to 2019, Percent

Source: Economic Research Service, USDA

Previous studies from the fields of economics and marketing have mainly centered attention on the determinants of store choice. Evidence from this rich literature suggests in large part that the choice of food stores is based on a variety of factors including prices, product variety, quality of meat and produce, distance from home, courteous services and degree of competition (Arnold, Oum, and Tigert, 1983; Smith, 2004; Smith, 2006; Hausman and Leibtag, 2007; Briesch, Chintagunta, and Fox, 2009; Richards, Hamilton and Yonezawa, 2016; Marshall and Pires, 2017; and Chenarides and Jaenicke, 2017). Store choice also has been shown to be influenced by household demographics and past purchase history (Staus, 2009) as well as by characteristics of the entire local food market (Feather, 2003; Kyureghian and Nayga, 2013; and Kyureghian, Nayga, and Bhattacharya, 2013), the degree of competition among food stores (Hausman and Leibtag, 2007), and prices offered by various outlet types (Volpe and Lavoie, 2008; Broda, Leibtag, and Weinstein, 2009; Basker and Noel, 2009; and Leibtag, Barker, and Dutko, 2010).. Additionally, previous studies have investigated the role that food access plays in food insecurity, malnutrition, and fruit and vegetable consumption, among other concerns (Rose and Richards, 2004; Bustillos et al., 2008; and Powell and Bao, 2009).

Taylor and Villas-Boas (2016) investigated choices of store outlets as a function of household attributes using a multinomial mixed logit model based on data acquired from the National Household Food Acquisition and Purchase Survey (FoodAPS). Household attributes included participation in the Supplemental Nutrition and Assistance Program (SNAP), household income at various levels of the Federal Poverty Line (FPL), and various measures of the food environment and food access—population density, the share of households living in rural and urban census tracts, the share of households living in a census block group identified as a food desert, and share of households without car access. The store outlets considered were supermarkets, superstores, grocery stores, convenience stores, and farmers' markets.

Moreover, based on data from a panel of 3,376 households collected from 11 randomly selected mid-sized counties in the United States, Fan (2017) analyzed the effect of improving food accessibility by way of subsidizing purchases of fruits and vegetables across food deserts and non-food deserts. The household panel was compiled from 174 food stores collected using scanning devices from Information Resources, Inc. (IRI InfoScan) over a period of 16 quarters from 2009 to 2012 in the 11 sample counties. The IRI InfoScan data provided weekly prices and quantities of various fruits and vegetables by food stores. Store characteristics came from Nielsen TDLinx store directory data, and census-tract level socio-demographic information were obtained from the 2008-2012 American Community Survey (ACS). Census-tract level food deserts indicators were compiled from the 2010 USDA Food Access Research Atlas (FARA, USDA, 2013).

The choice of store outlet, specifically convenience stores, club stores, dollar stores, drug stores, grocery stores, and mass merchandisers, in each census tract in a county was estimated using a random-coefficient discrete choice model, known as the BLP model (Berry, Levinsohn, and Pakes, 1995). This discrete choice model for food stores incorporating household heterogeneity was estimated to quantify the welfare impact of expanding access to fruits and vegetables in food deserts and to compare this welfare effect to the welfare effect associated with a subsidy to fruits and vegetable prices in food deserts. The principal conclusion was that expanding the availability of fruits and vegetables in the nearest stores of food deserts without changing prices did not affect appreciably store choice or enhance the welfare of the household panel. In contrast, price subsidy programs associated with fruits and vegetables in food deserts improved the welfare of household panelists.

Volpe, Kuhns, and Jaenicke (2017) examined the effect of store format and income on the *healthfulness* of food purchased based on a large nationwide sample of households as recorded by the Information Resources, Inc. (IRI) over the period between 2008 and 2012. The healthfulness measures used were based on the Low-Cost, Moderate-Cost, and Liberal Food Plans (2007) developed by the Center for Nutrition Policy and Promotion, USDA as well as the Healthy Eating Index developed by the USDA in 2005 (Carlson, Lino, and Fungwe, 2007). Correlations between store formats and the respective healthfulness measures as well as correlations between store formats and expenditure shares by food category were presented. The store formats in this study were supermarkets, drug stores, mass merchandisers, supercenters, convenience stores, dollar stores, and club stores. Despite the wealth of descriptive information provided, Volpe, Kuhns, and Jaenicke (2017) did not provide a formal econometric analysis.

Finally, Volpe, Jaenicke, and Chenarides (2018) investigated the relationship between store formats and the healthfulness of at-home food purchases. The store formats used in this study were supermarkets, drug stores, mass merchandisers, supercenters, club stores, convenience stores, and other stores. To investigate the healthfulness of household food purchases, based on the methodology developed by Volpe and Okrent (2012) a healthfulness score was assigned, hereafter called the USDA*Score*, to the shopping baskets of each household by quarter. The source of data for this analysis was the Nielsen Homescan Panel over the period between 2004 and 2010. The USDA*Score* is based on the differences between category-specific observed expenditure shares and USDA recommended expenditure shares. The principal goal was to investigate how store format decisions and other factors affect the household-specific USDA*Score*.

Because store-format choice and food-purchase healthfulness were hypothesized to be interrelated decisions, a simultaneous-equation system was developed consisting of eight reduced-form equations. Seven of the respective equations expressed store-format expenditure share as a function of prices measured by the publicly available data from the USDA Quarterly Food-at-Home Price Database, the food retail environment measured by counts of the number of supermarkets, convenience stores, and supercenter stores) and household demographics (namely household income, household size, race, employment status, education, presence of male and/or female household heads and participation in the Women's Infants and Children (WIC) program. The remaining equation expressed USDA *Score* as a function of store format shares, prices, market structure measured by the Herfindahl-Hirschman Index of food retailers, and the aforementioned household demographics. Empirical results pertaining to impacts on USDA Score were obtained for all households as well as by three household income levels, less than the 25th percentile of the sample, between the 25th percentile and the 75th percentile of the sample, and greater than the 75th percentile). The principal conclusion drawn from this analysis was that healthier food choices were associated with higher food expenditure shares at supermarkets and supercenters and lower food expenditure shares at drug stores and convenience stores. In addition, increased retail food industry concentration had a negative effect on shopping healthfulness.

II.2 Objective

Consumers/households currently face a number of choices beyond the traditional supermarket owing to the increasingly diverse U.S. retail food landscape. Despite the plethora of previous studies that largely focus on factors affecting store choice, one area of research that has received relatively little attention is how the magnitude of ***household food expenditures*** is impacted by store formats and store characteristics. In this light, the sole purpose of this study is to examine how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type. Whether traditional or nontraditional, store outlets differ in prices, product assortment, advertising strategies, and location (Volpe, Kuhns, and Jaenicke, 2017). The outlets considered in this study are grocery, convenience, discount, club, drug, and dollar store types. The source of data for this analysis is the Nielsen Homescan Panel over the period between 2011 and 2015. Specifically, we use a balanced panel of 28,109 households who participated in the survey for all five years from 2011 to 2015. The total number of observations available for analysis is 140,545. Through relationships with NielsenIQ and Nielsen, the Kilts Center for Marketing at the University of Chicago Booth School of Business provides this data set to academic researchers for a subscription fee (https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen).

As mentioned previously, prior works mainly highlighted store choice. To differentiate our study from the extant literature, we explore the factors which directly affect household food expenditure by store outlet. Indeed, Volpe, Jaenicke, and Chenarides (2018) estimated the impacts of *expenditure share* by store format, but in our study, we quantify the magnitude of the impact of socio-demographics, the retail food environment, and spending habits on household food and beverage expenditures by diverse store types. Hence, by analyzing factors that impact household food expenditure across the aforementioned six-store types, this study contributes to the economic literature. Another contribution is that our study also considers habitual persistence or spending habits, a dynamic property of household food expenditure. However, in the previously mentioned studies, habitual behavior was not included in the set of explanatory variables.

To further differentiate our study from previous studies, we employ a dynamic correlated random effect Tobit model to incorporate habitual purchasing behavior. As mentioned previously, we construct a panel data set with households as cross-sections over five annual periods, 2011 to 2015. The panel structure allows us to incorporate dynamic modeling by including lagged dependent variables as explanatory variables to account for spending habits. Another advantage of the use of this model is that we are in a position to handle corner solution problems. The dependent variables, which reflect household purchasing history according to store type, have zero values and hence are left censored.

II.3 Organization

This work is organized as follows. Initially, we provide definitions of the respective store outlets. Subsequently, we provide the theoretical framework and the empirical model for this study. Then we describe the Nielsen Homescan data, the construction of the balanced panel of households, and present descriptive statistics of model variables. Issues associated with the estimation of the dynamic correlated random effect Tobit model are discussed next. Following this discussion, the empirical results are presented. Finally, concluding remarks are made along with a discussion of study limitations and possibilities for further research.

II.4 Definitions of Store Types

While there are no universally accepted definitions and classifications of food retail store formats, throughout this study we use the store format names provided by Nielsen, the vendor responsible for the collection of the Homescan data. A traditional supermarket is a food retailer with greater than 9,000 square feet of selling space and at least $2 million in annual sales. Drug stores feature prescription-based pharmacies but generate at least 20 percent of their total sales from other categories, including general merchandise and food. Discount stores are mass merchandisers and typically large department stores (e.g. Target) that sell primarily general merchandise and nonperishables but also carry limited assortments of grocery products. Supercenters also have been known as hypermarkets and superstores are the largest formats, in terms of both square footage and product volume. Supercenters are hybrid stores that combine mass merchandisers with full supermarkets. These stores have a reputation among consumers for stressing low prices and convenience over consumer service (Carpenter and Moore, 2006). Convenience stores are the smallest of the major retail formats in terms of size and product offerings and feature a limited selection of staple foods as well as ready-to-eat, prepared foods (e.g., hot dogs). Additionally, convenience stores sell general merchandise and, in many locations, alcohol, and tobacco. Dollar stores range in size and product variety, placing emphasis on low prices and offering little in the way of customer service. As the name suggests, many products in these stores cost one dollar. Club stores, also referred to as warehouse or volume stores, are large-format outlets that specialize in selling food and selected general merchandise. The grocery line features foods and beverages in bulk for relatively low prices. A feature of this format unique in food retailing is that memberships must be paid in order to shop there.

II.5 Theoretical Framework

On the basis of household production theory, the expenditure function for any commodity is the product of derived demand for factor inputs and the corresponding price vector of factor inputs (Yen, 1993; Bryne, Capps, and Saha, 1996; and Nayga, 1998). Let the commodity in question be all food and beverages purchased by household h at store outlet k. Then as given by equation (1), household expenditure at store k may be written as

$$EX_{hk} = P_{hk}X_{hk} = g\left(P_{hk,}Y_h, W_h, D_h, E_h\right) \tag{1}$$

where X_i is the derived demand of factor inputs for household h at store outlet k, P_{hk} is the price vector of factor inputs paid by household h at store outlet k, W_h is a measure of the opportunity cost of time of household h, Y_i represents the income level of household h, D_h represents the set of socio-demographic characteristics of household h, and E_h represents the retail environment faced by household h.

Household heterogeneity typically is accounted for incorporating socio-demographic variables in the theoretical model. Hill and Lynchehaun (2002) identified various cultural and socio-economic factors influencing consumer preferences including age, ethnicity, income, education, gender, presence of children, marital status, region, and race. In particular, education reflects knowledge about health and nutrition (McCracken and Brandt, 1987; Nayga and Capps, 1992; Byrne, Capps, and Saha, 1996; Nayga, 1998; and Volpe, Kuhns, and Jaenicke 2017). Similar to Volpe, Janeicke, and Chenarides (2018), we include household income, household size, age, urbanization, race and ethnicity, region, and education in the set of socio-demographic variables in this study.

Additionally, in our theoretical model, we consider the potential importance of the retail environment in the household expenditure function. The retail environment represents the number of stores in the area in which the household lives; accessibility to store outlets may affect household production and consequently, household purchases of food and beverages. In this study, to address the impact of the retail environment, similar to Volpe, Jaenicke, and Chenarides (2018), we count the number of supermarkets and other grocery stores, convenience stores, drug stores, and warehouse club stores based on zip codes.

Past studies related to the choice of store outlet did not account for habitual purchasing behavior (Taylor and Villas-Boas, 2016; Fan, 2017; and Volpe, Jaenicke, and Chenarides, 2018). Habits refer to repetitive behavior in purchasing and consumption behavior (Ji and Wood, 2007). The habitual behavior of consumer purchasing patterns has been studied widely in the field of psychology (Bettman and Zins, 1977; Ehrenberg, 1988; and Ehrenberg 1991). This repeated purchasing behavior has been investigated in a wide range of products and services, including but not limited to potato chips, bread, tissue, laundry detergent, catsup, yogurt, sugar-sweetened beverages, and cigarettes (Deighton, Henderson, and Neslin, 1994; Motes and Woodside, 2001; Taylor, 2001; Khare and Inman, 2006; Zhen et al., 2011; Adamowicz and Swait, 2012; and Zhen et al. 2013).

To account for habitual purchasing behavior, we introduce a one-period lagged dependent variable in the model. (Mutlu and Garcia, 2006; and Rieger, Kuhlgatz, and Anders, 2016). As such, we augment the expenditure function given in equation (1) for household h for store outlet k in time period t as:

$$EX_{hkt} = h(EX_{hk,t-1}, P_{ht}, Y_{ht}, W_{ht}, D_{ht}, E_{ht}) \qquad (2)$$

$where\ EX_{hk,t-1}$ is the one-period lagged expenditure variable for household h for each store type k in time period t.

II.6 Empirical Model

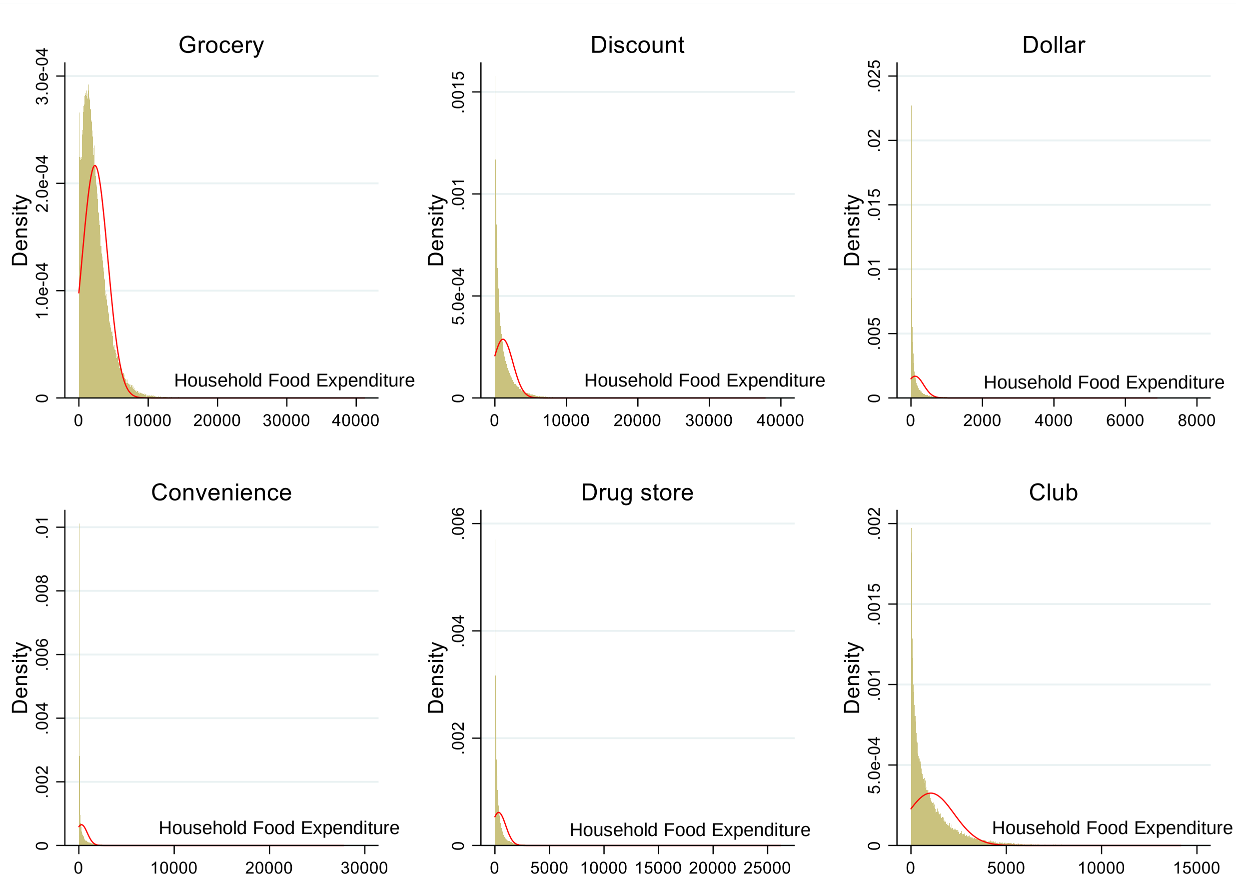
Given the focus of our research in analyzing the impacts of habitual spending behavior, the retail environment, and household heterogeneity by store outlets, we employ a dynamic correlated random effect Tobit model. This model specification allows us to deal with dynamics, panel data, and data censoring issues simultaneously accounting for household demographic variables and retail environment variables as explanatory variables. The model also accounts for potentially household-specific unobserved heterogeneity. As well, conventional fixed effect nonlinear models such as probit, logit, and Tobit models can produce biased estimates of structural parameters (Greene, 2004). The use of the dynamic correlated random effect Tobit model circumvents this deficiency and produces consistent estimates of structural parameters.

Owing to the number and heterogeneity of purchases of specific food items and beverages as well as the censored observations associated with household food and beverage expenditures by store outlets, we omit prices from the model. In the Nielsen data prices are derived as the ratio of expenditures to quantities purchased. By omitting prices from the model, we avoid making imputations of missing prices and we avoid the potential endogeneity of prices with household expenditures. Simply, we assume that the impact of the price is implicitly captured by the type of store outlet.

We transform the dependent variables which include zero-valued observations using the inverse hyperbolic sine (arcsinh) mechanism (Bellemare and Wichman, 2020). A notable problem with taking the logarithm of any variable is that it does not allow retaining zero-valued observations because ln (0) is undefined. As pointed out by Bellemare and Wichman (2020), "applied econometricians are typically loath to drop those observations for which the logarithm is undefined." Consequently, researchers often have resorted to ad hoc means of accounting for this situation when taking the natural logarithm of a variable, such as adding 1 to the variable prior to its transformation (MaCurdy and Pencavel, 1986). In recent years, the inverse hyperbolic sine (or arcsinh) transformation has grown in popularity among applied econometricians due to the fact that it is similar to the behavior of the logarithm function, it allows retaining zero-valued observations without any arbitrariness, and it often results in normal distributions (Burbidge et al., 1988; Yen and Jones, 1997; MacKinnon and Magee, 1990; Pence, 2006; Van den Heuvel et al., 2011; Bellemere, Barrett, and Just, 2013; Brown et al., 2015; and Bellemere and Wichman, 2020).

Figures II-3 and II-4 show the distributions of the original and transformed household food expenditure associated with each store type conditional on expenditures above zero (Horizontal axis indicates expenditure). In Figure II-3, only food expenditures from grocery stores follows a truncated normal distribution. In Figure II-4, after implementing the inverse hyperbolic transformation, all six dependent variables appear to follow normal distributions. In addition, the zero values of the dependent variables still remain as zeros with the inverse hyperbolic transformations. So, by transforming our dependent variables pertaining to household food expenditures, we can deal with corner solution issues in our data using the Tobit model.

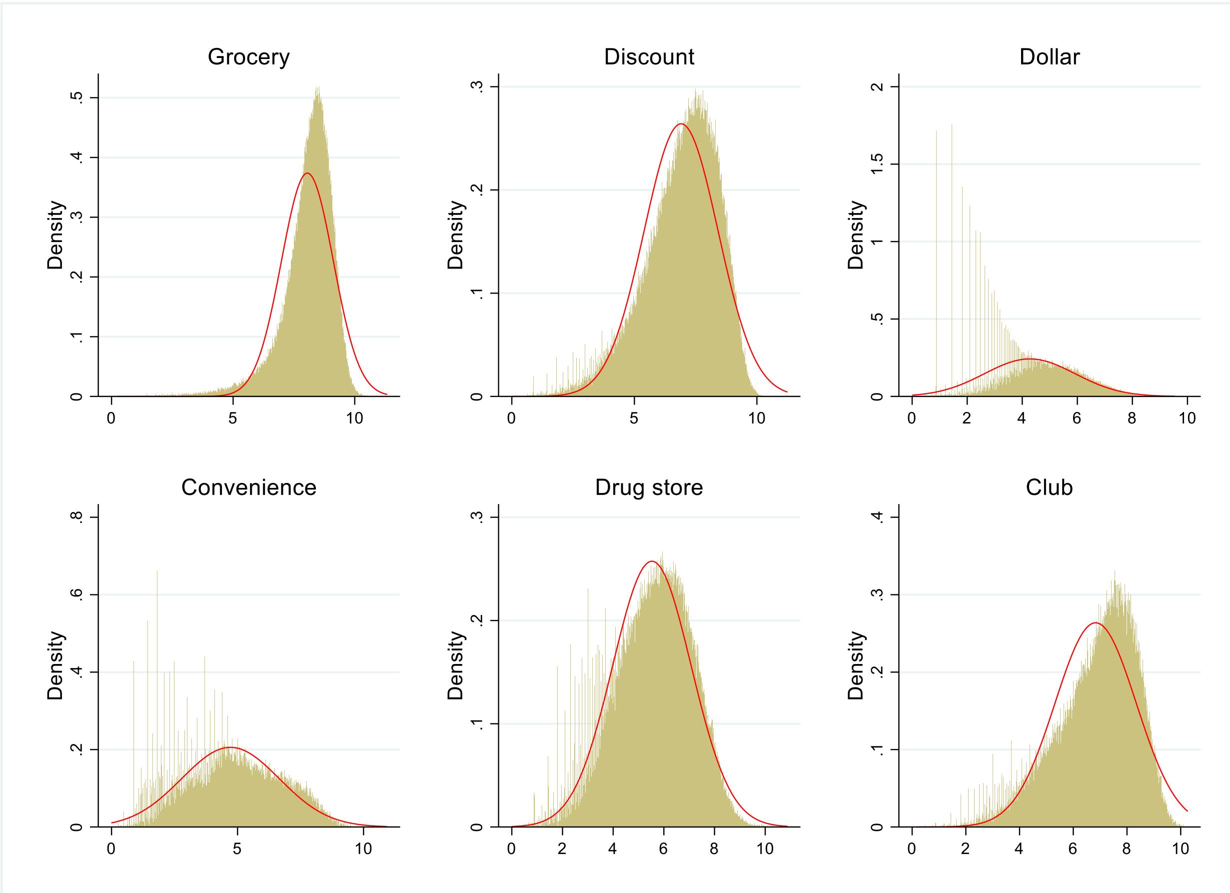
Figure II-3. Distributions of Household Food Expenditure Associated with the Six Store Types



Using equation (3), we express the transformed household food and beverage expenditure variables by store type based on the inverse hyperbolic sine method. We denote $\overline{EX}_{ht}^{k}$ as the dependent variables for household h, for store outlet k, and for time period t in our empirical model.

$$\overline{EX}_{ht}^{k} = arcsinh\left(EX_{ht}^{k}\right) = ln(EX_{ht}^{k} + \sqrt{{EX_{ht}^{k}}^{2} + 1}) \quad (3)$$

Figure II-4. Distributions of Transformed (inverse hyperbolic sine) Household Food Expenditure Associated with the Six Store Types



Initially we start from the definition of our latent variables, that is, the expenditure variables denoted in equation (4). This latent variable property is maintained after transforming the dependent variables with the inverse hyperbolic sine method, equation (5). Zero observations reflect the decision by households to not make food and/or beverage purchases over the course of at least one calendar year in a particular store outlet.

$$EX_{ht}^{k\,*} = EX_{ht}^{k} \qquad if\;\; EX_{ht}^{k} > 0 \qquad (4)$$

$$EX_{ht}^{k\,*} = 0 \qquad if\;\; EX_{ht}^{k} = 0$$

$$\overline{EX}_{ht}^{k\,*} = \overline{EX}_{ht}^{k} \qquad\qquad if\ \ \overline{EX}_{ht}^{k} > 0 \tag{5}$$
$$\overline{EX}_{ht}^{k\,*} = 0 \qquad\qquad if\ \ \overline{EX}_{ht}^{k} = 0$$

Our empirical model for each of the respective six store types k at year t for household h is described in equations (6) through (8). We start from basic random effect model described in equation (6). $\overline{EX}_{ht}^{k\,*}$ is annual food and beverage expenditure of household h at year t for store type k, censored at zero. $\overline{EX}_{h,t-1}^{k\,*}$ is one year lagged dependent variable capture dynamic spending behavior of consumers, and c_h^k is a random effect term associated with the household's unobserved heterogeneity. ε_{ht}^{k} is the idiosyncratic error term.

Producing consistent estimators with lagged dependent variables in the random effect Tobit model, controlling for the initial condition, is critical. Wooldridge (2005) suggested a general and tractable approach to overcome the initial condition problem2. Following Chamberlain (1980), we assume that the unobserved heterogeneity term, c_h^k, has a distribution conditional on time-averaged

2 The initial condition problem occurs when initial value of stochastic process is not observed. For example, consider following equation: $Y_{it} = \alpha + \beta Y_{it-1} + c_i + \varepsilon_{it}$. This equation contains a lagged dependent variable as a covariate. If we recursively rewrite this equation, then we can finally derive Y_{it} with Y_{i0} as a covariate. That said, defining Y_{i0} the initial value of the stochastic process is a difficult task. Wooldridge (2005) proposed a conditional maximum likelihood estimator that approximates the unobserved heterogeneity term, c_i, with the use of the initial observation, $y_{i,0}^{m}$, of the dataset and exogenous variables, d_i^m, to overcome the initial condition problem. Following Chamberlain (1980), we assume that the unobserved heterogeneity term has a distribution conditional on time-averaged continuous explanatory variables and the initial value of latent dependent variable. $y_{i,0}^{m}$ is the initial observation of the value of household expenditure on food and beverages in store outlet m in 2011, the initial calendar year of the data used in this analysis.

continuous explanatory variables3 (d_h^k) and the value of the latent dependent variable in the initial time period (2011) of our sample ($\overline{EX}_{h,0}^{k\ *}$). Then, upon substitution of the unobserved heterogeneity term, equation (7), into the base random effect model, equation (6), we subsequently derive equation (8), the correlated random effect model. Because we assumeu_h^k, the error term in the unobserved heterogeneity function (equation (7)), is random, equation (8) corresponds to a random effect model. The respective distributions of the error terms are given in equation (9).

z_{ht}^k is a vector of explanatory variables including the logarithm of household annual income, household size (number of household members), and number of club, convenience, drug, supercenter and grocery stores within the household's zip code area as well as indicator (dummy) variables related to age, education, race/ethnicity of the household head, urban/rural delineation, and region in which the household is located. In our correlated random effect model specification, we assume that u_h^k follows a standard normal distribution with variance $\sigma_u^{2,k}$. In equation (8), we can treat u_h^k as a random effect. This term corresponds to household unobserved heterogeneity. To estimate this random effect Tobit model, we employ the econometrics software package STATA version 15 using the command xttobit.

$$\overline{EX}_{ht}^{k\ *} = \alpha z_{ht}^k + \rho \overline{EX}_{h,t-1}^{k\ *} + c_h^k + \varepsilon_{ht}^k \tag{6}$$

$$c_h^k = \ \theta \overline{EX}_{h,0}^{k\ *} + \vartheta d_h^k + u_h^k \tag{7}$$

$$\overline{EX}_{ht}^{k\ *} = \alpha z_{ht}^k + \rho \overline{EX}_{h,t-1}^{k\ *} + \theta \overline{EX}_{h,0}^{k\ *} + \vartheta d_h^k + u_h^k + \varepsilon_{ht}^k \tag{8}$$

---

3 d_h^k is a vector of time-averaged continuous explanatory variables. In constructing the correlated random effect model, we only averaged continuous explanatory variables that are time-varying for each household. For example, for household income, we construct time-averaged variable by calculating $\overline{in}_h = \sum_{t=2011}^{2015} in_{ht}$, where in_{ht} is the annual income for household h in year t and t is time period 2011 to 2015.

$$with\ \varepsilon_{ht}^{k} \sim N\left(0, \sigma_{\varepsilon}^{2,k}\right), \qquad u_{h}^{k} \sim N\left(0, \sigma_{u}^{2,k}\right) \tag{9}$$

We address habitual spending behavior by adding a lagged dependent variable in the model. The coefficient of this variable, ρ, should be between 0 and 1. Statistical significance of this coefficient also confirms the existence of habitual spending behavior at certain store types. We jointly test significance of coefficients θ and ϑ using the Wald test to empirically check for household heterogeneity as denoted in equation (7). Likelihood ratio tests also are performed to compare the panel data model and the pooled data model. Another likelihood ratio test is designed to compare the correlated random effect model and the simple random effect model. 4

II.7 Marginal Effects

Marginal effects refer to changes in the dependent variables (expenditures) attributed to unit changes in the continuous explanatory variables. For discrete explanatory variables, marginal effects refer to changes in expenditures relative to base or reference categories. The estimated parameters are not the marginal effects.

Equation (10) shows the conditional expectation of the original (untransformed) dependent variables. We present the details of the derivation in Appendix. $E\left[EX_{ht}^{k} \middle| z_{ht}^{k}, EX_{ht}^{k} > 0\right]$ is conditional expectation when household food expenditure is greater than zero. The function Φ is cumulative distribution function of the normal distribution with mean zero and variance $\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}$. EX_{ht}^{k} refers to the original dependent variables, untransformed annual household

---

4 The pooled Tobit model is given by: $\overline{EX}_{ht}^{k\ *} = \alpha z_{ht}^{k} + \rho \overline{EX}_{h,t-1}^{k\ *} + \theta \overline{EX}_{h,0}^{k\ *} + \vartheta d_{h}^{k} + \upsilon_{ht}^{k}$ which ignores the panel data structure. As discussed previously, the correlated random effect model permits the correlation between explanatory variables and unobserved heterogeneity, but the simple random effect model does not.

expenditures on food and beverages by store format, k. z_{ht}^{k} is a vector of explanatory variables associated with continuous and binary variables as elements, and β_h^k refers to the estimated coefficients of the structural parameters.

$$E\left[EX_{ht}^{k}\middle|z_{ht}^{k}, EX_{ht}^{k} > 0\right] = \frac{1}{2}\exp\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}{2}\right) * \left[\exp\left(z_{ht}^{k}\beta_h^k\right)\left\{\frac{\Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{1 - \Phi\left(\frac{-z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}\right\} - \exp\left(-z_{ht}^{k}\beta_h^k\right)\left\{\frac{1 - \Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} - z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{\Phi\left(\frac{z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}\right\}\right] \tag{10}$$

If the explanatory variables are continuous variables, we can take derivative with respect to these variables to obtain marginal effects. In equation (11), we provide the expression of the marginal effects for continuous variables. ϕ is probability density function with mean zero and variance $\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}$. The notations for equation (11) are similar to the notations in equation (10). In equation (12), we represent the marginal effects for binary explanatory variables. We calculate these marginal effects by taking the difference between conditional expectation when $z_{ht}^{k} = 1$ and $z_{ht}^{k} = 0$. Equation (11) is expressed as

$$\frac{\partial E\left[EX_{ht}^{k} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right]}{\partial X_{ht}^{k}}$$

$$= \beta_h^k \frac{1}{2} \exp\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}{2}\right) \left[\exp\left(X_{ht}^k \beta_h^k\right) \left\{ \frac{\Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)} \right\} \right.$$

$$\left. - \frac{1}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}} \left\{ \frac{\phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right) - \Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right)^2} \right\} \right] \quad (11)$$

$if\ z_{ht}^k\ corresponds\ to\ countinuos\ variables$

$$\frac{\partial E\left[EX_{ht}^{k} \middle| z_{ht}^{k}, EX_{ht}^{k} > 0\right]}{\partial z_{ht}^{k}} = E\left[EX_{ht}^{k} \middle| z_{ht}^{k} = 1, E_{ht}^{k} > 0\right] - E\left[EX_{ht}^{k} \middle| z_{ht}^{k} = 0, EX_{ht}^{k} > 0\right] \quad (12)$$

$if\ z_{ht}^k\ corresponds\ to\ binary\ variables$

We employ a lagged dependent variable and the logarithm of household income in our explanatory variable set. Marginal effects for these variables need to be treated with care. As exhibited in Appendix, the marginal effect associated with the lagged dependent variables is given in equation (13).

$$\frac{\partial E\left[EX_{ht}^{k} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right]}{\partial EX_{ht}^{k}} = \frac{\partial E\left[EX_{ht}^{k} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right]}{\partial \overline{EX}_{h,t-1}^{k}{}^{*}} * \frac{\partial \overline{EX}_{h,t-1}^{k}{}^{*}}{\partial EX_{ht}^{k}} \quad (13)$$

To derive the marginal effect for income, we simply divide equation (11) by income.

Following McDonald and Moffitt (1980), for the Tobit model, the derivative of the unconditional expectation with respect to explanatory variables can be decomposed into two parts:

(1) the conditional marginal effect times the probability of non-zero household food expenditures at the various store outlets and (2) the conditional expectation times the change in probability of non-zero expenditures due to unit changes in the explanatory variables.

$$\frac{\partial E\left[EX_{ht}^{k\,*}\middle|z_{ht}^{k}\right]}{\partial z_{ht}^{k}} = \frac{\partial P\left[EX_{ht}^{k\,*} > 0\middle|z_{ht}^{k}\right]}{\partial z_{ht}^{k}} * E\left[EX_{ht}^{k\,*}\middle|z_{ht}^{k}, EX_{ht}^{k\,*} > 0\right] + P[EX_{ht}^{k\,*} > 0|z_{ht}^{k}] * \frac{\partial E[EX_{ht}^{k\,*}|z_{ht}^{k}, EX_{ht}^{k\,*} > 0]}{\partial z_{ht}^{k}} \quad (14)$$

We adopt this decomposition to explore the effects of explanatory variables on the probability of households to spend at various store formats, $\frac{\partial P(EX_{ht}^{k\,*}>0|z_{ht}^{k})}{\partial z_{ht}^{k}}$ in equation (14) as well as to explore the effects of explanatory variables on the magnitude of spending at particular store formats, $\frac{\partial E(EX_{ht}^{k\,*}|z_{ht}^{k},Y>0)}{\partial z_{ht}^{k}}$ in equation (14). Note that the effects on the magnitude of spending for various store formats is the same as the conditional expectation previously described in equation (10). With the assumption of a Tobit model, the probability is given by the linear combination of estimated coefficients associated with z_{ht} for store outlet k. The change in the probability is given the probability density function times the estimated coefficients divided by the estimate of the variance of the normal distribution.

We calculate the marginal effects associated with equations (11), (12), (13) and (14) with the use of the software package STATA15. Standard errors of the marginal effects are obtained using the delta method, as these are nonlinear combination of coefficients and the data (Bellemare and Wichman, 2020).

II.8 Data

The source of the data for this study is the Nielsen Homescan Panel covering the period between 2011 and 2015, the most recent data available to us at the time of this analysis. Volpe, Jaenicke, and Chenarides (2018) also use the Nielsen Homescan Panel for the period between 2004 and 2010. As such, we extend the time period of coverage concerning expenditures made by U.S. households at various store outlets. Hence, our analysis not only allows us to check on robustness of findings from the literature but also serves as a reference for future studies using more recent data.

Nielsen collects weekly surveys from more than 60,000 panelists every year in the entire United States. The Homescan data contain detailed information about quantities purchased and corresponding expenditures made by household by Universal Product Code (UPC) and by store type, Also, the Nielsen Homescan data incorporate a plethora of socio-demographic variables to account for household characteristics. We center attention on *annual* expenditures made by households for all food and beverage items. Finally, we consider those households that participate in each of the five years over the period between 2011 and 2015. In our balanced panel, 28,109 households participate in the survey for the five-year period from 2011 to 2015. Hence, the total number of observations available in our study is 140,545.

In Table II-1, we report the unconditional and conditional[5] means and standard deviations of household food and beverage expenditure expressed in dollars by store type. Not unexpectedly, household spending for food and beverages is highest in grocery stores, and lowest in dollar stores.

[5] The term conditional corresponds to only those expenditure values above zero. On the other hand, the term unconditional refers to zero values as well as non-zero expenditure values.

Not all households purchase food and beverages at all store outlets even over a calendar year. Therefore, zero values are evident in household expenditures for food and beverages across the respective store types. As such, household expenditures are left censored at zero. The number of zero observations and the degree of censoring of household expenditures on food and beverages are exhibited in Table II-2. The degree of censoring is defined as the number of zero observations times 100 divided by the number of observations.

The degree of censoring is greatest for convenience stores at roughly 70 percent. The censoring degree is lowest in grocery stores at approximately 1 percent. In discount stores, the degree of censoring is on the order of 5 to 8 percent; in drug stores, the degree of censoring is on the order of 18 to 24 percent. In dollar stores, the magnitude of censoring ranges from 23 to 45 percent. Finally, for club stores, the degree of censoring ranges from 31 to 64 percent.

Table II-1. Unconditional and Conditional Means and Standard Deviations of Household Food and Beverage Expenditure by Store Type, Nielsen Panel Data 2011 to 2015[a]

Store Type	Unconditional	Conditional
Club	573 [a]	1,001
	(1,018)	(1,176)
Convenience	81	267
	(345)	(588)
Dollar	74	114
	(189)	(226)
Grocery	2,177	2,197
	(1,750)	(1,745)
Drug	259	332
	(557)	(575)
Discount	1,004	1,072
	(1,302)	(1,318)
Across All Store Outlets	4,162	4,982

[a] All values are expressed in terms of dollars, and standard deviations are in parentheses.

Table II-2. Number of Zero Values and Degree of Censoring for Household Food and Beverage Expenditures by Store Type, Nielsen Panel Data 2011 to 2015

Store Type	Entire sample
Club	60,148[a]
	(42.8%)[b]
Convenience	98,239
	(70.0%)
Dollar	49,552
	(35.3%)
Grocery	1,286
	(0.9%)
Drug	28,067
	(20.0%)
Discount	8,802
	(6.3%)
Total number of observations	140,545

[a] Number of zero observations associated with household food and beverage expenditures.
[b] Degree of censoring expressed as a percent.
(Degree of censoring = number of zero observations*100/total number of observations)

Similar to Kyureghian and Nayga (2013), we use store density data to account for the retail environment. However, we examine store density by zip code, which represents smaller residential areas rather than by county level as was done by Kyureghian and Nayga (2013). These variables were obtained from Business Pattern Data (BPD hereafter) produced by the U.S. Census Bureau. The BPD contains data represent the number of stores categorized by North American Industry Classification System (NAICS). From these data, we obtained counts concerning four types of store outlets (grocery stores and supercenters, NAICS code 445110; warehouse club stores, NACIS code 452910; convenience stores, NAICS code 445120; and drug stores, NAICS code 446110). The data from the Nielsen Homescan Panel are available by zip code; consequently, we are able to augment the Nielsen Homescan Panel with the respective counts of store outlets from BPD. A shortcoming in this augmentation process is that the classifications of store formats from the Nielsen Homescan Panel and Business Pattern Data are different. Nevertheless, we provide a viable proxy for the retail environment based on counts of store outlets from BPD.

In order to identify differences in food and beverage expenditures by store outlets between households who live in urban and rural areas, we form urban and rural indicator (dummy) variables. Our dummy variables correspond to the six-category urban and rural classification scheme developed by the National Center for Health Statistics (NCHS). In our study, we form three dummy variables that represent category 1 through category 6 of the NCHS classification scheme. The dummy variable URBAN corresponds to NCHS urban and rural classification categories 1 and 2, the most densely populated areas, typically metropolitan areas. The dummy variable RURAL corresponds to NCHS urban and rural classification categories 5 and 6, rural areas with the least dense population. The dummy variable not classified as urban or rural corresponds to NCHS urban and rural classification categories 3 and 4. Then, we aligned these indicator variables with our Nielsen Homescan data based on zip code. The use of the NCHS classification scheme affords a richer consideration of the role of urban and rural areas in influencing household food and beverage expenditures by store outlet.

In Table II-3, the means and standard deviations of explanatory variables in the model are presented. The average number of grocery stores by zip code is slightly more than five not only for the entire sample. The average number of drug stores by zip code in this analysis is nearly four across the board. Similarly, on average the number of convenience stores by zip code is between two and three for each of the respective samples. Finally, the average number of club stores by zip code is between zero and one, again across the board. In sum, with respect to the number of store outlets by zip code, the average number of club stores, convenience stores, supercenter and grocery stores, and drug stores does not vary much among the entire.

Table II-3. Means and Standard Deviations of Explanatory Variables in the Tobit Random Effect Model for the Entire Sample

Variable Name	Variable Description	Entire sample
	Continuous Variables	
Household income	Household income	\$53,589 (\$26,601)
Household size	Number of household members	2.17 (1.15)
Cu	NAICS code 452910, # of warehouse club stores by zip code	0.54 (0.80)
Cv	NAICS code 445120, # of convenience stores by zip code	2.36 (2.98)
Sg	NAICS code 445110, # of supercenters and grocery stores by zip code	5.42 (7.19)
Dr	NAICS code 446110, # of drug stores by zip code	3.82 (3.59)
	Degree of Urbanization	
Urban	NCHS urban and rural classification categories 1 and 2	0.54 (0.50)
Not urban* or rural	NCHS urban and rural classification categories 3 and 4)	0.30 (0.46)
Rural	NCHS urban and rural classification categories 5 and 6	0.15 (0.36)
	Age	
Age<40	Age of household head below 40	0.01 (0.12)
40<Age<60*	Age of household head above 40 and below 60	0.18 (0.38)
Age>60	Age of household head over 60	0.82 (0.39)
	Education	
Under high school	Household head education less than high school	0.01 (0.10)
Graduate high school	Household head is a high school graduate	0.18 (0.39)
College experience	Household head had some college education but is not a college graduate	0.28 (0.45)
Graduate* college	Household head is a college graduate	0.53 (0.50)

Table II-3. Continued

Variable Name	Variable Description	Entire sample
	Race and ethnicity	
Nonhisp-* White	Household head is non-Hispanic white	0.82 (0.39)
Nonhisp-Black	Household head is non-Hispanic black	0.09 (0.29)
Nonhisp-Asian	Household head is non-Hispanic Asia.	0.03 (0.16)
Nonhisp-Other	Household head is non-Hispanic other	0.02 (0.14)
Hisp	Household head is Hispanic	0.04 (0.21)
	Region	
Ne	Household located in the New England region (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont)	0.05 (0.21)
Ma	Household located in the Middle Atlantic region, (New Jersey, New York, and Pennsylvania)	0.13 (0.34)
Enc	Household located in East North Central region (Illinois, Indiana, Michigan, Ohio, and Wisconsin)	0.18 (0.39)
Wnc	Household located in the West North Central region (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)	0.09 (0.29)
Sa	Household located in the South Atlantic region (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, and West Virginia)	0.20 (0.40)
Esc	Household located in East South Central region (Alabama, Kentucky, Mississippi, and Tennessee)	0.06 (0.23)
Wsc	Household located in the West South Central region (Arkansas, Louisiana, Oklahoma, and Texas)	0.10 (0.30)
Mt	Household located in the Mountain region (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming)	0.07 (0.26)
Pac*	Household located in the Pacific region (Alaska, California, Hawaii, Oregon, and Washington)	0.12 (0.33)
Number of observations		140,545

Standard deviations are in parentheses.

Superscript * associated with the variable name indicates the base category or reference category.

More than half of the entire sample households live in urban areas, and 15 percent live in rural areas. The base category or reference category with respect to degree of urbanization is the 'not urban or not rural' category. Additionally, in Table II-3, we present descriptive statistics concerning socio-demographic variables, namely household income, household size, age, education level, race/ethnicity of the household head, and region in which the household is located. Household income is reported by ranges in the Nielsen Homescan Panel. Similar to previous studies, we take the midpoint of each household income range as the income level of the household (Kyureghian and Nayga, 2013; Austin et al., 2017; and Senia, Dharmasena, and Capps, 2019). The mean value of household income for the entire sample is $53,589. The average household size for the entire sample is 2.17 members.

We employ three classifications of the age of the household head, less than 40, between 40 and 60, and over 60. For the entire sample, the proportion of households whose heads are less than 40 is roughly 1 percent; the proportion of households whose heads are between 40 and 60 is 18 percent; and the proportion of households whose heads are over 60 is 82 percent. This pattern is reversed in regard to households whose heads are between 40 and 60. Across the respective data samples, the number of households whose heads are less than 40 is around 1 percent. The data concerning age of the household head unequivocally are skewed toward older household heads. The base category is age of the household head between 40 and 60.

We consider four categories concerning the level of education of the household head—less than high school, high school graduate, some college experience, and college graduate. To illustrate, the proportion of households whose heads have a college degree is 53 percent for the entire sample. Further, the proportion of households wherein the highest level of education is a high school degree is 18 percent for the entire sample. Very few household heads in the respective data samples have less than a high-school education. The vast majority of household heads in the respective data samples have at least some college-level educations. The base category of education level corresponds to household heads with a college degree.

We employ five joint classifications of the race and ethnicity of the household head—non-Hispanic white, non-Hispanic black, non-Hispanic Asian, non-Hispanic other, and Hispanic. For the entire sample, the proportion of households whose heads are non-Hispanic white is 82 percent; the proportion of households whose heads are non-Hispanic black is 9 percent; the proportion of households whose heads are non-Hispanic Asian is 3 percent, and the proportion of households whose heads are non-Hispanic other races is 2 percent; and the proportion of Hispanic household heads is 4 percent. The base category of race/ethnicity is non-Hispanic white households.

We rely on nine categories concerning the region in which the household is located: (1) New England ((Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont); (2_ Middle Atlantic (New Jersey, New York, and Pennsylvania); (3) East North Central (Illinois, Indiana, Michigan, Ohio, and Wisconsin); (4) West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota**);** (5) South Atlantic (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, and West Virginia); (6) East South Central (Alabama, Kentucky, Mississippi, and Tennessee); (7) West South Central (Arkansas, Louisiana, Oklahoma, and Texas); (8) Mountain (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming); and (9) Pacific (Alaska, California, Hawaii, Oregon, and Washington). This delineation affords more detail concerning the impact of region on household food and beverage expenditures by store outlet. Across the board, roughly 20 percent of the households reside in the South Atlantic region, 18 percent reside in the East North Central region, 13 percent in the Middle Atlantic region, 12 percent in the Pacific region, 10 percent in the West South-Central region, 9 percent in the West North Central region, 7 percent in the Mountain region, 6 percent in the East South-Central region, and 5 percent in the New England region. The reference category is the Pacific region.

II.9 Potential Endogeneity of the Retail Environment Variables

There is a debate in the literature as to whether or not the retail environment variables are endogenous. This issue is important due to the fact that the endogeneity of explanatory variables leads to inconsistent parameter estimates. The endogeneity issue was addressed in several studies (Dunn et al., 2012; Kyureghian and Nayga, 2013; Ver Pleog et al., 2015; Handbury, Rahkovsky, and Schnell, 2016; and Allcott, Diamond, and Dubé, 2017). On the other hand, Currie et al. (2010) and Taylor and Villas-Boas (2016) argued that the endogeneity of retail environment variables does not lead to bias or inconsistency of parameter estimates.

Although previous works recognized potential endogeneity issues regarding retail environment variables, those studies just assumed the presence of endogeneity and estimated models with instrument variables. In this study, we formally test whether or not the retail environment variables suffer from the endogeneity problem. In order to account for the retail environment, we use variables that represent the number of stores, a metric of store density, by store type in the zip code area in which the household is located. As mentioned previously, we only incorporate four categories of store types, namely supercenters and grocery stores, club stores, drug stores, and convenience stores from the BPD data. Then, we test exogeneity of those four variables in each of the equations pertaining to the six store types. To carry out the Hausman test of endogeneity, we initially estimate each retail environment variable as a function of the remaining explanatory variables. Consistent with Hausman (1978) we incorporate the residuals from the first-stage estimation results in the full model. Subsequently, we test the null hypothesis that the coefficients associated with these residuals in the respective equations are all equal to zero; this null hypothesis is tantamount the exogeneity of the respective retail environment variables. A rejection of this null hypothesis then is statistical evidence that the set of retail environment variables are endogenous.

Table II-4. Results of the Hausman Endogeneity Chi-Squared Tests Associated with the Retail Environment Variables[a]

Model	Club	Convenience	Dollar	Discount	Grocery	Drug
Chi-squared statistic	1.60	8.12	0.50	0.34	3.45	4.41
p-value	0.80	0.09	0.97	0.99	0.49	0.35

[a] Chi-squared tests each with four degrees-of-freedom.

Results based on the Hausman test in Table II-4 indicate the lack of evidence of endogeneity of the retail environment variables. These results are consistent with the assumption of the lack of endogeneity of retail environment variables made by Currie et al. (2010) and Taylor and Villas-Boas (2016). However, unlike previous studies, we provide statistical evidence to support the claim that the set of retail environment variables indeed are exogenous.

II.10 Empirical Results

Maximum likelihood estimates of the respective parameters and standard errors in the various models are obtained with the use of the software package STATA Version 15. In Table II-5, we provide the parameter estimates, associated p-values, likelihood ratio and Wald tests, and goodness-of-fit metrics for the dynamic correlated random effect Tobit model for the entire sample. Additionally, in Table II-10, we provide the marginal effects for the entire sample as well as for data samples for the respective three income levels. In this study, we adopt a level of significance of 0.01 because of the sizable number of observations.

The parameter estimates associated with the standard deviation of the random effect term, σ_u^k, are statistically significant for all store types. As such, household unobserved heterogeneity plays a decisive role in food purchasing behavior. On the basis of likelihood ratio tests, the correlated random effect Tobit model is superior to the pooled Tobit model as well as the random effect Tobit model. The Wald test is the analogue of the conventional F-statistic in regression analysis. For each of models, the Wald tests supports the contention that at least one estimated coefficient is statistically different from zero. Alternatively, the Wald tests support the hypothesis that each model explains a significant amount of variation in household food and beverage expenditures across all store outlets.

We report two different goodness-of-fit metrics to determine the degree of explanatory power associated with each of the respective correlated random effect Tobit models. The first measure, labeled as pseudo R^2, is the square of the correlation of the ***unconditional*** expected value and the actual value of household food and beverage expenditures. For the entire sample, as exhibited in Table II-5, the Pseudo R^2 ranges from 0.212 (model for convenience stores) to 0.323 (model for grocery stores). Alternatively, we use the computation method to calculate the goodness-of-fit metric proposed by Veall and Zimmermann (1996) (V-Z hereafter). The V-Z Pseudo R^2 statistic is the square of the correlation of the ***conditional*** expected value and the actual value of household food and beverage expenditures. As shown in Table II-5, for the entire sample, the R-squared statistic ranges from 0.542 (model for convenience stores) to 0.778 (model for grocery stores). On the basis of these goodness-of-fit measures, the correlated random effect Tobit models explain a notable amount of the variability in household food and beverage expenditures for each store type.

We organize the ensuing discussion of the massive set of empirical results as follows. We initially focus on the entire sample. We initially discuss the statistically significant drivers. Subsequently, we present the impacts of household income, household size, age, urbanization, education, race and ethnicity, the number of club stores, convenience stores, grocery stores, and drug stores, and region, centering attention on the conditional marginal effects.

II.10.1 Estimation Results: Entire Sample

As exhibited in Table *II*-5, expenditures made in the previous year are positively related to current expenditures across all store outlets. The coefficients of the lagged dependent variable are statistically significant, ranging from 0.381 (dollar stores) to 0.669 (grocery stores). As such, these results support our hypothesis of habit persistence associated with food and beverage expenditures made by households. That is to say, we confirm the supposition of habitual spending across all store outlets. These results suggest that, within our data period 2011 to 2015, habitual spending behavior is a key factor in affecting nominal food and beverage expenditures across all store outlets.

As exhibited in Table II-5, household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. Household size is positively related to household expenditures made at discount stores, club stores, and dollar stores, but household size is negatively related to household expenditures made at drug stores. Relative to households in the 40-year-old to 60-year-old category, household expenditures made at grocery stores, dollar stores, and drug stores are higher for households 60 years of age and older. But relative to households in the 40-year-old to 60-year-old category, household expenditures made at discount stores are lower for households 60 years of age and older. No statistically significant differences are evident for households less than 40 years of age and for households in the 40-year-old to 60-year-old category.

For households located in urban areas, household food and beverage expenditures are higher in drug stores, but lower in discount stores, dollar stores, and convenience stores relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are higher in discount stores and dollar stores but lower in grocery stores, club stores, and drug stores relative to households located outside of urban and rural areas.

Relative to households who have graduated from college, households with less than a high school education expend more at discount stores but less at drug stores. Relative to households who have graduated from college, households with a high school education spend more on food and beverages at discount stores and dollar stores but less at grocery stores and drug stores. Relative to households who have graduated from college, households with some level of college experience expend more at discount stores and dollar stores.

Relative to non-Hispanic white households, non-Hispanic black households spend more on food and beverages at discount stores, club stores, dollar stores, convenience stores, and drug stores but less at grocery stores. Relative to non-Hispanic white households, non-Hispanic Asian households spend more on food and beverages at club stores but less at grocery stores. Relative to non-Hispanic white households, non-Hispanic non-black and non-Asian households spend more on food and beverages at dollar stores and convenience stores. Finally, relative to non-Hispanic white households, Hispanic households spend more on food and beverages at discount stores, dollar stores, convenience stores, and drug stores.

The number of club stores within the residence of households negatively impacts food and beverage expenditures made at grocery stores and drug stores but positively affects food and beverage expenditures made at club stores. The number of grocery stores and supercenters within the residence of households negatively impacts expenditures made at discount stores. On the other hand, the number of convenience stores and the number of drug stores within the residence of households are not statistically significant factors affecting expenditures made at any of the six store outlets.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are higher at grocery stores, convenience stores, and drug stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Middle Atlantic region are higher at grocery stores, dollar stores, and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East North Central region are higher at grocery stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West North Central region are higher at discount stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the South Atlantic region are higher at discount stores, dollar stores, and convenience stores but are lower at club stores.

Table II-5. Maximum Likelihood Parameter Estimates and the Associated p-values for the Explanatory Variables Based on the Entire Sample of Panel Households

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
Lagged dependent variable	0.669*	0.435*	0.509*	0.381*	0.549*	0.402*
	(0.006)	(0.005)	(0.006)	(0.005)	(0.009)	(0.005)
Income	0.000	0.023	0.059	-0.042	0.057	0.003
	(0.007)	(0.013)	(0.029)	(0.019)	(0.048)	(0.020)
Household size	0.003	0.027*	0.060*	0.032*	-0.019	-0.019*
	(0.002)	(0.004)	(0.010)	(0.007)	(0.016)	(0.007)
Age <40	0.008	-0.007	-0.116	-0.058	0.300	0.055
	(0.016)	(0.035)	(0.087)	(0.059)	(0.132)	(0.056)
Age >60	0.017*	-0.059*	0.074	0.080*	-0.018	0.095*
	(0.005)	(0.012)	(0.030)	(0.020)	(0.047)	(0.019)
Urban	0.012	-0.097*	-0.012	-0.103*	-0.397*	0.061*
	(0.005)	(0.011)	(0.030)	(0.020)	(0.045)	(0.018)
Rural	-0.030*	0.083*	-0.459*	0.105*	-0.052	-0.123*
	(0.006)	(0.015)	(0.042)	(0.027)	(0.060)	(0.025)
Under high school	-0.047	0.119*	0.001	0.173	-0.196	-0.192*
	(0.019)	(0.042)	(0.111)	(0.068)	(0.168)	(0.067)
Graduate high school	-0.031*	0.085*	-0.010	0.191*	-0.036	-0.062*
	(0.006)	(0.013)	(0.034)	(0.022)	(0.051)	(0.021)
College experienced	-0.009	0.047*	0.005	0.131*	0.098	0.003
	(0.005)	(0.010)	(0.026)	(0.017)	(0.040)	(0.016)
Non-Hispanic black	-0.034*	0.107*	0.322*	0.332*	0.326*	0.146*
	(0.007)	(0.018)	(0.046)	(0.031)	(0.069)	(0.028)
Non-Hispanic Asian	-0.041*	-0.018	0.210*	-0.054	-0.264	-0.097
	(0.013)	(0.030)	(0.075)	(0.054)	(0.128)	(0.048)
Non-Hispanic other	-0.034	0.070	0.080	0.151*	0.483*	0.034
	(0.014)	(0.031)	(0.075)	(0.051)	(0.117)	(0.048)
Hispanic	-0.013	0.083*	0.131	0.210*	0.247*	0.099*
	(0.010)	(0.023)	(0.058)	(0.040)	(0.092)	(0.037)
Number of Club stores	-0.028*	0.012	0.136*	-0.005	-0.060	-0.107*
	(0.007)	(0.014)	(0.029)	(0.021)	(0.052)	(0.021)
Number of Convenience stores	0.001	-0.003	-0.002	0.002	0.010	0.003
	(0.002)	(0.003)	(0.007)	(0.005)	(0.012)	(0.005)
Number of Grocery stores and Supercenters	0.003	-0.008*	0.006	-0.003	-0.009	0.001
	(0.001)	(0.002)	(0.006)	(0.004)	(0.010)	(0.004)
Number of Drug stores	0.002	0.002	-0.009	-0.008	0.027	0.006
	(0.002)	(0.004)	(0.008)	(0.006)	(0.015)	(0.006)

Table II-5. Continued

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
New England	0.032*	-0.062	-0.363*	-0.094	0.406*	0.176*
	(0.011)	(0.027)	(0.071)	(0.049)	(0.111)	(0.043)
Middle Atlantic	0.027*	-0.032	-0.433*	0.125*	1.183*	0.058
	(0.008)	(0.020)	(0.052)	(0.035)	(0.081)	(0.032)
East North Central	0.029*	-0.042	-0.196*	0.034	0.593*	0.037
	(0.008)	(0.019)	(0.049)	(0.033)	(0.077)	(0.030)
West North Central	-0.019	0.118*	-0.235*	-0.068	1.387*	-0.064
	(0.009)	(0.022)	(0.059)	(0.039)	(0.089)	(0.036)
South Atlantic	0.006	0.086*	-0.251*	0.125*	0.773*	0.074
	(0.008)	(0.018)	(0.047)	(0.033)	(0.076)	(0.029)
East South Central	0.011	0.091*	-0.303*	0.210*	0.234	0.022
	(0.011)	(0.026)	(0.068)	(0.045)	(0.106)	(0.041)
West South Central	0.015	0.095*	-0.211*	0.065	0.386*	-0.028
	(0.009)	(0.021)	(0.056)	(0.038)	(0.089)	(0.034)
Mountain	0.015	0.060	-0.070	-0.048	0.733*	-0.111*
	(0.010)	(0.023)	(0.059)	(0.041)	(0.095)	(0.037)
Constant	0.491*	0.675*	-6.118*	1.070*	-4.208*	-0.934*
	(0.041)	(0.095)	(0.260)	(0.170)	(0.377)	(0.152)
Initial value	0.224*	0.460*	0.686*	0.657*	0.849*	0.526*
	(0.005)	(0.005)	(0.007)	(0.006)	(0.011)	(0.005)
σ_u	0.217*	0.649*	1.802*	1.220*	2.448*	1.062*
	(0.005)	(0.007)	(0.017)	(0.011)	(0.028)	(0.011)
σ_e	0.586*	1.114*	2.107*	1.532*	3.035*	1.668*
	(0.002)	(0.003)	(0.007)	(0.004)	(0.013)	(0.004)
LR χ^2 test of the Correlated Random Effect Tobit Model vs the Pooled Tobit Model	378.9*	2,675.5*	5,227.7*	6,472.5*	4,288.8*	3,540.3*
LR χ^2 test of the Correlated Random Effect Tobit Model vs Random Effect Tobit Model	2,748.9*	9,085.7*	9,062.9*	12,033.9*	6,808.6*	9,791.9*
Wald χ^2 test	1,965.2*	9,730.3*	10,690.9*	12,334.3*	5,760.3*	10,201.6*
Pseudo R^2	0.323	0.233	0.233	0.221	0.212	0.215
V-Z R^2	0.778	0.702	0.749	0.678	0.542	0.603
Number of Observations	140,545	140,545	140,545	140,545	140,545	140,545
Number of Households	28,109	28,109	28,109	28,109	28,109	28,109

*p<0.01

Numbers in parentheses correspond to standard errors.

Table II-6. Mean, Minimum, and Maximum Predicted Probability to Purchase at Each Store Type for the Entire Sample and for the Three Income Subsamples

		Grocery	Discount	Club	Dollar	Convenience	Drug
	mean	0.999	0.988	0.656	0.732	0.323	0.904
Entire	min	0.857	0.356	0.056	0.231	0.067	0.356
	max	1	1	0.999	0.999	0.996	0.999
	mean	0.999	0.983	0.396	0.859	0.331	0.866
Low	min	0.798	0.443	0.004	0.173	0.068	0.272
	max	1	1	0.998	1	0.997	0.999
	mean	0.999	0.989	0.617	0.788	0.340	0.904
Mid	min	0.869	0.33	0.103	0.15	0.023	0.287
	max	1	1	0.999	0.999	0.996	0.999
	mean	0.999	0.985	0.789	0.612	0.301	0.916
High	min	0.856	0.404	0.158	0.146	0.061	0.343
	max	1	1	0.999	0.999	0.996	0.999

The predicted probability of purchasing food and beverages is given by $P\left(EX_{ht}^{k\ *} > 0 \middle| z_{ht}^{k}\right)$.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region are higher at discount stores and dollar stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West South Central region are higher at discount stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Mountain region are higher at convenience stores but are lower at drug stores. Without question, region is a key determinant of household food and beverage expenditures across the six store outlets.

The conditional marginal effects of the respective explanatory variables associated with household expenditures on food and beverages in the correlated random effect Tobit models across store outlets and across income categories are exhibited in Table II-7. We present the impacts of household income, household size, age, urbanization, education, race and ethnicity, the number of club stores, convenience stores, grocery stores, and drug stores, and region. In addition, we also report the marginal effects of these aforementioned explanatory variables concerning the probability to visit.

II.10.2 Entire Sample: Conditional Marginal Effects

As presented in Table II-7, note that changes in household income do not significantly affect the level of household expenditure for food and beverages. Unit increases in household size result in a $19.40 increase in household expenditures at discount stores, a $21.05 rise in household expenditures at club stores, and a $0.99 increase in household expenditures at dollar stores annually. But unit increases in household size result in a decline of $4.20 at drug stores annually, holding all other factors invariant.

Relative to households in the 40-year old to 60-year old category, household expenditures made at grocery stores, club stores, dollar stores, and drug stores are higher by $29.66, $25.92, $2.43, and $21.42 annually for households 60 years of age and older. But relative to households in the 40-year old to 60-year old category, household expenditures made at discount stores are lower by $42.17 annually for households 60 years of age and older.

For households located in urban areas, household food and beverage expenditures are higher in grocery stores and in drug stores by $21.07 and $13.87 annually, but lower in discount stores, dollar stores, and convenience stores by $69.75, $3.13, and $130.38 respectively on an annual basis relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are higher in discount stores and dollar stores by $59.42 and $3.19 annually but lower in grocery stores, club stores, and drug stores by

$54.28, $161.14, and $27.79 annually relative to households located outside of urban and rural areas.

Relative to households who have graduated from college, households with less than a high school education expend $85.21 more at discount stores but $43.41 less at drug stores annually. Relative to households who have graduated from college, households with a high school education spend $60.83 more and $5.83 more on food and beverages at discount stores and dollar stores but $55.83 less at grocery stores and $13.98 less at drug stores annually. Relative to households who have graduated from college, households with some level of college experience expend $33.61 more at discount stores and $4.01 more at dollar stores annually.

Relative to non-Hispanic white households, non-Hispanic black households spend $76.71, $113.05, $10.51, $107.13, and $32.94 more on food and beverages at discount stores, club stores, dollar stores, convenience stores, and drug stores respectively, but $60.94 less at grocery stores. Relative to non-Hispanic white households, Asian households spend $73.62 more on food and beverages at club stores but $73.12 less at grocery stores annually. Relative to non-Hispanic white households, non-Hispanic non-black and non-Asian households spend $4.63 and $158.54 more on food and beverages at dollar stores and convenience stores annually. Finally, relative to non-Hispanic white households, Hispanic households spend $59.35, $6.43, $81.24, and $22.26 more on food and beverages at discount stores, dollar stores, convenience stores, and drug stores, respectively on an annual basis.

For each unit increase in the number of club stores, household expenditures made at grocery stores decline by $49.72 and expenditures made at drug stores decline by $24.11. On the other hand, for each unit increase in club stores, expenditures made at club stores increases by

$47.61. For each unit increase in the number of grocery stores and supercenters, household expenditures made at discount stores decline by $5.99.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are higher $57.92, $133.50, and $39.75 annually at grocery stores, convenience stores, and drug stores respectively but are lower by $127.56 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Middle Atlantic region are higher by $48.80, $3.82, $388.64 annually at grocery stores, dollar stores, and convenience stores respectively but are lower by $152.22 annually at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East North Central region are higher by $51.26 at grocery stores and by $194.83 at convenience stores annually but are lower by $68.74 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West North Central region are higher by $84.81 at discount stores and by $455.45 at convenience stores annually but are lower by $82.56 annually at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the South Atlantic region are higher annually by $61.51, $3.81, and $253.88 at discount stores, dollar stores, and convenience stores but are lower annually by $88.36 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region are higher by $65.53 at discount stores and by $6.42 at dollar stores but are lower by $106.36 at club stores annually. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West South Central region are higher by $68.54 at discount stores and by $126.89 at convenience stores annually but are lower by $74.16 at club stores annually. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Mountain region are higher by \$240.86 at convenience stores but are lower by \$25.03 at drug stores annually.

II.10.3 Entire Sample: Marginal Effects Associated with the Probability of Purchasing

As exhibited in Table II-10, the probability of purchasing food and beverages at grocery stores are significantly higher for households who are greater than 60 years of age as well as households located in New England, the Mid-Atlantic, and the East North Central regions of the United States. But this probability is significantly lower for household located in rural areas, for households with at most a high school education, for non-Hispanic black households and for non-Hispanic Asian households. That said, the magnitude of the marginal effects associated with purchasing food and beverages at grocery stores albeit statically significant is negligible from a practical standpoint.

As presented in Table II-10, unit increases in household size lead to increases in the probability of purchasing by 0.03% in discount stores, 0.39% in club stores, 0.35% in dollar stores. But, in drug stores, unit increases in household size lead to decreases in the probability of purchasing by 0.10%.

For household head age over 60, the probability of purchasing at dollar stores and drug stores is higher by 0.87% and 0.49% respectively stores is higher relative to the age group between 40 to 60. But the probability of purchasing at discount stores is lower by 0.07% relative to households in the age group between 40 and 60.

Households who reside in urban areas have a higher probability to purchase at drug stores by 0.32%. Conversely, this probability of purchasing is lower by 0.12%, 1.11%, and 2.68% in discount stores, dollar stores and convenience stores respectively. Households who reside in rural areas have a lower probability to purchase at club stores by 2.97% and at drugstores by 0.63% than households who do not live in urban and rural areas. But, households who reside in rural areas have a higher probability to purchase at discount stores by 0.10% and at dollar stores by 1.13%.

Relative to households who have graduated from college, household heads who have less than a high school education have a higher probability to purchase at discount stores by 0.14%, but have a lower probability to purchase at drug stores by 0.99%. Similarly, relative to households who have graduated from college, household heads who have a high school degree have a higher probability to purchase at discount stores by 0.10%, but have a lower probability to purchase at drug stores by 0.32%. Relative to households who have graduated from college, households with some college education are more likely to purchase at discount stores and at dollar stores by 0.06% and 1.43% respectively.

Relative to non-Hispanic white households, non-Hispanic black households are more likely to purchase at discount stores by 0.13%, at club stores by 2.08%, at dollar stores by 3.61%, at convenience stores by 2.2%, and at drug stores by 0.75%. Non-Hispanic Asian households have a higher probability to purchase at club stores by 1.36%. Non-Hispanic other households have a higher probability to purchase at dollar stores by 1.36% and at convenience stores by 3.26% stores than non-Hispanic white households. Hispanic households are more likely to purchase at discount stores by 0.10%, at dollar stores by 2.28%, at convenience stores by 1.67%, and at drug stores by 0.51%.

Unit increases in the number of club stores increase the probability of purchasing at club stores by 0.88% decrease the likelihood of purchasing at drug stores by 0.55%. Unit increase in the number of grocery stores and supercenters negatively impacts the likelihood of purchasing at discount stores by 0.01%.

Relative to households who are located in the Pacific region, households who are located in New England are more likely to purchase at convenience stores by 2.74% and at drug stores by0.90%. However, they are less likely to purchase at club stores by 2.35%. Households located in the Mid-Atlantic region are more likely to purchase at dollar stores and at convenience stores by 1.36% and 7.99% respectively but are less likely to purchase at club stores by 2.80% relative to households located in the Pacific region. Households located in East North Central region have a higher probability to purchase at convenience stores by4.00%, but have a lower probability to purchase at club stores by 1.27% relative to households located in the Pacific region. For households who reside in the West North Central region, the probabilities to purchase at discount stores and convenience stores are higher by 0.14% and 9.36% respectively than households who reside in the Pacific region. But the probability to purchase at club stores is lower by 1.52%. Households located in South Atlantic region have higher probabilities to purchase at discount stores, at dollar stores, and at conveniences stores by 0.10%, 1.35%r, and 5.22% relative to households located in the Pacific region. That said, this probability is lower by 1.63% for club stores for households located in the South Atlantic region relatively to households located in the Pacific region. Households located in West South Central region have a higher probability to purchase at discount stores by 0.12% and at convenience stores by 2.61% but a lower probability to purchase at club stores by 1.37% relative to households located in the Pacific region. Households located in the Mountain region are more likely to purchase at convenience stores by 4.95% but are less likely to purchase at drug stores by -0.57% vis-vis households located in the Pacific region.

Table II-7. Conditional Marginal Effects of Household Food and Beverage Expenditures and Marginal Effects Associated with the Probability of Purchasing by Store Type Based on the Entire Sample of Panel Households

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Household income	0.000	0.000	0.0003	0.016	0.004	0.035	-0.005	-0.017	0.004	0.225	0.000	0.002
	(0.000)	(0.005)	(0.0002)	(0.009)	(0.002)	(0.017)	(0.002)	(0.008)	(0.003)	(0.190)	(0.001)	(0.002)
Household size	0.000	5.802	0.0003*	19.397*	0.004*	21.054*	0.004*	0.991*	-0.001	-6.401	-0.001*	-4.196*
	(0.000)	(3.312)	(0.0001)	(3.042)	(0.001)	(3.678)	(0.001)	(0.214)	(0.001)	(5.371)	(0.000)	(1.498)
Age<40	0.000	14.836	-0.0001	-4.791	-0.0075	-40.738	-0.0063	-1.781	0.0203	98.527	0.0028	12.384
	(0.000)	(28.944)	(0.0004)	(25.412)	(0.0056)	(30.632)	(0.0064)	(1.809)	(0.0089)	(43.760)	(0.0029)	(12.660)
Age>60	0.000*	29.658*	-0.0007*	-42.166*	0.0048	25.921	0.0087*	2.437*	-0.0012	-6.039	0.0049*	21.417*
	(0.000)	(9.634)	(0.0001)	(8.724)	(0.0019)	(10.562)	(0.0022)	(0.624)	(0.0032)	(15.466)	(0.001)	(4.366)
Urban	0.000	21.073	-0.0012*	-69.750*	-0.0008	-4.251	-0.0111*	-3.133*	-0.0268*	-130.382*	0.0032*	13.872*
	(0.000)	(8.518)	(0.0001)	(8.258)	(0.0019)	(10.480)	(0.0022)	(0.616)	(0.0031)	(17.315)	(0.0009)	(4.122)
Rural	-0.000*	-54.280*	0.0010*	59.642*	-0.0297*	-161.137*	0.0113*	3.193*	-0.0035	-17.216	-0.0063*	-27.790*
	(0.0000)	(11.595)	(0.0002)	(11.141)	(0.0027)	(15.793)	(0.0029)	(0.825)	(0.0041)	(19.859)	(0.0013)	(5.667)
Less than high school	-0.000	-83.340	0.0014*	85.207*	0.0000	0.223	0.0188	5.297	-0.0133	-64.520	-0.0099*	-43.417*
	(0.000)	(33.734)	(0.0005)	(30.384)	(0.0072)	(39.050)	(0.0074)	(2.081)	(0.0114)	(55.478)	(0.0034)	(15.167)
High school graduate	-0.000*	-55.831*	0.0010*	60.830*	-0.0007	-3.650	0.0207*	5.834*	-0.0025	-11.922	-0.0032*	-13.976*
	(0.000)	(10.130)	(0.0002)	(9.411)	(0.0022)	(11.786)	(0.0024)	(0.678)	(0.0035)	(16.863)	(0.0011)	(4.698)
College experienced	0.000	-16.898	0.0006*	33.611*	0.0003	1.858	0.0143*	4.013*	0.0066	32.306	0.0001	0.587
	(0.000)	(8.271)	(0.0001)	(7.486)	(0.0017)	(9.034)	(0.0019)	(0.529)	(0.0027)	(13.408)	(0.0008)	(3.714)
Non-Hispanic black	-0.000*	-60.939*	0.0013*	76.712*	0.0208*	113.054*	0.0361*	10.151*	0.0220*	107.134*	0.0075*	32.938*
	(0.000)	(13.153)	(0.0002)	(12.626)	(0.003)	(16.484)	(0.0033)	(0.952)	(0.0047)	(23.956)	(0.0014)	(6.327)
Non-Hispanic Asian	-0.000*	-73.118*	-0.0002	-12.856	0.0136*	73.620*	-0.0058	-1.644	-0.0178	-86.734	-0.005	-21.991
	(0.000)	(22.648)	(0.0004)	(21.588)	(0.0049)	(26.471)	(0.0058)	(1.645)	(0.0087)	(42.508)	(0.0025)	(10.826)
Non-Hispanic other	0.000	-60.696	0.0009	50.558	0.0052	28.280	0.0164*	4.625*	0.0326*	158.542*	0.0018	7.759
	(0.000)	(24.240)	(0.0004)	(21.974)	(0.0048)	(26.294)	(0.0055)	(1.553)	(0.0079)	(39.953)	(0.0025)	(10.911)
Hispanic	0.000	-22.642	0.0010*	59.347*	0.0085	46.169	0.0228*	6.431*	0.0167*	81.235*	0.0051*	22.261*
	(0.000)	(17.703)	(0.0003)	(16.627)	(0.0038)	(20.450)	(0.0043)	(1.217)	(0.0062)	(30.862)	(0.0019)	(8.309)

Table II-7. Continued

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Club stores	-0.000*	-49.715*	0.0002	8.843	0.0088*	47.614*	-0.0006	-0.165	-0.004	-19.603	-0.0055*	-24.109*
	(0.0000)	(12.877)	(0.0002)	(9.903)	(0.0019)	(10.435)	(0.0023)	(0.640)	(0.0035)	(16.998)	(0.0011)	(4.819)
Convenience stores	0.000	1.583	0.0000	-2.364	-0.0001	-0.655	0.0003	0.071	0.0007	3.264	0.0001	0.574
	(0.000)	(2.921)	(0.0000)	(2.256)	(0.0004)	(2.409)	(0.0005)	(0.144)	(0.0008)	(3.914)	(0.0002)	(1.081)
Grocery stores and Supercenters	0.000	5.425	-0.0001*	-5.988*	0.0004	1.993	-0.0003	-0.093	-0.0006	-3.002	0.0001	0.288
	(0.000)	(2.310)	(0.0000)	(1.791)	(0.0004)	(1.961)	(0.0004)	(0.117)	(0.0006)	(3.161)	(0.0002)	(0.850)
Drug stores	0.000	3.289	0.0000	1.436	-0.0006	-3.044	-0.0009	-0.247	0.0018	8.979	0.0003	1.335
	(0.000)	(3.549)	(0.0000)	(2.747)	(0.0005)	(2.934)	(0.0006)	(0.179)	(0.001)	(4.857)	(0.0003)	(1.315)
Ne	0.000*	57.923*	-0.0008	-44.697	-0.0235*	-127.559*	-0.0102	-2.872	0.0274*	133.498*	0.0090*	39.749*
	(0.000)	(20.147)	(0.0003)	(19.500)	(0.0046)	(25.435)	(0.0053)	(1.493)	(0.0075)	(37.670)	(0.0022)	(9.780)
Ma	0.000*	48.799*	-0.0004	-23.331	-0.0280*	-152.223*	0.0136*	3.823*	0.0799*	388.642*	0.003	13.148
	(0.000)	(14.770)	(0.0002)	(14.299)	(0.0034)	(19.031)	(0.0038)	(1.083)	(0.0055)	(37.455)	(0.0016)	(7.168)
Enc	0.000*	51.261*	-0.0005	-30.494	-0.0127*	-68.744*	0.0037	1.053	0.0400*	194.826*	0.0019	8.275
	(0.000)	(13.762)	(0.0002)	(13.311)	(0.0032)	(17.288)	(0.0036)	(1.009)	(0.0052)	(28.526)	(0.0015)	(6.681)
Wnc	0.000	-34.301	0.0014*	84.810*	-0.0152*	-82.561*	-0.0074	-2.070	0.0936*	455.453*	-0.0033	-14.457
	(0.000)	(16.422)	(0.0003)	(15.954)	(0.0038)	(20.807)	(0.0043)	(1.205)	(0.006)	(42.433)	(0.0018)	(8.020)
Sa	0.000	11.413	0.0010*	61.508*	-0.0163*	-88.361*	0.0135*	3.805*	0.0522*	253.879*	0.0038	16.650
	(0.000)	(13.610)	(0.0002)	(13.229)	(0.0031)	(16.908)	(0.0035)	(0.999)	(0.0051)	(30.219)	(0.0015)	(6.606)
Esc	0.000	19.602	0.0011*	65.531*	-0.0196*	-106.356*	0.0228*	6.421*	0.0158	76.943	0.0011	4.997
	(0.000)	(19.085)	(0.0003)	(18.453)	(0.0044)	(24.306)	(0.0049)	(1.379)	(0.0072)	(35.345)	(0.0021)	(9.276)
Wsc	0.000	27.010	0.0012*	68.543*	-0.0137*	-74.157*	0.0071	1.998	0.0261*	126.887*	-0.0015	-6.395
	(0.000)	(15.852)	(0.0003)	(15.368)	(0.0036)	(19.686)	(0.0041)	(1.154)	(0.006)	(30.309)	(0.0017)	(7.689)
Mt	0.000	27.014	0.0007	42.932	-0.0045	-24.561	-0.0052	-1.473	0.0495*	240.857*	-0.0057*	-25.025*
	(0.000)	(17.299)	(0.0003)	(16.691)	(0.0038)	(20.910)	(0.0045)	(1.263)	(0.0064)	(35.094)	(0.0019)	(8.429)

Superscript * indicates p-value <0.01

Numbers in parentheses correspond to standard errors.

The marginal effect of the probability to purchase at any store outlet is given by $\frac{\partial P(EX_{ht}^{k*}>0|z_{ht}^k)}{\partial z_{ht}^k}$, and the marginal effect of the conditional expectation of household food and beverage expenditures is expressed mathematically as $\frac{\partial E[EX_{ht}^k|z_{ht}^k, EX_{ht}^k > 0]}{\partial z_{ht}^k}$. All calculations pertaining to marginal effects are made at the sample means of the data.

II.11 Concluding Remarks

A number of choices is evident beyond traditional supermarkets or grocery stores owing to the increasingly diverse U.S. retail food landscape. Despite the plethora of previous studies that largely focus on factors affecting store choice, one area of research that has received relatively little attention is how the magnitude of household food and beverage expenditures is impacted by the type of store outlets. In this light, the purpose of this study is to examine how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type. The list of socio-demographic factors includes: (1) household income; (2) household size; (3) age; (4) urbanization; (5) education; (6) race and ethnicity; and (7) region. Characteristics of the retail environment relate to the number of club stores, the number of convenience stores, the number of grocery stores and supercenters and the number of drug stores within the zip code area of the of the household. Whether traditional or non-traditional, store outlets differ in prices, product assortment, advertising strategies, and location (Volpe, Kuhns, and Jaenicke, 2017). The outlets considered in this study are grocery, convenience, discount, club, drug, and dollar store types.

As mentioned previously, prior works mainly highlighted store choice. To differentiate our study from the extant literature, we explore the factors which directly affect household food expenditure by store outlet. Indeed, Volpe, Jaenicke, and Chenarides (2018) estimated the impacts of expenditure share by store format, but in our study, we quantify the magnitude of the impact of household socio-demographics, the retail food environment, and spending habits on food and beverage expenditures by diverse store types. Hence, by analyzing factors that impact household food expenditure across the aforementioned six store types, this study contributes to the economic literature. Another contribution is that our study also considers habitual persistence or spending habits, a dynamic property of household expenditure on food and beverages. However, in the previously mentioned studies, habitual behavior was not included in the set of explanatory variables.

To further differentiate our study from previous studies, we employ a dynamic correlated random effect Tobit model to incorporate habitual purchasing behavior. The source of data for this analysis is the Nielsen Homescan Panel over the period between 2011 and 2015. Specifically, we use a balanced panel of 28,109 households who participated in the survey for all five years from 2011 to 2015. The total number of observations available for analysis is 140,545. The panel structure allows us to incorporate dynamic modeling by including lagged dependent variables as explanatory variables to account for spending habits.

Another advantage of the use of this model is that we are in a position to handle corner solution problems. The dependent variables reflect household purchasing history according to store type and indeed have zero values; hence the dependent variables are left censored. A differentiated feature of our empirical analysis relates to transforming the dependent variables which include zero observations using the inverse hyperbolic sine (arcsinh) method (Bellemare and Wichman 2020). A notable problem with taking the logarithm of any variable is that it does not allow retaining zero-valued observations because the ln (0) is undefined. As pointed out by Bellemare and Wichman (2019), "applied econometricians are typically loath to drop those observations for which the logarithm is undefined." Consequently, researchers often have resorted to ad hoc means of accounting for this situation when taking the natural logarithm of a variable, such as adding 1 to the variable prior to its transformation (MaCurdy and Pencavel, 1986). In recent years, the inverse hyperbolic sine (or arcsinh) transformation has grown in popularity among applied econometricians due to the fact that it is similar to the behavior of the logarithm function, it allows retaining zero-valued observations without any arbitrariness, and it often results in normal distributions (Burbidge et al. 1988; Yen and Jones 1997; MacKinnon and Magee 1990; Pence 2006; Van den Heuvel et al. 2011; Bellemere, Barrett, and Just 2013; Brown et al. 2015; Bellemere and Wichman 2020).

The results support the supposition of habitual spending across all store outlets. These results suggest that, within the data period 2011 to 2015, habitual spending behavior is undoubtedly a key factor in affecting nominal food and beverage expenditures across all store formats. Household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. However, household size, age, urbanization, education, race and ethnicity, region, time-invariant socio-demographic variables, indeed are drivers of household food and beverage expenditures at the six store outlets across the income categories. This finding is in line with the hypothesis of underlying household heterogeneity and in agreement with the results of Bilsard, Stewart, and Jolliffe (2004) and of Taylor and Villas-Boas (2016).

Further, the number of convenience stores in the zip code area of households do not significantly influence the level of food and beverage expenditures across the respective store outlets and across the respective income categories. The same result is true for drug stores but for a single exception. In the high-income sample, the number of drug stores in the zip code area negatively impacts food and beverage expenditures made at dollar stores. In the entire sample and in the mid-income sample, the number of club stores negatively impacts household expenditures made at grocery stores and drug stores. But this finding is not the case within the low-income sample and within the high-income sample. In addition, in the entire sample, the number of grocery stores and supercenters in the zip code area negatively impacts household food and beverage expenditures made at discount stores. Nevertheless, this finding is not the case in each of the respective income sub-samples.

Bottom line, evidence exists to support the hypothesis that the retail environment plays a limited role in affecting household expenditures for food and beverages across store outlets and across income sub-samples. This result differs from previous findings by Kyureghian and Nayga (2013) and by Taylor and Villas-Boas (2016), but this result is in alignment with the work by Ver Ploeg and Wilde (2018).

The findings in this study make several contributions to the current economic literature. First, we provide a detailed view that describes household spending behavior across six store types for three income classifications. Second, the construction and estimation of dynamic random effect Tobit models constitute the first attempt in the literature dealing with household food and expenditure by store outlets for various income classifications. Third, we use a novel method to deal with problems in data (zero observations and extreme values) through the inverse hyperbolic sine transformation. Fourth, we derive the accompanying expressions for calculating conditional marginal effects and the marginal effects associated with the probability of purchasing food and beverages on the basis of the inverse hyperbolic sine transformation.

Future research in this area may center attention on specific household food and beverage expenditures rather than the aggregate, for example, fresh fruits and vegetables or meat products. Particularly for low-income households, we are in position to investigate nutrition intake of households associated with the six store types by income level. As such, this research may uncover a link between store type and nutrition intake, especially useful for policies dealing with various food assistance programs. Although this research covers the period 2011 to 2015, this study establishes a baseline. Our study can be replicated using more recent data to determine the robustness of our findings. Without question, because today's food retail environment is considerably diverse, more work is needed to understand the role of store outlets in affecting dietary quality in America across various income sub-samples.

CHAPTER III

DO HOUSEHOLD FOOD EXPENDITURES BY STORE TYPE DIFFER BY INCOME LEVEL?

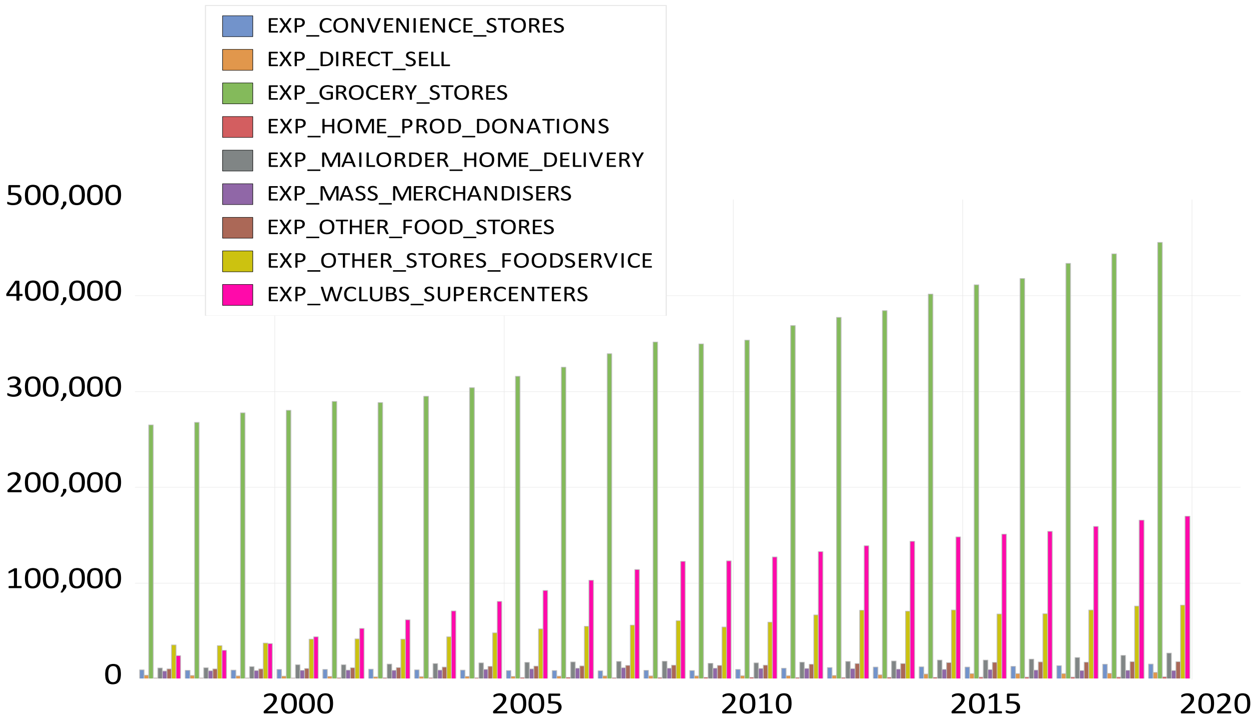
III.1 Introduction

Without question, the food retail environment has changed over the past few decades (Capps and Griffin, 1998; Goldman and Hino, 2005; and Volpe, Kuhns, and Jaenicke, 2017). Over the past 25 years, a number of nontraditional store formats—including supercenters (such as Wal-Mart), dollar stores, and club stores—have gained market share and prominence in the retail food landscape. As exhibited in Figure III-1, the Economic Research Service (ERS) breaks down nominal food expenditures into nine categories: (1) convenience stores; (2) grocery stores; (3) mail order/home delivery; (4) mass merchandisers; (5) warehouse clubs/supercenters; (6) direct sales; (7) other food stores; (8) other stores foodservice; and (9) donations. In particular, over the period 1997 to 2019, nominal expenditures from convenience store were $10.93 billion on average; currently $15.75 billion; from grocery stores $$347.83 billion on average; currently $455.73 billion; from mail order/home delivery $17.80 billion on average; currently $27.00 billion; from mass merchandisers $9.92 billion on average; currently $8.78 billion; from warehouse clubs/supercenters $106.53 billion on average; currently $169.90 billion.

As shown in Figure III-2, shares of nominal food at home expenditures over the period 1997 to 2019 were as follows: (1) convenience stores, 1.96% on average, ranging from 1.51% to 2.58%; currently 2.02%; (2) grocery stores, 61.91% on average, ranging from 57.97% to 71.63%; currently 58.31%; (3) mail order/home delivery, 3.15% on average, ranging from 2.80% to 3.61%; currently 3.45%; (4) mass merchandisers, 1.82% on average, ranging 1.12% to 2.27%; currently

1.12%; and (5) warehouse/supercenters, 17.70% on average, ranging from 6.61% to 21.78%; currently 21.74%. Accounting for about 80% of at-home food expenditures, the major outlets unequivocally are grocery stores and warehouse clubs/supercenters. That said, other longstanding outlets such as convenience stores, discount stores, and dollar stores have expanded their food offerings to better attract grocery shoppers (Volpe, Kuhns, and Jaenicke, 2017).

Figure III-1. Breakdown of Nominal Food at Home Expenditures, 1997 to 2019 (Millions of Dollars) (2)



Source: Economic Research Service, USDA.

Figure III-2. Share of Nominal Food at Home Expenditures, 1997 to 2019, Percent (2)

Source: Economic Research Service, USDA

Previous studies from the fields of economics and marketing have mainly centered attention on the determinants of store choice. Evidence from this rich literature suggests in large part that the choice of food stores is based on a variety of factors including prices, product variety, quality of meat and produce, distance from home, courteous services and degree of competition (Arnold, Oum, and Tigert, 1983; Smith, 2004; Smith, 2006; Hausman and Leibtag, 2007; Briesch, Chintagunta, and Fox, 2009; Richards, Hamilton and Yonezawa, 2016; Marshall and Pires, 2017; and Chenarides and Jaenicke, 2017). Store choice also has been shown to be influenced by household demographics and past purchase history (Staus, 2009) as well as by characteristics of the entire local food market (Feather, 2003; Kyureghian and Nayga, 2013; and Kyureghian, Nayga, and Bhattacharya, 2013), the degree of competition among food stores (Hausman and Leibtag, 2007), and prices offered by various outlet types (Volpe and Lavoie, 2008; Broda, Leibtag, and Weinstein, 2009; Basker and Noel, 2009; and Leibtag, Barker, and Dutko, 2010).. Additionally, previous studies have investigated the role that food access plays in food insecurity, malnutrition, and fruit and vegetable consumption, among other concerns (Rose and Richards, 2004; Bustillos et al., 2008; and Powell and Bao, 2009).

Taylor and Villas-Boas (2016) investigated choices of store outlets as a function of household attributes using a multinomial mixed logit model based on data acquired from the National Household Food Acquisition and Purchase Survey (FoodAPS). Household attributes included participation in the Supplemental Nutrition and Assistance Program (SNAP), household income at various levels of the Federal Poverty Line (FPL), and various measures of the food environment and food access—population density, the share of households living in rural and urban census tracts, the share of households living in a census block group identified as a food desert, and share of households without car access. The store outlets considered were supermarkets, superstores, grocery stores, convenience stores, and farmers' markets.

Moreover, based on data from a panel of 3,376 households collected from 11 randomly selected mid-sized counties in the United States, Fan (2017) analyzed the effect of improving food accessibility by way of subsidizing purchases of fruits and vegetables across food deserts and non-food deserts. The household panel was compiled from 174 food stores collected using scanning devices from Information Resources, Inc. (IRI InfoScan) over a period of 16 quarters from 2009 to 2012 in the 11 sample counties. The IRI InfoScan data provided weekly prices and quantities of various fruits and vegetables by food stores. Store characteristics came from Nielsen TDLinx store directory data, and census-tract level socio-demographic information were obtained from the 2008-2012 American Community Survey (ACS). Census-tract level food deserts indicators were compiled from the 2010 USDA Food Access Research Atlas (FARA, USDA, 2013).

The choice of store outlet, specifically convenience stores, club stores, dollar stores, drug stores, grocery stores, and mass merchandisers, in each census tract in a county was estimated using a random-coefficient discrete choice model, known as the BLP model (Berry, Levinsohn, and Pakes, 1995). This discrete choice model for food stores incorporating household heterogeneity was estimated to quantify the welfare impact of expanding access to fruits and vegetables in food deserts and to compare this welfare effect to the welfare effect associated with a subsidy to fruits and vegetable prices in food deserts. The principal conclusion was that expanding the availability of fruits and vegetables in the nearest stores of food deserts without changing prices did not affect appreciably store choice or enhance the welfare of the household panel. In contrast, price subsidy programs associated with fruits and vegetables in food deserts improved the welfare of household panelists.

Volpe, Kuhns, and Jaenicke (2017) examined the effect of store format and income on the *healthfulness* of food purchased based on a large nationwide sample of households as recorded by the Information Resources, Inc. (IRI) over the period between 2008 and 2012. The healthfulness measures used were based on the Low-Cost, Moderate-Cost, and Liberal Food Plans (2007) developed by the Center for Nutrition Policy and Promotion, USDA as well as the Healthy Eating Index developed by the USDA in 2005 (Carlson, Lino, and Fungwe, 2007). Correlations between store formats and the respective healthfulness measures as well as correlations between store formats and expenditure shares by food category were presented. The store formats in this study were supermarkets, drug stores, mass merchandisers, supercenters, convenience stores, dollar stores, and club stores. Despite the wealth of descriptive information provided, Volpe, Kuhns, and Jaenicke (2017) did not provide a formal econometric analysis.

Finally, Volpe, Jaenicke, and Chenarides (2018) investigated the relationship between store formats and the healthfulness of at-home food purchases. The store formats used in this study were supermarkets, drug stores, mass merchandisers, supercenters, club stores, convenience stores, and other stores. To investigate the healthfulness of household food purchases, based on the methodology developed by Volpe and Okrent (2012) a healthfulness score was assigned, hereafter called the USDA*Score*, to the shopping baskets of each household by quarter. The source of data for this analysis was the Nielsen Homescan Panel over the period between 2004 and 2010. The USDA*Score* is based on the differences between category-specific observed expenditure shares and USDA recommended expenditure shares. The principal goal was to investigate how store format decisions and other factors affect the household-specific USDA*Score*.

Because store-format choice and food-purchase healthfulness were hypothesized to be interrelated decisions, a simultaneous-equation system was developed consisting of eight reduced-form equations. Seven of the respective equations expressed store-format expenditure share as a function of prices measured by the publicly available data from the USDA Quarterly Food-at-Home Price Database, the food retail environment measured by counts of the number of supermarkets, convenience stores, and supercenter stores) and household demographics (namely household income, household size, race, employment status, education, presence of male and/or female household heads and participation in the Women's Infants and Children (WIC) program. The remaining equation expressed USDA *Score* as a function of store format shares, prices, market structure measured by the Herfindahl-Hirschman Index of food retailers, and the aforementioned household demographics. Empirical results pertaining to impacts on USDA Score were obtained for all households as well as by three household income levels, less than the 25th percentile of the sample, between the 25th percentile and the 75th percentile of the sample, and greater than the 75th percentile). The principal conclusion drawn from this analysis was that healthier food choices were associated with higher food expenditure shares at supermarkets and supercenters and lower food expenditure shares at drug stores and convenience stores. In addition, increased retail food industry concentration had a negative effect on shopping healthfulness.

III.2 Objective

Consumers/households currently face a number of choices beyond the traditional supermarket owing to the increasingly diverse U.S. retail food landscape. Despite the plethora of previous studies that largely focus on factors affecting store choice, one area of research that has received relatively little attention is how the magnitude of ***household food expenditures*** is impacted by store formats and store characteristics. In this light, the sole purpose of this study is to examine how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type and by income level. Whether traditional or nontraditional, store outlets differ in prices, product assortment, advertising strategies, and location (Volpe, Kuhns, and Jaenicke, 2017). The outlets considered in this study are grocery, convenience, discount, club, drug, and dollar store types. The source of data for this analysis is the Nielsen Homescan Panel over the period between 2011 and 2015. Specifically, we use a balanced panel of 28,109 households who participated in the survey for all five years from 2011 to 2015. The total number of observations available for analysis is 140,545. Through relationships with NielsenIQ and Nielsen, the Kilts Center for Marketing at the University of Chicago Booth School of Business provides this data set to academic researchers for a subscription fee (https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen).

As mentioned previously, prior works mainly highlighted store choice. To differentiate our study from the extant literature, we explore the factors which directly affect household food expenditure by store outlet. Indeed, Volpe, Jaenicke, and Chenarides (2018) estimated the impacts of *expenditure share* by store format, but in our study, we quantify the magnitude of the impact of socio-demographics, the retail food environment, and spending habits on household food and beverage expenditures by diverse store types. Hence, by analyzing factors that impact household food expenditure across the aforementioned six-store types, this study contributes to the economic literature. Another contribution is that our study also considers habitual persistence or spending habits, a dynamic property of household food expenditure. However, in the previously mentioned studies, habitual behavior was not included in the set of explanatory variables.

To further differentiate our study from previous studies, we employ a dynamic correlated random effect Tobit model to incorporate habitual purchasing behavior. As mentioned previously, we construct a panel data set with households as cross-sections over five annual periods, 2011 to 2015. The panel structure allows us to incorporate dynamic modeling by including lagged dependent variables as explanatory variables to account for spending habits. Another advantage of the use of this model is that we are in a position to handle corner solution problems. The dependent variables, which reflect household purchasing history according to store type, have zero values and hence are left censored.

We estimate separate dynamic correlated random effect Tobit models for sub-samples in accordance with household income level (low, middle, and high). Because households typically have different shopping baskets by income level (Taylor and Villas-Boas 2016; and Volpe, Jaenicke, and Chenarides 2018), we can compare and contrast our findings across income levels for each of the six-store types considered in our study.

III.3 Organization

This work is organized as follows. Initially, we provide definitions of the respective store outlets. Subsequently, we provide the theoretical framework and the empirical model for this study. Then we describe the Nielsen Homescan data, the construction of the balanced panel of households, and present descriptive statistics of model variables. Issues associated with the estimation of the dynamic correlated random effect Tobit model are discussed next. Following this discussion, the empirical results are presented. Finally, concluding remarks are made along with a discussion of study limitations and possibilities for further research.

III.4 Definitions of Store Types

While there are no universally accepted definitions and classifications of food retail store formats, throughout this study we use the store format names provided by Nielsen, the vendor responsible for the collection of the Homescan data. A traditional supermarket is a food retailer with greater than 9,000 square feet of selling space and at least $2 million in annual sales. Drug stores feature prescription-based pharmacies but generate at least 20 percent of their total sales from other categories, including general merchandise and food. Discount stores are mass merchandisers and typically large department stores (e.g. Target) that sell primarily general merchandise and nonperishables but also carry limited assortments of grocery products. Supercenters also have been known as hypermarkets and superstores are the largest formats, in terms of both square footage and product volume. Supercenters are hybrid stores that combine mass merchandisers with full supermarkets. These stores have a reputation among consumers for stressing low prices and convenience over consumer service (Carpenter and Moore, 2006). Convenience stores are the smallest of the major retail formats in terms of size and product offerings and feature a limited selection of staple foods as well as ready-to-eat, prepared foods (e.g., hot dogs). Additionally, convenience stores sell general merchandise and, in many locations, alcohol, and tobacco. Dollar stores range in size and product variety, placing emphasis on low prices and offering little in the way of customer service. As the name suggests, many products in these stores cost one dollar. Club stores, also referred to as warehouse or volume stores, are large-format outlets that specialize in selling food and selected general merchandise. The grocery line features foods and beverages in bulk for relatively low prices. A feature of this format unique in food retailing is that memberships must be paid in order to shop there.

III.5 Theoretical Framework

On the basis of household production theory, the expenditure function for any commodity is the product of derived demand for factor inputs and the corresponding price vector of factor inputs (Yen, 1993; Bryne, Capps, and Saha, 1996; and Nayga, 1998). Let the commodity in question be all food and beverages purchased by household h at store outlet k. Then as given by equation (1), household expenditure at store k may be written as

$$EX_{hk} = P_{hk}X_{hk} = g\left(P_{hk,}Y_h, W_h, D_h, E_h\right) \tag{1}$$

where X_i is the derived demand of factor inputs for household h at store outlet k, P_{hk} is the price vector of factor inputs paid by household h at store outlet k, W_h is a measure of the opportunity cost of time of household h, Y_i represents the income level of household h, D_h represents the set of socio-demographic characteristics of household h, and E_h represents the retail environment faced by household h.

Household heterogeneity typically is accounted for incorporating socio-demographic variables in the theoretical model. Hill and Lynchehaun (2002) identified various cultural and socio-economic factors influencing consumer preferences including age, ethnicity, income, education, gender, presence of children, marital status, region, and race. In particular, education reflects knowledge about health and nutrition (McCracken and Brandt, 1987; Nayga and Capps, 1992; Byrne, Capps, and Saha, 1996; Nayga, 1998; and Volpe, Kuhns, and Jaenicke 2017). Similar to Volpe, Janeicke, and Chenarides (2018), we include household income, household size, age, urbanization, race and ethnicity, region, and education in the set of socio-demographic variables in this study.

Additionally, in our theoretical model, we consider the potential importance of the retail environment in the household expenditure function. The retail environment represents the number of stores in the area in which the household lives; accessibility to store outlets may affect household production and consequently, household purchases of food and beverages. In this study, to address the impact of the retail environment, similar to Volpe, Jaenicke, and Chenarides (2018), we count the number of supermarkets and other grocery stores, convenience stores, drug stores, and warehouse club stores based on zip codes.

Past studies related to the choice of store outlet did not account for habitual purchasing behavior (Taylor and Villas-Boas, 2016; Fan, 2017; and Volpe, Jaenicke, and Chenarides, 2018). Habits refer to repetitive behavior in purchasing and consumption behavior (Ji and Wood, 2007). The habitual behavior of consumer purchasing patterns has been studied widely in the field of psychology (Bettman and Zins, 1977; Ehrenberg, 1988; and Ehrenberg 1991). This repeated purchasing behavior has been investigated in a wide range of products and services, including but not limited to potato chips, bread, tissue, laundry detergent, catsup, yogurt, sugar-sweetened beverages, and cigarettes (Deighton, Henderson, and Neslin, 1994; Motes and Woodside, 2001; Taylor, 2001; Khare and Inman, 2006; Zhen et al., 2011; Adamowicz and Swait, 2012; and Zhen et al. 2013).

To account for habitual purchasing behavior, we introduce a one-period lagged dependent variable in the model. (Mutlu and Garcia, 2006; and Rieger, Kuhlgatz, and Anders, 2016). As such, we augment the expenditure function given in equation (1) for household h for store outlet k in time period t as:

$$EX_{hkt} = h(EX_{hk,t-1}, P_{ht}, Y_{ht}, W_{ht}, D_{ht}, E_{ht}) \tag{2}$$

$where\ EX_{hk,t-1}$ is the one-period lagged expenditure variable for household h for each store type k in time period t.

III.6 Empirical Model

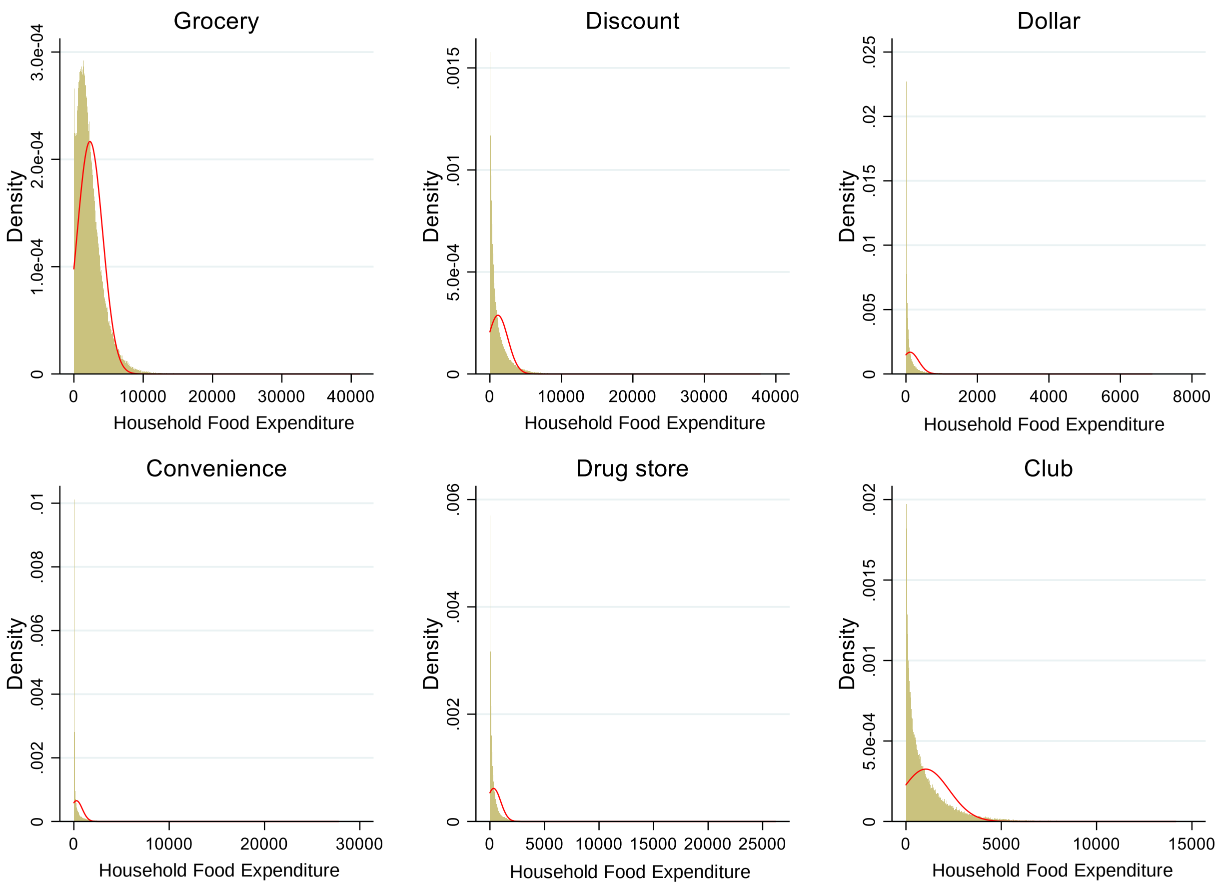
Given the focus of our research in analyzing the impacts of habitual spending behavior, the retail environment, and household heterogeneity by store outlets, we employ a dynamic correlated random effect Tobit model. This model specification allows us to deal with dynamics, panel data, and data censoring issues simultaneously accounting for household demographic variables and retail environment variables as explanatory variables. The model also accounts for potentially household-specific unobserved heterogeneity. As well, conventional fixed effect nonlinear models such as probit, logit, and Tobit models can produce biased estimates of structural parameters (Greene, 2004). The use of the dynamic correlated random effect Tobit model circumvents this deficiency and produces consistent estimates of structural parameters.

Owing to the number and heterogeneity of purchases of specific food items and beverages as well as the censored observations associated with household food and beverage expenditures by store outlets, we omit prices from the model. In the Nielsen data prices are derived as the ratio of expenditures to quantities purchased. By omitting prices from the model, we avoid making imputations of missing prices and we avoid the potential endogeneity of prices with household expenditures. Simply, we assume that the impact of the price is implicitly captured by the type of store outlet.

We transform the dependent variables which include zero-valued observations using the inverse hyperbolic sine (arcsinh) mechanism (Bellemare and Wichman, 2020). A notable problem with taking the logarithm of any variable is that it does not allow retaining zero-valued observations because ln(0) is undefined. As pointed out by Bellemare and Wichman (2020), "applied econometricians are typically loath to drop those observations for which the logarithm is undefined." Consequently, researchers often have resorted to ad hoc means of accounting for this situation when taking the natural logarithm of a variable, such as adding 1 to the variable prior to its transformation (MaCurdy and Pencavel, 1986). In recent years, the inverse hyperbolic sine (or arcsinh) transformation has grown in popularity among applied econometricians due to the fact that it is similar to the behavior of the logarithm function, it allows retaining zero-valued observations without any arbitrariness, and it often results in normal distributions (Burbidge et al., 1988; Yen and Jones, 1997; MacKinnon and Magee, 1990; Pence, 2006; Van den Heuvel et al., 2011; Bellemere, Barrett, and Just, 2013; Brown et al., 2015; and Bellemere and Wichman, 2020).

Figures III-3 and III-4 show the distributions of the original and transformed household food expenditure associated with each store type conditional on expenditures above zero (Horizontal axis indicates expenditure). In Figure III-3, only food expenditures from grocery stores follows a truncated normal distribution. In Figure III-4, after implementing the inverse hyperbolic transformation, all six dependent variables appear to follow normal distributions. In addition, the zero values of the dependent variables still remain as zeros with the inverse hyperbolic transformations. So, by transforming our dependent variables pertaining to household food expenditures, we can deal with corner solution issues in our data using the Tobit model.

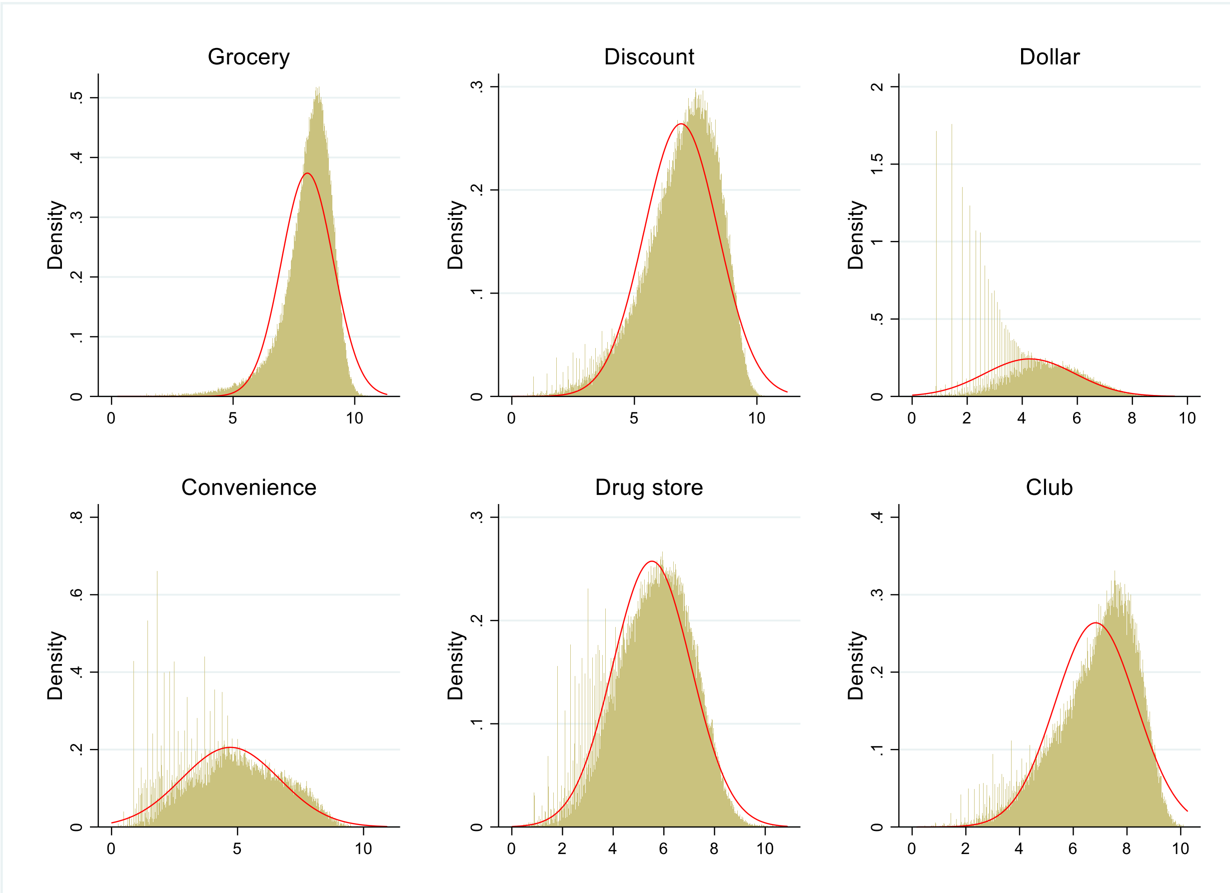
Figure III-3. Distributions of Household Food Expenditure Associated with the Six Store Types (2)



Using equation (3), we express the transformed household food and beverage expenditure variables by store type based on the inverse hyperbolic sine method. We denote $\overline{EX}_{ht}^{k}$ as the dependent variables for household h, for store outlet k, and for time period t in our empirical model.

$$\overline{EX}_{ht}^{k} = arcsinh\left(EX_{ht}^{k}\right) = ln(EX_{ht}^{k} + \sqrt{{EX_{ht}^{k}}^{2} + 1}) \qquad (3)$$

Figure III-4. Distributions of Transformed (inverse hyperbolic sine) Household Food Expenditure Associated with the Six Store Types (2)



Initially we start from the definition of our latent variables, that is, the expenditure variables denoted in equation (4). This latent variable property is maintained after transforming the dependent variables with the inverse hyperbolic sine method, equation (5). Zero observations reflect the decision by households to not make food and/or beverage purchases over the course of at least one calendar year in a particular store outlet.

$$
\begin{aligned}
EX_{ht}^{k\ *} &= EX_{ht}^{k} && if\ \ EX_{ht}^{k} > 0 \\
EX_{ht}^{k\ *} &= 0 && if\ \ EX_{ht}^{k} = 0
\end{aligned}
\qquad (4)
$$

$$\overline{EX}_{ht}^{k\,*} = \overline{EX}_{ht}^{k} \qquad if\ \overline{EX}_{ht}^{k} > 0$$
$$\overline{EX}_{ht}^{k\,*} = 0 \qquad if\ \overline{EX}_{ht}^{k} = 0 \tag{5}$$

Our empirical model for each of the respective six store types k at year t for household h is described in equations (6) through (8). We start from basic random effect model described in equation (6). $\overline{EX}_{ht}^{k\,*}$ is annual food and beverage expenditure of household h at year t for store type k, censored at zero. $\overline{EX}_{h,t-1}^{k\,*}$ is one year lagged dependent variable capture dynamic spending behavior of consumers, and c_h^k is a random effect term associated with the household's unobserved heterogeneity. ε_{ht}^k is the idiosyncratic error term.

Producing consistent estimators with lagged dependent variables in the random effect Tobit model, controlling for the initial condition, is critical. Wooldridge (2005) suggested a general and tractable approach to overcome the initial condition problem[6]. Following Chamberlain (1980), we assume that the unobserved heterogeneity term, c_h^k, has a distribution conditional on time-averaged

[6] The initial condition problem occurs when initial value of stochastic process is not observed. For example, consider following equation: $Y_{it} = \alpha + \beta Y_{it-1} + c_i + \varepsilon_{it}$. This equation contains a lagged dependent variable as a covariate. If we recursively rewrite this equation, then we can finally derive Y_{it} with Y_{i0} as a covariate. That said, defining Y_{i0} the initial value of the stochastic process is a difficult task. Wooldridge (2005) proposed a conditional maximum likelihood estimator that approximates the unobserved heterogeneity term, c_i, with the use of the initial observation, $y_{i,0}^m$, of the dataset and exogenous variables, d_i^m, to overcome the initial condition problem. Following Chamberlain (1980), we assume that the unobserved heterogeneity term has a distribution conditional on time-averaged continuous explanatory variables and the initial value of latent dependent variable. $y_{i,0}^m$ is the initial observation of the value of household expenditure on food and beverages in store outlet m in 2011, the initial calendar year of the data used in this analysis.

continuous explanatory variables[7] (d_h^k) and the value of the latent dependent variable in the initial time period (2011) of our sample ($\overline{EX}_{h,0}^{k\ *}$). Then, upon substitution of the unobserved heterogeneity term, equation (7), into the base random effect model, equation (6), we subsequently derive equation (8), the correlated random effect model. Because we assume u_h^k, the error term in the unobserved heterogeneity function (equation (7)), is random, equation (8) corresponds to a random effect model. The respective distributions of the error terms are given in equation (9).

z_{ht}^k is a vector of explanatory variables including the logarithm of household annual income, household size (number of household members), and number of club, convenience, drug, supercenter and grocery stores within the household's zip code area as well as indicator (dummy) variables related to age, education, race/ethnicity of the household head, urban/rural delineation, and region in which the household is located. In our correlated random effect model specification, we assume that u_h^k follows a standard normal distribution with variance $\sigma_u^{2,k}$. In equation (8), we can treat u_h^k as a random effect. This term corresponds to household unobserved heterogeneity. To estimate this random effect Tobit model, we employ the econometrics software package STATA version 15 using the command xttobit.

$$\overline{EX}_{ht}^{k\ *} = \alpha z_{ht}^k + \rho \overline{EX}_{h,t-1}^{k\ \ *} + c_h^k + \varepsilon_{ht}^k \qquad (6)$$

$$c_h^k = \theta \overline{EX}_{h,0}^{k\ *} + \vartheta d_h^k + u_h^k \qquad (7)$$

$$\overline{EX}_{ht}^{k\ *} = \alpha z_{ht}^k + \rho \overline{EX}_{h,t-1}^{k\ \ *} + \theta \overline{EX}_{h,0}^{k\ *} + \vartheta d_h^k + u_h^k + \varepsilon_{ht}^k \qquad (8)$$

---

[7] d_h^k is a vector of time-averaged continuous explanatory variables. In constructing the correlated random effect model, we only averaged continuous explanatory variables that are time-varying for each household. For example, for household income, we construct time-averaged variable by calculating $\overline{in}_h = \sum_{t=2011}^{2015} in_{ht}$, where in_{ht} is the annual income for household h in year t and t is time period 2011 to 2015.

$$with\ \varepsilon_{ht}^{k} \sim N\left(0, \sigma_{\varepsilon}^{2,k}\right), \qquad u_{h}^{k} \sim N\left(0, \sigma_{u}^{2,k}\right) \tag{9}$$

We address habitual spending behavior by adding a lagged dependent variable in the model. The coefficient of this variable, ρ, should be between 0 and 1. Statistical significance of this coefficient also confirms the existence of habitual spending behavior at certain store types. We jointly test significance of coefficients θ and ϑ using the Wald test to empirically check for household heterogeneity as denoted in equation (7). Likelihood ratio tests also are performed to compare the panel data model and the pooled data model. Another likelihood ratio test is designed to compare the correlated random effect model and the simple random effect model. [8]

III.7 Marginal Effects

Marginal effects refer to changes in the dependent variables (expenditures) attributed to unit changes in the continuous explanatory variables. For discrete explanatory variables, marginal effects refer to changes in expenditures relative to base or reference categories. The estimated parameters are not the marginal effects.

Equation (10) shows the conditional expectation of the original (untransformed) dependent variables. We present the details of the derivation in Appendix. $E\left[EX_{ht}^{k} \middle| z_{ht}^{k}, EX_{ht}^{k} > 0\right]$ is conditional expectation when household food expenditure is greater than zero. The function Φ is cumulative distribution function of the normal distribution with mean zero and variance $\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}$. EX_{ht}^{k} refers to the original dependent variables, untransformed annual household

---

[8] The pooled Tobit model is given by: $\overline{EX}_{ht}^{k\,*} = \alpha z_{ht}^{k} + \rho \overline{EX}_{h,t-1}^{k\,*} + \theta \overline{EX}_{h,0}^{k\,*} + \vartheta d_{h}^{k} + \upsilon_{ht}^{k}$ which ignores the panel data structure. As discussed previously, the correlated random effect model permits the correlation between explanatory variables and unobserved heterogeneity, but the simple random effect model does not.

expenditures on food and beverages by store format, k. z_{ht}^{k} is a vector of explanatory variables associated with continuous and binary variables as elements, and β_h^k refers to the estimated coefficients of the structural parameters.

$$E\left[EX_{ht}^{k}\middle|z_{ht}^{k}, EX_{ht}^{k} > 0\right] = \frac{1}{2}\exp\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}{2}\right)$$
$$* \left[\exp\left(z_{ht}^{k}\beta_h^k\right)\left\{\frac{\Phi\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k} + z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}{1 - \Phi\left(\frac{-z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}\right\}\right. \tag{10}$$
$$\left. - \exp\left(-z_{ht}^{k}\beta_h^k\right)\left\{\frac{1 - \Phi\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k} - z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}{\Phi\left(\frac{z_{ht}^{k}\beta_h^k}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}\right\}\right]$$

If the explanatory variables are continuous variables, we can take derivative with respect to these variables to obtain marginal effects. In equation (11), we provide the expression of the marginal effects for continuous variables. ϕ is probability density function with mean zero and variance $\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}$. The notations for equation (11) are similar to the notations in equation (10). In equation (12), we represent the marginal effects for binary explanatory variables. We calculate these marginal effects by taking the difference between conditional expectation when $z_{ht}^{k} = 1$ and$z_{ht}^{k} = 0$. Equation (11) is expressed as

$$\frac{\partial E\left[EX_{ht}^k \middle| X_{ht}^k, EX_{ht}^k > 0\right]}{\partial X_{ht}^k}$$

$$= \beta_h^k \frac{1}{2} \exp\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}{2}\right) \left[\exp\left(X_{ht}^k \beta_h^k\right) \left\{ \frac{\Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)} \right\} \right.$$

$$\left. - \frac{1}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}} \left\{ \frac{\phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right) - \Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right)^2} \right\} \right] \quad (11)$$

$if\ z_{ht}^k\ corresponds\ to\ countinuos\ variables$

$$\frac{\partial E\left[EX_{ht}^k \middle| z_{ht}^k, EX_{ht}^k > 0\right]}{\partial z_{ht}^k} = E\left[EX_{ht}^k \middle| z_{ht}^k = 1, E_{ht}^k > 0\right] - E\left[EX_{ht}^k \middle| z_{ht}^k = 0, EX_{ht}^k > 0\right] \quad (12)$$

$if\ z_{ht}^k\ corresponds\ to\ binary\ variables$

We employ a lagged dependent variable and the logarithm of household income in our explanatory variable set. Marginal effects for these variables need to be treated with care. As exhibited in Appendix, the marginal effect associated with the lagged dependent variables is given in equation (13).

$$\frac{\partial E\left[EX_{ht}^k \middle| X_{ht}^k, EX_{ht}^k > 0\right]}{\partial EX_{ht}^k} = \frac{\partial E\left[EX_{ht}^k \middle| X_{ht}^k, EX_{ht}^k > 0\right]}{\partial \overline{EX}_{h,t-1}^{k\ *}} * \frac{\partial \overline{EX}_{h,t-1}^{k\ *}}{\partial EX_{ht}^k} \quad (13)$$

To derive the marginal effect for income, we simply divide equation (11) by income.

Following McDonald and Moffitt (1980), for the Tobit model, the derivative of the unconditional expectation with respect to explanatory variables can be decomposed into two parts:

(1) the conditional marginal effect times the probability of non-zero household food expenditures at the various store outlets and (2) the conditional expectation times the change in probability of non-zero expenditures due to unit changes in the explanatory variables.

$$\frac{\partial E\left[EX_{ht}^{k\,*}\middle|z_{ht}^{k}\right]}{\partial z_{ht}^{k}} = \frac{\partial P\left[EX_{ht}^{k\,*} > 0\middle|z_{ht}^{k}\right]}{\partial z_{ht}^{k}} * E\left[EX_{ht}^{k\,*}\middle|z_{ht}^{k}, EX_{ht}^{k\,*} > 0\right] + P[EX_{ht}^{k\,*} > 0|z_{ht}^{k}] * \frac{\partial E[EX_{ht}^{k\,*}|z_{ht}^{k}, EX_{ht}^{k\,*} > 0]}{\partial z_{ht}^{k}} \tag{14}$$

We adopt this decomposition to explore the effects of explanatory variables on the probability of households to spend at various store formats, $\frac{\partial P(EX_{ht}^{k\,*}>0|z_{ht}^{k})}{\partial z_{ht}^{k}}$ in equation (14) as well as to explore the effects of explanatory variables on the magnitude of spending at particular store formats, $\frac{\partial E(EX_{ht}^{k\,*}|z_{ht}^{k},Y>0)}{\partial z_{ht}^{k}}$ in equation (14). Note that the effects on the magnitude of spending for various store formats is the same as the conditional expectation previously described in equation (10). With the assumption of a Tobit model, the probability is given by the linear combination of estimated coefficients associated with z_{ht} for store outlet k. The change in the probability is given the probability density function times the estimated coefficients divided by the estimate of the variance of the normal distribution.

We calculate the marginal effects associated with equations (11), (12), (13) and (14) with the use of the software package STATA15. Standard errors of the marginal effects are obtained using the delta method, as these are nonlinear combination of coefficients and the data (Bellemare and Wichman, 2020).

III.8 Data

The source of the data for this study is the Nielsen Homescan Panel covering the period between 2011 and 2015, the most recent data available to us at the time of this analysis. Volpe, Jaenicke, and Chenarides (2018) also use the Nielsen Homescan Panel for the period between 2004 and 2010. As such, we extend the time period of coverage concerning expenditures made by U.S. households at various store outlets. Hence, our analysis not only allows us to check on robustness of findings from the literature but also serves as a reference for future studies using more recent data.

Nielsen collects weekly surveys from more than 60,000 panelists every year in the entire United States. The Homescan data contain detailed information about quantities purchased and corresponding expenditures made by household by Universal Product Code (UPC) and by store type, Also, the Nielsen Homescan data incorporate a plethora of socio-demographic variables to account for household characteristics. We center attention on *annual* expenditures made by households for all food and beverage items. Finally, we consider those households that participate in each of the five years over the period between 2011 and 2015. In our balanced panel, 28,109 households participate in the survey for the five-year period from 2011 to 2015. To make comparisons by income level, we focus on three levels of household income, low, middle, and high. The low-income sample corresponds to those households whose annual income below is $25,000. The middle-income sample corresponds to those households whose annual income is above $25,000 but below $70,000. The high-income sample corresponds to those households whose annual income is above $70,000. We follow this segmentation of household income based on the work by Allcott, Diamond, and Dubé (2017).

In Table III-1, we report the unconditional and conditional[9] means and standard deviations of household food and beverage expenditure expressed in dollars by store type and by income level. Spending in grocery stores, club stores, and drug stores are positively associated with household income. But the reverse is the case for dollar stores. Household spending for food and beverages in convenience and discount stores is similar across the respective income levels.

Not all households purchase food and beverages at all store outlets even over a calendar year. Therefore, zero values are evident in household expenditures for food and beverages across the respective store types. As such, household expenditures are left censored at zero. The number of zero observations and the degree of censoring of household expenditures on food and beverages are exhibited in Table III-2. The degree of censoring is defined as the number of zero observations times 100 divided by the number of observations. The total number of observations in the low-income sample is 19,365; the total number in the mid-income sample is 66,540; and the total number in the high-income sample is 54,640.

[9] The term conditional corresponds to only those expenditure values above zero. On the other hand, the term unconditional refers to zero values as well as non-zero expenditure values.

Table III-1. Unconditional and Conditional Means and Standard Deviations of Household Food and Beverage Expenditure by Store Type and Income, Nielsen Panel Data 2011 to 2015[a]

Store Type	Unconditional			Conditional		
	Low-Income sample	Mid-Income sample	High-Income sample	Low-Income sample	Mid-Income sample	High-Income sample
Club	197	458	855	550	849	1,258
	(545)	(861)	(1,235)	(800)	(1,022)	(1,335)
Convenience	69	88	77	222	278	266
	(268)	(378)	(328)	(446)	(634)	(554)
Dollar	129	84	43	168	120	76
	(260)	(198)	(138)	(286)	(228)	(171)
Grocery	1,600	2,093	2,478	1,624	2,112	2,517
	(1,431)	(1,671)	(1,881)	(1,428)	(1,666)	(1,888)
Drug	216	260	269	283	325	332
	(589)	(558)	(537)	(662)	(607)	(575)
Discount	894	1,069	967	965	1,131	1,024
	(1,188)	(1,337)	(1,299)	(1,207)	(1,349)	(1,308)
Across All Store Outlets	3,108	4,043	4,679	3,813	4,815	5,474

[a] All values are expressed in terms of dollars, and standard deviations are in parentheses.

Table III-2. Number of Zero Values and Degree of Censoring for Household Food and Beverage Expenditures by Store Type and Income, Nielsen Panel Data 2011 to 2015

Store Type	Low-income Sample	Mid-income Sample	High-income Sample
Club	12,455	30.722	16,971
	(64.3%)	(46.2%)	(31.1%)
Convenience	13,397	45,579	39,263
	(69.2%)	(68.5%)	(71.9%)
Dollar	4,546	20,260	24,746
	(23.5%)	(30.5%)	(45.3%)
Grocery	290	633	363
	(1.5%)	(1.0%)	(0.7%)
Drug	4,640	13.241	10,186
	(24.0%)	(19.9%)	(18.7%)
Discount	1,435	3,680	3,687
	(7.4%)	(5.5%)	(6.8%)
Total number of observations	19,365	66,540	54,640

[a] Number of zero observations associated with household food and beverage expenditures.
[b] Degree of censoring expressed as a percent.
(Degree of censoring = number of zero observations*100/total number of observations)

The magnitude of the censoring varies by household income. The degree of censoring is greatest for convenience stores at roughly 70 percent by household income level. The censoring degree is lowest in grocery stores at approximately 1 percent by household income level. In discount stores, the degree of censoring is on the order of 5 to 8 percent; in drug stores, the degree of censoring is on the order of 18 to 24 percent. In dollar stores, the magnitude of censoring ranges from 23 to 45 percent. Finally, for club stores, the degree of censoring ranges from 31 to 64 percent. More variation associated with the degree of censoring by income level is evident for drug stores, dollar stores, and club stores.

Similar to Kyureghian and Nayga (2013), we use store density data to account for the retail environment. However, we examine store density by zip code, which represents smaller residential areas rather than by county level as was done by Kyureghian and Nayga (2013). These variables were obtained from Business Pattern Data (BPD hereafter) produced by the U.S. Census Bureau. The BPD contain data represent the number of stores categorized by North American Industry Classification System (NAICS). From these data, we obtained counts concerning four types of store outlets (grocery stores and supercenters, NAICS code 445110; warehouse club stores, NACIS code 452910; convenience stores, NAICS code 445120; and drug stores, NAICS code 446110). The data from the Nielsen Homescan Panel are available by zip code; consequently, we are able to augment the Nielsen Homescan Panel with the respective counts of store outlets from BPD. A shortcoming in this augmentation process is that the classifications of store formats from the Nielsen Homescan Panel and Business Pattern Data are different. Nevertheless, we provide a viable proxy for the retail environment based on counts of store outlets from BPD.

In order to identify differences in food and beverage expenditures by store outlets between households who live in urban and rural areas, we form urban and rural indicator (dummy) variables. Our dummy variables correspond to the six-category urban and rural classification scheme developed by the National Center for Health Statistics (NCHS). In our study, we form three dummy variables that represent category 1 through category 6 of the NCHS classification scheme. The dummy variable URBAN corresponds to NCHS urban and rural classification categories 1 and 2, the most densely populated areas, typically metropolitan areas. The dummy variable RURAL corresponds to NCHS urban and rural classification categories 5 and 6, rural areas with the least dense population. The dummy variable not classified as urban or rural corresponds to NCHS urban and rural classification categories 3 and 4. Then, we aligned these indicator variables with our Nielsen Homescan data based on zip code. The use of the NCHS classification scheme affords a richer consideration of the role of urban and rural areas in influencing household food and beverage expenditures by store outlet.

In Table III-3, the means and standard deviations of explanatory variables in the model are presented. The average number of grocery stores by zip code is slightly more than five for across the three income levels considered. The average number of drug stores by zip code in this analysis is nearly four across the board. Similarly, on average the number of convenience stores by zip code is between two and three for each of the respective samples. Finally, the average number of club stores by zip code is between zero and one, again across the board. In sum, with respect to the number of store outlets by zip code, the average number of club stores, convenience stores, supercenter and grocery stores, and drug stores does not vary much among the respective samples of household delineated by income in this analysis.

Table III-3. Means and Standard Deviations of Explanatory Variables in the Tobit Random Effect Model for the Various Samples by Income Category

Variable Name	Variable Description	Low-income sample	Mid-income Sample	High-income sample
	Continuous Variables			
Household income	Household income	$15,961 ($8,081)	$42,620 ($15,537)	$81,019 ($11,646)
Household size	Number of household members	1.58 (0.92)	2.07 (1.11)	2.51 (1.17)
Cu	NAICS code 452910, # of warehouse club stores by zip code	0.54 (0.79)	0.53 (0.80)	0.55 (0.82)
Cv	NAICS code 445120, # of convenience stores by zip code	2.48 (3.12)	2.32 (2.99)	2.35 (2.89)
Sg	NAICS code 445110, # of supercenters and grocery stores by zip code	5.56 (7.96)	5.21 (6.74)	5.60 (7.29)
Dr	NAICS code 446110, # of drug stores by zip code	3.79 (3.74)	3.69 (3.38)	3.97 (3.74)
	Degree of Urbanization			
Urban	NCHS urban and rural classification categories 1 and 2	0.44 (0.50)	0.51 (0.50)	0.63 (0.48)
Not urban* or rural	NCHS urban and rural classification categories 3 and 4)	0.32 (0.47)	0.32 (0.47)	0.28 (0.45)
Rural	NCHS urban and rural classification categories 5 and 6	0.24 (0.43)	0.17 (0.38)	0.10 (0.30)
	Age			
Age<40	Age of household head below 40	0.01 (0.10)	0.01 (0.12)	0.01 (0.12)
40<Age<60*	Age of household head above 40 and below 60	0.12 (0.32)	0.16 (0.36)	0.23 (0.42)
Age>60	Age of household head over 60	0.88 (0.33)	0.84 (0.37)	0.77 (0.42)
	Education			
Under high school	Household head education less than high school	0.04 (0.19)	0.01 (0.10)	0.00 (0.04)
Graduate high school	Household head is a high school graduate	0.35 (0.48)	0.22 (0.42)	0.07 (0.26)
College experience	Household head had some college education but is not a college graduate	0.34 (0.47)	0.33 (0.47)	0.20 (0.40)
Graduate* college	Household head is a college graduate	0.28 (0.45)	0.44 (0.47)	0.73 (0.44)

Table III-4. Continued

Variable Name	Variable Description	Low-income sample	Mid- income Sample	High-income sample
	Race and ethnicity			
Nonhisp-* White	Household head is non-Hispanic white	0.84 (0.36)	0.83 (0.37)	0.79 (0.41)
Nonhisp-Black	Household head is non-Hispanic black	0.09 (0.28)	0.09 (0.28)	0.09 (029)
Nonhisp-Asian	Household head is non-Hispanic Asia.	0.01 (0.10)	0.02 (0.13)	0.05 (0.21)
Nonhisp-Other	Household head is non-Hispanic other	0.03 (0.16)	0.02 (0.14)	0.02 (0.14)
Hisp	Household head is Hispanic	0.03 (0.18)	0.04 (0.20)	0.05 (0.22)
	Region			
Ne	Household located in the New England region (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont)	0.04 (0.21)	0.04 (0.20)	0.05 (0.22)
Ma	Household located in the Middle Atlantic region, (New Jersey, New York, and Pennsylvania)	0.12 (0.33)	0.12 (0.33)	0.14 (0.35)
Enc	Household located in East North Central region (Illinois, Indiana, Michigan, Ohio, and Wisconsin)	0.20 (0.40)	0.19 (0.39)	0.17 (0.38)
Wnc	Household located in the West North Central region (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)	0.10 (0.30)	0.10 (0.30)	0.08 (0.27)
Sa	Household located in the South Atlantic region (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, and West Virginia)	0.20 (0.40)	0.20 (0.40)	0.19 (0.40)
Esc	Household located in East South Central region (Alabama, Kentucky, Mississippi, and Tennessee)	0.06 (0.24)	0.06 (0.24)	0.05 (0.21)
Wsc	Household located in the West South Central region (Arkansas, Louisiana, Oklahoma, and Texas)	0.10 (0.30)	0.10 (0.30)	0.10 (0.30)
Mt	Household located in the Mountain region (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming)	0.08 (0.27)	0.07 (0.26)	0.07 (0.25)
Pac*	Household located in the Pacific region (Alaska, California, Hawaii, Oregon, and Washington)	0.10 (0.30)	0.11 (0.31)	0.15 (0.35)
Number of observations		19,365	66,540	54,640

Standard deviations are in parentheses.

Superscript * associated with the variable name indicates the base category or reference category.

Roughly 44 percent of households live in urban areas in the low-income sample, about 50 percent in the middle-income sample, and close to 63 percent in the high- income sample. On the other hand, the percentage of households living in rural areas decreases with increases in household income. Approximately 24 percent of households live in rural areas in the low-income sample, about 18 percent in the middle-income sample, and close to 10 percent in the high-income sample. The base category or reference category with respect to degree of urbanization is the ‘not urban or not rural’ category.

Additionally, in Table III-3 we present descriptive statistics concerning socio-demographic variables, namely household income, household size, age, education level, race/ethnicity of the household head, and region in which the household is located. Household income is reported by ranges in the Nielsen Homescan Panel. Similar to previous studies, we take the midpoint of each household income range as the income level of the household (Kyureghian and Nayga, 2013; Austin et al., 2017; and Senia, Dharmasena, and Capps, 2019). For the three income segments, the mean values of household income are \$15,961, \$42,620, and \$81,019 respectively. On average, household size rises from 1.58 to 2.51 members as income increases.

We employ three classifications of the age of the household head, less than 40, between 40 and 60, and over 60. The lowest (highest) proportion of households whose heads are over 60 is for the high-income (low-income) sample. This pattern is reversed in regard to households whose heads are between 40 and 60. Across the respective data samples, the number of households whose heads are less than 40 is around 1 percent. The data concerning age of the household head unequivocally are skewed toward older household heads. The base category is age of the household head between 40 and 60.

We consider four categories concerning the level of education of the household head—less than high school, high school graduate, some college experience, and college graduate. As presented in Table III-3, the level of education of household heads is positively associated with household income. Roughly 34 percent of households whose heads have a college degree are in the low-income sample, about 44 percent are in the middle-income sample, and slightly more than 73 percent are in the high-income sample. Further, the proportion of households wherein the highest level of education is a high school degree is 35 percent for the low-income sample, about 22 percent for the middle-income sample, and slightly more than 7 percent for in the high-income sample. Very few household heads in the respective data samples have less than a high-school education. The vast majority of household heads in the respective data samples have at least some college-level educations. The base category of education level corresponds to household heads with a college degree.

We employ five joint classifications of the race and ethnicity of the household head—non-Hispanic white, non-Hispanic black, non-Hispanic Asian, non-Hispanic other, and Hispanic. The proportion of non-Hispanic households decreases from 84 percent to 79 percent with increases in household income. But the proportion on non-Hispanic Asian households rises from 1 percent to 5 percent across the respective income samples. As well, the proportion of Hispanic households increases from 3 percent to 5 percent across the respective income samples. The base category of race/ethnicity is non-Hispanic white households.

We rely on nine categories concerning the region in which the household is located: (1) New England ((Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont); (2_ Middle Atlantic (New Jersey, New York, and Pennsylvania); (3) East North Central (Illinois, Indiana, Michigan, Ohio, and Wisconsin); (4) West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota**);** (5) South Atlantic (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, and West Virginia); (6) East South Central (Alabama, Kentucky, Mississippi, and Tennessee); (7) West South Central (Arkansas, Louisiana, Oklahoma, and Texas); (8) Mountain (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming); and (9) Pacific (Alaska, California, Hawaii, Oregon, and Washington). This delineation affords more detail concerning the impact of region on household food and beverage expenditures by store outlet. Across the board, roughly 20 percent of the households reside in the South Atlantic region, 18 percent reside in the East North Central region, 13 percent in the Middle Atlantic region, 12 percent in the Pacific region, 10 percent in the West South-Central region, 9 percent in the West North Central region, 7 percent in the Mountain region, 6 percent in the East South-Central region, and 5 percent in the New England region. The reference category is the Pacific region.

III.9 Potential Endogeneity of the Retail Environment Variables

There is a debate in the literature as to whether or not the retail environment variables are endogenous. This issue is important due to the fact that the endogeneity of explanatory variables leads to inconsistent parameter estimates. The endogeneity issue was addressed in several studies (Dunn et al., 2012; Kyureghian and Nayga, 2013; Ver Pleog et al., 2015; Handbury, Rahkovsky, and Schnell, 2016; and Allcott, Diamond, and Dubé, 2017). On the other hand, Currie et al. (2010) and Taylor and Villas-Boas (2016) argued that the endogeneity of retail environment variables does not lead to bias or inconsistency of parameter estimates.

Although previous works recognized potential endogeneity issues regarding retail environment variables, those studies just assumed the presence of endogeneity and estimated models with instrument variables. In this study, we formally test whether or not the retail environment variables suffer from the endogeneity problem. In order to account for the retail environment, we use variables that represent the number of stores, a metric of store density, by store type in the zip code area in which the household is located. As mentioned previously, we only incorporate four categories of store types, namely supercenters and grocery stores, club stores, drug stores, and convenience stores from the BPD data. Then, we test exogeneity of those four variables in each of the equations pertaining to the six store types. To carry out the Hausman test of endogeneity, we initially estimate each retail environment variable as a function of the remaining explanatory variables. Consistent with Hausman (1978) we incorporate the residuals from the first-stage estimation results in the full model. Subsequently, we test the null hypothesis that the coefficients associated with these residuals in the respective equations are all equal to zero; this null hypothesis is tantamount the exogeneity of the respective retail environment variables. A rejection of this null hypothesis then is statistical evidence that the set of retail environment variables are endogenous.

Table III-5. Results of the Hausman Endogeneity Chi-Squared Tests Associated with the Retail Environment Variables[a] (Income level)

Model	Club	Convenience	Dollar	Discount	Grocery	Drug
Chi-squared statistic	1.60	8.12	0.50	0.34	3.45	4.41
p-value	0.80	0.09	0.97	0.99	0.49	0.35

[a] Chi-squared tests each with four degrees-of-freedom.

Results based on the Hausman test in Table III-4 indicate the lack of evidence of endogeneity of the retail environment variables. These results are consistent with the assumption of the lack of endogeneity of retail environment variables made by Currie et al. (2010) and Taylor and Villas-Boas (2016). However, unlike previous studies, we provide statistical evidence to support the claim that the set of retail environment variables indeed are exogenous.

III.10 Empirical Results

Maximum likelihood estimates of the respective parameters and standard errors in the various models are obtained with the use of the software package STATA Version 15. In Table III-5-7, we provide the parameter estimates, associated p-values, likelihood ratio and Wald tests, and goodness-of-fit metrics for the dynamic correlated random effect Tobit model for the respective samples delineated by the three income categories. Additionally, in Tables III-8-10, we provide the marginal effects for data samples for the respective three income levels. In this study, we adopt a level of significance of 0.01 because of the sizable number of observations.

The parameter estimates associated with the standard deviation of the random effect term, σ_u^k, are statistically significant for all store types and for all income levels. As such, household unobserved heterogeneity plays a decisive role in food purchasing behavior. On the basis of likelihood ratio tests, the correlated random effect Tobit model is superior to the pooled Tobit model as well as the random effect Tobit model. The Wald test is the analogue of the conventional F-statistic in regression analysis. For each of the respective 24 models, the Wald tests supports the contention that at least one estimated coefficient is statistically different from zero. Alternatively, the Wald tests support the hypothesis that each model explains a significant amount of variation in household food and beverage expenditures across all store outlets and across all income levels.

We report two different goodness-of-fit metrics to determine the degree of explanatory power associated with each of the respective correlated random effect Tobit models. The first measure, labeled as pseudo R^2, is the square of the correlation of the ***unconditional*** expected value and the actual value of household food and beverage expenditures. Alternatively, we use the computation method to calculate the goodness-of-fit metric proposed by Veall and Zimmermann (1996) (V-Z hereafter). The V-Z Pseudo R^2 statistic is the square of the correlation of the ***conditional*** expected value and the actual value of household food and beverage expenditures. As shown in Table III-5, the R-squared statistic ranges from 0.542 (model for convenience stores) to 0.778 (model for grocery stores). This range of the pseudo R^2 and the V-Z Pseudo R^2 statistics are evident in sub-sample estimation by income level across store outlets. On the basis of these goodness-of-fit measures, the correlated random effect Tobit models explain a notable amount of the variability in household food and beverage expenditures for each store type and for each income level.

We organize the ensuing discussion of the massive set of empirical results as follows. We focus on the low-income sample, the mid-income sample, and the high-income sample of panel households. Within each of these aforementioned four samples, we initially discuss the statistically significant drivers. Subsequently, we present the impacts of household income, household size, age, urbanization, education, race and ethnicity, the number of club stores, convenience stores, grocery stores, and drug stores, and region, centering attention on the conditional marginal effects.

III.10.1 Estimation Results: Low-Income Sample

As discussed previously, the low-income sample corresponds to those households whose annual income below is $25,000. As exhibited in Table III-5, the estimated coefficients associated with the lagged dependent variable are statistically significant across all store outlets, ranging from 0.422 (drug stores) to 0.721 (grocery stores). As such, for the low-income sample, this finding confirms the supposition of habitual spending across all store outlets.

Additionally, as exhibited in Table III-5, in the low-income sample, household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. However, household size is positively related to household expenditures made at discount stores. Relative to households in the 40-year-old to 60-year-old category, household expenditures made at discount stores are lower for households 60 years of age and older.

In the low-income sample, for households located in urban areas, household food and beverage expenditures are lower at discount stores and convenience stores relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are lower at club stores relative to households located outside of urban and rural areas.

Relative to households who have graduated from college, households with a high school education spend more on food and beverages at discount stores and dollar stores but less at convenience stores and drug stores. In the low-income sample, few differences in expenditures made by households for food and beverages across store outlets are evident concerning the education level of the household.

Similarly, in the low-income sample, few differences are evident in expenditures for food and beverages by race and ethnicity. Exceptions include the following. Relative to non-Hispanic white households, non-Hispanic black households spend more on food and beverages at club stores and dollar stores. Relative to non-Hispanic white households, Hispanic households spend more on food and beverages at discount stores and dollar stores.

Further, the number of club stores, the number of convenience stores, the number of grocery store and supercenters, and the number of drug stores within the residence of households do not significantly affect food and beverage expenditures made at any of the six store outlets.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are higher at drug stores only. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Middle Atlantic region and in the West North Central region are higher at convenience stores only. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region and in the South Atlantic region are higher at discount stores only. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East North Central region, the West South Central region, and in the Mountain region are not statistically different across the respective store outlets. In the low-income sample, region is a statistically significant determinant of household food and beverage expenditures in four regions relative to the Pacific region. No differences in household expenditures are evident at grocery stores, club stores, and dollar stores across the respective regions in the low-income sample.

Table III-6. Maximum Likelihood Parameter Estimates and the Associated p-values for the Explanatory Variables Based on the Low-Income Sample of Panel Households

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
Lagged dependent variable	0.721*	0.457*	0.530*	0.443*	0.603*	0.422*
	(0.018)	(0.012)	(0.021)	(0.013)	(0.025)	(0.014)
Income	-0.012	0.014	-0.007	-0.002	0.101	0.029
	(0.013)	(0.023)	(0.067)	(0.027)	(0.078)	(0.035)
Household size	0.016	0.040*	0.090	0.009	-0.016	-0.008
	(0.007)	(0.014)	(0.048)	(0.018)	(0.048)	(0.022)
Age <40	0.084	-0.185	-0.439	-0.163	0.162	0.214
	(0.058)	(0.118)	(0.436)	(0.150)	(0.387)	(0.186)
Age >60	0.041	-0.141*	0.359	-0.012	-0.231	0.068
	(0.019)	(0.040)	(0.142)	(0.052)	(0.136)	(0.064)
Urban	-0.001	-0.162*	0.149	-0.005	-0.351*	0.076
	(0.015)	(0.034)	(0.122)	(0.046)	(0.121)	(0.054)
Rural	-0.031	0.063	-0.693*	0.030	-0.062	-0.156
	(0.017)	(0.038)	(0.148)	(0.052)	(0.136)	(0.062)
Under high school	-0.007	0.119	-0.147	0.102	-0.163	-0.281
	(0.033)	(0.069)	(0.255)	(0.088)	(0.245)	(0.111)
Graduate high school	-0.031	0.087*	0.039	0.123*	-0.276	-0.181*
	(0.015)	(0.033)	(0.119)	(0.044)	(0.118)	(0.053)
College experienced	-0.009	0.018	0.057	0.098	-0.012	0.006
	(0.015)	(0.031)	(0.108)	(0.040)	(0.108)	(0.049)
Non-Hispanic black	0.020	0.004	0.866*	0.178*	0.365	0.135
	(0.023)	(0.051)	(0.183)	(0.069)	(0.182)	(0.082)
Non-Hispanic Asian	0.043	-0.066	-0.122	-0.258	-0.474	-0.278
	(0.061)	(0.138)	(0.479)	(0.189)	(0.526)	(0.219)
Non-Hispanic other	-0.049	0.070	0.550	0.101	0.466	0.135
	(0.037)	(0.080)	(0.276)	(0.105)	(0.273)	(0.129)
Hispanic	-0.014	0.200*	0.451	0.332*	0.361	0.247
	(0.034)	(0.075)	(0.265)	(0.099)	(0.262)	(0.120)
Number of Club stores	-0.028	0.008	0.201	-0.019	0.121	-0.048
	(0.024)	(0.041)	(0.122)	(0.048)	(0.136)	(0.063)
Number of Convenience stores	-0.000	0.006	0.004	0.000	-0.016	0.020
	(0.005)	(0.009)	(0.027)	(0.010)	(0.030)	(0.013)
Number of Grocery stores and Supercenters	0.000	-0.005	0.012	0.010	-0.037	-0.004
	(0.004)	(0.007)	(0.023)	(0.008)	(0.024)	(0.010)
Number of Drug stores	-0.000	0.009	0.023	0.008	-0.016	0.017
	(0.006)	(0.011)	(0.036)	(0.013)	(0.038)	(0.017)
New England	-0.037	0.022	-0.062	0.013	0.190	0.336*
	(0.035)	(0.080)	(0.296)	(0.111)	(0.292)	(0.129)

Table III-5. Continued

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
Middle Atlantic	0.013	0.056	-0.403	0.009	0.791*	0.118
	(0.026)	(0.060)	(0.220)	(0.083)	(0.218)	(0.097)
East North Central	0.002	0.015	-0.274	-0.002	0.221	0.068
	(0.024)	(0.055)	(0.199)	(0.075)	(0.202)	(0.089)
West North Central	-0.036	0.154	-0.447	-0.042	1.134*	0.106
	(0.028)	(0.064)	(0.237)	(0.088)	(0.230)	(0.104)
South Atlantic	-0.036	0.160*	-0.477	0.097	0.282	0.135
	(0.024)	(0.055)	(0.198)	(0.076)	(0.202)	(0.089)
East South Central	-0.012	0.233*	-0.541	0.121	-0.135	0.194
	(0.033)	(0.073)	(0.276)	(0.100)	(0.272)	(0.119)
West South Central	-0.042	0.147	-0.397	0.012	0.152	0.009
	(0.028)	(0.063)	(0.230)	(0.087)	(0.233)	(0.103)
Mountain	-0.025	0.157	-0.127	-0.107	0.338	-0.018
	(0.029)	(0.066)	(0.233)	(0.091)	(0.242)	(0.108)
Constant	0.622*	0.855*	-6.554*	-0.540	-4.017*	-0.639
	(0.129)	(0.284)	(1.069)	(0.387)	(1.019)	(0.459)
Initial value	0.178*	0.420*	0.909*	0.550*	0.797*	0.524*
	(0.016)	(0.012)	(0.027)	(0.014)	(0.029)	(0.015)
σ_u	0.210*	0.670*	2.462*	0.987*	2.275*	1.112*
	(0.023)	(0.020)	(0.070)	(0.024)	(0.072)	(0.031)
σ_e	0.702*	1.196*	2.776*	1.325*	2.983*	1.779*
	(0.006)	(0.008)	(0.030)	(0.010)	(0.035)	(0.013)
LR χ^2 test of the Correlated Random Effect Tobit Model vs the Pooled Tobit Model	17.9*	336.6*	666.3*	709.9*	502.9*	460.4*
LR χ^2 test of the Correlated Random Effect Tobit Model vs the Random Effect Tobit Model	279.6*	1,132.6*	1,174.2*	1,430.2*	877.4*	1,279.8*
Wald χ^2 test	137.0*	1,194.5*	1,153.4*	1,611.2*	749.5*	1,306.9*
Pseudo R^2	0.298	0.233	0.225	0.234	0.215	0.217
V-Z R^2	0.749	0.697	0.672	0.717	0.548	0.601
Number of Observations	19,365	19,365	19,365	19,365	19,365	19,365
Number of Households	3,873	3,873	3,873	3,873	3,873	3,873

*p<0.01

Numbers in parentheses correspond to standard errors.

III.10.2 Estimation Results: Mid-Income Sample

As discussed earlier, the middle-income sample corresponds to those households whose annual income is above $25,000 but below $70,000. As exhibited in Table III-6, the estimated coefficients associated with the lagged dependent variable are statistically significant in all store type models, ranging from 0.395 (dollar stores) to 0.671 (grocery stores). As such, these results confirm the supposition of habitual spending on food and beverages across all store outlets.

As exhibited in Table III-6, household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. Household size is positively related to household expenditures made at discount stores, club stores, and dollar stores.

Relative to households in the 40-year old to 60-year old category, household expenditures made at drug stores are higher for households 60 years of age and older. No other differences by age are evident for food and beverage expenditures across the respective store outlets.

For households located in urban areas, household food and beverage expenditures are higher in grocery stores, but lower in discount stores, dollar stores, and convenience stores relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are higher in discount stores but lower at grocery stores, club stores, and drug stores relative to households located outside of urban and rural areas.

Relative to households who have graduated from college, households with less than a high school education expend more at dollar stores. Relative to households who have graduated from college, households with a high school education spend more on food and beverages at discount stores and dollar stores but less at grocery stores. Relative to households who have graduated from college, households with some level of college experience expend more at discount stores and dollar stores.

Relative to non-Hispanic white households, non-Hispanic black households spend more on food and beverages at discount stores, club stores, dollar stores, and convenience stores but less at grocery stores. No differences are evident between Asian households and non-Hispanic white households concerning the level of food and beverage expenditures across the six store outlets. Relative to non-Hispanic white households, non-Hispanic non-black and non-Asian households spend more on food and beverages at convenience stores. Finally, relative to non-Hispanic white households, Hispanic households spend more on food and beverages at discount stores and dollar stores in the mid-income sample.

The number of club stores within the residence of households negatively impacts food and beverage expenditures made at grocery stores and drug stores but positively affects food and beverage expenditures made at club stores. The number of grocery stores and supercenters within the residence of households negatively impacts expenditures made at discount stores. The number of convenience stores and the number of drug stores within the residence of households are not statistically significant factors affecting food and beverage expenditures made at any of the six store outlets.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Middle Atlantic region are higher at dollar stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East North Central region are higher at convenience stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West North Central region are higher at discount stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the South Atlantic region are higher at discount stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region are higher at dollar stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West South Central region are higher at discount stores and convenience stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Mountain region are higher at convenience stores. No differences in household spending on food and beverages are evident for grocery stores and drug stores with respect to region. But region plays a statistically significant role in affecting household food and beverage expenditures in discount stores, club stores, dollar stores and convenience stores.

Table III-7. Maximum Likelihood Parameter Estimates and Associated p-values for the Explanatory Variables Based on the Mid-Income Sample of Panel Households

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
Lagged dependent variable	0.671*	0.454*	0.519*	0.395*	0.562*	0.416*
	(0.008)	(0.007)	(0.009)	(0.007)	(0.013)	(0.007)
Income	0.007	0.027	0.078	-0.058	0.047	-0.027
	(0.009)	(0.016)	(0.040)	(0.025)	(0.064)	(0.026)
Household size	0.002	0.029*	0.053*	0.041*	0.011	-0.016
	(0.003)	(0.006)	(0.016)	(0.010)	(0.023)	(0.010)
Age <40	0.014	0.014	-0.161	-0.063	0.369	-0.022
	(0.023)	(0.047)	(0.134)	(0.078)	(0.183)	(0.079)
Age >60	0.028*	-0.017	0.020	0.021	0.069	0.118*
	(0.008)	(0.018)	(0.050)	(0.029)	(0.070)	(0.029)
Urban	0.022*	-0.092*	-0.022	-0.094*	-0.355*	0.058
	(0.007)	(0.016)	(0.047)	(0.027)	(0.064)	(0.026)
Rural	-0.028*	0.072*	-0.456*	0.080	-0.105	-0.158*
	(0.009)	(0.020)	(0.063)	(0.035)	(0.082)	(0.034)
Under high school	-0.046	0.096	0.158	0.259*	-0.189	-0.186
	(0.028)	(0.059)	(0.167)	(0.094)	(0.242)	(0.097)
Graduate high school	-0.032*	0.071*	0.015	0.172*	0.023	-0.033
	(0.008)	(0.017)	(0.048)	(0.028)	(0.067)	(0.028)
College experienced	-0.015	0.045*	0.004	0.109*	0.089	0.004
	(0.007)	(0.014)	(0.039)	(0.023)	(0.055)	(0.023)
Non-Hispanic black	-0.047*	0.107*	0.294*	0.325*	0.449*	0.083
	(0.011)	(0.024)	(0.073)	(0.042)	(0.099)	(0.041)
Non-Hispanic Asian	-0.027	0.053	0.312	0.055	-0.162	-0.020
	(0.023)	(0.050)	(0.143)	(0.087)	(0.221)	(0.084)
Non-Hispanic other	-0.032	0.058	0.063	0.168	0.561*	-0.045
	(0.020)	(0.042)	(0.116)	(0.068)	(0.164)	(0.069)
Hispanic	-0.005	0.140*	0.150	0.160*	0.128	0.088
	(0.015)	(0.033)	(0.095)	(0.056)	(0.138)	(0.055)
Number of Club stores	-0.033*	0.022	0.129*	-0.020	-0.105	-0.163*
	(0.011)	(0.019)	(0.047)	(0.029)	(0.075)	(0.031)
Number of Convenience stores	-0.000	-0.005	-0.018	0.003	0.018	-0.003
	(0.002)	(0.004)	(0.011)	(0.007)	(0.017)	(0.007)
Number of Grocery stores and Supercenters	0.004	-0.009	0.005	-0.000	0.005	0.001
	(0.002)	(0.004)	(0.009)	(0.005)	(0.014)	(0.006)
Number of Drug stores	0.005	-0.001	-0.004	0.003	0.019	0.001
	(0.003)	(0.005)	(0.013)	(0.008)	(0.021)	(0.009)
New England	0.043	-0.021	-0.411*	-0.051	0.389	0.137
	(0.017)	(0.039)	(0.117)	(0.069)	(0.166)	(0.065)

Table III-6. Continued

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
Middle Atlantic	0.028	0.016	-0.443*	0.160*	1.273*	0.050
	(0.013)	(0.029)	(0.086)	(0.050)	(0.122)	(0.048)
East North Central	0.025	-0.031	-0.186	0.061	0.795*	0.059
	(0.012)	(0.026)	(0.079)	(0.046)	(0.113)	(0.044)
West North Central	-0.028	0.125*	-0.306*	-0.100	1.401*	-0.105
	(0.014)	(0.031)	(0.093)	(0.054)	(0.129)	(0.052)
South Atlantic	0.006	0.096*	-0.276*	0.118	0.889*	0.059
	(0.012)	(0.026)	(0.077)	(0.046)	(0.113)	(0.044)
East South Central	0.002	0.090	-0.303*	0.191*	0.329	0.008
	(0.016)	(0.035)	(0.106)	(0.061)	(0.151)	(0.059)
West South Central	0.006	0.117*	-0.210	0.050	0.620*	-0.081
	(0.014)	(0.030)	(0.090)	(0.053)	(0.130)	(0.051)
Mountain	0.008	0.066	-0.021	0.023	0.920*	-0.085
	(0.014)	(0.032)	(0.094)	(0.057)	(0.136)	(0.054)
Constant	0.572*	0.187	-6.624*	0.284	-6.023*	-1.263*
	(0.119)	(0.265)	(0.812)	(0.465)	(1.098)	(0.445)
Initial value	0.221*	0.441*	0.721*	0.615*	0.818*	0.515*
	(0.007)	(0.007)	(0.010)	(0.008)	(0.016)	(0.008)
σ_u	0.223*	0.614*	1.941*	1.142*	2.413*	1.053*
	(0.008)	(0.010)	(0.027)	(0.014)	(0.039)	(0.015)
σ_e	0.593*	1.065*	2.241*	1.464*	2.986*	1.659*
	(0.002)	(0.004)	(0.010)	(0.006)	(0.019)	(0.006)
LR χ^2 test of the Correlated Random Effect Tobit Model vs the Pooled Tobit Model	193.7*	1176.9*	2455.9*	2883.2*	2062.6*	1640.9*
LR χ^2 test of the Correlated Random Effect Tobit Model vs the Random Effect Tobit Model	1,244.6*	3,970.8*	4,219.8*	5,411.5*	3,206.9*	4,479.3*
Wald χ^2 test	949.1*	4,280.8*	4,893.1*	5,744.9*	2,729.3*	4,677.2*
Pseudo R^2	0.321	0.237	0.229	0.223	0.214	0.216
V-Z R^2	0.774	0.709	0.726	0.677	0.549	0.608
Number of Observations	66,540	66,540	66,540	66,540	66,540	66,540
Number of Households	13,308	13,308	13,308	13,308	13,308	13,308

*p<0.01

Numbers in parentheses correspond to standard errors.

III.10.3 Estimation Results: High-Income Sample

As previously mentioned, the high-income sample corresponds to those households whose annual income is above $70,000. As exhibited in Table III-7, the coefficients associated with the lagged dependent variable are statistically significant in across all store outlet models, ranging from 0.349 (dollar stores) to 0.639 (grocery stores). As such, these results once again confirm the supposition of habitual spending across all store outlets. Indeed, habitual spending or habit persistence is a key factor in affecting nominal food and beverage expenditures across all store types.

In the high-income sample, as exhibited in Table III-7, household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. Household size is positively related to household expenditures made at discount stores, club stores, and dollar stores. Relative to households in the 40-year-old to 60-year-old category, household expenditures made at dollar stores and drug stores are higher for households 60 years of age and older. But relative to households in the 40-year-old to 60-year-old category, household expenditures for food and beverages made at discount stores are lower for households 60 years of age and older.

For households located in urban areas, household food and beverage expenditures are lower in discount stores, dollar stores, and convenience stores relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are higher in discount stores and dollar stores but lower in club stores relative to households located outside of urban and rural areas.

Relative to households who have graduated from college, households with less than a high school education expend but less at grocery stores. Relative to households who have graduated from college, households with a high school education spend more on food and beverages at discount stores and dollar stores but less at grocery stores. Relative to households who have graduated from college, households with some level of college experience expend more at dollar stores in the high-income sample.

Relative to non-Hispanic white households, non-Hispanic black households spend more on food and beverages at discount stores, club stores, dollar stores, and drug stores but less at grocery stores. Relative to non-Hispanic white households, Asian households spend less at grocery stores. Relative to non-Hispanic white households, expenditures made by non-Hispanic non-black and non-Asian households are statistically the same across the six store outlets in the high-income sample. Finally, relative to non-Hispanic white households, Hispanic households spend more on food and beverages at dollar stores.

The number of club stores within the residence of households positively affects food and beverage expenditures made at club stores. The number of grocery stores and supercenters within the residence of households negatively impacts expenditures made at discount stores. The number of convenience stores and the number of drug stores within the residence of households are not statistically significant factors affecting expenditures made at any of the six store outlets.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are higher at grocery stores, convenience stores and drug stores but are lower at discount stores, club stores, and dollar stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Middle Atlantic region are higher at convenience stores but are lower at discount stores and club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East North Central region are higher at grocery stores and convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West North Central region are higher at discount stores and convenience stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the South Atlantic region are higher at convenience stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region are higher at dollar stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West South Central region are higher at grocery stores but are lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Mountain region are higher at convenience stores but are lower at drug stores. Without question, region is a key determinant of household food and beverage expenditures across the six store outlets.

The mean, minimum, and maximum predicted probability to purchase at each store outlet for the entire sample and across the three income categories is presented in Table III-8. On average, for the entire sample of panel households, the probability is highest for grocery stores, discount stores, and drug stores, and the probability is lowest for convenience stores. This pattern is similar for the mid-income and high-income subsamples. But, based on the low-income sample, on average the probability is highest for grocery stores, discount stores, drug stores, and dollar stores and lowest for club stores and convenience stores.

Table III-8. Maximum Likelihood Parameter Estimates and Associated p-values for the Explanatory Variables Based on the High-income Sample of Panel Households

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
Lagged dependent variable	0.639*	0.404*	0.512*	0.349*	0.513*	0.377*
	(0.009)	(0.007)	(0.008)	(0.009)	(0.016)	(0.008)
Income	0.007	0.029	0.165	-0.140	-0.111	0.075
	(0.020)	(0.044)	(0.078)	(0.075)	(0.169)	(0.065)
Household size	0.001	0.022*	0.059*	0.047*	-0.049	-0.023
	(0.003)	(0.007)	(0.013)	(0.013)	(0.026)	(0.010)
Age <40	-0.024	0.020	-0.023	-0.061	0.182	0.122
	(0.024)	(0.059)	(0.114)	(0.112)	(0.219)	(0.088)
Age >60	0.002	-0.087*	0.079	0.158*	-0.064	0.077*
	(0.007)	(0.018)	(0.036)	(0.035)	(0.073)	(0.028)
Urban	0.006	-0.073*	-0.028	-0.107*	-0.433*	0.064
	(0.007)	(0.019)	(0.039)	(0.038)	(0.076)	(0.029)
Rural	-0.026	0.113*	-0.332*	0.257*	0.106	-0.014
	(0.011)	(0.030)	(0.064)	(0.059)	(0.117)	(0.046)
Under high school	-0.277*	0.188	0.227	-0.103	-1.185	-0.001
	(0.062)	(0.154)	(0.294)	(0.294)	(0.716)	(0.231)
Graduate high school	-0.030*	0.085*	-0.035	0.220*	-0.059	-0.023
	(0.011)	(0.029)	(0.058)	(0.054)	(0.114)	(0.044)
College experienced	-0.005	0.045	-0.006	0.112*	0.097	-0.015
	(0.007)	(0.019)	(0.036)	(0.035)	(0.072)	(0.028)
Non-Hispanic black	-0.040*	0.143*	0.254*	0.428*	0.153	0.227*
	(0.011)	(0.029)	(0.059)	(0.057)	(0.116)	(0.044)
Non-Hispanic Asian	-0.050*	-0.048	0.184	-0.031	-0.270	-0.111
	(0.015)	(0.040)	(0.080)	(0.081)	(0.170)	(0.060)
Non-Hispanic other	-0.033	0.085	-0.037	0.156	0.340	0.095
	(0.021)	(0.053)	(0.103)	(0.100)	(0.210)	(0.080)
Hispanic	-0.020	-0.001	0.090	0.264*	0.343	0.070
	(0.013)	(0.035)	(0.071)	(0.069)	(0.143)	(0.054)
Number of Club stores	-0.020	0.001	0.131*	0.018	-0.064	-0.064
	(0.010)	(0.022)	(0.038)	(0.039)	(0.084)	(0.033)
Number of Convenience stores	0.003	-0.005	0.012	0.004	0.011	0.002
	(0.002)	(0.005)	(0.009)	(0.009)	(0.020)	(0.007)
Number of Grocery stores and Supercenters	0.003	-0.009	0.006	-0.016	-0.013	0.003
	(0.002)	(0.004)	(0.007)	(0.007)	(0.016)	(0.006)
Number of Drug stores	-0.000	0.004	-0.019	-0.028*	0.052	0.007
	(0.003)	(0.006)	(0.011)	(0.011)	(0.024)	(0.009)
New England	0.043*	-0.115*	-0.403*	-0.196	0.500*	0.168*
	(0.016)	(0.043)	(0.089)	(0.088)	(0.175)	(0.065)

Table III-7. Continued

Explanatory Variable	Grocery	Discount	Club	Dollar	Convenience	Drug
Middle Atlantic	0.029	-0.100*	-0.454*	0.128	1.210*	0.057
	(0.011)	(0.031)	(0.065)	(0.062)	(0.127)	(0.047)
East North Central	0.039*	-0.064	-0.213*	-0.017	0.459*	-0.004
	(0.011)	(0.030)	(0.062)	(0.060)	(0.125)	(0.045)
West North Central	-0.009	0.115*	-0.132	-0.057	1.463*	-0.075
	(0.014)	(0.037)	(0.077)	(0.074)	(0.147)	(0.056)
South Atlantic	0.018	0.057	-0.198*	0.147	0.814*	0.075
	(0.011)	(0.029)	(0.060)	(0.058)	(0.120)	(0.044)
East South Central	0.023	0.056	-0.276*	0.289*	0.257	-0.022
	(0.016)	(0.044)	(0.092)	(0.086)	(0.183)	(0.067)
West South Central	0.042*	0.061	-0.199*	0.114	0.191	0.027
	(0.013)	(0.034)	(0.070)	(0.068)	(0.143)	(0.052)
Mountain	0.032	0.029	-0.159	-0.163	0.625*	-0.182*
	(0.014)	(0.039)	(0.078)	(0.078)	(0.159)	(0.059)
Constant	1.187*	2.168*	-2.843	8.685*	2.402	-0.241
	(0.290)	(0.784)	(1.637)	(1.580)	(3.152)	(1.190)
Initial value	0.250*	0.496*	0.599*	0.757*	0.905*	0.540*
	(0.008)	(0.008)	(0.009)	(0.011)	(0.019)	(0.008)
σ_u	0.207*	0.679*	1.493*	1.415*	2.532*	1.048*
	(0.007)	(0.011)	(0.022)	(0.020)	(0.046)	(0.017)
σ_e	0.530*	1.141*	1.847*	1.732*	3.122*	1.639*
	(0.002)	(0.005)	(0.008)	(0.009)	(0.023)	(0.007)
LR χ^2 test of the Correlated Random Effect Tobit Model vs the Pooled Tobit Model	193.6*	1156.6*	1812.6*	2659.2*	1670.6*	1421.3*
LR χ^2 test of the Correlated Random Effect Tobit Model vs the Random Effect Tobit Model	1,400.4*	4,046.2*	3,403.9*	5,013.6*	2,746.4*	4,121.5*
Wald χ^2 test	1,020.1*	4,291.1*	4,272.9*	4,675.0*	2,292.0*	4,261.1*
Pseudo R^2	0.337	0.228	0.236	0.211	0.209	0.212
V-Z R^2	0.788	0.695	0.754	0.617	0.529	0.596
Number of Observations	54,640	54,640	54,640	54,640	54,640	54,640
Number of Households	10,928	10,928	10,928	10,928	10,928	10,928

*p<0.01

Numbers in parentheses correspond to standard errors.

Table III-9. Mean, Minimum, and Maximum Predicted Probability to Purchase at Each Store Type for the Entire Sample and for the Three Income Subsamples

		Grocery	Discount	Club	Dollar	Convenience	Drug
	mean	0.999	0.988	0.656	0.732	0.323	0.904
Entire	min	0.857	0.356	0.056	0.231	0.067	0.356
	max	1	1	0.999	0.999	0.996	0.999
	mean	0.999	0.983	0.396	0.859	0.331	0.866
Low	min	0.798	0.443	0.004	0.173	0.068	0.272
	max	1	1	0.998	1	0.997	0.999
	mean	0.999	0.989	0.617	0.788	0.340	0.904
Mid	min	0.869	0.33	0.103	0.15	0.023	0.287
	max	1	1	0.999	0.999	0.996	0.999
	mean	0.999	0.985	0.789	0.612	0.301	0.916
High	min	0.856	0.404	0.158	0.146	0.061	0.343
	max	1	1	0.999	0.999	0.996	0.999

The predicted probability of purchasing food and beverages is given by $P\left(EX_{ht}^{k\,*} > 0 \middle| z_{ht}^{k}\right)$.

The conditional marginal effects of the respective explanatory variables associated with household expenditures on food and beverages in the correlated random effect Tobit models across store outlets and across income categories are exhibited in Tables III-9-11. We present the impacts of household income, household size, age, urbanization, education, race and ethnicity, the number of club stores, convenience stores, grocery stores, and drug stores, and region. In addition, we also report the marginal effects of these aforementioned explanatory variables concerning the probability to visit.

III.10.4 Low-Income Sample: Conditional Marginal Effects

As exhibited in Table III-9, household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. However, with unit changes in household size, expenditures made at discount stores change by $26.19. Relative to households in the 40-year-old to 60-year-old category, household expenditures made at discount stores are lower by $91.61 for households 60 years of age and older.

In the low-income sample, for households located in urban areas, household food and beverage expenditures are lower by $105.11 at discount stores and by $$74.53 at convenience stores relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are lower by $248.92 at club stores relative to households located outside of urban and rural areas. Relative to households who have graduated from college, households with a high school education spend $56.37 more on food and beverages at discount stores and $7.00 more at dollar stores but $58.65 less at convenience stores and $32.81 less at drug stores annually.

Relative to non-Hispanic white households, non-Hispanic black households spend $311.33 more on food and beverages at club stores and $10.11 more at dollar stores annually. Relative to non-Hispanic white households, Hispanic households spend $130.12 more on food and beverages at discount stores and $18.84 more at dollar stores annually. Unit changes in the number of club stores, convenience, grocery stores and supercenters, and drug stores have no statistically significant effects on the magnitude of food and beverage expenditures at the six store types.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are higher by $61.08 at drug stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Middle Atlantic region and in the West North Central region are higher by $167.91 and by $240.71 at convenience stores respectively. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region and in the South Atlantic region are higher by $151.25 and $104.04 at discount stores respectively.

III.10.5 Low-Income Sample: Marginal Effects Associated with the Probability of Purchasing

As shown in Table III-9, in the low-income sample, the marginal effects of household income, household size, age, degree of urbanization, educational level, race/ethnicity and region of purchasing food and beverages at grocery stores are not statistically significant. Household income also not have any statistically significant impact on the probability of purchasing at any of the six store outlets. Unit changes in household size is positively linked to increasing the probability of purchasing at discount stores by 0.06%.

Relative to household age group between 40 and 60, households who are in the over 60 age group have a lower probability of purchasing at discount stores by 0.21%. Households who reside in urban area have a lower probability to purchase food and beverages at discount stores by -0.24% and at convenience stores by 2.46% relative to households located outside urban and rural areas. Households located in rural area have a lower probability to purchase at club stores by 4.15%.

Households with a high school education are more likely to purchase at discount stores by 0.13% and at dollar stores by 0.97%, but they are less likely to purchase food and beverages at drug stores by1.14%. Relative to non-Hispanic white households, non-Hispanic black households have a higher probability to purchase food and beverages at club store by 5.19%. Hispanic households have a higher likelihood of purchasing at discount stores by 0.30% and at dollar stores by 2.62%. The number of clubs, convenience stores, grocery stores and supercenters, and drug stores, the proxy for the retail environment, does not significantly affect the probability to purchasing food and beverages at any of the six store types.

Households located in New England region have a higher probability of purchasing at drug stores by 2.13% relative to households located in the Pacific region. The probability of purchasing food and beverages is higher at convenience stores by 5.54% and by 7.95% for households located in the Mid-Atlantic region and in the West North Central region than for households located in the Pacific region. Similarly, the probability of purchasing food and beverages is higher at discount stores by 0.24% and by 0.35% for households located in the South Atlantic region and in the East South Central region relative to households located in the Pacific region.

Table III-10. Conditional Marginal Effects of Household Food and Beverage Expenditures and Marginal Effects Associated with the Probability of Purchasing by Store Type Based on the Low-Income Sample of Panel Households

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Household income	0.0000	-0.010	0.0002	0.010	-0.0004	-0.013	-0.0002	-90.001	0.0071	0.302	0.0018	0.022
	(0.0000)	(0.010)	(0.0003)	(0.016)	(0.004)	(0.118)	(0.0021)	(0.011)	(0.0055)	(0.237)	(0.0022)	(0.027)
Household size	0.0000	21.704	0.0006*	26.192*	0.0054	32.464	0.0007	0.486	-0.0011	-3.323	-0.0005	-1.511
	(0.0000)	(8.997)	(0.0002)	(9.162)	(0.0029)	(17.818)	(0.0014)	(1.014)	(0.0034)	(10.218)	(0.0014)	(4.054)
Age<40	0.0002	111.212	-0.0028	-120.176	-0.0263	-157.576	-0.0129	-9.249	0.0114	34.465	0.0135	38.861
	(0.0002)	(77.432)	(0.0018)	(76.682)	(0.0261)	(159.055)	(0.0118)	(8.507)	(0.0271)	(82.215)	(0.0118)	(33.841)
Age>60	0.0001	53.953	-0.0021*	-91.607*	0.0215	129.127	-0.0009	-0.671	-0.0162	-48.979	0.0043	12.417
	(0.0001)	(25.656)	(0.0006)	(26.255)	(0.0085)	(55.366)	(0.0041)	(2.925)	(0.0095)	(29.840)	(0.0041)	(11.694)
Urban	0.0000	-1.064	-0.0024*	-105.114*	0.0089	53.388	-0.0004	-0.297	-0.0246*	-74.533*	0.0048	13.743
	(0.0000)	(19.757)	(0.0005)	(22.101)	(0.0073)	(44.863)	(0.0036)	(2.598)	(0.0085)	(28.550)	(0.0034)	(9.886)
Rural	-0.0001	-41.551	0.001	40.866	-0.0415*	-248.942*	0.0023	1.676	-0.0043	-13.080	-0.0099	-28.399
	(0.0000)	(22.491)	(0.0006)	(24.788)	(0.0089)	(67.480)	(0.0041)	(2.944)	(0.0095)	(28.836)	(0.0039)	(11.395)
Less than high school	0.0000	-8.965	0.0018	77.068	-0.0088	-52.666	0.0081	5.788	-0.0114	-34.641	-0.0178	-51.074
	(0.0001)	(43.246)	(0.001)	(44.822)	(0.0153)	(91.968)	(0.007)	(5.025)	(0.0172)	(52.244)	(0.007)	(20.234)
High school graduate	-0.0001	-40.834	0.0013*	56.365*	0.0023	14.037	0.0097*	7.001*	-0.0194	-58.645	-0.0114*	-32.805*
	(0.0000)	(20.468)	(0.0005)	(21.637)	(0.0071)	(42.907)	(0.0035)	(2.498)	(0.0083)	(26.777)	(0.0034)	(9.780)
College experienced	0.0000	-11.411	0.0003	11.687	0.0034	20.336	0.0077	5.552	-0.0009	-2.593	0.0004	1.118
	(0.0000)	(19.962)	(0.0005)	(20.184)	(0.0064)	(38.877)	(0.0031)	(2.259)	(0.0076)	(22.993)	(0.0031)	(8.959)
Non-Hispanic black	0.0001	26.883	0.0001	2.789	0.0519*	311.327*	0.0141	10.112*	0.0256	77.480	0.0085	24.510
	(0.0001)	(30.024)	(0.0008)	(33.125)	(0.0109)	(84.188)	(0.0055)	(3.937)	(0.0127)	(40.716)	(0.0052)	(14.896)
Non-Hispanic Asian	0.0001	57.696	-0.001	-43.017	-0.0073	-43.771	-0.0204	-14.672	-0.0332	-100.627	-0.0176	-50.458
	(0.0002)	(81.277)	(0.0021)	(89.357)	(0.0287)	(172.418)	(0.0149)	(10.729)	(0.0368)	(112.919)	(0.0138)	(39.804)
Non-Hispanic other	-0.0001	-64.437	0.0011	45.182	0.0329	197.647	0.008	5.726	0.0327	98.929	0.0085	24.519
	(0.0001)	(49.794)	(0.0012)	(52.248)	(0.0165)	(104.422)	(0.0083)	(5.955)	(0.0191)	(60.144)	(0.0081)	(23.444)
Hispanic	0.0000	-18.248	0.0030*	130.115*	0.027	162.054	0.0262*	18.841*	0.0253	76.600	0.0156	44.952
	(0.0001)	(45.411)	(0.0011)	(48.811)	(0.0158)	(98.926)	(0.0078)	(5.641)	(0.0183)	(56.918)	(0.0076)	(21.878)

Table III-9. Continued

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Club stores	-0.0001	-37.767	0.0001	5.369	0.012	72.117	-0.0015	-1.094	0.0085	25.641	-0.003	-8.697
	(0.0001)	(31.822)	(0.0006)	(26.576)	(0.0073)	(45.397)	(0.0038)	(2.714)	(0.0095)	(29.229)	(0.004)	(11.501)
Convenience stores	0.0000	-0.413	0.0001	4.097	0.0002	1.351	0.0000	0.004	-0.0011	-3.388	0.0012	3.567
	(0.0000)	(6.768)	(0.0001)	(5.693)	(0.0016)	(9.806)	(0.0008)	(0.580)	(0.0021)	(6.433)	(0.0008)	(2.429)
Grocery stores and Supercenters	0.0000	0.313	-0.0001	-3.475	0.0007	4.343	0.0008	0.583	-0.0026	-7.822	-0.0003	-0.744
	(0.0000)	(5.097)	(0.0001)	(4.295)	(0.0014)	(8.386)	(0.0006)	(0.444)	(0.0017)	(5.259)	(0.0006)	(1.810)
Drug stores	0.0000	-0.323	0.0001	5.905	0.0014	8.116	0.0006	0.450	-0.0011	-3.482	0.0011	3.155
	(0.0000)	(8.487)	(0.0002)	(7.147)	(0.0021)	(12.844)	(0.001)	(0.740)	(0.0026)	(8.036)	(0.0011)	(3.032)
Ne	-0.0001	-49.086	0.0003	14.426	-0.0037	-22.284	0.001	0.731	0.0133	40.254	0.0213*	61.079*
	(0.0001)	(46.887)	(0.0012)	(52.219)	(0.0177)	(106.456)	(0.0087)	(6.288)	(0.0204)	(62.265)	(0.0082)	(23.638)
Ma	0.0000	17.423	0.0008	36.165	-0.0241	-144.690	0.0007	0.512	0.0554*	167.910*	0.0075	21.480
	(0.0001)	(35.133)	(0.0009)	(39.172)	(0.0131)	(82.684)	(0.0065)	(4.698)	(0.0153)	(54.069)	(0.0061)	(17.689)
Enc	0.0000	2.733	0.0002	9.668	-0.0164	-98.292	-0.0001	-0.093	0.0155	46.892	0.0043	12.378
	(0.0001)	(31.962)	(0.0008)	(35.593)	(0.0119)	(73.581)	(0.0059)	(4.273)	(0.0142)	(43.622)	(0.0056)	(16.094)
Wnc	-0.0001	-48.111	0.0023	100.120	-0.0268	-160.556	-0.0033	-2.402	0.0795*	240.707*	0.0067	19.180
	(0.0001)	(37.383)	(0.001)	(41.726)	(0.0142)	(89.286)	(0.0069)	(4.991)	(0.0161)	(62.924)	(0.0066)	(18.880)
Sa	-0.0001	-48.370	0.0024*	104.039*	-0.0286	-171.399	0.0077	5.505	0.0198	59.946	0.0085	24.474
	(0.0001)	(32.178)	(0.0008)	(35.972)	(0.0118)	(76.788)	(0.006)	(4.297)	(0.0142)	(44.086)	(0.0056)	(16.161)
Esc	0.0000	-15.482	0.0035*	151.254*	-0.0324	-194.459	0.0095	6.847	-0.0094	-28.564	0.0123	35.246
	(0.0001)	(43.199)	(0.0011)	(47.867)	(0.0166)	(104.802)	(0.0079)	(5.687)	(0.019)	(57.830)	(0.0075)	(21.714)
Wsc	-0.0001	-56.192	0.0022	95.815	-0.0238	-142.572	0.0009	0.662	0.0107	32.327	0.0006	1.629
	(0.0001)	(37.254)	(0.001)	(41.271)	(0.0138)	(86.082)	(0.0068)	(4.912)	(0.0163)	(49.762)	(0.0065)	(18.637)
Mt	-0.0001	-33.224	0.0024	101.908	-0.0076	-45.718	-0.0084	-6.049	0.0237	71.719	-0.0011	-3.226
	(0.0001)	(38.814)	(0.001)	(43.123)	(0.014)	(84.232)	(0.0072)	(5.174)	(0.0169)	(52.627)	(0.0068)	(19.533)

Superscript * indicates p-value <0.01

Numbers in parentheses correspond to standard errors.

The marginal effect of the probability to purchase at any store outlet is given by $\frac{\partial P(EX_{ht}^{k*}>0|z_{ht}^{k})}{\partial z_{ht}^{k}}$, and the marginal effect of the conditional expectation of household food and beverage expenditures is expressed mathematically as $\frac{\partial E[EX_{ht}^{k}|z_{ht}^{k}, EX_{ht}^{k}>0]}{\partial z_{ht}^{k}}$. All calculations pertaining to marginal effects are made at the sample means of the data.

III.10.6 Mid-Income Sample: Conditional Marginal Effects

As exhibited in Table III-10, household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. With unit increases in household size, expenditures made at discount stores, club stores, and dollar stores rise by $22.18, $19.75, and $1.54 annually.

Relative to households in the 40-year-old to 60-year-old category, household expenditures made at grocery stores and at drug stores are higher by $47.79 and $26.17 respectively for households in the 60 years of age and older category. For households located in urban areas, household food and beverage expenditures are higher by $37.53 annually at grocery stores, but lower by $70.12, $3.53, and $111.32 annually at discount stores, at dollar stores, and at convenience stores respectively relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are higher by $55.06 annually at discount stores but lower by $48.62, $169.22, and $35.19 annually at grocery stores, at club stores, and at drug stores relative to households located outside of urban and rural areas.

Relative to households who have graduated from college, households with less than a high school education expend $9.72 more at dollar stores annually. Relative to households who have graduated from college, households with a high school education spend $54.03 more on food and beverages at discount stores and $6.45 more at dollar stores annually but $55.94 less at grocery stores annually. Relative to households who have graduated from college, households with some college experience expend $33.99 more at discount stores annually and $4.11 more at dollar stores annually but $25.18 less at grocery stores annually.

Relative to non-Hispanic white households, non-Hispanic black households spend $81.20, $108.91, $12.23, and $140.75 more on food and beverages at discount stores, at club stores, at dollar stores, and at convenience stores respectively but $47.35 less at grocery stores. Relative to non-Hispanic white households, non-Hispanic other households spend $175.77 more on food and beverages at convenience stores annually. Finally, relative to non-Hispanic white households, Hispanic households spend $106.90 more on food and beverages at discount stores and $6.03 more at dollar stores in the mid-income sample.

The number of club stores within the residence of households negatively impacts food and beverage expenditures made at grocery stores and drug stores but positively affects food and beverage expenditures made at club stores. With each unit increase in the number of club stores, household expenditures made at grocery stores decline by $57.61 and household expenditures made at drug stores decline by $36.33 annually, but household expenditures made at club stores rises by $47.65 annually. With each unit increase in the number of grocery stores and supercenters, household expenditures decline by $6.87 annually at discount stores but household expenditures rise by $7.45 annually.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are lower by $152.50 annually at club stores. Relative to households located in the Pacific region, annual food and beverage expenditures made by households located in the Middle Atlantic region are higher by $6.00 at dollar stores and by $398.78 at convenience stores but are lower by $164.18 at club stores. Relative to households located in the Pacific region, annual food and beverage expenditures made by households located in the East North Central region are higher by $249.21 at convenience stores. Relative to households located in the Pacific region, annual food and beverage expenditures made by households located in the West North Central region are higher by $95.34 at discount stores and by $439.00 at convenience stores but are lower by $113.46 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the South Atlantic region are higher by $73.35 annually at discount stores and by $278.62 annually at convenience stores but are lower by $102.48 annually at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region are $7.18 higher annually at dollar stores but are $112.44 lower annually at club stores. Relative to households located in the Pacific region, annual food and beverage expenditures made by households located in the West South Central region are higher by $88.79 at discount stores and by $194.32 at convenience stores. Relative to households located in the Pacific region, annual food and beverage expenditures made by households located in the Mountain region are higher by $288.23 at convenience stores.

III.10.7 Mid-Income Sample: Marginal Effects Associated with the Probability of Purchasing

As exhibited in Table III-10, the probability of purchasing food and beverages at grocery stores are significantly higher for households who are greater than 60 years of age and households located in urban areas. But this probability is significantly lower for household located in rural areas, for households with at most a high school education, and for non-Hispanic black households. As well, with unit increases in the number of club stores the probability of purchasing food and beverages declines at grocery stores. That said, the magnitude of the marginal effects associated with purchasing food and beverages at grocery stores albeit statically significant is negligible from a practical standpoint.

Household income is not statistically significant factor affecting probability to visit all store types. For households that fall in the 60 and over age category, the probability of purchasing food and beverages is higher at drug stores 0.60% relative to households that fall in the age 40 to 60 category. Households located in urban areas are less likely to purchase at discount stores, dollar stores, and convenience stores by 0.10%, 0.94%, and 2.47% relative to households not located in rural or urban areas. Household located in rural area have lower probabilities of purchasing at club stores and drug stores by 3.04% and 0.81% respectively, but have a higher probability of purchasing at discount stores by 0.08%.

Relative to household who graduated college, households whose education level is less than high school are more likely to make purchases at dollar stores by 2.58%. Households who have a high school education have higher probabilities to make purchases at discount stores and dollar stores by 0.07% and 1.71% respectively. Households who have some college experience are more likely to make purchases at discount stores and dollar stores by 0.05% and 1.09%.

Relative to non-Hispanic white households, non-Hispanic black households have higher probabilities to purchase food and beverages at discount stores, at club stores, at dollar stores, and at convenience stores by 0.11%, 1.96%, 3.25%, and 3.12%. Non-Hispanic other households have a higher likelihood to make purchases at convenience stores by 3.90% relative to non-Hispanic white households. No differences are evident in the likelihood of purchasing food and beverages between non-Hispanic Asian households and non-Hispanic white households. Hispanic households have a higher probability to make purchases of food and beverages at discount stores and at dollar stores by 0.15% and 1.60% respectively than do non-Hispanic white households.

With unit increases in the number of club stores in the zip code area of the residence of the household, the likelihood of purchasing food and beverages at drug stores declines by 0.83%, but the likelihood of purchasing probability at club stores increases by 0.86%. Changes in the numbers of convenience stores, grocery stores and supercenters, and drug stores have no statistically significant impact on the likelihood of purchasing food and beverages at the respective store outlets.

Relative to households residing in the Pacific region, households located in the West North Central region, the South Atlantic region, and the West South Central region are more likely to make purchases at discount stores by 0.13%, 0.10%, and 0.12% respectively. Households located in New England, the Mid-Atlantic region, the West North Central region, the South Atlantic region, and the East South Central region have lower probabilities of purchasing food and beverages at club stores by 2.74%, 2.95%, 2.04%, 1.84%, and 2.02% respectively relative to households located in the Pacific region. Households residing in the Mid-Atlantic region and the East South Central region have higher likelihoods to make purchases at dollar stores by 1.59% and 1.91% respectively than do households who live in the Pacific region. The probabilities of purchasing food and beverages at convenience stores are higher for households located in the Mid-Atlantic region by 8.85%, in the East North Central region by 5.53%, in the West North Central region by 9.74%, in the South Atlantic region by 6.18%, in the West South Central region by 4.31%, and in the Mountain region by 6.39% than for households located in the Pacific region.

Table III-11. Conditional Marginal Effects of Household Food and Beverage Expenditures and Marginal Effects Associated with the Probability of Purchasing by Store Type Based on the Mid-Income Sample of Panel Households

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Household income	0.0000	0.006	0.0003	0.018	0.0052	0.062	-0.0058	-0.025	0.0033	0.163	-0.0014	-0.022
	(0.0000)	(0.007)	(0.0002)	(0.011)	(0.0027)	(0.032)	(0.0025)	(0.011)	(0.0045)	(0.222)	(0.0013)	((0.022)
Household size	0.0000	4.226	0.0003*	22.178*	0.0035*	19.750*	0.0041*	1.538*	0.0007	3.335	-0.0008	-3.617
	(0.0000)	(4.778)	(0.0001)	(4.493)	(0.0011)	(6.155)	(0.001)	(0.360)	(0.0016)	(7.253)	(0.0005)	(2.148)
Age<40	0.0000	23.606	0.0001	10.699	-0.0107	-59.741	-0.0063	-2.361	0.0257	115.722	-0.0011	-4.815
	(0.0000)	(40.053)	(0.0005)	(36.084)	(0.0089)	(49.632)	(0.0078)	(2.940)	(0.0127)	(58.451)	(0.004)	(17.504)
Age>60	0.0000*	47.785*	-0.0002	-12.664	0.0013	7.391	0.0021	0.804	0.0048	21.709	0.0060*	26.173*
	(0.0000)	(14.506)	(0.0002)	(13.402)	(0.0033)	(18.503)	(0.0029)	(1.089)	(0.0049)	(21.974)	(0.0015)	(6.549)
Urban	0.0000*	37.528*	-0.0010*	-70.116*	-0.0015	-8.278	-0.0094*	-3.527*	-0.0247*	-111.322*	0.003	12.882
	(0.0000)	(12.159)	(0.0002)	(11.985)	(0.0031)	(17.283)	(0.0027)	(1.021)	(0.0045)	(22.668)	(0.0013)	(5.819)
Rural	-0.0000*	-48.617*	0.0008*	55.055*	-0.0304*	-169.222*	0.008	3.020	-0.0073	-32.765	-0.0081*	-35.185*
	(0.0000)	(15.844)	(0.0002)	(15.507)	(0.0042)	(25.180)	(0.0035)	(1.316)	(0.0057)	(25.876)	(0.0018)	(7.691)
Less than high school	-0.0001	-80.545	0.001	73.053	0.0105	58.598	0.0258*	9.720*	-0.0132	-59.384	-0.0095	-41.280
	(0.0000)	(49.146)	(0.0006)	(44.867)	(0.0111)	(62.067)	(0.0093)	(3.521)	(0.0168)	(76.087)	(0.005)	(21.613)
High school graduate	-0.0000*	-55.937*	0.0007*	54.029*	0.001	5.679	0.0171*	6.453*	0.0016	7.254	-0.0017	-7.248
	(0.0000)	(13.375)	(0.0002)	(12.673)	(0.0032)	(17.820)	(0.0028)	(1.048)	(0.0047)	(21.051)	(0.0014)	(6.145)
College experienced	0.0000	-25.176	0.0005*	33.988*	0.0003	1.627	0.0109*	4.111*	0.0062	27.920	0.0002	0.885
	(0.0000)	(11.436)	(0.0001)	(10.533)	(0.0026)	(14.321)	(0.0022)	(0.851)	(0.0038)	(17.459)	(0.0012)	(5.077)
Non-Hispanic black	-0.0001*	-80.799*	0.0011*	81.204*	0.0196*	108.910*	0.0325*	12.229*	0.0312*	140.747*	0.0042	18.365
	(0.0000)	(18.990)	(0.0003)	(18.547)	(0.0048)	(27.531)	(0.0042)	(1.599)	(0.0069)	(33.689)	(0.0021)	(9.028)
Non-Hispanic Asian	0.0000	-47.354	0.0006	40.182	0.0208	115.519	0.0055	2.060	-0.0113	-50.910	-0.001	-4.377
	(0.0000)	(39.809)	(0.0005)	(38.433)	(0.0095)	(53.399)	(0.0087)	(3.283)	(0.0153)	(69.323)	(0.0043)	(18.742)
Non-Hispanic other	0.0000	-54.721	0.0006	44.247	0.0042	23.258	0.0167	6.296	0.0390*	175.770*	-0.0023	-10.028
	(0.0000)	(34.544)	(0.0004)	(31.892)	(0.0077)	(43.074)	(0.0068)	(2.565)	(0.0114)	(54.042)	(0.0035)	(15.372)
Hispanic	0.0000	-8.651	0.0015*	106.901*	0.01	55.499	0.0160*	6.028*	0.0089	40.067	0.0045	19.598
	(0.0000)	(26.651)	(0.0004)	(25.425)	(0.0063)	(35.372)	(0.0056)	(2.106)	(0.0096)	(43.439)	(0.0028)	(12.340)

Table III-10. Continued

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Club stores	-0.0000*	-57.614*	0.0002	16.962	0.0086*	47.645*	-0.002	-0.770	-0.0073	-32.866	-0.0083*	-36.333*
	(0.0000)	(18.580)	(0.0002)	(14.725)	(0.0031)	(17.529)	(0.0029)	(1.088)	(0.0052)	(23.652)	(0.0016)	(6.985)
Convenience stores	0.0000	-0.498	-0.0001	-4.012	-0.0012	-6.555	0.0003	0.098	0.0013	5.773	-0.0001	-0.559
	(0.0000)	(4.247)	(0.0000)	(3.377)	(0.0007)	(4.075)	(0.0007)	(0.246)	(0.0012)	(5.399)	(0.0004)	(1.570)
Grocery stores and Supercenters	0.0000	7.448*	-0.0001	-6.867*	0.0003	1.712	0.0000	-0.007	0.0003	1.492	0.0001	0.268
	(0.0000)	(3.345)	(0.0000)	(2.667)	(0.0006)	(3.264)	(0.0005)	(0.201)	(0.001)	(4.313)	(0.0003)	(1.228)
Drug store	0.0000	8.008	0.0000	-0.754	-0.0003	-1.656	0.0003	0.125	0.0013	6.064	0.0001	0.223
	(0.0000)	(5.197)	(0.0001)	(4.150)	(0.0009)	(4.967)	(0.0008)	(0.309)	(0.0015)	(6.734)	(0.0004)	(1.930)
Ne	0.0001	75.348	-0.0002	-15.855	-0.0274*	-152.504*	-0.0051	-1.905	0.0271	121.983	0.007	30.463
	(0.0000)	(30.061)	(0.0004)	(29.591)	(0.0078)	(44.212)	(0.0068)	(2.577)	(0.0115)	(53.272)	(0.0033)	(14.447)
Ma	0.0000	49.244	0.0002	11.814	-0.0295*	-164.176*	0.0159*	5.999*	0.0885*	398.779*	0.0026	11.209
	(0.0000)	(22.262)	(0.0003)	(21.932)	(0.0058)	(33.261)	(0.005)	(1.896)	(0.0084)	(53.625)	(0.0025)	(10.704)
Enc	0.0000	42.630	-0.0003	-23.652	-0.0124	-69.135	0.0061	2.302	0.0553*	249.210*	0.003	13.131
	(0.0000)	(20.360)	(0.0003)	(20.057)	(0.0053)	(29.482)	(0.0046)	(1.735)	(0.0079)	(42.576)	(0.0022)	(9.787)
Wnc	0.0000	-48.929	0.0013*	95.344*	-0.0204*	-113.464*	-0.01	-3.774	0.0974*	438.998*	-0.0054	-23.365
	(0.0000)	(23.800)	(0.0003)	(23.548)	(0.0062)	(35.059)	(0.0054)	(2.030)	(0.0089)	(57.776)	(0.0026)	(11.523)
Sa	0.0000	9.795	0.0010*	73.345*	-0.0184*	-102.483*	0.0117	4.420	0.0618*	278.615*	0.003	13.015
	(0.0000)	(20.248)	(0.0003)	(20.069)	(0.0051)	(29.143)	(0.0046)	(1.728)	(0.0078)	(43.965)	(0.0022)	(9.734)
Esc	0.0000	3.068	0.001	68.753	-0.0202*	-112.439*	0.0191*	7.178*	0.0229	103.137	0.0004	1.825
	(0.0000)	(27.148)	(0.0004)	(26.765)	(0.0071)	(39.941)	(0.0061)	(2.291)	(0.0105)	(48.199)	(0.003)	(13.054)
Wsc	0.0000	10.867	0.0012*	88.792*	-0.014	-77.819	0.005	1.885	0.0431*	194.319*	-0.0041	-18.020
	(0.0000)	(23.471)	(0.0003)	(23.200)	(0.006)	(33.687)	(0.0053)	(1.989)	(0.009)	(44.549)	(0.0026)	(11.290)
Mt	0.0000	13.81	0.0007	50.22	-0.0014	-7.748	0.0023	0.873	0.0639*	288.228*	-0.0043	-18.883
	(0.0000)	(25.075)	(0.0003)	(24.662)	(0.0063)	(34.913)	(0.0056)	(2.124)	(0.0095)	(50.613)	(0.0028)	(12.076)

Superscript * indicates p-value <0.01

Numbers in parentheses correspond to standard errors.

The marginal effect of the probability to purchase at any store outlet is given by $\frac{\partial P(EX_{ht}^{k*}>0|z_{ht}^k)}{\partial z_{ht}^k}$, and the marginal effect of the conditional expectation of household food and beverage expenditures is expressed mathematically as $\frac{\partial E[EX_{ht}^k|z_{ht}^k, EX_{ht}^k > 0]}{\partial z_{ht}^k}$. All calculations pertaining to marginal effects are made at the sample means of the data.

III.10.8 High-Income Sample: Conditional Marginal Effects

In the high-income sample, as exhibited in Table III-11, household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets. With unit increases in household size, food and beverage expenditures made at discount stores, club stores, and dollar stores rise by $15.37, $35.09, and $1.03 annually. Relative to households in the 40-year-old to 60-year-old category, expenditures made at dollar stores and drug stores are higher by $3.48 and $18.88 annually, but expenditures for food and beverages made at discount stores are lower by $60.06 annually for households 60 years of age and older.

For households located in urban areas, household food and beverage expenditures are lower by $50.63, $2.36, and $171.67 respectively at discount stores, dollar stores, and convenience stores relative to households located outside of urban and rural areas. For households located in rural areas, household food and beverage expenditures are higher by $78.14 in discount stores and by $5.65 at dollar stores, but lower by $197.34 at club stores relative to households located outside of urban and rural areas.

Relative to households who have graduated from college, households with less than a high school education expend $567.56 less annually at grocery stores. Relative to households who have graduated from college, households with a high school education spend $59.04 more on food and beverages at discount stores and $4.84 more at dollar stores but $60.88 less at grocery stores annually. Relative to households who have graduated from college, households with some level of college experience expend $2.47 more at dollar stores in the high-income sample.

Relative to non-Hispanic white households, non-Hispanic black households spend $99.29, $151.19, $9.42, and $55.92 more on food and beverages at discount stores, club stores, dollar stores, and drug stores respectively but $82.10 less at grocery stores annually. Relative to non-Hispanic white households, Asian households spend $103.07 less at grocery stores on a yearly basis. Relative to non-Hispanic white households, Hispanic households spend $5.82 more on food and beverages at dollar stores.

With each unit increase in the number of club stores, expenditures made at club stores increase by $77.74 annually. With each unit increase in the number of grocery stores and supercenters, expenditures made at discount stores fall by $6.51 annually.

Relative to households located in the Pacific region, food and beverage expenditures made by households located in the New England region are higher by $88.39, $198.48, and $41.39 at grocery stores, convenience stores and drug stores respectively but are lower by $79.57 at discount stores and by $239.58 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Middle Atlantic region are higher by $480.22 at convenience stores but are lower by $69.38 at discount stores and by $269.60 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East North Central region are higher by $182.14 at convenience stores but are lower by $126.40 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West North Central region are higher by $79.67 at discount stores and by $580.39 at convenience stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the South Atlantic region are higher by $322.87 at convenience stores but are lower by $117.40 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the East South Central region are higher by $6.36 at dollar stores but are lower by $163.75 at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the West South Central region are $85.91 higher at grocery stores but are $118.52 lower at club stores. Relative to households located in the Pacific region, food and beverage expenditures made by households located in the Mountain region are higher by $247.80 at convenience stores but are lower by $44.79 at drug stores.

III.10.9 High-Income Sample: Marginal Effects Associated with the Probability of Purchasing

As exhibited in Table III-11, household income has no statistically significant impact on the likelihood of making purchases of food and beverages at any of the six outlets based on the high-income sample of panel households. With unit increases in household size, the likelihood of purchasing increases by 0.03% at discount stores, by 0.34% at club stores, and by 0.56% at dollar stores respectively. Relative to households in the 40 to 60 age category, households over 60 years of age are more likely to purchase food and beverages at dollar stores by 1.89% and at drug stores by 0.36%, but are less likely to purchase at club stores by 0.12%.

Household living in urban areas have lower probabilities of purchasing food and beverages at discount stores by 0.10%, at dollar stores by 1.28%, and at convenience stores by 2.77% than do households located outside urban and rural areas. Conversely, households residing in rural areas have higher likelihoods of purchasing at discount stores by 0.15%) and at dollar stores by 3.07%, but are less likely to purchase at club stores by 1.89%.

Relative to households who have a college degree, household heads who did not graduate from high school have lower probability of purchasing food and beverages at grocery store by 0.03%. Household heads who graduated from high school are more likely to purchase at discount stores by 0.11% and at dollar stores by 2.63%. Household heads with some level of college have a higher probability of purchasing food and beverages at dollar stores by 1.34% relative to households who have a college degree.

Relative to non-Hispanic white households, non-Hispanic black households have higher probability to purchase at discount stores by 0.19%, at club stores by 1.45%, at dollar stores by 5.12%, and at drug stores by 1.07%. Non-Hispanic black households and non-Hispanic Asian households have lower probabilities to purchase food and beverages at grocery stores than non-Hispanic white households. But this difference in the respective likelihoods is less than 0.01%. Hispanic households are more likely to purchase at dollar stores by 3.16% compared to non-Hispanic white households.

With unit increases in the number of club stores in the zip codes where households are located, the likelihood of purchasing food and beverages at club stores increases by 0.74%. With unit increases in the number of drug stores in the zip codes are located, the probability of purchasing at dollar stores falls by 0.34%. However, changes in number of convenience stores and changes in the number of grocery stores and supercenters have no impact on the probability of purchasing food and beverages at any of the six stores outlets considered in this.

Relative to households located in Pacific region, households located in New England, the East North Central region, and the West South Central region are more likely to purchase food and beverages at grocery stores. But the difference in this likelihood is less than 0.01%. Household residing in New England and the Mid-Atlantic region have lower probabilities of making food and beverage purchases at discount stores by 0.15% and 0.13% respectively, but households residing in the West North Central region have a higher probability of making purchases at discount stores by 0.15% relative to households residing in the Pacific region. Relative to households located in the Pacific region, the probabilities of making purchases at club stores are lower for households located in New England by 2.29%, the Mid-Atlantic region by 2.58%, the East North Central region by 1.21%, the South Atlantic region by 1.12%, the East South Central region by 1.57%, and the West South Central region by 1.13%. Relative to households located in the Pacific region, for households located in the East South Central region the likelihood of purchasing food and beverages at dollar stores is lower by 3.45%. The probabilities of purchasing food and beverages at convenience stores are higher for households located in New England by 3.20%, in the Mid-Atlantic region by 7.75%, in the East North Central region by 2.94%, in the West North Central region by 9.37%, in the South Atlantic region by 5.21%, and in the Mountain region by 4.00% relative to households located in the Pacific region. Household located in New England have a higher probability of purchasing food and beverages at drug stores by 0.79%, households located in the Mountain region have a lower probability of purchasing by 0.86%.

Table III-12. Conditional Marginal Effects of Household Food and Beverage Expenditures and Marginal Effects Associated with the Probability of Purchasing by Store Type Based on the High-Income Sample of Panel Households

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Household income	0.0000	0.005	0.0004	0.020	0.0094	0.111	-0.0167	-0.070	-0.0071	-0.560	0.0035	0.064
	(0.0000)	(0.016)	(0.0006)	(0.030)	(0.0044)	(0.053)	(0.009)	(0.038)	(0.0108)	(0.856)	(0.0031)	(0.064)
Household size	0.0000	1.772	0.0003*	15.370*	0.0034*	35.094*	0.0056*	1.026*	-0.0031	-19.494	-0.0011	-5.773
	(0.0000)	(5.382)	(0.0001)	(4.668)	(0.0007)	(7.877)	(0.0015)	(0.281)	(0.0017)	(10.730)	(0.0005)	(2.478)
Age<40	0.0000	-48.540	0.0003	13.729	-0.0013	-13.644	-0.0073	-1.344	0.0117	72.177	0.0057	30.024
	(0.0000)	(49.254)	(0.0008)	(40.560)	(0.0065)	(67.589)	(0.0133)	(2.457)	(0.014)	(87.335)	(0.0042)	(21.811)
Age>60	0.0000	4.162	-0.0012*	-60.060*	0.0045	47.063	0.0189*	3.478*	-0.0041	-25.433	0.0036*	18.882*
	(0.0000)	(15.000)	(0.0002)	(12.839)	(0.0021)	(21.631)	(0.0042)	(0.787)	(0.0047)	(29.174)	(0.0013)	(6.889)
Urban	0.0000	12.389	-0.0010*	-50.626*	-0.0016	-16.496	-0.0128*	-2.364*	-0.0277*	-171.672*	0.003	15.695
	(0.0000)	(14.599)	(0.0003)	(13.300)	(0.0022)	(23.368)	(0.0045)	(0.840)	(0.0049)	(36.246)	(0.0014)	(7.151)
Rural	0.0000	-53.922	0.0015*	78.142*	-0.0189*	-197.336*	0.0307*	5.650*	0.0068	41.905	-0.0006	-3.353
	(0.0000)	(23.213)	(0.0004)	(20.949)	(0.0036)	(38.753)	(0.007)	(1.311)	(0.0075)	(46.621)	(0.0022)	(11.400)
Less than high school	-0.0003*	-567.562*	0.0025	130.391	0.0129	135.008	-0.0123	-2.257	-0.0759	-470.219	-0.0001	-0.347
	(0.0001)	(127.928)	(0.002)	(106.636)	(0.0167)	(174.570)	(0.0352)	(6.474)	(0.0459)	(289.548)	(0.0109)	(56.908)
High school graduate	-0.0000	-60.883*	0.0011*	59.036*	-0.002	-20.516	0.0263*	4.843*	-0.0038	-23.459	-0.0011	-5.740
	(0.0000)	(23.398)	(0.0004)	(20.192)	(0.0033)	(34.439)	(0.0065)	(1.210)	(0.0073)	(45.474)	(0.0021)	(10.810)
College experienced	0.0000	-9.922	0.0006	31.290	-0.0003	-3.383	0.0134*	2.470*	0.0062	38.495	-0.0007	-3.822
	(0.0000)	(15.044)	(0.0002)	(12.874)	(0.0021)	(21.615)	(0.0042)	(0.768)	(0.0046)	(29.057)	(0.0013)	(6.890)
Non-Hispanic black	-0.0000*	-82.102*	0.0019*	99.292*	0.0145*	151.190*	0.0512*	9.424*	0.0098	60.765	0.0107*	55.921*
	(0.0000)	(22.247)	(0.0004)	(20.059)	(0.0034)	(35.718)	(0.0068)	(1.290)	(0.0074)	(46.479)	(0.0021)	(10.841)
Non-Hispanic Asian	-0.0000*	-103.066*	-0.0006	-33.536	0.0104	109.198	-0.0037	-0.676	-0.0173	-107.070	-0.0052	-27.315
	(0.0000)	(30.635)	(0.0005)	(27.520)	(0.0045)	(47.689)	(0.0097)	(1.777)	(0.0109)	(68.611)	(0.0028)	(14.870)
Non-Hispanic other	0.0000	-68.130	0.0011	58.916	-0.0021	-22.012	0.0186	3.425	0.0218	135.045	0.0045	23.424
	(0.0000)	(43.228)	(0.0007)	(36.766)	(0.0058)	(61.052)	(0.012)	(2.214)	(0.0134)	(84.547)	(0.0038)	(19.654)
Hispanic	0.0000	-40.154	0.0000	-0.965	0.0051	53.600	0.0316*	5.818*	0.022	136.124	0.0033	17.349
	(0.0000)	(27.676)	(0.0005)	(24.547)	(0.004)	(42.210)	(0.0083)	(1.537)	(0.0091)	(58.776)	(0.0025)	(13.215)

Table III-11. Continued

	Grocery		Discount		Club		Dollar		Convenience		Drug	
	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure	Probability to purchase	Expenditure
Club stores	0.0000	-41.033	0.0000	0.962	0.0074*	77.743*	0.0022	0.398	-0.0041	-25.514	-0.003	-15.876
	(0.0000)	(21.072)	(0.0003)	(15.377)	(0.0022)	(23.020)	(0.0046)	(0.851)	(0.0054)	(33.472)	(0.0015)	(8.084)
Convenience stores	0.0000	5.721	-0.0001	-3.238	0.0007	7.147	0.0005	0.088	0.0007	4.202	0.0001	0.463
	(0.0000)	(4.843)	(0.0001)	(3.552)	(0.0005)	(5.365)	(0.0011)	(0.196)	(0.0013)	(7.871)	(0.0004)	(1.845)
Grocery stores and Supercenters	0.0000	5.377	-0.0001	-6.510*	0.0003	3.647	-0.0019	-0.357	-0.0009	-5.319	0.0001	0.778
	(0.0000)	(3.929)	(0.0001)	(2.900)	(0.0004)	(4.380)	(0.0009)	(0.162)	(0.001)	(6.498)	(0.0003)	(1.490)
Drug stores	0.0000	-0.881	0.0000	2.426	-0.0011	-11.256	-0.0034*	-0.622*	0.0033	20.746	0.0004	1.844
	(0.0000)	(5.773)	(0.0001)	(4.238)	(0.0006)	(6.419)	(0.0013)	(0.238)	(0.0015)	(9.800)	(0.0004)	(2.200)
Ne	0.0000*	88.394*	-0.0015*	-79.574*	-0.0229*	-239.584*	-0.0234	-4.311	0.0320*	198.481*	0.0079*	41.392*
	(0.0000)	(32.650)	(0.0006)	(29.749)	(0.0051)	(53.471)	(0.0105)	(1.938)	(0.0112)	(73.489)	(0.0031)	(16.000)
Ma	0.0000	60.511	-0.0013*	-69.382*	-0.0258*	-269.601*	0.0153	2.814	0.0775*	480.215*	0.0027	14.038
	(0.0000)	(23.583)	(0.0004)	(21.490)	(0.0037)	(39.605)	(0.0074)	(1.372)	(0.0081)	(75.355)	(0.0022)	(11.552)
Enc	0.0000*	79.530*	-0.0009	-44.600	-0.0121*	-126.398*	-0.002	-0.368	0.0294*	182.136*	-0.0002	-0.947
	(0.0000)	(22.698)	(0.0004)	(20.623)	(0.0035)	(37.203)	(0.0072)	(1.320)	(0.008)	(53.810)	(0.0021)	(11.131)
Wnc	0.0000	-19.305	0.0015*	79.665*	-0.0075	-78.616	-0.0068	-1.247	0.0937*	580.387*	-0.0035	-18.405
	(0.0000)	(28.062)	(0.0005)	(25.626)	(0.0044)	(45.820)	(0.0089)	(1.636)	(0.0094)	(89.639)	(0.0026)	(13.827)
Sa	0.0000	36.087	0.0008	39.548	-0.0112*	-117.396*	0.0176	3.245	0.0521*	322.871*	0.0035	18.561
	(0.0000)	(22.120)	(0.0004)	(20.172)	(0.0034)	(35.658)	(0.007)	(1.287)	(0.0077)	(60.671)	(0.0021)	(10.843)
Esc	0.0000	46.765	0.0007	39.018	-0.0157*	-163.746*	0.0345*	6.356*	0.0165	102.116	-0.001	-5.414
	(0.0000)	(33.516)	(0.0006)	(30.439)	(0.0052)	(54.947)	(0.0103)	(1.915)	(0.0117)	(73.543)	(0.0031)	(16.481)
Wsc	0.0000*	85.914*	0.0008	42.197	-0.0113*	-118.521*	0.0136	2.502	0.0123	75.888	0.0013	6.740
	(0.0000)	(26.012)	(0.0005)	(23.684)	(0.004)	(42.001)	(0.0081)	(1.503)	(0.0092)	(57.545)	(0.0024)	(12.724)
Mt	0.0000	65.409	0.0004	19.736	-0.009	-94.356	-0.0195	-3.593	0.0400*	247.796*	-0.0086*	-44.793*
	(0.0000)	(29.575)	(0.0005)	(26.777)	(0.0045)	(46.782)	(0.0094)	(1.728)	(0.0102)	(69.363)	(0.0028)	(14.610)

Superscript * indicates p-value <0.01

Numbers in parentheses correspond to standard errors.

The marginal effect of the probability to purchase at any store outlet is given by $\frac{\partial P(EX_{ht}^{k*}>0|z_{ht}^k)}{\partial z_{ht}^k}$, and the marginal effect of the conditional expectation of household food and beverage expenditures is expressed mathematically as $\frac{\partial E[EX_{ht}^k|z_{ht}^k, EX_{ht}^k > 0]}{\partial z_{ht}^k}$. All calculations pertaining to marginal effects are made at the sample means of the data.

III.11 Concluding Remarks

A number of choices is evident beyond traditional supermarkets or grocery stores owing to the increasingly diverse U.S. retail food landscape. Despite the plethora of previous studies that largely focus on factors affecting store choice, one area of research that has received relatively little attention is how the magnitude of household food and beverage expenditures is impacted by the type of store outlets. In this light, the purpose of this study is to examine how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type and by income level. The list of socio-demographic factors includes: (1) household income; (2) household size; (3) age; (4) urbanization; (5) education; (6) race and ethnicity; and (7) region. Characteristics of the retail environment relate to the number of club stores, the number of convenience stores, the number of grocery stores and supercenters and the number of drug stores within the zip code area of the of the household. Whether traditional or non-traditional, store outlets differ in prices, product assortment, advertising strategies, and location (Volpe, Kuhns, and Jaenicke, 2017). The outlets considered in this study are grocery, convenience, discount, club, drug, and dollar store types.

As mentioned previously, prior works mainly highlighted store choice. To differentiate our study from the extant literature, we explore the factors which directly affect household food expenditure by store outlet. Indeed, Volpe, Jaenicke, and Chenarides (2018) estimated the impacts of expenditure share by store format, but in our study, we quantify the magnitude of the impact of household socio-demographics, the retail food environment, and spending habits on food and beverage expenditures by diverse store types. Hence, by analyzing factors that impact household food expenditure across the aforementioned six store types, this study contributes to the economic literature. Another contribution is that our study also considers habitual persistence or spending habits, a dynamic property of household expenditure on food and beverages. However, in the previously mentioned studies, habitual behavior was not included in the set of explanatory variables.

To further differentiate our study from previous studies, we employ a dynamic correlated random effect Tobit model to incorporate habitual purchasing behavior. The source of data for this analysis is the Nielsen Homescan Panel over the period between 2011 and 2015. Specifically, we use a balanced panel of 28,109 households who participated in the survey for all five years from 2011 to 2015. The total number of observations available for analysis is 140,545. The panel structure allows us to incorporate dynamic modeling by including lagged dependent variables as explanatory variables to account for spending habits.

Another advantage of the use of this model is that we are in a position to handle corner solution problems. The dependent variables reflect household purchasing history according to store type and indeed have zero values; hence the dependent variables are left censored. A differentiated feature of our empirical analysis relates to transforming the dependent variables which include zero observations using the inverse hyperbolic sine (arcsinh) method (Bellemare and Wichman 2020). A notable problem with taking the logarithm of any variable is that it does not allow retaining zero-valued observations because the ln (0) is undefined. As pointed out by Bellemare and Wichman (2019), "applied econometricians are typically loath to drop those observations for which the logarithm is undefined." Consequently, researchers often have resorted to ad hoc means of accounting for this situation when taking the natural logarithm of a variable, such as adding 1 to the variable prior to its transformation (MaCurdy and Pencavel, 1986). In recent years, the inverse hyperbolic sine (or arcsinh) transformation has grown in popularity among applied econometricians due to the fact that it is similar to the behavior of the logarithm function, it allows retaining zero-valued observations without any arbitrariness, and it often results in normal distributions (Burbidge et al. 1988; Yen and Jones 1997; MacKinnon and Magee 1990; Pence 2006; Van den Heuvel et al. 2011; Bellemere, Barrett, and Just 2013; Brown et al. 2015; Bellemere and Wichman 2020).

Importantly, we estimate separate dynamic correlated random effect Tobit models for sub-samples in accordance with household income level. Because households typically have different shopping baskets by income level (Taylor and Villas-Boas 2016; Volpe, Jaenicke, and Chenarides 2018), we compare and contrast our findings across income levels for each of the six store types considered in our study. We consider three distinct income categories—low, middle, and high. The low-income sample corresponds to those households whose annual income below is $25,000. The middle-income sample corresponds to those households whose annual income is above $25,000 but below $70,000. The high-income sample corresponds to those households whose annual income is above $70,000. We follow this segmentation of household income based on the work by Allcott, Diamond, and Dubé (2017). Hence, we estimate 24 different dynamic correlated random effect Tobit models, covering six store types and four data samples.

The results support the supposition of habitual spending across all store outlets. These results suggest that, within the data period 2011 to 2015, habitual spending behavior is undoubtedly a key factor in affecting nominal food and beverage expenditures across all store formats. This finding also holds across the three respective income sub-samples. Household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets even across the various income sub-samples. However, household size, age, urbanization, education, race and ethnicity, region, time-invariant socio-demographic variables, indeed are drivers of household food and beverage expenditures at the six store outlets across the income categories. This finding is in line with the hypothesis of underlying household heterogeneity and in agreement with the results of Bilsard, Stewart, and Jolliffe (2004) and of Taylor and Villas-Boas (2016).

Further, the number of convenience stores in the zip code area of households do not significantly influence the level of food and beverage expenditures across the respective store outlets and across the respective income categories. The same result is true for drug stores but for a single exception. In the high-income sample, the number of drug stores in the zip code area negatively impacts food and beverage expenditures made at dollar stores. In the entire sample and in the mid-income sample, the number of club stores negatively impacts household expenditures made at grocery stores and drug stores. But this finding is not the case within the low-income sample and within the high-income sample. In addition, in the entire sample, the number of grocery stores and supercenters in the zip code area negatively impacts household food and beverage expenditures made at discount stores. Nevertheless, this finding is not the case in each of the respective income sub-samples.

Bottom line, evidence exists to support the hypothesis that the retail environment plays a limited role in affecting household expenditures for food and beverages across store outlets and across income sub-samples. This result differs from previous findings by Kyureghian and Nayga (2013) and by Taylor and Villas-Boas (2016), but this result is in alignment with the work by Ver Ploeg and Wilde (2018).

The findings in this study make several contributions to the current economic literature. First, we provide a detailed view that describes household spending behavior across six store types for three income classifications. Second, the construction and estimation of dynamic random effect Tobit models constitute the first attempt in the literature dealing with household food and expenditure by store outlets for various income classifications. Third, we use a novel method to deal with problems in data (zero observations and extreme values) through the inverse hyperbolic sine transformation. Fourth, we derive the accompanying expressions for calculating conditional marginal effects and the marginal effects associated with the probability of purchasing food and beverages on the basis of the inverse hyperbolic sine transformation.

Future research in this area may center attention on specific household food and beverage expenditures rather than the aggregate, for example, fresh fruits and vegetables or meat products. Particularly for low-income households, we are in position to investigate nutrition intake of households associated with the six store types by income level. As such, this research may uncover a link between store type and nutrition intake, especially useful for policies dealing with various food assistance programs. Although this research covers the period 2011 to 2015, this study establishes a baseline. Our study can be replicated using more recent data to determine the robustness of our findings. Without question, because today's food retail environment is considerably diverse, more work is needed to understand the role of store outlets in affecting dietary quality in America across various income sub-samples.

CHAPTER IV

THE EFFECT OF IMMIGRATION POLICY REGIME CHANGE ON STATE-LEVEL WIC PROGRAM PARTICIPATION RATES

IV.1 Introduction

Within five days of taking office, President Trump issued a series of executive orders that promised major changes to the U.S. immigration system. These executive orders demonstrated the Trump administration's focus to make changes in border security and interior enforcement. Concerning border security[10], the construction of barriers along the southern border and zero-tolerance to all individuals crossing the border illegally were the predominant changes taken. In another executive order[11], a new interior enforcement regime was mentioned, expanding the classes of non-citizens who are priorities for removal and directing agencies to execute U.S. immigration laws against "all removable aliens." With this regime change, the Trump administration abandoned the prosecutorial discretion guideline under the Obama administration, wherein non-citizens prioritized for removal were only those who had criminal convictions, who recently crossed the border illegally, or who had been ordered removed. With changes in immigration enforcement, the Trump administration made policies that were disadvantageous to

---

[10] Trump, D. (2017,January 25). Presidential executive order on enhancing Public Safety in the Interior of the United States. White House Press Office. Available online: https://www.whitehouse.gov/presidential-actions/executive-order-enhancing-public-safety-interior-united-states/

[11] Trump, D. (2017,January 25). Presidential executive order on Border Security and Immigration Enforcement Improvements. White House Press Office. Available online: https://www.whitehouse.gov/presidential-actions/executive-order-border-security-immigration-enforcement-improvements/

non-citizens who hold a legal visa or permanent residence status and to unauthorized immigrants. For example, aliens who applied for adjustment of status or extension of stay who receive public benefits, such as Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance to Needy Families (TANF), Medicaid, can be denied their application by United States Citizenship and Immigration Service (USCIS) due to Inadmissibility on Pubic Charge Grounds final rule[12]. Also, those aliens are inadmissible to the United States and ineligible to become a lawful permanent resident (Green Card).

Scholars have studied how immigration policy change affects the fear of deportation of non-citizens. Hispanic families have deportation fear due to their immigration status, affecting their food security status, school enrollment, and access to social benefits (Berk and Schur, 2001; Jefferies, 2014; Sullivan and Enriquez, 2015; Becerra, 2016). Unlike other social programs such as TANF, SNAP, and Medicaid, the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and the National School Lunch Program (NSLP) have no eligibility restrictions based on applicants' immigration status or legal status. Immigrants face fewer barriers to WIC and NSLP programs relative to other social programs (Vericker et al., 2010). Vargas and Pirog (2016) reported decreases in the participation rate in WIC attributed to increases of deportation fear.

Because of recent changes in immigration policy, fear of deportation or losing legal immigration status of non-citizens has been growing (Hing, 2018; Torres et al., 2018; Tummala-Narra, 2019; Alif et al., 2019, Fleming, 2019). On the basis of the UCLA Luskin Los Angeles

---

[12] Inadmissibility on Public Charge Grounds final rule was begun in February 24th in 2020. However, USCIS stopped applying this rule on and after March 9, 2021.

County Quality of Life Index,[13] more than one-third of Los Angeles County residents were concerned about deportation of their immigrant's friends and family members, and almost half of the county residents believed that a new federal health law proposed under the Trump administration may make them hard to access health care programs. The fear of deportation clearly affects decision-making of non-citizens as to whether to participate in social benefits and food assistance programs during the Trump administration (Bleich and Fleischhacker, 2019; Callaghan et al., 2019, Laird et al., 2019). Non-citizens avoid revealing their status information to the government authority because revealing this information may increase their risk to be deported.

This effect may be larger in the Hispanic community because almost half of immigrants are from Mexico, the Caribbean, and South America[14], and more than 30 percent of the population of undocumented immigrants is of Hispanic ethnicity[15]. Watson (2014) and Alsan and Yang (2019) reported decreases in Hispanic participation of social benefit programs after implementing specific immigration policies. Specifically, Alsan and Yang (2019) detected direct and indirect effects of immigration policy changes of Hispanics. The direct effect is the effect from immigration policy change within the non-citizen Hispanics population. But immigration policy changes also may affect citizens Hispanic households (indirect effect) because of concern about their non-citizen Hispanic neighbors. Callaghan et al. (2019) also highlight that participation in health care

---

[13] UCLA Luskin Los Angeles County Quality of Life Index is a project of the Los Angeles initiative at the UCLA Luskin school of Public Affairs in partnership with the California Endowment and the public opinion research firm Fairbank, Maslin, Maulin, Metz & Associates. This annual survey is based on interviews from 1,600 residents of Los Angeles County from February. 28 to March. 12 2017.

[4] Pew Research Center, https://www.pewresearch.org/?p=290738

[15] Pew Research Center, https://www.pewresearch.org/?p=290738

programs in Texas by Hispanics has decreased due to immigration policies enacted by the Trump administration. Another study deals with decreases in SNAP participation after implementation of immigration policy changes (Laird et al., 2019).

IV.2 Objectives

This work deals with investigating how fear of deportation from immigration policy changes affects participation in food assistance programs of non-citizens. We raise two research questions. First, does immigration policy change affect noncitizens' public benefit participation rate.? Despite many articles [16] from the popular press which addressed negative impacts of immigration policy regime change during the Trump administration on non-citizen households' public benefit participation, these claims have not yet been substantiated. In fact, the decreasing pattern in public benefits participation rate may be caused by other factors, such as changes in income, employment status, or immigration policy. Therefore, a systematic analysis done via regression analysis concerning public benefit participation incorporating relevant factors is needed to identify and assess the impact of immigration policy change. Second, how does immigration policy change affect Hispanics' public benefits participation.? The impact of immigration policy change may vary by race and ethnicity of non-citizens. Hispanics occupy a large portion of the non-citizen population in United States. As well, more than 30 percent of undocumented immigrants are Hispanics.

---

[16] Torbati, Yebaneh. 2018. "Exclusive: Trump administration may target immigrants who use food aid, other benefits." *Reuters*, in press.

Thrush, Glenn. 2018. "Spooked by Trumps proposals, Immigrants abandon public nutrition services." *The Ner York Times*, in press.

To identify the fear effect associated with public benefit participation, we focus exclusively on changes in WIC program participation rates for a couple of reasons. First, regardless of the specific immigration status of non-citizen, all non-citizens who meet income and categorical requirements are eligible to participate in the WIC program. But, to be approved for other social benefits (SNAP, Medicaid, TANF, and SSI), non-citizen applicants have to be 'qualified-aliens'. Second, any change in the WIC participation rate after immigration policy changes can be considered as a fear effect. Non-citizens who have benefited from participation in the WIC program are not targeted by any immigration policy change after the Trump Administration. The only policy revised by the Trump administration that related to use of social benefits is the Public Charge rule implemented by USCIS (United States Citizenship and Immigration Services). But the WIC program is not considered concerning the revised Public Charge rule[17]. So, although actual policy changes in the Trump administration are not related to the WIC program, decreases in the WIC participation rate by non-citizens after immigration policy changes may reflect the effect of fear of deportation from non-citizens.

To address the previously mentioned research questions, we estimate the change in immigration policy pre- versus post- Trump administration in state-level WIC participation rates by citizenship and ethnicity. We use the Triple Difference (Difference-in-Difference-in-Difference) methodology to compare the program participation rate for non-citizen Hispanic households to the participation rate for non-citizen non-Hispanic and citizen non-Hispanic households before versus after the Trump administration. We use data from the CPS-ASEC

[17] Department Homeland Security (DHS) are considering public benefits in the Public Charge rule as listed: SNAP, Medicaid, TANF, and SSI,

(Current Population Survey- Annual Social and Economic Supplement), publicly accessible at the IPUMS (Integrated Public Use Microdata Series)[18] website, to estimate differences of WIC participation rates of non-citizens between the second Obama administration and the Trump administration (2013-2018). The CPS-ASEC data provide repeated cross-sections surveyed every March by different panelists in each year from 2013 to 2018. The CPS-ASEC data also provide socio-demographic information, including income and citizenship, region up to the county level, social benefit participation in various programs (e.g., SNAP, Medicaid, TANF, SSI, and WIC), and employment status of survey participants.

We hypothesize that if eligible non-citizens express fear of deportation from immigration policy changes, those individuals forfeit participating in food assistance programs. Moreover, we investigate whether fear of deportation affects specific ethnic groups. As previously mentioned, Hispanics may be more prone to fear of deportation than other ethnic groups. Furthermore, we identify how fear of deportation affects non-citizens by different immigration status. Reactions of non-citizens to immigration policy changes may vary by their legal status (legal and illegal immigrants) because government authorities such as ICE (U.S. Immigration and Customs Enforcement) have the authority to remove undocumented immigrants. We hypothesize that non-citizens who possess immigration status that does not guarantee stable residence in the United States may have deeper fears of deportation from recent immigration policy changes.

---

[18] Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D030.V8.0

This study extends the existing literature by estimating causal effects of immigration policy regime change on WIC participation of non-citizens. Because immigration policy has been changed under the Trump administration, identifying recent trends in the WIC participation rate of non-citizens is important. Second, our research design investigating the fear of deportation on WIC participation by immigration status and ethnicity provides a clearer understanding of WIC participation by non-citizens.

IV.3 Organization

In next section, we provide information about WIC program eligibility and immigration status. In sections three and four, we discuss theoretical frameworks and empirical strategies to estimate impacts of fear change on WIC participation of Hispanic non-citizens by comparing pre- and post-policy regime changes. In fifth section, we present our estimation results, and we provide robustness checks. In last section, we provide a summary of the research as well as suggestions for further research

IV.4 WIC Eligibility and Immigration status

The WIC program is the third largest food assistance program in the United State (Oliveira and Frazao, 2015). All applicants have to satisfy four eligibility requirements to get this benefit: categorical, income, residential, and nutrition risk requirements. All eligibility requirements and regulations are issued and monitored by USDA FNS (U.S. Department of Agriculture, Food and Nutrition Service).

Categorical requirement relates to the target group of people that may apply and be selected as participants. The WIC program is designed for serving certain categories of women, infants, and children.[19] The residential requirement is about checking the residency of applicants. Applicants also must live in the state in which they apply. To pass the income requirement, individuals must have household income below the 185% poverty threshold. The income guideline is the poverty guideline updated annually by the Department of Health and Human Services (DHHS). Certain applicants who already have other social benefits such as SNAP, Medicaid, TANF, and certain other state-administered programs are eligible to participate in WIC automatically. For the nutrition risk requirement, applicants must be seen by a health professional such as a physician, nurse, or nutritionist. In most cases, this is done in the WIC clinic at no cost to the applicant.

The eligibility criteria concerning other social benefits have strict regulation requirements for unqualified and undocumented non-citizens. Initially, immigrants were not eligible for social benefits due to the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA, PL 104-193) enacted in 1996. After then, several other relevant laws[20] were enacted to restrict

[19] Categories of women include pregnant women during pregnancy and up to 6 weeks after the birth of infant or the end of pregnancy, postpartum women who in six months duration after the birth of infant or end of pregnancy, and breastfeeding women who has infants age up to 1. Category for infants is defined as the infants whose age up to 1. Lastly, children age up to 5 are eligible to receive WIC benefits.

[20] Balanced Budget Act of 1997 (Public Law 105-33), Agricultural Research, Extension and Education Reform Act of 1998 (Public Law 105-185), Noncitizen Benefit Clarification and Other Technical Amendments Act of 1998 (Public Law 105-306), Trafficking Victims Protection Act of 2000 (Public Law 106-386), Food Stamp Reauthorization Act of 2002 (Public Law 107-171), SSI Extension for Elderly and Disables Refugees Act (Public Law 110-328), Food and Nutrition Act of 2008 (Public Law 88-525), and Children’s Health Insurance Program Reauthorization Act of 2009 (Public Law 111-3).

legal immigrants to access some programs and undocumented immigrants to access most of government-funded programs, such as SNAP, TANF, and Medicaid.

Before PRWORA was signed into law, legal immigrants were allowed to access most of U.S. social programs. However, in PRWORA, the distinction of "qualified-aliens" and "unqualified-aliens" was established. The definition of "qualified-aliens" includes legal permanent residents, asylees, refugees, aliens paroled into the United States for at least one year, aliens granted conditional entry, battered alien spouses, battered alien children, alien parents of battered children, alien children of battered parents who fit certain criteria, Cuban and Haitian entrants, and victims of a severe form of trafficking. Even if aliens are classified as "qualified-aliens", aliens are ineligible to receive federal public benefits such as, SSI (Supplemental Security Income), TANF, and Medicaid for their first five years in the United States as qualified aliens.

But exceptions were made concerning the WIC program and the National School Lunch Program (NSLP). For these programs, any alien can apply and get the benefits if they satisfy all categorical, income, residence, and nutritional risk eligibility requirements. WIC and NSLP are the federal programs which use the poverty guideline, updated annually for eligibility.

In Table IV-1, we describe eligibility by immigration status for six public benefit programs such as WIC, NSLP, SNAP, Medicaid, TANF, and SSI. Illegal immigrants are able to access WIC and NSLP. WIC and NSLP benefits were not restricted to access by illegal and "unqualified-aliens" immigrants by PRWORA. For the case of the WIC program, even if program applicants are illegal immigrants, state agencies or clinics do not have any obligation to report personal information of all applicants including immigration status to federal authorities such as USDA FNS who administer WIC program or immigration authorities (7 CFR. § 246.26). In some cases, WIC state agencies can disclose confidential information of WIC applicants and participants. But state agencies need authorization of applicants and participants to reveal this information by federal WIC regulation (7 CFR. § 246.26, paragraph (h)).

In addition, public education (K-12) has been available for illegal immigrants (Plyer v Doe, 457 U.S. 202 1982). So, the eligibility guideline of NSLP relates to all students attending public school including illegal immigrants from households whose income meets the income eligibility criteria of the NSLP. Due to the Family Educational Rights and Privacy Act (FERPA), schools are forbidden to disclose confidential information of students to non-school persons including immigration authorities (20 U.S.C. § 1232g; 34 C.F.R. Part 99).

Table IV-1. Eligibility by Immigration Status.

Programs	Illegal immigrant	Legal Immigrant	Permanent resident	Naturalized citizen	Native citizen
WIC	√	√	√	√	√
NSLP	√	√	√	√	√
SNAP		√*	√	√	√
Medicaid		√*	√	√	√
TANF		√*	√	√	√
SSI		√*	√	√	√

[1]Illegal and legal immigrants well as permanent residents are categorized as immigrants.

[2]Naturalized citizens are citizens who are not native and born in U.S. but achieve citizenship by the naturalization process.

[3]Native citizens are citizens who are native and born in the United States.

[4]Asterisk(*) indicates that only qualified aliens can receive benefits such as SNAP, Medicaid, TANF, and SSI.

IV.5 Theoretical Framework

Specific partitions of the non-citizen population may have different weights related to fear of deportation. Undocumented immigrants may be more concerned about fear of deportation than other non-citizens. On the other hand, non-citizens that hold a relatively stable legal status such as permanent resident or non-immigrant visa also may express fear about proceeding to achieve permanent residency or naturalization to be a U.S. citizen. Pursuing next-level legal status by a non-citizen is a common path to reside in the United States with stable status (Bruno, 2014).

Previous studies (Vargas and Pirog, 2016; Alsan and Yang, 2019) defined and regarded fear as simply deportation fear. However, this definition may be problematic in misinterpreting the behavior of different people in different visa statuses within the non-citizen population. The non-citizen population includes all people who hold permanent residence status, who hold non-immigrant visas and who are undocumented (unauthorized) immigrants. But no data exist which provide detailed information about the legal status of non-citizens because there are no accurate devices to separate these categories of the non-citizen population.

Therefore, we define fear following equation (1) which indicates that fear is a weighted sum of deportation and disadvantage. Deportation fear can be defined as the fear that comes from the fact that they may be deported by participating in the WIC program. Disadvantage fear is formed by anxiety that they may have the disadvantage to proceed to naturalize as a citizen or achieve permanent residence status (green card) after obtaining social benefits.

$$fear_i = w_i^p \, Deportation_i + w_i^d \, Disadvantage_i \qquad (1)$$
$$where \; w_i^p + w_i^d = 1$$

The subscript i represents the household head. The superscripts p and d indicate deportation and disadvantage. The variable $fear_i$ indicates the overall level of fear by household head i. $Deportation_i$ and $Disadvantage_i$ are deportation and disadvantage fear variables of household head i. w_i^p and w_i^d are weights that assigned by the household head i and the sum of these weights is unity. Through our empirical analysis, our main goal is identifying the magnitude and sign of coefficient representing $fear_i$ of non-citizens after the change in immigration policy implemented by the Trump administration.

IV.6 Theoretical Model

We adopt the Moffit (1983) model as well as the Alsan and Yang (2019) model of non-participation in social programs. Unlike Alsan and Yang (2019), we account for citizenship (C_j) of the household head. In the extant literature (Watson 2014; Vargas and Pirog 2016; Alsan and Yang 2019), two separate models, using county level data, historically have been constructed according to the citizenship status of the household head. These models have centered attention on the effects of deportation fear within a household with the same citizenship status. But the goal of this research is to detect the impact of immigration policy regime change by comparing the periods pre- and post- Trump administration focusing not only on citizenship but also ethnicity. We address policy and unobserved differences among states by using state level data and incorporating fixed effects in our empirical model.

We start from their model to formalize how immigration policy change can lead to changes in WIC participation. Potential participants maximize expected household utility by placing weights on each set of family members- citizens and non-citizens. The expected utility of household welfare from WIC participation is a weighted sum of household members' welfare gains by their citizenship and eligibility of each household member. We can derive the expected utility function formed as expressed in equation (2).

$$EU_{ijs} = \lambda_i\big(Y_j - \pi_{js}(p_{ijs}) * C_j\big) + \lambda_c(Y_j) + \lambda_{nc}\left(Y_j - \pi_{js}(p_{ijs})\right) + \lambda_{ce}\left(Y_j + P_{ijs}(B_{ce})\right) + \lambda_{nce}\left(Y_j + P_{ijs}(B_{nce}) - \pi_{js}(p_{ijs})\right) \tag{2}$$

Subscript j refers to households, and subscript s refers to household location (e.g. state). Subscript i corresponds to the household head, subscript c indicates citizen household member but not eligible to participate in the WIC program, nc indicates non-citizen household member with no eligibility, ce indicates household member with citizenship and eligibility, and nce indicates non-citizen household member but eligible to participate in the WIC program. Lambda coefficients are weights assigned to household members by citizenship and WIC eligibility. These weights show the influence of each household member in the decision-making process of the household. The sum of all weights is unity ($\lambda_i + \lambda_c + \lambda_{nc} + \lambda_{ce} + \lambda_{nce} = 1$). π_{js} represents the probability to have a deportation or disadvantage owing to immigration status and is an increasing function by WIC participation. If a household member has non-citizen status, their utility decreases due to the presence of deportation or disadvantage probability (π_{js}). For example, undocumented immigrants (illegal immigrants), have the fear of deportation, whereas non-citizens who hold green cards (permanent resident status) have the fear of disadvantage to proceed to naturalize as a citizen. P_{ijs} is an indicator variable reflecting the decision by the household head to participate or to not participate in the WIC program. If the household head decides to participate in the WIC program, this indicator variable takes on the value of one, but if the household head decides to not participate in the WIC program the indicator variable takes on the value of zero. B_{ce} and B_{nce} are the per capita benefits to eligible citizen (ce) and eligible non-citizen (nce) household members. C_i represents the citizenship status of the household head. If household head is a citizen, C_i takes on the value of one, but if the household head is not a citizen, then C_i takes on the value of zero.

The specific description of each variable is illustrated in Table 2.

The first term of equation (2), $\lambda_i\left(Y_j - \pi_{js}\left(p_{ijs}\right) * C_i\right)$, is associated with the expected utility of household head. Expected utility is determined by total household income and the probability of deportation and disadvantage. If the household head is non-citizen (C_i=1), the expected utility of the household decreases. The second term, $\lambda_c\left(Y_j\right)$, is expected utility for non-citizen household members who are not eligible for the WIC program. Also, because they are non-eligible to participate in the WIC program, no benefits are possible by participating in the WIC program. The third term, $\lambda_{nc}\left(Y_j - \pi_{js}\left(p_{ijs}\right)\right)$, is the expected utility for WIC non-eligible and non-citizen household members. Even though these household members are non-eligible for participation in the WIC program, their expected utility decreases ($-\pi_{js}\left(p_{ijs}\right)$) because there is a non-zero probability of deportation and disadvantage due to their immigration status. The fourth term, $\lambda_{ce}\left(Y_j + P_{ijs} * B_{ce}\right)$, is the expected utility for eligible and citizen household members. Because they are citizens, there is no probability of deportation and disadvantage. In addition, they receive benefits from WIC participation (B_{ce}). The last term in equation (2), $\lambda_{nce}\left(Y_j + P_{ijs} * B_{nce} - \pi_{js}\left(p_{ijs}\right)\right)$, represents the expected utility for eligible non-citizen household members. Theyy receive benefits from participating in the WIC program due to their eligibility, but these non-citizens have a non-zero probability of deportation and disadvantage.

Table IV-2. Variable Description

Variables	Description
EU_{ijs}	Expected household utility
Y_j	Total household income
C_i	Indicator of citizenship (0 citizen and 1 non-citizen) of household head
π_{js}	Probability of deportation or disadvantage
p_{ijs}	Program participation of household j, 1 participation and 0 non-participation
B_{ce}	Per capita benefit from participation given to WIC eligible member who is citizen
B_{nce}	Per capita benefit from participation given to WIC eligible member who is non-citizen
λ_i	Welfare weights head i gives to his/her own utility
λ_c	Welfare weights that head i gives to WIC non-eligible member who is citizen
λ_{nc}	Welfare weights that head i gives to WIC non-eligible member who is non-citizen
λ_{ce}	Welfare weights that head i gives to WIC eligible member who is citizen
λ_{nce}	Welfare weights that head i gives to WIC eligible member who is non-citizen

Let the probability of disadvantage fear by participating in the WIC program be denoted as $\pi_{js}(1)$. Then $\pi_{js}(0)$ indicates the probability of disadvantage fear when not participating in the WIC program. The household head decides to participate in the WIC program when the expected utility of participation exceeds the expected utility on non-participation ($EU_{ijs}[\pi_{js}(1)] - EU_{ijs}[\pi_{js}(0)] > 0$). We assume that there are no differences between per capita benefits from WIC participation by citizenship ($B_{ce} = B_{nce}$). Then, using this assumption and the relationship between expected utilities, we can re-write equation (2) and derive the inequalities given in equations (3) and (4).

$$\begin{aligned} Y_j + (\lambda_{ce} + \lambda_{nce})B_{ce} - \left(\lambda_i * C_j + \lambda_{nc} + \lambda_{nce}\right)\pi_{js}(1) \\ > Y_j + (\lambda_i * C_j + \lambda_{nc} + \lambda_{nce})\pi_{js}(0) \end{aligned} \tag{3}$$

$$\left[\frac{(\lambda_{ce} + \lambda_{nce})}{(\lambda_i * C_j + \lambda_{nc} + \lambda_{nce})}\right] B_{ce} > \pi_{js}(1) - \pi_{js}(0) \tag{4}$$

Equation (3) compares total benefits post- participation and pre- participation in the WIC program. Finally, equation (4) shows the difference of deportation or disadvantage probability when participating in the WIC program and the deportation or disadvantage probability when not participating in the WIC program is less than $\left[\frac{(\lambda_{ce}+\lambda_{nce})}{(\lambda_i*C_j+\lambda_{nc}+\lambda_{nce})}\right] B_{ce}$. The difference of probabilities $(\pi_{js}(1) - \pi_{js}(0))$ represents the disadvantage of non-citizen household members if the household participates in the WIC program compared to no participation ($\Delta\pi_{js}(p_{ijs}) = \beta D_s + \epsilon_{js}$). D_s is a location-specific immigration policy change and ϵ_{js} is error term that follows an F-distribution (Moffit, 1983; Alsan and Yang, 2019).

We let the threshold term $\left[\frac{(\lambda_{ce}+\lambda_{nce})}{(\lambda_i*C_j+\lambda_{nc}+\lambda_{nce})}\right] B_{ce}$ be labeled as r_s. Then, we can express the average household threshold, r_s, to participate in the program within the state as $\overline{r_s}$. Inequality (4) can be re-written as $\overline{r_s} - \beta D_s > \epsilon_s$. Since ϵ_s is a random variable and follows an F-distribution, the probability that the random variable ϵ_s is less than $\overline{r_s} - \beta D_s$ can be expressed using the cumulative distribution function (CDF), $P(\overline{r_s} - \beta D_s > \epsilon_s) = F(\overline{r_s} - \beta D_s)$. This probability indicates the average household participation share associated with the WIC program in each state.

Let S_s the share of participation rate associated with the WIC program in state s[21], we can derive equation (5). Equation (5) then serves as our theoretical model to apply in the empirical work.

$$S_s = F(r_s - \beta D_s) \tag{5}$$

The participation rate is an increasing function of the amount of benefits (B_{ce}, B_{nce}) and the weight assigned to all program eligible household members regardless of citizenship. As well, the participation share depends upon immigration policy (D_s) and on weights assigned to all non-citizen household members. If the household head is a non-citizen, the threshold r_s will decrease. Consequently, the state-level program participation rate (S_s) will decrease.

IV.7 Empirical Strategy

We use the triple difference (Difference-in-Difference-in-Difference or DDD) estimator proposed by Imbens and Wooldridge (2007) to exploit different reactions after changes in immigration policy between citizen and non-citizens as well as between Hispanics and non-Hispanics. The Triple Difference estimator has been widely used to measure impacts of policy effects (Gruber, 1994; Gruber and Poterba, 1994; Yelowitz, 1995; Ravallion, 2007; Hornbeck, 2010; Kleven et al., 2013; Muehlenbachs et al., 2015; Hoynes et al., 2016; Nilsson, 2017).

The Triple Difference estimator is an expanded version of the Difference-in-Difference (DD) estimator. In the simple case of the DD estimator, two time periods are evident, namely the

---

[21] The share of participation for the non-citizen population is calculated by dividing the number of non-citizen WIC participants by the total number of non-citizens in each state. The share of participation for the citizen population is calculated by dividing the number of citizen WIC participants by the total number of citizens in each state.

Obama second administration (pre-immigration policy, 2013-2016) and the Trump Administration (post-immigration policy, 2017-2018) There are two groups, a control group and a treatment group. The sole purpose of the DD estimator is to capture the difference between the control and the treatment group after the change in immigration policy. The Triple Difference (DDD) estimator adds another dimension dealing with citizens/non-citizens. The control group corresponds to citizens and the treatment group corresponds to non-citizens. Finally, we add other groups related to ethnicity, Hispanics and non-Hispanics. Therefore, the chief goal in our empirical analysis is to capture the change in the WIC participation rate of non-citizens and Hispanics after the change in immigration policy.

Generally, the literature associated with DD or DDD estimators attempts to capture policy effects using indicator or dummy variables to represent control and treatment groups, as well as to represent pre- and post- changes in immigration policy. In our model, the indicator variable associated with pre- and post-Trump administration is zero before 2017 and one after 2017. The indicator variable related to citizenship is zero for citizens and one for non-citizens, and the indicator variable for ethnicity is zero for Hispanics and one for non-Hispanics.

IV.7.1 Empirical Model

Equation (6) depicts the empirical model based on state-level data for the covering the years 2013-2018.

$$\begin{aligned} Y_{scht} = \alpha_0 + \alpha_1 I_t^{post} + \alpha_2\left(I_t^{post} * I_c\right) + \alpha_3\left(I_t^{post} * I_c * I_h\right) + \alpha_4\left(I_t^{post} * I_h\right) \\ + \alpha_5(I_c * I_h) + \alpha_6 IN_{scht} + \alpha_7 FS_{scht} + \alpha_8 ES_{scht} + \theta_c + \kappa_h + \mu_s \\ + \phi_t + \epsilon_{scht} \end{aligned} \tag{6}$$

where s indicates the state, c indicates citizenship, h indicates ethnicity, and t represents the yearly time period; the dependent variable Y_{scht} corresponds to the share of WIC program participation among eligible households. In each year and each state, there are four groups of households; citizen Hispanic, citizen non-Hispanic, non-citizen Hispanic, and non-citizen non-Hispanic. So, given 50 states, two types of citizenship and two types of ethnicities, and six years (2013-2018), the number of observations available for analysis is 1,200. I_t^{post} is an indicator variable with zero during the pre-policy regime period (2013-2016), and with one during the post-policy regime period (2017-2018). I_c and I_h are indicator variables for citizenship and ethnicity. I_c is equal to zero for the citizen Hispanic and for the citizen non-Hispanic group in each state and year. I_c is equal to one for the non-citizen Hispanic and for the non-citizen non-Hispanic in each state and year. I_h is equal to zero for the citizen non-Hispanic and for the non-citizen non-Hispanic group in each state and year. I_h is equal to 1 for the citizen Hispanic and for the non-citizen Hispanic groups in each state and year. IN_{scht}, FS_{scht}, and ES_{scht} correspond to average log-transformed income, average household size, and share of employed household heads by the aforementioned four groups in each state and year.

We also incorporate fixed effect dummy variables in our model to capture unobserved factors that may influence state-level WIC participation rates. The fixed effect (θ_c) accounts for the differential effect associated with citizens and non-citizens attributed to other policies or changes in economic conditions. The fixed effect (κ_h) captures unobserved changes in behavior associated with Hispanic and non-Hispanics. The fixed effect (μ_s) accounts for state-specific government policies. The fixed effect (ϕ_t) captures different policies that affect WIC participation rates such as State Vendor Authorization. These fixed effect variables also capture differentiated effects associated with immigration policy due to the fact that each state's immigration laws are different.

IV.7.2 Identification

Our main interest lies with the estimation results for coefficient the α_3 in equation (6) which shows the difference in participation rate between pre- and post-policy regime change by citizenship and ethnicity of the household head. The estimated coefficient α_3 can be expressed as equation (7), and this expression shows how we identify the triple difference estimator in our model. $\bar{Y}_{s,c,h,post}$ is the outcome when indicator variable I_c, I_t^{post}, I_h are equal to one. $\bar{Y}_{s,nc,h,post}$ is the outcome when indicator variable I_t^{post}, I_h are equal to one and I_c is equal to zero.

$$\hat{\alpha}_3 = \{(\bar{Y}_{s,c,h,post} - \bar{Y}_{s,c,h,pre}) - (\bar{Y}_{s,nc,h,post} - \bar{Y}_{s,nc,h,pre})\} \\ - \{(\bar{Y}_{s,c,nh,post} - \bar{Y}_{s,c,nh,pre}) - (\bar{Y}_{s,nc,nh,post} - \bar{Y}_{s,nc,nh,pre})\} \quad (7)$$

The first term in equation (7), $(\bar{Y}_{s,c,h,post} - \bar{Y}_{s,c,h,pre}) - (\bar{Y}_{s,nc,h,post} - \bar{Y}_{s,nc,h,pre})$, is the DD estimator if we focus only on Hispanics and use citizenship as control group. The second term$(\bar{Y}_{s,c,nh,post} - \bar{Y}_{s,c,nh,pre}) - (\bar{Y}_{s,nc,nh,post} - \bar{Y}_{s,nc,nh,pre})$ is a DD estimator if we focus only on non-Hispanics. The DDD estimator can be calculated by differencing those two terms. Because the trends in WIC participation rates by Hispanics and non-Hispanics are different, we incorporate citizenship indicator and Hispanic indicator in our model at the same time to capture policy effects on state-level WIC participation rates.

The coefficient α_2 in equation (6) is the DD estimator ($(\bar{Y}_{s,c,post} - \bar{Y}_{s,c,pre}) - (\bar{Y}_{s,nc,post} - \bar{Y}_{s,nc,pre})$) that compares pre- and post-reaction by citizenship, not considering ethnicity. If coefficient α_2 is statistically significant, then state-level WIC participation rates of two groups (non-citizen Hispanics, and non-citizen non-Hispanics) are influenced by the change in immigration policy during the Trump administration. If the coefficient α_2 and α_3 both are statistically significant, these results represent the fact that all non-citizens are affected by change in immigration policy. If the coefficient α_2 is statistically significant but α_3 is not statistically significant, only non-citizen non-Hispanics are affected by the change in immigration policy. If the coefficient α_3 is the only statistically significant coefficient, only non-citizen Hispanics are influenced by the change in immigration policy.

IV.8 Data

As mentioned previously, we use publicly accessible CPS- ASEC (Current Population Survey, Annual Social and Economic Supplement) data from IPUMS-CPS database managed by University of Minnesota. The CPS is a monthly U.S. household survey conducted by the U.S. Census Bureau and the Bureau of Labor Statistics. Over 65,000 households participate in this survey. This survey gathers information about employment status and socio-demographics of the U.S population. ASEC is a supplement of CPS dealing with special topics collected in March every year. The ASEC data also gather information on topic of detailed employment status and social benefits participation status. This supplemental data contains information about WIC participation and citizenship status. The data also include individual socio-demographics and other social benefits participation. Other data such as the SIPP (Survey of Income and Program Participation) and PSID (Panel Study of Income Dynamics) also contain information about citizenship and WIC participation of household. But the SIPP does not provide annual data after 2014 yet, and PSID only collects data from 11,000 families, and consequently contains fewer observations than the ASEC. Currently, the most up to date year of ASEC data released is the wave from April 2018 to March 2019 which is suitable for our analysis that allows the comparison of changes in immigration policy change attributed to the Trump administration.

We focus on data period between 2013 to 2018 to compare before (2013-2016) and after (2017-2018) the immigration policy regime change. We construct state level data to estimate the previously described empirical model. The CPS-AEPC data contain surveyed information of each household member.

We calculate the state-level WIC participation rate, the dependent variable in the empirical model, by dividing the number of WIC participants by the eligible population of WIC participants in each state. To derive state-level household income, we average household income levels for each state using the state FIPS[22] code variable which assigns a unique number for each state. We construct the share of employed individuals in each state by dividing the total number of employed individuals by the number of individuals in the state. Household size corresponds to the average household size in each state.

IV.8.1 Measuring WIC Participation Rate

As mentioned previously, we use publicly accessible CPS- ASEC (Current Population Survey, Annual Social and Economic Supplement) data from IPUMS-CPS database managed by University of Minnesota. The CPS is a monthly U.S. household survey conducted by the U.S. Census Bureau and the Bureau of Labor Statistics. Over 65,000 households participate in this survey. This survey gathers information about employment status and socio-demographics of the U.S population. ASEC is a supplement of CPS dealing with special topics collected in March every

---

[22] State FIPS (Federal Information Processing Standard) code are developed by National Institute of Standard and Technology (NIST). State FIPS codes are the assigned unique number to each state. We use this code to calculate average household income by state in STATA 15 using the "by" command.

year. The ASEC data also gather information on topic of detailed employment status and social benefits participation status. This supplemental data contains information about WIC participation and citizenship status. The data also include individual socio-demographics and other social benefits participation. Other data such as the SIPP (Survey of Income and Program Participation) and PSID (Panel Study of Income Dynamics) also contain information about citizenship and WIC participation of household. But the SIPP does not provide annual data after 2014 yet, and PSID only collects data from 11,000 families, and consequently contains fewer observations than the ASEC. Currently, the most up to date year of ASEC data released is the wave from April 2018 to March 2019 which is suitable for our analysis that allows the comparison of changes in immigration policy change attributed to the Trump administration.

We focus on data period between 2013 to 2018 to compare before (2013-2016) and after (2017-2018) the immigration policy regime change. We construct state level data to estimate the previously described empirical model. The CPS-AEPC data contain surveyed information of each household member.

We calculate the state-level WIC participation rate, the dependent variable in the empirical model, by dividing the number of WIC participants by the eligible population of WIC participants in each state. To derive state-level household income, we average household income levels for each state using the state FIPS[23] code variable which assigns a unique number for each state. We construct the share of employed individuals in each state by dividing the total number of employed

[23] State FIPS (Federal Information Processing Standard) code are developed by National Institute of Standard and Technology (NIST). State FIPS codes are the assigned unique number to each state. We use this code to calculate average household income by state in STATA 15 using the "by" command.

individuals by the number of individuals in the state. Household size corresponds to the average household size in each state.

Figure IV-1. Distribution of State-Level WIC Participation Rate Across Ethnicity and Citizenship

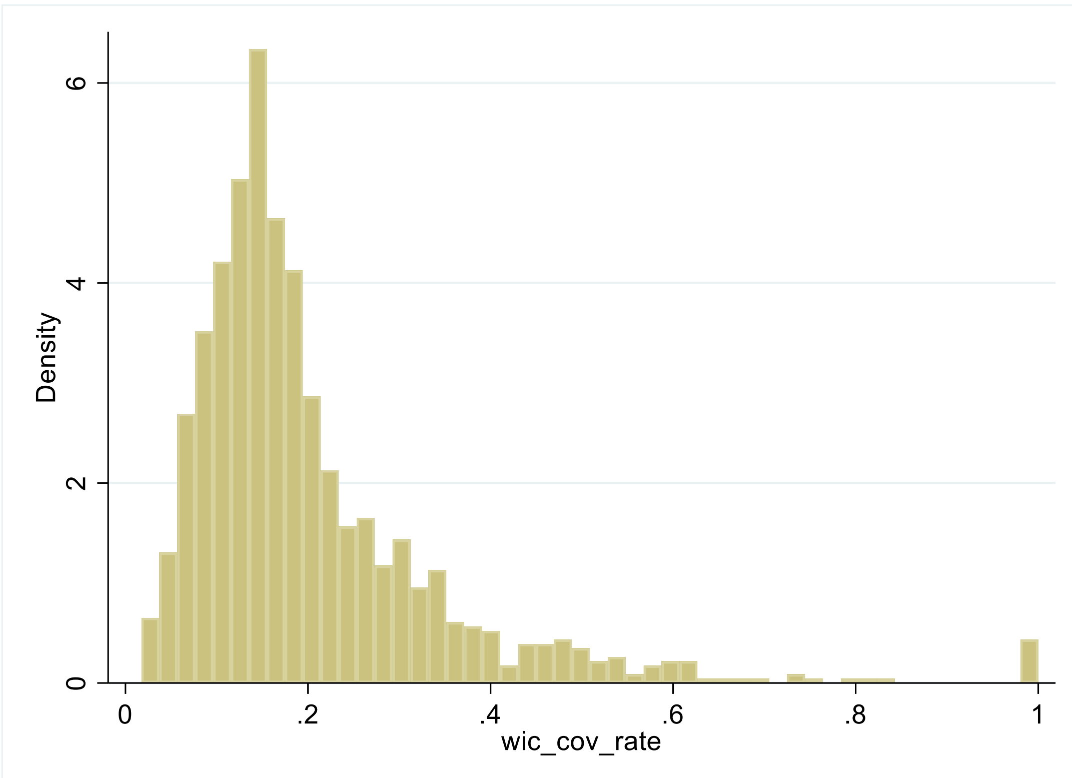


Figure Given that CPS-ASEC are self-reported, there are three potential concerns in survey process and responses: (1) mis-representation of citizenship; (2) mis-representation of WIC participation, and (3) refusal to participate in the survey by non-citizens. Mis-representation of citizenship may be attributed to the increasing fear of deportation that comes from strengthening interior enforcement. However, as mentioned by Sommers (2010), the CPS-ASEC never ask about specific legal status for non-citizens only for citizens. This fact may relieve stress of the non-citizen household concerning the reporting of their true citizenship. Further, previous research (Schmidley and Robinson, 2003) concluded that the citizen variable in CPS-ASEC is reliable for tracking the non-citizen population. At the state level, similar demographic composition in the CPS-ASEC data was found compared to the ACS (American Community Survey). The sample size of the ACS data is larger than that of the CPS-ASEC data.

To control for the truncation problem in state-level WIC participation rates, we estimate the inverse mills ratio through the use of a binary Probit model. The dependent variable is equal to 1 if state-level WIC participation is above zero, and equal to 0 if state-level WIC participation is zero. Then, we calculate the inverse mills ratio and incorporate this variable in our empirical model to calculate inverse the mills ratio, we estimate the Probit model using as explanatory variables average household income, average household size, and average employment status.

IV.8.2 Descriptive Statistics

In Table IV-3, the means and standard deviations of the respective variables are exhibited according to citizenship status, ethnicity, and before and after the changes in immigration policy made by the Trump administration. Average WIC participation rates decreased during the Trump administration for both citizens and non-citizens. The WIC participation rate of citizens was 16.1% before 2017 and 14.9% after 2017. For non-citizens, the WIC participation rate was 19.6% before 2017 and 15% after 2017. Further, the WIC participation rate for Hispanic citizens changed from 17.9% to 17.6%, and the WIC participation rate for Hispanic non-citizens changed from 23% to 16.4%. Finally, the WIC participation rate for non-Hispanic citizens changed from 14.4% to 12.3%, and for non-Hispanic non-citizens, the WIC participation rate changed from 16.4% to 13.6%.

Average household size was larger for non-citizens relative to citizens. Average household income was higher for citizens relative to non-citizens. Average employment rates increased for citizens and non-citizens after 2017 compared to before 2017. Moreover, for citizens, household income rose after 2017 compared to before 2017. But the reverse was true for non-citizens. In particular, average household income for Hispanic citizens increased from $42,481 before 2017 to $46,566 after 2017. As well, average household income for non-Hispanic citizens increased from $50,276 before 2017 to $55,390 after 2017. However, for Hispanic non-citizens, average household income decreased from $44,791 to $44,317. Average household income for non-Hispanic non-citizens decreased from $47,244 to $45,406.

Figure IV-2 the average WIC participation rate across all states by citizenship and ethnicity. The WIC participation rate of non-citizens and Hispanics is higher than non-citizens and non-Hispanics, citizens and Hispanics, and citizens and non-Hispanics. During the second term of the Obama administration (2013-2016), both graph in Figure 2 have similar decreasing pattern. But during the Trump administration (2017-2018), participation rate for non-citizen and Hispanic group has been decreased. However, participation rate for another group of households has been increased between 2017 and 2018. Figure 3 show more detailed pattern in WIC participation rate change between pre- and post- Trump administration by citizenship and ethnicity. All Panels in Figure 3 show decreasing pattern in pre-Trump administration period (2013-2016). However, WIC participation rate for Non-citizen & Hispanic group was decreased, but participation rate for other groups were increased. These results suggest that there is a break point between two periods, second Obama administration period (2013-2016) and first Trump administration period (2017-2018).

Table IV-3. Means and Standard Deviations of WIC Participation Rate, Household Size, Household Income, and Employment Status by Citizenship Before and After 2017

	Citizens		Non-citizens	
Variables	Before 2017	After 2017	Before 2017	After 2017
Whole sample				
WIC coverage rate	0.161	0.149	0.161	0.149
	(10.4)	(11.3)	(10.4)	(11.3)
HH size	3.667	3.577	3.667	3.577
	(0.454)	(0.472)	(0.454)	(0.472)
HH income ($)	46,611	50,978	46,611	50,978
	(13,600)	(13,147)	(13,600)	(13,147)
Employment (share)	0.636	0.654	0.636	0.654
	(0.128)	(0.127)	(0.128)	(0.127)
Hispanics				
WIC coverage rate	0.179	0.176	0.179	0.176
	(0.138)	(0.151)	(0.138)	(0.151)
HH size	3.801	3.658	3.801	3.658
	(0.559)	(0.594)	(0.559)	(0.594)
HH income ($)	42,481	46,566	42,481	46,566
	(14,871)	(13,144)	(14,871)	(13,144)
Employment (share)	0.651	0.661	0.651	0.661
	(0.168)	(0.171)	(0.168)	(0.171)
Non-Hispanics				
WIC coverage rate	0.144	0.123	0.144	0.123
	(4.5)	(3.9)	(4.5)	(3.9)
Household size	3.532	3.497	3.617	3.842
	(0.254)	(0.287)	(0.257)	(1.000)
HH income ($)	50,726	55,390	47,244	45,406
	(10,749)	(11,632)	(50,333)	(27,125)
Employment (share)	0.620	0.647	0.617	0.649
	(0.067)	(0.059)	(0.257)	(0.249)

Figure IV-2. State-Level WIC Participation Rate from 2013 to 2018

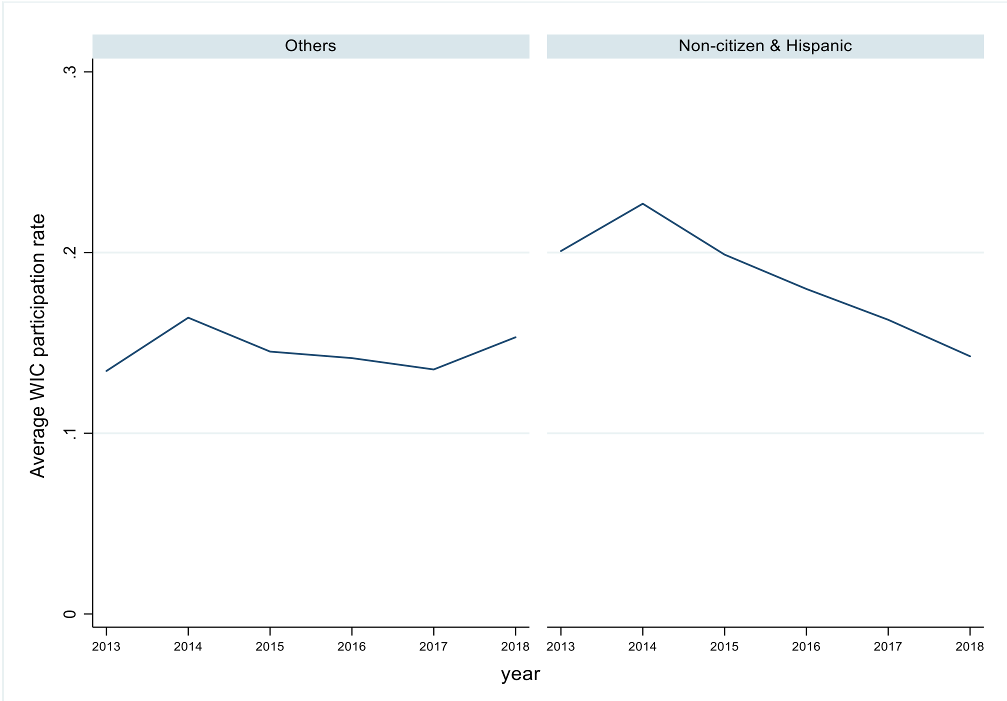
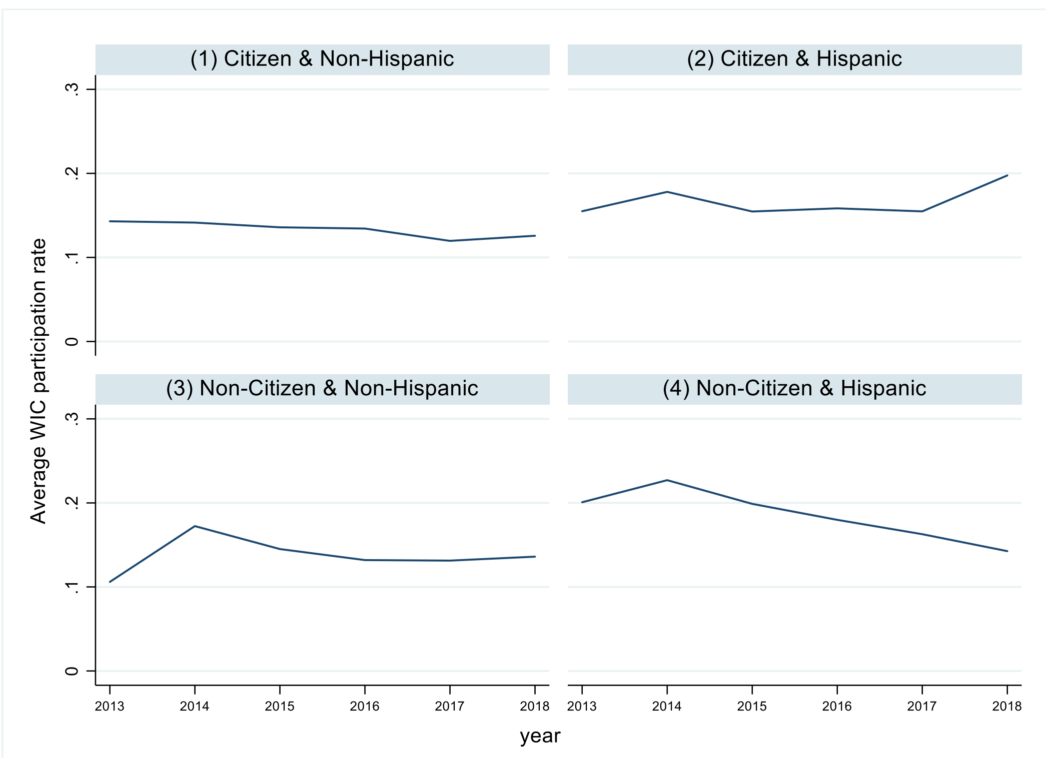


Figure IV-3. WIC Participation Rate by Citizenship and Ethnicity from 2013 to 2018



IV.9 Empirical Results

IV.9.1 Test Parallel trend Assumption

Key assumption of DD and DDD estimation is that trends in outcomes in the pre- treatment period have been the same in both the control and treatment group in the period before policy change (Wooldridge, 2010). This assumption, the so-called "parallel trend assumption", is necessary condition for identifying causal impact. In the period of pre-policy change, the difference between the treatment and control group is constant over time. In our study, the treatment group corresponds to non-citizen Hispanic households.

We test the parallel trend assumption by estimating modified version of equation (6) using data from 2013 to 2016. Instead of post policy regime change dummy variable (I_t^{post}) in equation (6), we incorporate a time trend variable in our equation.

$$\begin{aligned} Y_{scht} = \gamma_0 + \gamma_1 T_t + \gamma_2(T_t * I_c) + \gamma_3(T_t * I_c * I_h) + \gamma_4(T_t * I_h) + \gamma_5(I_c * I_h) \\ + \gamma_6 IN_{scht} + \gamma_7 FS_{scht} + \gamma_8 ES_{scht} + \theta_c + \kappa_h + \mu_s + \phi_t + \epsilon_{scht} \end{aligned} \tag{8}$$

So, we estimate equation (8) for testing parallel trend assumption. T_t is the time trend during the pre-policy change period. For our analysis, T_t is one in 2013 and six in 2016. Other notations are the same as those conveyed earlier in discussing equation (6).

Table 4 presents the results of OLS regression for testing the parallel trend assumption. The key coefficient is the interaction of the trend, non-citizen dummy, and Hispanic dummy variables. If this coefficient is statistically significant, then difference between the treatment and control groups is not constant over time. Our estimation results show that this estimated coefficient is not statistically significant, hence is no specific trend in non-citizen Hispanics before the policy change period.

Table IV-4. Results of Parallel Assumption Test

Explanatory Variable	WIC participation rate (dependent variable)
Trend	0.001
	(0.005)
Trend×Noncitizen × Hispanic	0.007
	(0.005)
Log household income	-0.046***
	(0.017)
Family size	0.066***
	(0.010)
Employment	-0.022
	(0.033)
Fixed effect	State, year, citizenship, and ethnicity
Adjusted R squared	0.215
N	800

[1]* 90%, ** 95%, *** 99% significance level.

[2]We incorporate state, year, citizenship, and Hispanic fixed effect in our model.

Source: STATA version 15.

IV.9.2 Triple Difference estimation Results

In Table 5, we present the DDD estimation results of immigration policy change concerning state-level WIC program participation rates. For estimation, we use STATA version 15. We estimate equation (6) with Ordinary Least Squares (OLS). To begin, the model explains roughly 20% of the variability in state-level WIC participation rates. The WIC participation rates are inversely related to average household income. Higher levels of household income result in lower WIC participation rates. The reverse is true concerning household size. Holding other factors invariant, increases in household size are linked to higher state-level WIC participation rates. Increases in state-level employment rates are negatively related to state-level WIC participation rates. But this effect is not statistically different from zero. The coefficient associated with the interaction term of noncitizens, Hispanics and post-policy regime change is statistically different from zero. Specifically, WIC participation rates of non-citizens and Hispanics are 8.6% lower than other groups of households.

Table IV-5. Triple Difference Estimation Results (state level data)

Explanatory Variable	WIC participation rate (dependent variable)
Post	-0.002
	(0.020)
Noncitizen × post	0.021
	(0.025)
Noncitizen × Hispanic × post	-0.086**
	(0.034)
Log household income	-0.026**
	(0.010)
Household size	0.063***
	(0.009)
Employment	-0.016
	(0.029)
Inverse mills ratio	-0.066***
	(0.024)
Fixed effects	State ,year, citizenship, and ethnicity
Adjusted R squared	0.195
N	1,200

[1]* 90%, ** 95%, *** 99% significance level.

[2]To calculate inverse mills ratio, we estimate a Probit model with average household income, average household size, and average employment rate as explanatory variables.

[3]The fixed effects relate to state, year, citizenship, and ethnicity in the model. Given the plethora of indicator or dummy variables associated with these fixed effects (e.g., 49 dummy variables for state), we do not present these estimated coefficients and associated standard errors. These results are available upon request.

Source: Stata version 15.

IV.10 Check on Robustness

IV.10.1 Test for placebo effects

We test whether placebo effects are evident. The purpose of this test is to find out whether another break point exist in the sample before policy change. Another purpose of this test is ensuring immigration policy change after 2017 is the only break point in our data. To test this effect, we set up three placebo break point variables (break points in year 2014, 2015, and 2016). Break point variables are indicator variable that represent placebo effects. The placebo break point in 2014 takes on the value of one in year 2014 to 2016 and the value of zero in year 2013. The placebo break point in 2015 takes on the value of one in year 2015 to 2016 and the value of zero in years 2013 and 2014. Finally, the placebo break point in 2016 takes on the value of one in year 2016 and the value of zero in years 2013 through 2015. We estimate our empirical model (equation 6) by replacing the post-policy change variable (I_t^{post}) with the placebo break point variables. Therefore, we estimate three models with placebo break point variables. The first test estimates the model with the placebo break point in 2014. The second test estimates the model with placebo break point in 2015. The third test estimates the model with the placebo break point in 2016.

Each column in Table 6 shows the estimation results concerning the placebo effect. All placebo break points have no effect on state-level WIC participation rates. As exhibited in Table 6, the coefficients associated with the interaction of non-citizen, Hispanic, and the placebo break points in 2014, 2015, and 2016 are not statistically significant.

Table IV-6. Placebo Test Results

Break point	(1) 2014	(2) 2015	(3) 2016
Post	-0.006	0.012	0.015
	(0.022)	(0.020)	(0.022)
Noncitizen × Post	0.022	-0.005	-0.019
	(0.029)	(0.026)	(0.030)
Noncitizen × Hispanic × Post	-0.005	0.011	-0.002
	(0.042)	(0.036)	(0.041)
Log household income	-0.048***	-0.047***	-0.048***
	(0.017)	(0.017)	(0.017)
Household size	0.057***	0.058***	0.056***
	(0.011)	(0.011)	(0.011)
Employment	-0.025	-0.026	-0.029
	(0.033)	(0.033)	(0.033)
Fixed effect		State, year, citizenship, and ethnicity	
R squared	0.214	0.212	0.216
N	800	800	800

[1]* 90%, ** 95%, *** 99% significance level.

[2]To calculate inverse mills ratio, we estimate a Probit model with average household income, average household size, and average employment rate as explanatory variables.

[3]The fixed effects relate to state, year, citizenship, and ethnicity in the model. Given the plethora of indicator or dummy variables associated with these fixed effects (e.g., 49 dummy variables for state), we do not present these estimated coefficients and associated standard errors. These results are available upon request.

Source: STATA version 15.

IV.10.2 Randomized inference test

To further confirm our DDD estimation results are robust, we also perform a randomized inference test. The randomized inference (RI) test was initially proposed by Fisher (1935) and updated by Rosenbaum (2002) to conduct exact tests for experiments. Recently, the RI test has been widely implemented in the literature using DD and DDD methods (Chakrabarti et al., 2018; Alsan and Yang, 2019). The randomized Inference test considers all possible random assignments that could have happened described by I_t^{post}. With the Randomized Inference test, interest exists concerning the significance or non-significance of the estimated coefficient (α_3) considering all situations by reassigning the timing variable (I_t^{post}).

We use "ritest" command (Hess, 2019) in STATA 15 to perform the randomized inference test. The "ritest" command estimates the p-value on the basis of Monte Carlo simulations. In this study, we test for the coefficient of interaction on non-citizens, Hispanics, and post-policy regime change. We set up 2,000 permutations on the variable I_t^{post} which indicate post-policy regime change. As exhibited in Table 7, the coefficient associated with the interaction term is statistically significant at the 95% confidence level. Thirty-three test statistics are produced from this test procedure. This result suggests that the effect of post-policy regime change estimated and reported in Table 5 is unlikely to be observed by chance.

Table IV-7. Permutation Test Result

Coefficient	Estimation result	c	N	p=c/n	Standard error	95% Confidence Interval	
α_3	-0.086	33	2000	0.017	0.003	0.011	0.023

Note: Confidence interval is with respect to p=c/n.

Source: STATA version 15.

IV.10.3 Triple Difference Estimate Excluding States with Notable Hispanic Populations

We estimate the Triple Difference estimator (equation 6) with the data excluding states which have high Hispanic populations. In 2019, the states with notable segments of Hispanic population are California, New York, Texas, and, Florida. We exclude data for these four states in our sample and re-estimate equation (6). Our hypothesis is that non-citizen Hispanic households located in states of relatively lower Hispanic population have lower WIC participation rates due to the post-policy regime change.

Column (1) in Table 8 shows the estimation results of the DDD model based on data excluding the most populous states of Hispanics (California, New York, Texas, and, Florida). The coefficient associated with the interaction on non-citizens, Hispanics, and post-policy regime change indicator variables is -0.091 and statistically significant. The coefficient associated with the interaction on non-citizens and the post-policy regime change indicator variables is 0.031, but not statistically significant. These results suggest that, even in states with relatively lower Hispanic populations, the average WIC participation rate of non-citizen, Hispanic households is lower due to the post-policy regime change.

IV.10.4 Triple Difference Estimation Using Data Exclusively for Female Household Heads

Again, to check on robustness, we also estimate equation (6) using only data corresponding to female household heads. Initially, we assume that household participation in WIC program is decided by the female household head. To form the data corresponding to the female household head, we average household level data by citizenship, ethnicity, state, year, and gender.

We find similar results with those reported in column (1) and those reported in Table 5. As exhibited in Table 8, in column (2) the coefficient of the interaction of non-citizens, Hispanics, and post-policy regime change indicator variables is -0.091 and statistically significant. The coefficient of the interaction of non-citizens and post-policy regime change indicator variables is

0.031, but not significant. These results suggest that our previously reported findings (Table 5) indeed are robust.

Table IV-8. Robustness Check

	State level		
	(1)	(2)	(3)
	Hispanic	Female	Tobit
Post	-0.000	0.000	-0.002
	(0.023)	(0.21)	(0.022)
Noncitizen × Post	0.021	0.031	0.002
	(0.028)	(0.026)	(0.027)
Noncitizen × Hispanic × Post	-0.091**	-0.091**	-0.078**
	(0.039)	(0.036)	(0.038)
Log household income	-0.025**	-0.010	-0.014
	(0.011)	(0.007)	(0.011)
Household size	0.062***	0.034***	0.084***
	(0.010)	(0.007)	(0.009)
Employment	-0.015	-0.055***	-0.039
	(0.031)	(0.020)	(0.032)
Mills ratio	-0.069**	-0.064***	
	(0.029)	(0.021)	
Fixed effect	State, year, citizenship, and ethnicity		
R squared	0.189	0.216	
N	744	1,170	1,240

[1]* 90%, ** 95%, *** 99% significance level.

[2]Column (2) shows estimation results with the data excluding four states (California, Texas, Florida, and New York) that have high Hispanic populations. Column (3) shows estimation results with the data for female household heads only. Column (4) shows estimation results by utilizing the Tobit model with same data used to produce the empirical results reported in Table 5.

Source: STATA version 15.

IV.10.5 Estimation of the Tobit Model

As another check on robustness of results with the state-level data, we estimate Tobit model to apply another technique to capture data censoring issue on WIC participation rate rather than including the inverse mills ratio. That is, we re-estimate equation (6) without the inverse mill's ratio, but employ a Tobit with the same data.

Column (3) in Table 8 shows results from Tobit model estimation. The coefficient associated with the interaction of non-citizens, Hispanics, and post-policy regime change is -0.078 and statistically significant. But the coefficient associated with the interaction of non-citizens and Hispanics is 0.002, but not statistically significant. These results suggest that impact of the policy regime change on state-level WIC participation rates only affect non-citizen Hispanic households. Also, these results are in alignment with those obtained from the DDD estimation. Therefore, our model specification and estimation results from the Triple Difference estimation are robust.

IV.10.6 Event Study of State-Level WIC Participation Rate

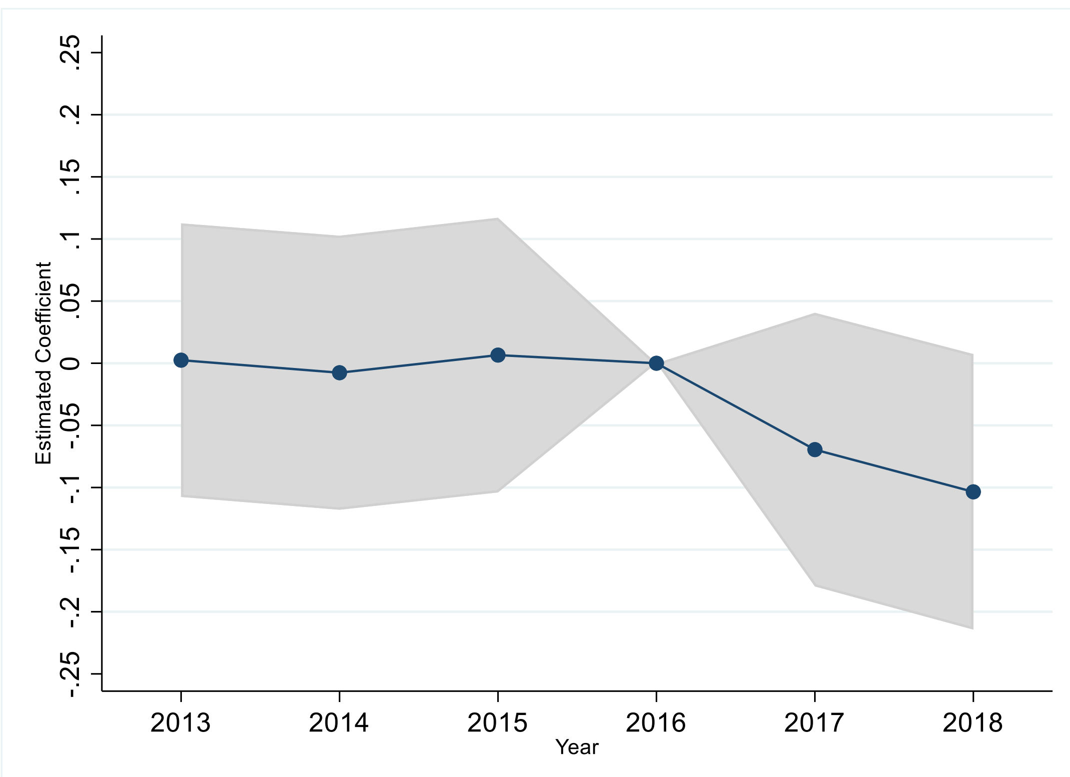
We also employ an event study specification as yet another check on the robustness of our empirical results. Instead of I_t^{post}, we incorporate a series of time dummy variables for each year. The base or reference year is 2016. In equation (10), $I_{t=y}$ is an indicator for each year (2013 ~ 2018) other than the year 2016. The coefficient, δ_2^y, traces the WIC participation rate for eligible non-citizens before and after policy regime change relative to eligible citizens. The coefficient, δ_3^y, traces the WIC participation rate for eligible non-citizens and Hispanics before and after policy regime change relative to all other groups of households.

$$Y_{scht} = \delta_0 + \sum_{y \neq 2016} \delta_1^y (I_{t=y}) + \sum_{y \neq 2016} \delta_2^y (I_{t=y} * I_c) + \sum_{y \neq 2016} \delta_3^y (I_{t=y} * I_c * I_h) \tag{10}$$

$$+ \delta_4 IN_{scht} + \delta_5 FS_{scht} + \delta_6 ES_{scht} + \theta_s + \kappa_t + \mu_n + \phi_h + \epsilon_{scht}$$

In Figure 3, we plot the coefficients (δ_3^y) associated with interactions of each time indicator, for non-citizens and Hispanics. The gray area shows the 95% confidence interval for each of the respective coefficients. Relative to 2016, in 2017 and 2018, the estimated coefficients gradually decreased. Estimated coefficients before 2016 are not statistically different from zero, tantamount to no distinguished effect during the period 2013-2016. But for years 2017 and 2018, state-level WIC participation rates for non-citizens and Hispanics were significantly lower relative to 2016.

Figure IV-4. Event Study of State-Level WIC Participation Rates



IV.11 Summary and Suggestions for Future Research

In this study, we test the hypothesis that state-level WIC participation rates of Hispanic non-citizens is reduced after the immigration policy change implemented by the Trump administration beginning in 2017. To test this hypothesis, we use the Triple Difference estimate method, along with CPS-AEPC data from the IPUMS system. We also perform various robustness checks to confirm our estimation results. In doing so, we show that our results hold under various situations and model specifications.

Importantly, we find that state-level WIC participation rates of Hispanic non-citizens are significantly lower after the immigration policy change implemented by the Trump administration. But this finding only holds for Hispanic non-citizens, not non-Hispanic non-citizens. Specifically, the state-level WIC participation rate for Hispanic non-citizens, all other factors invariant, is lower by 8.6% relative to all other groups (Hispanic citizens, non-Hispanic citizens, and non-Hispanic non-citizens). This finding provides quantitative evidence concerning the ongoing debate about the impact of the immigration policy changes under the Trump administration.

But this analysis is done at the state-level only. Future work should replicate this analysis using household level data, that is not aggregating over households. In doing so, we are in position to estimate binary qualitative choice models such as the logit model or the Probit model. The dependent variable in this case would be a dummy variable, say 1 if the household participates in the WIC program and 0 if the household does not. With this methodology, we could estimate four different qualitative choice models for four groups of households—Hispanic non-citizens; Hispanic citizens; non-Hispanic non-citizens; and non-Hispanic citizens. Controlling for other factors such as region (state, county), household income, household size, employment status, gender of household head, and so on, we would be able to determine if the probability of participating in the WIC program fell as a result of the change in immigration policy implemented by the Trump administration. In addition, with these qualitative choice models we would be in position to profile various households who participate in the WIC program. That said, we also would be in position to profile those households who are not participating in the WIC program.

Moreover, future work could center attention on other assistance programs such as SNAP, Medicaid, and TANF. Unlike the WIC program, these programs require participants to be a citizen or to hold a legally effective immigration status who have resided in the United States more than five years. As well, we would be in position to discern whether participation in SNAP, Medicaid, and TANF affects the likelihood of participating in the WIC program.

CHAPTER V

SUMMARY, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER RESEARCH

A number of choices is evident beyond traditional supermarkets or grocery stores owing to the increasingly diverse U.S. retail food landscape. Despite the plethora of previous studies that largely focus on factors affecting store choice, one area of research that has received relatively little attention is how the magnitude of household food and beverage expenditures is impacted by the type of store outlets. In this light, the purpose of this study is to examine how socio-demographic factors, spending habits, and characteristics of the retail food environment affect household expenditure across all food and beverage categories by store type and by income level. The list of socio-demographic factors includes: (1) household income; (2) household size; (3) age; (4) urbanization; (5) education; (6) race and ethnicity; and (7) region. Characteristics of the retail environment relate to the number of club stores, the number of convenience stores, the number of grocery stores and supercenters and the number of drug stores within the zip code area of the of the household. Whether traditional or non-traditional, store outlets differ in prices, product assortment, advertising strategies, and location (Volpe, Kuhns, and Jaenicke, 2017). The outlets considered in this study are grocery, convenience, discount, club, drug, and dollar store types.

As mentioned previously, prior works mainly highlighted store choice. To differentiate our study from the extant literature, we explore the factors which directly affect household food expenditure by store outlet. Indeed, Volpe, Jaenicke, and Chenarides (2018) estimated the impacts of expenditure share by store format, but in our study, we quantify the magnitude of the impact of household socio-demographics, the retail food environment, and spending habits on food and beverage expenditures by diverse store types. Hence, by analyzing factors that impact household food expenditure across the aforementioned six store types, this study contributes to the economic literature. Another contribution is that our study also considers habitual persistence or spending habits, a dynamic property of household expenditure on food and beverages. However, in the previously mentioned studies, habitual behavior was not included in the set of explanatory variables.

To further differentiate our study from previous studies, we employ a dynamic correlated random effect Tobit model to incorporate habitual purchasing behavior. The source of data for this analysis is the Nielsen Homescan Panel over the period between 2011 and 2015. Specifically, we use a balanced panel of 28,109 households who participated in the survey for all five years from 2011 to 2015. The total number of observations available for analysis is 140,545. The panel structure allows us to incorporate dynamic modeling by including lagged dependent variables as explanatory variables to account for spending habits.

Another advantage of the use of this model is that we are in a position to handle corner solution problems. The dependent variables reflect household purchasing history according to store type and indeed have zero values; hence the dependent variables are left censored. A differentiated feature of our empirical analysis relates to transforming the dependent variables which include zero observations using the inverse hyperbolic sine (arcsinh) method (Bellemare and Wichman 2020). A notable problem with taking the logarithm of any variable is that it does not allow retaining zero-valued observations because the ln (0) is undefined. As pointed out by Bellemare and Wichman (2019), "applied econometricians are typically loath to drop those observations for which the logarithm is undefined." Consequently, researchers often have resorted to ad hoc means of accounting for this situation when taking the natural logarithm of a variable, such as adding 1 to the variable prior to its transformation (MaCurdy and Pencavel, 1986). In recent years, the inverse hyperbolic sine (or arcsinh) transformation has grown in popularity among applied econometricians due to the fact that it is similar to the behavior of the logarithm function, it allows retaining zero-valued observations without any arbitrariness, and it often results in normal distributions (Burbidge et al. 1988; Yen and Jones 1997; MacKinnon and Magee 1990; Pence 2006; Van den Heuvel et al. 2011; Bellemere, Barrett, and Just 2013; Brown et al. 2015; Bellemere and Wichman 2020).

Importantly, we estimate separate dynamic correlated random effect Tobit models for sub-samples in accordance with household income level. Because households typically have different shopping baskets by income level (Taylor and Villas-Boas 2016; Volpe, Jaenicke, and Chenarides 2018), we compare and contrast our findings across income levels for each of the six store types considered in our study. We consider three distinct income categories—low, middle, and high. The low-income sample corresponds to those households whose annual income below is \$25,000. The middle-income sample corresponds to those households whose annual income is above \$25,000 but below \$70,000. The high-income sample corresponds to those households whose annual income is above \$70,000. We follow this segmentation of household income based on the work by Allcott, Diamond, and Dubé (2017). Hence, we estimate 24 different dynamic correlated random effect Tobit models, covering six store types and four data samples.

The results support the supposition of habitual spending across all store outlets. These results suggest that, within the data period 2011 to 2015, habitual spending behavior is undoubtedly a key factor in affecting nominal food and beverage expenditures across all store formats. This finding also holds across the three respective income sub-samples. Household income is not a statistically significant factor affecting household food and beverage expenditures in any of the respective store outlets even across the various income sub-samples. However, household size, age, urbanization, education, race and ethnicity, region, time-invariant socio-demographic variables, indeed are drivers of household food and beverage expenditures at the six store outlets across the income categories. This finding is in line with the hypothesis of underlying household heterogeneity and in agreement with the results of Bilsard, Stewart, and Jolliffe (2004) and of Taylor and Villas-Boas (2016).

Further, the number of convenience stores in the zip code area of households do not significantly influence the level of food and beverage expenditures across the respective store outlets and across the respective income categories. The same result is true for drug stores but for a single exception. In the high-income sample, the number of drug stores in the zip code area negatively impacts food and beverage expenditures made at dollar stores. In the entire sample and in the mid-income sample, the number of club stores negatively impacts household expenditures made at grocery stores and drug stores. But this finding is not the case within the low-income sample and within the high-income sample. In addition, in the entire sample, the number of grocery stores and supercenters in the zip code area negatively impacts household food and beverage expenditures made at discount stores. Nevertheless, this finding is not the case in each of the respective income sub-samples.

Bottom line, evidence exists to support the hypothesis that the retail environment plays a limited role in affecting household expenditures for food and beverages across store outlets and across income sub-samples. This result differs from previous findings by Kyureghian and Nayga (2013) and by Taylor and Villas-Boas (2016), but this result is in alignment with the work by Ver Ploeg and Wilde (2018).

The findings in this study make several contributions to the current economic literature. First, we provide a detailed view that describes household spending behavior across six store types for three income classifications. Second, the construction and estimation of dynamic random effect Tobit models constitute the first attempt in the literature dealing with household food and expenditure by store outlets for various income classifications. Third, we use a novel method to deal with problems in data (zero observations and extreme values) through the inverse hyperbolic sine transformation. Fourth, we derive the accompanying expressions for calculating conditional marginal effects and the marginal effects associated with the probability of purchasing food and beverages on the basis of the inverse hyperbolic sine transformation.

Future research in this area may center attention on specific household food and beverage expenditures rather than the aggregate, for example, fresh fruits and vegetables or meat products. Particularly for low-income households, we are in position to investigate nutrition intake of households associated with the six store types by income level. As such, this research may uncover a link between store type and nutrition intake, especially useful for policies dealing with various food assistance programs. Although this research covers the period 2011 to 2015, this study establishes a baseline. Our study can be replicated using more recent data to determine the robustness of our findings. Without question, because today's food retail environment is considerably diverse, more work is needed to understand the role of store outlets in affecting dietary quality in America across various income sub-samples.

In the third essay, we test the hypothesis that state-level WIC participation rates of Hispanic non-citizens are lower after the immigration policy change implemented by the Trump administration beginning in 2017. To test this hypothesis, we use the Triple Difference estimate method, along with CPS-AEPC data from the IPUMS system. We also perform various robustness checks to confirm our estimation results. In doing so, we show that our results hold under various situations and model specifications.

Importantly, we find that state-level WIC participation rates of Hispanic non-citizens are significantly lower after the immigration policy change implemented by the Trump administration. But this finding only holds for Hispanic non-citizens, not non-Hispanic non-citizens. Specifically, the state-level WIC participation rate for Hispanic non-citizens, all other factors invariant, is lower by 8.6% relative to all other groups (Hispanic citizens, non-Hispanic citizens, and non-Hispanic non-citizens). This finding provides quantitative evidence concerning the ongoing debate about the impact of the immigration policy changes under the Trump administration.

But this analysis is done at the state-level only. Future work should replicate this analysis using household level data, that is not aggregating over households. In doing so, we are in position to estimate binary qualitative choice models such as the logit model or the Probit model. The dependent variable in this case would be a dummy variable, say 1 if the household participates in the WIC program and 0 if the household does not. With this methodology, we could estimate four different qualitative choice models for four groups of households—Hispanic non-citizens; Hispanic citizens; non-Hispanic non-citizens; and non-Hispanic citizens. Controlling for other factors such as region (state, county), household income, household size, employment status, gender of household head, and so on, we would be able to determine if the probability of participating in the WIC program fell as a result of the change in immigration policy implemented by the Trump administration. In addition, with these qualitative choice models we would be in position to profile various households who participate in the WIC program. That said, we also would be in position to profile those households who are not participating in the WIC program.

Moreover, future work could center attention on other assistance programs such as SNAP, Medicaid, and TANF. Unlike the WIC program, these programs require participants to be a citizen or to hold a legally effective immigration status who have resided in the United States more than five years. As well, we would be in position to discern whether participation in SNAP, Medicaid, and TANF affects the likelihood of participating in the WIC program.

REFERENCES

Anamonics, W.L., and J.D. Swait. 2012. Are Food Choices Really Habitual? Integrating Habits, Variety-Seeking, and Compensatory Choice in a Utility-Maximizing Framework. *American Journal of Agricultural Economics* 95(1): 17-41.

Alif, A., S. Nelson, A. Stefancic, R. Ahmed, and S. Okazaki. 2020. Documentation Status and Psychological Distress among New York City Community College Students. *Cultural diversity and ethnic minority psychology* 26(1):11-21.

Allcott, H., R. Diamond, and J.P. Dubé. 2017. *The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States,* National Bureau of Economic Research.

Alsan, M., and C. Yang. 2018. *Fear and The Safety Net: Evidence from secure communities*. National Bureau of Economic Research.

Arnold, S., T. Oum, and D. Tigert. 1983. Determinant Attributes in Retail Patronage: Seasonal, Temporal, Regional, and International Comparisons. *Journal of Marketing Research* 20(2): 149-157.

Austin, S., K. Yu, S.H. Liu, F. Dong, and N. Tefft. 2017. Household Expenditures on Dietary Supplements Sold for Weight Loss, Muscle Building, and Sexual Function: Disproportionate Burden by Gender and Income. *Preventive Medicine Reports* 6: 236-241.

Basker, E., and M. Noel. 2009. The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry. *Journal of Economics and Management Strategy* 18: 977-1009.

Becerra, D. 2016. Anti-Immigration Policies and Fear of Deportation: A Human Rights Issue. *Journal of Human Rights and Social Work* 1(3):109–19.

Bellemare, M.F., C.B. Barrett, and D.R. Just. 2013. The Welfare Impacts of Commodity Price Volatility: Evidence from Rural Ethiopia. *American Journal of Agricultural Economics* 95(4): 877–899.

Bellemare, M.F., and C.J. Wichman. 2020. Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics* 82(1): 50–61.

Berk, M, and C. Schur. 2001. The Effect of Fear on Access to Care among Undocumented Latino Immigrants. *Journal of Immigrant Health* 3(3):151-156.

Berry, S., J. Levinsohn, and A. Pakes. 1995. Automobile Prices in Market Equilibrium. *Econometrica* 63 (4): 841-890.

Bettman, J.R., and M.A. Zins. 1977. Constructive Processes in Consumer Choice. *Journal of Consumer Research* 4(2): 75-85.

Bleich, S. N., and S. Fleischhacker. 2019. Hunger or Deportation: Implications of the Trump Administration's Proposed Public Charge Rule. *Journal of Nutrition Education and Behavior* 51(4): 505–509.

Briesch, R.A., P.K. Chintagunta, and E. J. Fox. 2009. How Does Assortment Affect Grocery Store Choice? *Journal of Marketing* 46(2): 176-189.

Broda, C., E. Leibtag, and D. Weinstein. 2009. The Role of Prices in Measuring the Poor's Living Standard. *Journal of Economic Perspectives* 23(2): 77–97.

Brown, S., W.H. Greene, M.N. Harris, and K. Taylor. 2015. An Inverse Hyperbolic Sine Heteroskedastic Latent Class Panel Tobit Model: An Application to Modelling Charitable Donations. *Economic Modelling* 50: 228–236.

Bruno, A. 2014. Unauthorized Aliens in The United States: Policy Discussion. *Current Politics and Economics of the United States, Canada and Mexico* 16(2):1-24.

Burbidge, J. B., L. Magee, and A. L. Robb. 1988. Alternative Transformations to Handle Extreme Values of the Dependent Variable. *Journal of the American Statistical Association* 83(401): 123–127.

Bustillos, B., J.R. Sharkey, J. Anding, and A. McIntosh. 2008. Availability of More Healthful Food Alternatives in Traditional, Convenience, and Nontraditional Types of Food Stores in Two Rural Texas Counties. *Journal of the American Dietetic Association* 109(5):883-889.

Byrne, P.J., O. Capps, and A. Saha. 1996. Analysis of Food-Away-From-Home Expenditure Patterns for U.S. Households, 1982–89. *American Journal of Agricultural Economics* 78(3): 614-627.

Callaghan, T., D. J. Washburn, K. Nimmons, D. Duchicela, A. Gurram, and J. Burdine. 2019. Immigrant Health Access in Texas: Policy, Rhetoric, and Fear in the Trump Era. *BMC Health Services Research* 19(1):1–8.

Capps Jr, O., and J.M. Griffin. 1998. Effect of a Mass Merchandiser on Traditional Food Retailers. *Journal of Food Distribution Research* 29(1): 1-7.

Carlson, A., M. Lino, and T. Fungwe. 2007. *The Low-Cost, Moderate-Cost, and Liberal Food Plans*. CNPP-20, U.S. Department of Agriculture, Center for Nutrition Policy and Promotion.

Carpenter, J.M., and M. Moore. 2006. Consumer Demographics, Store Attributes, and Retail Format Choice in the US Grocery Market. *International Journal of Retail & Distribution Management* 34(6):434-452.

Chakrabarti, S., A. Kishore, and D. Roy. 2018. Effectiveness of Food Subsidies in Raising Healthy Food Consumption: Public Distribution of Pulses in India. *American Journal of Agricultural Economics* 100(5):1427–49.

Chamberlain, G. 1980. Analysis of Covariance with Qualitative Data. *The Review of Economic Studies* 47(1): 225-238.

Chenarides, L. and E. C. Jaenicke. 2017. Store Choice and Consumer Behavior in Food Deserts: An Empirical Application of the Distance Metric Method. Department of Agricultural Economics, Sociology, and Education, Penn State University, Working Paper.

Currie, J., S. DellaVigna, E. Moretti, and V. Pathania. 2010. The Effect of Fast Food Restaurants on Obesity and Weight Gain. *American Economic Journal: Economic Policy* 2(3): 32-63.

Deighton, J., C.M. Henderson, and S.A. Neslin. 1994. The Effects of Advertising on Brand Switching and Repeat Purchasing. *Journal of Marketing Research* 31(1): 28-43.

Dunn, R.A., W.R. Dean, C.M. Johnson, A. Leidner, and J.R. Sharkey. 2012. The Effect of Distance and Cost on Fruit and Vegetable Consumption in Rural Texas. *Journal of Agricultural and Applied Economics* 44(4): 491-500.

Ehrenberg, A.S. 1991. New Brands and the Existing Market. *Market Research Society Journal.* 33(4): 1-16.

Fan, L. 2017. Access to Fruits and Vegetables and Consumer Welfare in Urban Food Deserts. Available at SSRN: https://ssrn.com/abstract=3068147.

Feather, P. 2003. Valuing Food Store Access: Policy Implications for the Food Stamp Program. *American Journal of Agricultural Economics* 85(1): 162–172.

Fisher, R. A. 1935. The Design of Experiments. Edinburgh: Oliver & Boyd.

Fleming, P. J., W. D. Lopez, H. Mesa, R. Rion, E. Rabinowitz, R. Bryce, and M. Doshi. 2019. A Qualitative Study on The Impact of The 2016 US Election on The Health of Immigrant Families in Southeast Michigan. *BMC Public Health* 19(1): 1-9.

Goldman, A., and H. Hino. 2005. Supermarkets vs. Traditional Retail Stores: Diagnosing the Barriers to Supermarkets' Market Share Growth in an Ethnic Minority Community. *Journal of Retailing and Consumer Services* 12(4): 273-284.

Greene, W. 2004. Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model. *Econometric Reviews* 23(2): 125-147.

Gruber, J. 1994. The Incidence of Mandated Maternity Benefits. *American Economic Review* 84(3):622–641.

Gruber, J. and J. Poterba. 1994. Tax Incentives and The Decision to Purchase Health Insurance: Evidence from The Self-Employed. *Quarterly Journal of Economics* 109(3):701–733.

Handbury, J., I. Rahkovsky, and M. Schnell. 2015. *Is the Focus on Food Deserts Fruitless? Retail Access and Food Purchases Across the Socioeconomic Spectrum*. National Bureau of Economic Research.

Hausman, Jerry A. 1978. Specification Tests in Econometrics. *Econometrica* 46(6):1251-1271.

Hausman, J. and E. Leibtag. 2007. Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart. *Journal of Applied Econometrics* 22(7): 1157-1177.

Heß, S. 2017. Randomization Inference with Stata: A Guide and Software. *The Stata Journal*, 17(3):630-651.

Heuvel, F.P. van den, P.W. de Langen, K.H. van Donselaar, and J.C. Fransoo. 2011. Identification of Employment Concentration and Specialization Areas : Theory and Application. *Beta Working Paper* 354 (August): 26p.

Hill, H. and F. Lynchehaun. 2002. Organic Milk: Attitudes and Consumption Patterns. *British Food Journal* 104(7):526-542.

Hing, B. O. 2017. Entering the Trump Ice Age: Contextualizing The New Immigration Enforcement Regime. *Texas A&M Law Reveiw* 5: 253.

Hornbeck, R. 2010. Barbed Wire: Property Rights and Agricultural Development. *Quarterly Journal of Economics* 125(2):767–810.

Hoynes, H., D. W. Schanzenbach, and D. Almond. 2016. Long-run Impacts of Childhood Access to The Safety Net. *American Economic Review* 106(4):903–34.

Imbens, G. W., and J. M. Wooldridge. 2007. What's New in Econometrics? Lecture 10 Difference-in-Differences Estimation. NBER Summer Institute.

Jefferies, J. 2014. Fear of Deportation in High School: Implications for Breaking The Circle of Silence Surrounding Migration Status. *Journal of Latinos and Education* 13(4):278-295.

Ji, M.F., and W. Wood. 2007. Purchase and Consumption Habits: Not Necessarily What You Intend. *Journal of Consumer Psychology* 17(4): 261–276.

Khare, A., and J.J. Inman. 2006. Habitual Behavior in American Eating Patterns: The Role of Meal Occasions. *Journal of Consumer Research* 32(4): 567-575.

Kilts Center for Marketing at the University of Chicago Booth School of Business (https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen).

Kleven, H. J., C. Landais, and E. Saez. 2013. Taxation and international migration of superstars: Evidence from the european football market. *American Economic Review* 103(5):1892–1924.

Kyureghian, G., R. Nayga, and S. Bhattacharya. 2013. The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effects Analysis. *Applied Economic Perspectives and Policy* 35(1): 69–88.

Kyureghian, G., and R.M. Nayga. 2013. Food Store Access, Availability, and Choice When Purchasing Fruits and Vegetables. *American Journal of Agricultural Economics* 95(5): 1280-1286.

Laird, J., I. Santelli, J. Waldfogel, and C. Wimer. 2019. Forgoing Food Assistance out of Fear: Simulating the Child Poverty Impact of a Making SNAP a Legal Liability for Immigrants. *Socius: Sociological Research for a Dynamic World* 5:1-8.

Leibtag, E., C. Barker, and P. Dutko. 2010. *How Much Lower Are Prices at Discount Stores?* ERR-105. United States Department of Agriculture. Economic Research Service.

MacKinnon, J.G. and L. Magee. 1990. Transforming the Dependent Variable in Regression Models. *International Economic Review* 31(2): 315–339.

MaCurdy, T.E. and J.H. Pencavel. 1986. Testing Between Competing Models of Wage and Employment Determination in Unionized Markets. *Journal of Political Economy* 94(3, Part 2), S3–S39.

Marshall, G. and T. Pires. 2017. Measuring the Impact of Travel Costs on Grocery Shopping. *The Economic Journal* 128(614): 2538–2557.

McCracken, V.A., and J.A. Brandt. 1987. Household Consumption of Food-Away-From-Home: Total Expenditure and by Type of Food Facility. *American Journal of Agricultural Economics* 69(2): 274-284.

McDonald, J.F., and R.A. Moffitt. 1980. The Uses of Tobit Analysis. *The Review of Economics and Statistics* 62(2): 318-321.

Moffitt, R. 1983. An Economic Model of Welfare Stigma. *American Economic Review* 73(5): 1023–1035.

Motes, W.H., and A.G. Woodside. 2001. Purchase Experiments of Extra-ordinary and Regular Influence Strategies Using Artificial and Real Brands. *Journal of Business Research* 53(1): 15-35.

Muehlenbachs, L., E. Spiller, and C. Timmins. 2015. The Housing Market Impacts of Shale Gas Development. *American Economic Review* 105(12):3633–3659.

Mutlu, S., and A. Gracia. 2006. Spanish Food Expenditure Away from Home (FAFH): By Type of Meal. *Applied Economics* 38(9): 1037-1047.

Nayga, R.M., and O. Capps, Jr. 1992. Determinants of Food Away from Home Consumption: An Update. *Agribusiness* 8(6): 549-559.

Nayga, R.M. 1998. A Sample Selection Model for Prepared Food Expenditures. *Applied Economics* 30(3): 345-352.

Nilsson, J. P. 2017. Alcohol Availability, Prenatal Conditions, and Long-Term Economic Outcomes. *Journal of Political Economy* 125(4):1149–1207.

Oh, M., H. H. Jensen, and I. Rahkovsky. 2016. Did Revisions to the WIC Program Affect Household Expenditures on Whole Grains.? *Applied Economic Perspectives and Policy* 38 (4):578–598.

Oliveira, V., and E. Frazão. 2015. *The WIC Program: Background, Trends, and Economic Issues, 2015 Edition.* United States Department of Agriculture. Economic Research Service.

Pence, K.M. 2006. The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving. *Contributions in Economic Analysis & Policy* 5(1).

Powell, L.M., and Y. Bao. 2009. Food Prices, Access to Food Outlets, and Child Weight. *Economics and Human Biology* 7(1):64-72.

Rahkovsky, I. and S. Snyder. 2015. *Food Choices and Store Proximity*. ERR-195. United States Department of Agriculture. Economic Research Service.

Ravallion, M. 2007. *Evaluating Anti-Poverty Programs*. Handbook of Development Economics 4:3787–3846.

Richards T. J., S. F. Hamilton, and K. Yonezawa. 2017. Variety and the Cost of Search in Supermarket Retailing. *Review of Industrial Organization* 50(3): 263-285.

Rieger, J., C, Kuhlgatz, and S. Anders. 2016. Food Scandals, Media Attention and Habit Persistence among Desensitized Meat Consumers. *Food Policy* 64: 82-92.

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.

Rosenbaum, P. R. 2002. *Observational Studies*. 2nd ed. New York: Springer.

Schmidley, A. D., and J. G. Robinson. 2003. *Measuring The Foreign-Born Population in The United States with The Current Population Survey: 1994-2002*. Population Division, US Bureau of the Census.

Senia, M., S. Dharmasena, and O. Capps. 2019. Can Dietary Fiber Intake Be Increased through Nutritional Education and through Subsidies on Selected Food Products? *Agricultural and Resource Economics Review* 48(3): 448-472.

Smith, H. 2004. Supermarket Choice and Supermarket Competition in Market Equilibrium. *Review of Economic Studies* 71: 235-263.

Smith, H. 2006. Store Characteristics in Retail Oligopoly. *The RAND Journal of Economics* 37(2): 416-430.

Sommers, B. D. 2010. Targeting in Medicaid: The Costs and Enrollment Effects of Medicaid's Citizenship Documentation Requirement. *Journal of Public Economics* 94 (1–2):174–82.

Staus, A. 2009. Determinants of Store Type Choice in the Food Market for Fruits and Vegetables. *International Journal of Arts and Sciences* 3(2): 138–174.

Sullivan, M. J., and R. Enriquez. 2015. The Impact of Interior Immigration Enforcement on Mixed-Citizenship Families. *American Journal of Public Health* 105(10):42–47.

Taylor, G.A. 2001. Coupon Response in Services. *Journal of Retailing* 77(1): 139-151.

Taylor, R., and S.B. Villas-Boas. 2016. Food Store Choices of Poor Households: A Discrete Choice Analysis of the National Household Food Acquisition and Purchase Survey (FoodAPS). *American Journal of Agricultural Economics* 98(2): 513-532.

Torres, S. A., C. D. Santiago, K. K. Walts, and M. H. Richards. 2018. Immigration Policy, Practices, and Procedures: The Impact on The Mental Health of Mexican and Central American Youth and Families. *American Psychologist* 73(7):843-854.

Tummala-Narra, P. 2020. The Fear of Immigrants. *Psychoanalytic Psychology* 37(1):50-61.

Vargas, E. D., and M. A. Pirog. 2016. Mixed-Status Families and WIC Uptake: The Effects of Risk of Deportation on Program Use. *Social Science Quarterly* 97(3):555–572.

Veall, M.R., and K.F. Zimmermann. 1996. Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *Journal of Economic Surveys* 10(3): 241-259.

Ver Ploeg, M., and P.E. Wilde. 2018. How Do Food Retail Choices Vary Within and Between Food Retail Environments? *Food Policy* 79: 300-308.

Vericker, T., K. Fortuny, K. Finegold, and S. B. Ozdemir. 2010. *Effects of Immigration on WIC and NSLP Caseloads.* the Urban Institute.

Volpe, R.J., and N. Lavoie. 2008. The Effect of Wal-Mart Supercenters on Grocery Prices in New England. *Review of Agricultural Economics* 30: 4-26.

Volpe, R.J., and A. Okrent. 2012. *Assessing the Healthfulness of Consumers' Grocery Purchases*. EIB-102. United States Department of Agriculture. Economic Research Service.

Volpe, R., A. Kuhns, and T. Jaenicke. 2017. *Store Formats and Patterns in Household Grocery Purchases.* EIB-167. United States Department of Agriculture. Economic Research Service.

Volpe, R., E.C. Jaenicke, and L. Chenarides. 2018. Store Formats, Market Structure, and Consumers' Food Shopping Decisions. *Applied Economic Perspectives and Policy* 40(4): 672-694.

Watson, T. 2014. Inside the Refrigerator: Immigration Enforcement and Chilling in Immi- grant Medicaid Participation. *American Economic Journal: Economic Policy* 6(3):313–338.

Wooldridge, J.M. 2005. Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity. *Journal of Applied Econometrics* 20(1): 39-54.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT press.

Yelowitz, A. S. 1995. The Medicaid Notch, Labor Supply, and Welfare Participation: Evidence from Eligibility Expansions. *The Quarterly Journal of Economics* 110(4):909–939.

Yen, S.T. 1993. Working Wives and Food Away from Home: The Box-Cox Double Hurdle Model. *American Journal of Agricultural Economics* 75(4): 884-895.

Yen, S.T., and A.M. Jones. 1997. Household Consumption of Cheese: An Inverse Hyperbolic Sine Double-Hurdle Model with Dependent Errors. *American Journal of Agricultural Economics* 79(1): 246–251.

Zhen, C., E.A. Finkelstein, J.M. Nonnemaker, S.A. Karns, and J.E. Todd. 2013. Predicting the Effects of Sugar-Sweetened Beverage Taxes on Food and Beverage Demand in a Large Demand System. *American Journal of Agricultural Economics* 96(1): 1-25.

Zhen, C., M.K. Wohlgenant, S. Karns, and P. Kaufman. 2011. Habit Formation and Demand for Sugar-Sweetened Beverages. *American Journal of Agricultural Economics* 93(1): 175-193.

APPENDIX

DERIVING THE CONDITIONAL EXPECTATION OF THE DEPENDENT VARIABLE FROM THE INVERSE HYPERBOLIC SINE TRANSFORMATION

$$\overline{EX}_{ht}^{k\ *} = arcsinh\left(EX_{ht}^{k}\right) = ln(EX_{ht}^{k} + \sqrt{EX_{ht}^{k\ 2} + 1})$$

$$\overline{EX}_{ht}^{k\ *} = \alpha z_{ht}^{k} + \rho \overline{EX}_{h,t-1}^{k\ \ *} + c_{h}^{k} + \varepsilon_{ht}^{k}$$

$$c_{h}^{k} = \ \theta \overline{EX}_{h,0}^{k\ *} + \vartheta d_{h}^{k} + u_{h}^{k}$$

$$\overline{EX}_{ht}^{k\ *} = \alpha z_{ht}^{k} + \rho \overline{EX}_{h,t-1}^{k\ \ *} + \theta \overline{EX}_{h,0}^{k\ *} + \gamma d_{h}^{k} + u_{h}^{k} + \varepsilon_{ht}^{k}$$

$$with\ \ \varepsilon_{ht}^{k} \sim N\left(0, \sigma_{\varepsilon}^{2,k}\right), \qquad u_{h}^{k} \sim N\left(0, \sigma_{u}^{2,k}\right)$$

$$\overline{EX}_{ht}^{k\ *} = ln\left(\overline{EX}_{ht}^{k\ *} + \sqrt{\overline{EX}_{ht}^{k\ *2} + 1}\right) = X_{ht}^{k}\beta_{h}^{k} + \epsilon_{ht}^{k}$$

$$where\ X_{ht}^{k} : a\ vector\ of\ explanatory\ variables$$

$$\beta_{h}^{k} : a\ vector\ of\ coefficients$$

$$\epsilon_{ht}^{k} = u_{h}^{k} + \varepsilon_{ht}^{k}\ \ and\ \ \epsilon_{ht}^{k} \sim N(0, \sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k})$$

$$EX_{ht}^{k} = \frac{1}{2}\exp\left(X_{ht}^{k}\beta_{h}^{k} + \epsilon_{ht}^{k}\right) - \frac{1}{2}\exp\ (X_{ht}^{k}\beta_{h}^{k} + \epsilon_{ht}^{k})^{-1}$$

$$\begin{aligned}
E\left[EX_{ht}^{k} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right] &= E\left[\frac{1}{2}\exp\left(X_{ht}^{k}\beta_{h}^{k} + \epsilon_{ht}^{k}\right) - \frac{1}{2}\exp\left(X_{ht}^{k}\beta_{h}^{k} + \epsilon_{ht}^{k}\right)^{-1} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right] \\
&= E\left[\frac{1}{2}\exp\left(X_{ht}^{k}\beta_{h}^{k} + \epsilon_{ht}^{k}\right) \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right] \\
&\qquad - E\left[\frac{1}{2}\exp\left(X_{ht}^{k}\beta_{h}^{k} + \epsilon_{ht}^{k}\right)^{-1} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right] \\
&= \frac{1}{2}\exp\left(X_{ht}^{k}\beta_{h}^{k}\right) E\left[\exp\left(\epsilon_{ht}^{k}\right) \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right] \\
&\qquad - \frac{1}{2}\exp\left(-X_{ht}^{k}\beta_{h}^{k}\right) E\left[\exp\left(-\epsilon_{ht}^{k}\right) \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right]
\end{aligned}$$

since the exponential function is an increasing and nonnegative function

$$= \frac{1}{2}\exp\left(X_{ht}^{k}\beta_{h}^{k}\right) E\left[\exp\left(\epsilon_{ht}^{k}\right) \middle| \exp\left(\epsilon_{ht}^{k}\right) > \exp\left(-X_{ht}^{k}\beta_{h}^{k}\right)\right]$$

$$- \frac{1}{2}\exp\left(-X_{ht}^{k}\beta_{h}^{k}\right) E\left[\exp\left(-\epsilon_{ht}^{k}\right) \middle| \exp\left(-\epsilon_{ht}^{k}\right) < \exp\left(X_{ht}^{k}\beta_{h}^{k}\right)\right]$$

since $\exp\left(\epsilon_{ht}^{k}\right) \sim LN\left(0, \sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}\right)$,

$$E\left[\exp\left(\epsilon_{ht}^{k}\right) \middle| \exp\left(\epsilon_{ht}^{k}\right) > \exp\left(-X_{ht}^{k}\beta_{h}^{k}\right)\right] = \exp\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}{2}\right)\left\{\frac{\Phi\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k} + X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}{1 - \Phi\left(\frac{-X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}\right\}$$

$$E\left[\exp\left(-\epsilon_{ht}^{k}\right) \middle| \exp\left(-\epsilon_{ht}^{k}\right) < \exp\left(X_{ht}^{k}\beta_{h}^{k}\right)\right] = \exp\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}{2}\right)\left\{\frac{1 - \Phi\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k} - X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}{\Phi\left(\frac{X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}\right\}$$

$$= \frac{1}{2}\exp\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}{2}\right)\left[\exp\left(X_{ht}^{k}\beta_{h}^{k}\right)\left\{\frac{\Phi\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k} + X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}{1 - \Phi\left(\frac{-X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}\right\}\right.$$

$$\left. - \exp\left(-X_{ht}^{k}\beta_{h}^{k}\right)\left\{\frac{1 - \Phi\left(\frac{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k} - X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}{\Phi\left(\frac{X_{ht}^{k}\beta_{h}^{k}}{\sqrt{\sigma_{\varepsilon}^{2,k} + \sigma_{u}^{2,k}}}\right)}\right\}\right]$$

Deriving the first derivative of the conditional expectation with respect to continuous explanatory variables:

$$\frac{\partial E\left[EX_{ht}^{k} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right]}{\partial X_{ht}^{k}}$$

$$= \beta_h^k \frac{1}{2} \exp\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}{2}\right) \left[\exp\left(X_{ht}^k \beta_h^k\right) \left\{ \frac{\Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)} \right\} \right.$$

$$\left. - \frac{1}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}} \left\{ \frac{\phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right) - \Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right) \phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right)^2} \right\} \right]$$

Deriving the first derivative of the conditional expectation with respect to discrete explanatory variables:

$$\frac{\partial E\left[EX_{ht}^{k} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right]}{\partial X_{ht}^{k}} = E\left[EX_{ht}^{k} \middle| X_{ht}^{k} = 1, EX_{ht}^{k} > 0\right] - E\left[EX_{ht}^{k} \middle| X_{ht}^{k} = 0, EX_{ht}^{k} > 0\right]$$

Deriving the first derivative of the conditional expectation with respect to the lagged dependent variable:

$$\frac{\partial E\left[EX_{ht}^{k} \middle| X_{ht}^{k}, EX_{ht}^{k} > 0\right]}{\partial \overline{EX}_{h,t-1}^{k}{}^{*}} * \frac{\partial \overline{EX}_{h,t-1}^{k}{}^{*}}{\partial EX_{ht}^{k}}$$

$$= \left[\beta_h^k \frac{1}{2} \exp\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}{2}\right) \left[\exp\left(X_{ht}^k \beta_h^k\right) \left\{ \frac{\Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)} \right\} \right.\right.$$

$$\left.\left. - \frac{1}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}} \left\{ \frac{\phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right) - \Phi\left(\frac{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k} + X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right) \phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)}{\left(1 - \Phi\left(\frac{-X_{ht}^k \beta_h^k}{\sqrt{\sigma_\varepsilon^{2,k} + \sigma_u^{2,k}}}\right)\right)^2} \right\} \right]\right]$$

$$* \left(\frac{1}{\sqrt{EX_{ht}^{k}{}^{2} + 1}}\right)$$