International Food and Agribusiness Management Review
Volume 15, Issue 4, 2012

Food Store Density, Nutrition Education, Eating Habits and Obesity

Alessandro Bonanno[ᶲa] and Stephan J. Goetz[b]

[a]*Assistant Professor of Agricultural Economics, Department of Agricultural Economics, Sociology and Education, The Pennsylvania State University, 207-D Armsby Building, University Park, Pennsylvania, 16802, USA*

[b]*Director, The Northeast Regional Center for Rural Development, Professor of Agricultural and Regional Economics, The Pennsylvania State University, 207-C Armsby Building, University Park, Pennsylvania, 16802 USA*

Abstract

Food retailers and restaurants are under scrutiny for their alleged effects on diets and obesity, although no clear evidence of a causal relationship exists. Furthermore, because no prior study controls for nutrition education and the dynamic nature of the underlying phenomena, existing estimates quantifying these relationships could be biased. Using state-level data for the continental U.S. we evaluate how the density of different food stores and per-capita expenditures on SNAP (nutrition) Education impact eating habits and (indirectly) adult obesity, controlling for endogeneity of store locations and con-sumption dynamics. Our results caution against using large-scale policies regulating the food environment and highlight the need to control for nutrition education and process dynamics to obtain unbiased estimates. Implications for the agribusiness sector are discussed.

Keywords: food store density, fruits and vegetables consumption, adult obesity, nutrition education, endogeneity bias, omitted variables bias

[ᶲ]Corresponding author: Tel: +1.814.863.8633
Email: abonanno@psu.edu

+The IFAMR is a non-profit publication. The additional support provided from this issue advertiser, Novus International helps us remain open access and dedicated to serving management, scholars, and policy makers worldwide.

Introduction, Background, and Objectives

In 2007-2008, one third of the U.S. adult population was classified as obese (Flegal et al. 2010).[1] Obesity results from consuming excess calories relative to need,[2] and rising obesity rates are associated with reduced physical labor due to technological change; increased agricultural productivity and food availability; and lower food prices (Philipson and Posner 1999; Lakdawalla and Philipson 2002). Technological progress in food processing has also reduced assembly/preparation time for meals and increased the availability of calorie-dense foods (Cutler, Glaeser, and Shapiro 2003). As a result of these trends, the agribusiness sector as a whole is under scrutiny as it may be contributing to the worsening of consumer diets and the obesity epidemic. Given the large scale of this problem, public policy makers are seeking to mitigate its dimension (Kuchler, Tegene, and Harris 2005). The policy debate assessing what tools may be most effective to curtail the "obesity epidemic" has considered, *inter alia*, taxing high-calorie carbonated soft drinks (Todd and Zhen 2010), nutrition labeling (Arsenault 2010) and regulating food access–i.e., the types of food outlets to which consumers are exposed.[3]

As higher consumption of particular food categories (such as fruit, as in Lin and Morrison, 2002) is associated with lower body weight, the effect of food outlets on obesity is likely to be indirect, through diets: for example, if certain kinds of food outlets facilitate the consumption of fruits and vegetables, they may mitigate the growing obesity phenomenon. To date, however, research findings in this area are mixed. Rose and Richards (2004) found ease of supermarket access to be associated with increased daily consumption of fruits and vegetables among low-income individuals (SNAP recipients), while Cummins et al. (2005) found no significant changes associated with the entry of a large-scale food retailer. Some studies have looked at the direct impact of food outlets on obesity rates or related consequences. Morland, Diez Roux, and Wing (2006) consider different outlets simultaneously, finding a negative relationship between the presence of supermarkets, overweight and obesity, and the opposite for grocery and convenience stores. Chen, Florax, and Snyder (2009), using geo-referenced micro data from the Indianapolis urban area, found a negative relationship between the density of grocery stores and Body Mass Index (BMI). Courtemanche and Carden (2011) found that the opening of one Wal-Mart Supercenter (henceforth WMSC) per 100,000 residents leads to a 0.24 point increase in BMI and an increase of 2.3% in the likelihood of being obese. This suggests that the real income-increasing effect of lower prices may not translate into consumption of healthier foods.

Researchers have also examined the role of restaurants on diets, as meals consumed in these establishments are usually less healthy than homemade meals (Lin and Frazão 1997; Chou, Grossman, and Saffer 2004). In particular, fast food restaurants are associated with higher consumption of fat, sodium and soft drinks (Bowman and Vinyard 2004) and lower consumption of fruits and vegetables (Powell et al. 2007). Also, using different databases and

[1] The BMI or Body Mass Index is the ratio of an individual's weight in Kilograms and height squared, in meters. The U.S. Center of Disease Control and Prevention (CDC) classifies the BMI of adults as: Underweight <18.5; Normal weight 18.5-24.9; Overweight: 25–29.9; Obese: >30. In 2007-08, 33.8% of adults were obese.
[2] For a discussion of the dynamics between energy stored, appetite, metabolism and the factors impacting these relationships, see Egger and Swinburn (1997).
[3] See White (2007) for a detailed review of the literature on food access and obesity.

empirical approaches, previous studies have established positive relationships between restaurant density and adult BMI (Chou, Grossman, and Saffer 2004; Chen, Florax, and Snyder 2009) or obesity among children (Currie et al. 2009). In response, policymakers have proposed zoning laws in urban areas to limit fast-food access (the CDC has a "Zoning to encourage healthy eating" program), mandatory caloric labels on restaurant menus and other strategies, all with limited effectiveness.[4]

Given the strong advocacy in the public media (see the examples reported in Collins and Baker, 2009) for regulating food outlet locations (in particular fast food restaurants) one would expect the existing empirical evidence to show a clear causal relationship between food store location, diets and obesity, and to account for sources of potential bias in the estimated impacts. However, several important aspects have been disregarded in prior research.

In the first place, most prior studies (Courtemanche and Carden 2011), being one of the exceptions, have largely disregarded issues of causality, failing to account for confounding factors (*e.g.*, endogeneity of store location) which could bias the estimated effect of particular food outlets on diets (fruits and vegetables consumption) and/or obesity rates/BMI. Such a bias could be particularly marked in the case of fast-food restaurants, as they may locate in neighborhoods where consumers are more likely to engage in unhealthy food choices and eating behaviors.[5] The existing evidence of causal effects from micro-level studies is mixed. Using number of highway exits at the county level as instruments for fast-food restaurant density, Dunn (2008) found that a 10% increase in fast-food restaurants increases BMI by 0.33 points. Using interstate highways in rural areas as instruments for restaurant density, Anderson and Matsa (2011) found no causal link between food consumption at fast-food and full service restaurants and obesity, indicating that consumers who eat more often at restaurants may offset calories by eating less on other occasions. Similar mixed findings emerge at the aggregate level: although a positive correlation exists between the presence of fast food restaurants and adult obesity at the state-level (Maddock 2004), serious doubt has been cast on whether the relationship is causal (Collins and Baker 2009).

Second, consumers' nutrition knowledge can influence the relationship between food stores and eating habits or obesity. Evidence exists that nutrition education results in higher intakes of fruit and vegetables (see, e.g., the results of nutrition interventions among older adults reported by Sahyoun, Pratt, and Anderson 2004) and that nutrition knowledge spills over across family members (in particular from mothers to preschool children, as shown by Variyam et al. 1999). Effective nutrition education could discourage consumers from patronizing fast food restaurants or increase patronage of fruit and vegetables stores as they seek healthier food products. The effect of nutrition education on the dietary choices of consumers patronizing larger stores is less clear; as these types of outlets offer access to a multitude of food items, some consumers may

---

[4] In New York City, for example, Elbel et al. (2009) found that labels increase awareness of calories but do not alter food choices.

[5] Areas with less-privileged individuals are characterized by limited access to "high quality" food stores (see for example Moore and Diez Roux, 2006; Powell et al. 2007), leading to poorer diets and higher obesity. Fast food outlets may locate disproportionately in low-income areas; to the extent that such areas contain more obese individuals, higher obesity rates may be a cause of fast food stores locating in a community, rather than a consequence.

still adopt unhealthy diets even in the presence of higher nutrition education expenditures, as they will be exposed to both healthy and unhealthy choices in the same store. Nonetheless, failure to control for the effect of nutrition knowledge or education on eating habits or obesity may result in omitted variables bias.

Finding variables that capture nutrition education is a challenge, as objective measures are needed in place of self-rated health knowledge measures which may be a weak determinant of consumption of fruits and vegetables (see for example Schroeter, House, and Lorence, 2009). Alternatively, one would need to capture a series of repeated controlled experiments or a large-scale nutrition education campaign. The USDA's Supplemental Nutritional Assistance Program-Nutrition Education (SNAP-Ed) is an example of the latter.[6] SNAP-Ed seeks to improve diets of low-income individuals, it is implemented to varying degrees across states,[7] and it has grown from $6.61 mn in FY 1992 to over $380 mn in FY 2010, providing variation in the implementation across time and space. At least one study (McGeary 2009) finds that increased federal SNAP-Ed outlays may help to mitigate adult obesity.

Last, as most studies linking food stores' presence, eating habits and obesity use cross sectional data, they cannot account for the dynamic aspects of the process generating eating habits and obesity. On average, consumers tend to prefer those foods they consume habitually (Mela 1999). This indicates that there is some persistence in eating habits over time, and that greater availability of certain foods may lead consumers to associate them with "the norm." Also, as energy-dense foods on average tend to be liked more than others (with taste rating higher than health and variety), habits involving the repeated consumption of these products are more likely to develop (Drewnowski and Specter 2004). Furthermore, as obesity can be an outcome of sustained energy imbalance over time (Egger and Swinburn 1997), current obesity levels are a function of present and past eating (and other) habits, as well as other features such as consumers' socio-demographic profiles and the food available to them. In sum, failure to account for the dynamic aspects of these relationships may result in model misspecifications and therefore in biased estimates.

The objective of this paper is to assess the existence of the biases discussed above and, more specifically, to: 1) analyze whether an aggregate, causal effect of the density of different food outlets on eating habits exists; 2) assess the indirect aggregate effect of food outlets on BMI via their impact on eating habits and; 3) determine whether investing in nutrition education could

[6] The Expanded Food and Nutrition Education Program (EFNEP) by the National Institute of Food and Agriculture is another example of federally funded program aimed at improving diets and nutrition education levels of limited-resource individuals. EFNEP activities target adults or youth audiences and aim to inform and train the recipients in different aspects of nutrition, diets, food preparation, food sourcing and health; program participants are selected through referral programs using several channels including SNAP and WIC offices, churches, local business, etc. Program delivery uses peer educators and volunteers trained by county extension professional and a variety of delivery methods. Given the heterogeneity of implementation we opted for not using it as proxy for nutrition education.

[7] The SNAP-Ed program is an optional program of nutrition education that State Agencies can deliver to SNAP recipients as part of their program operations. State agencies submit an annual SNAP-Ed plan to the Food and Nutrition System of the USDA (some States have multiyear plans), highlighting the budget and the proposed activities for the following year. Federal funds cover 50% of the program costs. The goal of SNAP-Ed is to increase the consumption of fruits and vegetables, whole grains, and fat-free or low-fat milk products, physical activity and to maintain a balanced caloric intake (USDA 2010).

substitute for regulation of food outlet locations in improving eating habits and obesity rates. We use state-level data for the Contiguous U.S. on the percent of adults eating at least five servings of fruit and vegetables daily as a proxy (albeit incomplete)[8] for eating habits, and the share of adult population with BMI of 30 or above, as a measure of obesity rates. Food outlet density is measured as per-capita grocery stores, fruit and vegetable stores, full and limited-service restaurants, and WMSCs. Our proxy for public expenditure on nutrition education programs are the inflation-adjusted per-capita SNAP-Ed expenditures.[9] We use supply-side drivers of store locations as instrumental variables and to account for possible endogenity of food store density. Also, to further reduce bias in the estimates, we account for the dynamic nature of the process generating eating habits and obesity.

Understanding the role of the food environment vs. nutrition education in expanding the share of adult population engaging in healthy eating habits has clear policy implications and is relevant for the agribusiness sector as a whole. Food retailers and food service companies, as well as many food manufacturers, are under scrutiny for their potential roles in shaping diets and in contributing to the obesity epidemic. This study seeks to provide additional evidence on whether policies aimed at regulating the food environment (i.e., the location of food retailers and restaurants) are likely to achieve the intended goals. Furthermore, analyzing the effect of nutrition education on fruit and vegetables consumption, vis-à-vis that of food outlets, could prove useful to agribusiness firms (especially food retailers) by allowing them to propose alternative "policy recipes" to direct regulation counteracting the obesity epidemic. Last, the quantification of the impact of nutrition education expenditure on the share of adults eating fruits and vegetables five or more times per day could prove useful to fruit and vegetables producers and processors as it may illustrate whether nutrition education can be used to expand their consumer base.

Empirical Methods

The Model

We posit a linear relationship between consumption of fruits and vegetables, food access (i.e., food store density), and other exogenous control variables. Let $FV5_{it}$ be the share of the adult population consuming fruit and vegetables at least five times a day in state i at time t (our proxy for healthy eating), NEd_{it} is a proxy measure of the average exposure to nutrition education of individuals in state i at time t, and FA_{jit} a proxy for the average consumer's access to the j-th type of food outlet in state i at time t. We consider:

[8] Fruit and vegetables consumption is, at best, an incomplete proxy for the adoption of healthy diets, as it does not take into account the consumption of other food groups (consumers who consume more servings of fruits and vegetables may also consume more "unhealthy" foods containing, for example, high values of sodium and fat). Furthermore, the concept of a "healthy" diet can differ among different subgroups of the population. Readers should keep in mind these caveats, which also hold for the interpretation of the results.
[9] All monetary variables are measured in real terms.

$$FV5_{it} = \gamma_0 + \gamma_{NEd} NEd_{it} + \sum_j \gamma_{FAj} FA_{jit} + \sum_h \gamma_{OHh} OH_{hit} + \sum_k \gamma_{SDk} SD_{kit} + \sum_m \gamma_{Enm} En_{mit} + \sum_l \gamma_{Rl} REG_{li} + \sum_t \gamma_{Tt} T_t + \varepsilon_{FVit} \quad (1)$$

where *OH* are variables capturing other habits and behaviors, *SD* are socio-demographic characteristics capturing heterogeneity in consumers' tastes, *En* are variables capturing environmental characteristics impacting eating habits, *REG* are regional fixed-effects capturing the heterogeneity of diets across areas, and *T* are year indicators controlling for changes in diets over time; the γs are parameters to be estimated, and ε_{FVit} is an idiosyncratic error term. Equation (1) ignores the dynamic aspects of the process generating eating habits, discussed above. Maintaining the assumptions of linearity, by virtue of the mechanics of the geometric distributed lag model[10] we can rewrite equation (1) to include the effect of lagged eating habits ($FV5_{it-1}$) as:

$$FV5_{it} = \gamma_0 + \gamma_{FV5t-1} FV5_{it-1} + \gamma_{NEd} NEd_{it} + \sum_j \gamma_{FAj} FA_{jit} + \sum_h \gamma_{OHh} OH_{hit} + \sum_k \gamma_{SDk} SD_{kit} + \sum_l \gamma_{Rl} REG_{li} + \sum_t \gamma_{Tt} T_t + \sum_m \gamma_{Enm} En_{mit} + \varepsilon_{FVit} \quad (2)$$

so that one can calculate both short-sun and long-run effects of explanatory variables on $FV5_{it}$. Consider a variable Z_j; while its short-run impact on $FV5_{it}$ is $\frac{\partial FV5_{it}}{\partial Z_j}^{SR} = \gamma_Z$, following the logic of the geometric distributed lag models (Greene, 2003), the long-run parameters (and implicitly the long-run marginal effects) of Z_j on $FV5_{it-1}$ can be measured as $\frac{\partial FV5_{it}}{\partial Z_j}^{LR} = \frac{\gamma_Z}{1-\gamma_{FV5t-1}}$. It should be noted that the long-run marginal effects refer to an indefinite future period in which the market reaches long-run equilibrium.

Let ObI_{it} represent the incidence of adult obesity in area *i* at time *t* and the relationship between obesity rates and its determinants be:

$$ObI_{it} = \delta_0 + \delta_{FV5} FV5_{it} + \sum_h \delta_{OHh} OH^{Ob}_{hit} + \sum_k \delta_{SDk} SD^{Ob}_{kit} + \sum_l \delta_{Rl} REG_{li} + \sum_t \delta_{Tt} T_t + \varepsilon_{ObIit} \quad (3)$$

where the groups of explanatory variables are described above, the δs are parameters to be estimated, and ε_{ObIit} is an idiosyncratic error term. Equation (3) states that obesity is a function of eating and other habits controlling also for socio-demographic characteristics: the subscript *Ob* indicates that the subsets of *OH* and *SD* entering the obesity equation are specific for that equation.

[10] See the Appendix for a brief illustration of this model. The interested reader will find a more detailed discussion of geometric lag models and other distributed lag models in Greene (2003).

One can combine the estimated coefficients of equations (1) and (3) to obtain the indirect marginal effects of food access and nutrition education on adult obesity incidence as:

$$\frac{\partial ObI_{it}}{\partial NEd_{it}} = \frac{\partial ObI_{it}}{\partial FV5_{it}} \frac{\partial FV5_{it}}{\partial NEd_{it}} = \delta_{FV5}\gamma_{NEd} \text{ and } \frac{\partial ObI_{it}}{\partial FA_{jit}} = \frac{\partial ObI_{it}}{\partial FV5_{it}} \frac{\partial FV5_{it}}{\partial FA_{jit}} = \delta_{FV5}\gamma_{FAj},$$

respectively. The intuition behind these measures is the following: given the assumptions of our model, a marginal change in nutrition education (access to outlet-type *j*) will lead to a marginal change in the incidence of adults consuming fruits and vegetables equal to γ_{NEd} (γ_{FAj}); since a marginal change in $FV5_{it}$ leads to a change in the incidence of adult obesity, a marginal change in NEd_{it} (FA_{jit}) will have an indirect impact on adult obesity equal to $\delta_{FV5}\gamma_{NEd}$ ($\delta_{FV5}\gamma_{FAj}$).

Similarly, the marginal effects on *ObI* of those variables impacting both the incidences of adult obesity and adults consuming fruit and vegetables at least five times a day, are illustrated below. Using a socio-demographic variable common to both equations (such as income and education, represented below as SD_k) one has:

$$\frac{\partial ObI_{it}}{\partial SD_{kit}} = \left.\frac{\partial ObI_{it}}{\partial SD_{kit}}\right|_{FV5_{it}} + \frac{\partial ObI_{it}}{\partial FV5_{it}} \frac{\partial FV5_{it}}{\partial SD_{kit}} = \delta_{SDk} + \delta_{FV5}\gamma_{SDk}.$$

where the first term in the mid-portion of the equation represents the direct effect of SD_k on adult obesity incidence, while the second captures the indirect effect through *FV5,* similar to the indirect effects of nutrition education and food access illustrated above. That is, these marginal effects will account for the fact that other habits (OH) and socio-demographic (SD) factors impact both eating habits and body weight.

Again, using the assumptions of the geometric distributed lag model and maintaining a linear functional form, an alternative specification of equation (3) that includes the effect of previous obesity values (ObI_{it-1}) on current values is:

$$\begin{aligned} ObI_{it} = {} & \delta_0 + \delta_{Ob-1}ObI_{it-1} + \delta_{FV5}FV5_{it} + \\ & \sum_h \delta_{OHh}OH_{hit} + \sum_k \delta_{SDk}SD_{kit} + \sum_l \delta_{Rl}REG_{li} + \sum_t \delta_{Tt}T_t + \varepsilon_{ObIit} \end{aligned} \tag{4}$$

where ObI_{it-1}represents the one-period lagged value of adult obesity in area *i*. Following the same logic discussed above for the other marginal effects and the calculation of long-run parameters, using the estimated coefficients of equations (2) and (4), the long-run marginal effects of the variation in food access, nutrition education and demographics on adult obesity rates are:

$$\frac{\partial ObI_{it}}{\partial FA_{jit}}^{LR} = \frac{\delta_{FV5}}{1-\delta_{Ob-1}} \frac{\gamma_{FAj}}{1-\gamma_{FV5t-1}}; \frac{\partial ObI_{it}}{\partial NEd_{it}}^{LR} = \frac{\delta_{FV5}}{1-\delta_{Ob-1}} \frac{\gamma_{NEd}}{1-\gamma_{FV5t-1}};$$

$$\frac{\partial ObI_{it}}{\partial SD_{kit}}^{LR} = \frac{\delta_{SDk}}{1-\delta_{Ob-1}} + \frac{\delta_{FV5}}{1-\delta_{Ob-1}} \frac{\gamma_{SDk}}{1-\gamma_{FV5t-1}}.$$

Data and Variables Description

The main data used are state-level aggregates from the "Prevalence and Trends Data" from the Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) survey,[11] available from the CDC website. "Adult obesity" is the state-level percent of adult population whose BMI is 30 and above, while the percent of "Adults who have consumed fruits and vegetables five or more times per day" (FV5 hereafter) is used as a proxy for healthy eating habits. The first variable is obtained from the "Weight Classification by Body Mass Index (BMI)" series, which presents the state-level share of adult population that is neither overweight nor obese (BMI $\leq$ 24.9), Overweight (25.0 $\leq$ BMI < 29.9) and Obese (BMI $\geq$ 30.0). The state-level aggregates are obtained based on individuals' BMIs calculated from self-reported values of height and weight (BMI>100 are discarded). Similarly FV5 is obtained via state averages of the number of survey respondents who were found to consume more than five servings of fruits and vegetables daily, with values imputed from individuals' answers to a series of six questions regarding frequency of consumption of fruit and vegetables.[12]

As noted, food store density is measured as the number of food store establishments divided by population. Food store establishment data are from the County Business Pattern, U.S. Bureau of Labor Statistics (BLS). The industries included are: NAICS 4451: Grocery Stores, NAICS 44523: Fruit and Vegetables Stores;[13] NAICS 7222: Limited Service Restaurants (proxy for fast-food) and NAICS 722: Food Services and Drinking Places. The difference between NAICS 722 and NAICS 7222 establishments represents full-service restaurants. State-level numbers of WMSCs are from the company's annual shareholder reports. State-level population is from the Population Estimates Program (PEP).

Nutrition education is measured as real state-level per capita annual federal outlays on SNAP-Nutrition Education program (SNAP-Ed). State-level federal outlays, in current dollars were obtained from public use data from the Food and Nutrition Service of the United States Department of Agriculture (FNS – USDA).[14] These amounts represent 50% of the expenditure in programs of nutrition education proposed by State Agencies in each state, i.e., the amount that is federally funded and excluding the expenditure of each state. To account for inflation and evaluate real expenditure, we divided the outlays by general Consumer Price Index (CPI) from

---

[11]The BRFSS is an on-going telephone health survey system tracking health conditions and risk behaviors among the U.S. population and collects data on 1) individuals' habits, such as smoking, physical activity, alcohol consumption; 2) health status and health prevention measures, e.g., whether the respondents had high blood pressure, high cholesterol levels, access to healthcare, etc., and 3) respondents' socio-demographic characteristics such a children in the household, age and gender. All of these characteristics are self-reported.

[12] Consumption of fruits and vegetables is recorded separately in the BRFSS through six questions, two on the frequency (daily, weekly, monthly or annually) of consumption of fruits (fruit juices and fruit, excluding juices, respectively) and four on the consumption of vegetables (green salad, potatoes - excluding chips and fries, carrots, and servings of other vegetables). The reported values are converted into daily servings consumed by each group, and then used to create summary indexes of fruits and vegetables consumed per day and an indicator variable capturing whether an individual consumed five or more servings of fruits and vegetables per day.

[13] According to the official Census Definition (http://www.census.gov/eos/www/naics/) . The establishments included in this industry are specialty food stores primarily engaged in retailing fresh fruits and vegetables, excluding roadside stands and electronic, direct or mail sales. Farmers markets are not included in this classification.

[14] We thank Alice Lockett, at the FNS-USDA for providing data on State-level Federal Outlays for the SNAP-Ed program.

the BLS, and by state-level population, to control for the different sizes of each state and for consistency with the other variables in the models. As such, the coefficeints associated with this variable measure how the program could perform if it reached the entire state-population.

Controls used in both equations are the percent of population that is physically active five or more times per week for at least 30 minutes (Phys Act); with a college degree or higher (>College); without children (No Child); and female (% Female) from the BRFSS data. Socio-demographic characteristics such as population belonging to different ethnic groups (% Black; % Other ethnicities) and average age (Age) come from the PEP; real per-capita personal income (Income) is obtained by dividing state-level total income from the American Community Survey by state-level population and by General CPI. The percent of adult population currently smoking (Smoke) and married (% Married), from the BRFSS, are included only in the obesity equation.[15] We use two environmental characteristics that can have an impact on eating habits. The firs variable is the general CPI, which accounts for price levels affecting real income and quality of food purchased. The second is the annual average state-level temperature (Temp) from the Earth System Research Laboratory (ESRL) capturing differences in diets due to geography, the likelihood of engaging in outdoor activities, or in activities which may result in different lifestyles and therefore diets (including visiting farm stands). The data cover the years 1998-2006[16] for 47 continental states,[17] for a total of 423 observations. U.S. Census Area fixed-effects and year dummies are used to account for unobservables and the panel nature of the data. A summary of the variables used in the estimation is provided in Table 1, along with descriptive statistics.

Estimation and Identification

An instrumental variable estimation method, the Generalized Method of Moments, or GMM (Hansen, 1982) is used to obtain unbiased estimated and to control for the endogeneity of some of the explanatory variables. This method, as illustrated by Hansen, can be seen as a generalization of other IV methods, including two-stage least squares. Given a vector of exogenous variables Z, in the case of linear models the GMM estimator aims to find the vector of coefficients θ that satisfies the following moment conditions $E[Z'(Y-X'\theta)]=0$ or $E[Z'e]=0$, or in other words, the vector of coefficients for which the errors obtained from the model are orthogonal to the vector of exogenous variables Z. Note that Z contains *all* of the exogenous variables in X and at least as many other exogenous variables as the number of endogenous variables. The GMM estimator solves

$$\min_{\theta}[Z'(Y-X'\theta)]'W[Z'(Y-X'\theta)],$$

[15] Smoking habits and marital status are usually controlled for in studies related to obesity; controlling for these variables is not as common in studies of fruit and vegetables consumption. For example Cummins *et al.* (2005) do not control for smoking habits and marital status; Rose and Richards (2004) who did not include smoking as control variable, include instead single parent households, a variable not available in our data.

[16] Although the state-level BRFSS Prevalence and Trends Data are available from 1984, we limit our analysis to the period 1998-2006 to avoid problems arising from changes in the Census' industry classification system, which switched from the SIC 1987 to the NAICS 1997. Observations for FV5 incidence and other regressors, not available for some years, were recovered using linear interpolation.

[17] Utah was excluded from the sample because of missing observations in the BRFSS data.

where W is a weight matrix (usually $Z'Z$), which is solved by

$$\theta^{GMM} = ((X'Z)W(Z'X))^{-1}(X'Z)Y$$ [18]

An illustration of the identification strategy follows. As the food store location decision is in part driven by demand-side factors impacting consumers' eating habits not captured by the other control variables in the model, FA is likely to be correlated with the errors in the FV5 equation.[19] Since food retailers and food service establishments locate preferentially where pre-existing infrastructures provide ease of transportation and implementation of logistics structure, our identification strategy uses historical information on infrastructure to capture exogenous (to diets) variation in store-density. We use state-level miles of federal highways in 1950 (U.S. Department of Transportation, Federal Highway Administration, 1950)[20], segmented by rural and urban areas, as well as the percent of federal highways in each state in urban and rural locations as instruments. The portion of rural and non-rural land in each state and the square miles of land are from the Gazetteer of counties (U.S. Bureau of Census, 2001). Furthermore, we control for market potential and proxies for land prices, *e.g.*, state-level population density, total land available and the percent of land in natural parks and preserves; also we use the maximum state-level corporate income tax rate (from the U.S. tax foundation) to capture the characteristics of the business environment across states. Our identification strategy for WMSCs[21] follows the notion that the company's expansion into food retailing capitalizes on converting its mass merchandize Discount Stores (DSs) into supercenters (see Bonanno, 2010); therefore the current number of WMSCs is regressed on the lagged number of DSs (similar to Basker and Noal (2009) who use historical values instead), the average distance from food distribution centers based on Holmes' store location database (Holmes 2010) regional fixed-effects and year dummies. The predicted number of stores is divided by total population (in hundreds of thousands) and used in lieu of the actual ones in the estimation.[22]

Endogeneity bias may also affect the obesity equation estimates, since unobservables impacting adult obesity rates could be correlated with fruit and vegetable consumption (e.g., consumption of other food groups). To resolve this issue, the excluded variables from the FV5 equations are used as instruments for $FV5_{it}$ in the obesity equation. In equation (2) such variables are CPI, average temperatures, per-capita real SNAP-Ed expenditure and the instrument for WMSC; the same variables are also used in equation (4), along with the appropriately instrumented lagged FV5.[23]

---

[18] For a more detailed illustration of the GMM estimator see Wooldridge (2002), Chapter 14.

[19] If, for example, an area is characterized by a higher demand for unhealthy high-calorie food, limited service restaurants will be more likely to find such areas to be profitable and locate there.

[20] We use historical highway density measures instead of contemporary ones to mitigate issues of spurious correlation which may arise if the objective of structural interventions to improve the capillarity of highway systems had, as a goal, that of attracting more businesses in a given area.

[21] Specifically, WMSCs locations may be correlated with particular socio-demographic profile, which may in turn be correlated with poorer diets (*e.g*., high poverty rates, as in Goetz and Swaminathan, 2006; or share of population receiving food stamps as in Bonanno, 2010).

[22] As different instruments were used to control for the endogeneity of WMSCs, we opted to run a separate first-stage OLS regression where both the lagged number of discount stores and the inverse distance from the company's food distribution centers showed positive and statistically significant coefficients, and the R-squared was 0.5996.

[23] The identifying assumptions for the ObI equations is that the exclusion restrictions used (*i.e.* the variables from the FV5 equations) explain FV5 but not obesity incidence and they are uncorrelated with the errors. The second point can be tested

Table 1. Variables used in the estimation

Variable	Description	Source	Mean	St. dev.
Ob	% of adult population with BMI > 30	BRFSS	21.82	3.50
FV5	% of adult population eating fruits and vegetables 5 or more times daily	BRFSS	23.44	3.75
Grocery	Establishments in NAICS 4451 / 1,000 people	BLS / PEP	0.34	0.09
FV stores	Establishments in NAICS 44523 / 100,000 people	BLS / PEP	0.95	0.66
Lim Serv. Res	(NAICS 722 – NAICS 7222) Establishments / 1,000 people	BLS / PEP	0.80	0.09
Full Serv. Res	Establishments in NAICS 7222 / 1,000 people	BLS / PEP	1.05	0.26
WMSCs	Number of WM Supercenters / 100,000 people	Wal-Mart Inc / PEP	0.49	0.46
SNAP-Ed	Per Capita Expenditure in SNAP-Ed / CPI	ERS / BLS	0.27	0.26
Phys Act	% adults: 30 + minutes of physical activity five or more days / week	BRFSS	74.50	5.13
Smoke	% respondents: currently smoke	BRFSS	22.57	2.93
No Child	% respondents: no children in their household	BRFSS	59.97	3.02
>College	% respondents: highest grade or year of school completed is “College or higher”	BRFSS	29.05	5.50
Income	Real per capita income (Total income / CPI / total population)	ACS / PEP	16.66	2.32
% Female	% of respondents being female	BRFSS	51.72	0.80
% Black	African American population / total population (%)	PEP	7.38	6.62
% Other ethnicities	Population other than White Caucasian or Black / total population (%)	PEP	28.66	2.97
% Married	% of adult population being married	BRFSS	60.34	3.24
Age	Average age	PEP	36.44	1.18
Temp	Average state-level annual temperature	ESRL	52.16	7.63
CPI	General Consumer Price Index	BLS	1.80	0.13

for (see footnote 24). Validating the first point is of particular importance given Courtemanche and Carden’s (2011) result that Wal-Mart’s presence is associated with higher obesity and McGeary’s (2009) findings that SNAP-Ed is negatively related to it. To that end, we included per capita real SNAP-Ed expenditure, CPI, average temperatures and the WMSC instrument as explanatory variables in a specification of the ObI equation where FV5 was excluded from the explanatory variables, and tested for their statistical significance. In the first place the coefficients of SNAP-Ed and temperatures were not statistically significant and, in spite those of those of CPI and WMSC being (individually) significant, jointly they were not. We re-estimated the ObI equation including FV5 in the model both via OLS and GMM (using temperature and SNAP-Ed as instruments for FV5), leaving CPI and WMSC as explanatory variables: the coefficients for both variables were not statistically significant. Repeating the same exercise including the instrumented lag of FV5 led to the same conclusions. This indicates that all these variables, including the WMSC instrument and SNAP-Ed, can be used as instruments in the ObI equation.

Lastly,[24] equations (2) and (4) contain lagged dependent variables on the RHS, which are endogenous by construction. In each of the two equations, the lagged dependent variable is regressed on lagged exogenous variables entering the respective equation, and the predicted values used in place of the actual ones. All equations are estimated both via Ordinary Least Squares (OLS) and (GMM).[25] An illustration of the moment conditions necessary to hold for the parameters of each equation to be identified is discussed in Appendix 2. Equation (1) was estimated with and without the inclusion of per-capita SNAP-Ed. Once estimates of the parameters are obtained, the impact of FA, SNAP-Ed and other relevant demographic variables on obesity are calculated. Estimation was performed using STATA v. 11.

Empirical Results

Table 2 presents the empirical results of different specifications of equations (1) and (2). The first two columns show OLS estimates without and with SNAP-Ed expenditure; respectively, the third and fourth columns are the GMM estimates. The models show similar goodness of fit (between 0.5274, for the model without SNAP-Ed, estimated via GMM, to 0.5835, for the model with SNAP-Ed, estimated via OLS).[26]

The magnitude and significance of the FA coefficients' estimates change once SNAP-Ed expenditure is controlled for, supporting the intuition that without controlling for nutrition education, the estimated impact of the built environment on eating habits (or obesity) may be biased. Also, the coefficients change considerably after accounting for FA endogeneity. The estimated grocery store coefficients are positive in all models, but in none of the models are they statistically significant suggesting that a positive correlation exists between access to grocery stores and FV5 incidence; however, this relationship is weak and in all likelihood non-causal, which may explain why previous more limited-scale studies show mixed findings (e.g., Rose and Richards' (2004) and Cummings et al. (2005) report conflicting findings).

[24] The use of a linearized geometric distributed lag model requires that the errors of the estimated model not be serially correlated. Tests for first order serial correlation, using Durbin's *h* statistics, show that, in all cases the errors were free from serial correlation.

[25] The orthogonality of the over-identifying instruments to the errors of the second stage regressions is evaluated via Hansen's (1982) test. The *J*-statistic of this test is distributed as chi-squared with degrees of freedom equal to the number of over-identifying instruments. A non-significant test statistic indicates that the instruments are valid. To evaluate instead the power of the instruments, we use Staiger and Stock's (1997) rule of thumb: if the *F*-statistic for the joint significance of the instruments' coefficients in the first stage equations exceeds 10 one can rule out weak instruments problems.

[26] The instruments used to correct for endogeneity of FA do not violate the orthogonality condition (the *p*-values of Hansen's (1982) *J*-tests are 0.2970 and 0.1649, respectively, in the model with and without SNAP-Ed).The larger *p*-value obtained when accounting for SNAP-Ed, suggests that controlling for this variable leads to more reliable results. Furthermore, the value of the *F*-statistics for the joint significance of the instruments' coefficients in the first stage equations exceed the "rule of thumb" value of 10, suggesting that the instruments used are not weak (Staiger and Stock, 1997). A more in-depth look at the first stage regression results shows the R-squared for the first-stage FA equations ranging from circa 0.6 (limited-service restaurants) to 0.84 (fruits and vegetables stores) and that remarkable similarities appear in the results. For example, maximum corporate tax rate has a negative and significant effect on the density of three out of the four types of stores while a positive relationship exists with population density (the one exception being full service restaurants). The proxy for land availability seems to impact primarily the store density measures with the exception of limited service restaurants, which are, conversely, heavily impacted by highway density (percent of federal highways in urban and rural areas), much more so than other outlets (for example grocery stores and fruits and vegetables stores are only weakly impacted by these variables).

Table 2. Estimated coefficients – Equation 1: incidence of consuming five servings of fruits and vegetables per day.

	OLS		**GMM**	
	NO SNAP-Ed	With SNAP-Ed	NO SNAP-Ed	With SNAP-Ed
Grocery	3.2273	2.7901	4.4157	3.0192
	(2.3454)	(2.3224)	(6.2735)	(6.0118)
FVstores	0.8349**	1.1348***	0.6645*	0.8953**
	(0.3268)	(0.3364)	(0.4058)	(0.4077)
LimRes	-1.5921	-0.8313	-6.8123	-4.2693
	(2.1214)	(2.1104)	(4.4663)	(4.5650)
FullRes	-2.5394***	-2.8996***	3.0450*	1.6750
	(0.9027)	(0.8994)	(1.8373)	(1.8789)
WMSCs	-0.9640***	-1.0324***	-0.7796***	-0.8967***
	(0.2217)	(0.2201)	(0.2439)	(0.2453)
SNAP-Ed		1.8211***		1.3709**
		(0.5685)		(0.5830)
Phys Act	0.1124**	0.1039**	0.0899	0.0885
	(0.0499)	(0.0494)	(0.0752)	(0.0730)
No Child	0.0582	0.0621	0.0572	0.0554
	(0.0770)	(0.0762)	(0.1031)	(0.0996)
>College	0.2400***	0.2686***	0.1721**	0.1972**
	(0.0568)	(0.0569)	(0.0878)	(0.0858)
PC Income	0.2778**	0.2560**	0.5165***	0.4553**
	(0.1127)	(0.1116)	(0.1892)	(0.1872)
%Race	-0.2465***	-0.2832***	-0.2793***	-0.2960***
	(0.0746)	(0.0747)	(0.0806)	(0.0799)
% Black	-0.08243**	-0.09438**	-0.1248***	-0.1212***
	(0.0407)	(0.0404)	(0.0361)	(0.0352)
% Fem	-0.0625	-0.0537	0.6452	0.5308
	(0.3890)	(0.3845)	(0.4658)	(0.4567)
Age	-0.0160	-0.1030	-0.4510	-0.4221
	(0.2459)	(0.2445)	(0.3435)	(0.3342)
Temp	0.0226	0.0415	0.0885	0.0846
	(0.0361)	(0.0362)	(0.0566)	(0.0552)
CPI	11.348***	10.003**	11.899***	11.333***
	(4.3538)	(4.3242)	(4.4064)	(4.1458)
Constant	-10.2790	-5.8101	-36.2370	-30.0000
	(22.7260)	(22.5080)	(27.8550)	(26.7270)
R-squared	0.5727	0.5835	0.5274	0.5542
Hansen's J [$\chi^2_{(4)}$]			6.4976	4.9069
p-value J			0.1649	0.2970

Note. *, **, and *** represent 10, 5 and 1% significance levels, respectively. Standard errors in parenthesis. Regional fixed-effects and time dummy coefficients excluded for brevity.

The estimated coefficients of F&V stores are positive and significant, across specifications and estimation methods. After controlling for SNAP-Ed expenditure, the estimated coefficient for F&V stores increases by approximately 30% (0.835 to 1.135 in the OLS results; 0.665 to 0.895 in the GMM). Since the sample average state-level density of FV stores is about one store per 100,000 people, the results indicate that doubling access to F&V stores raises the FV5 incidence by 0.66 % to about 0.9%.

The OLS estimates of the full and limited service restaurants coefficients are negative; the former (−2.539 and −2.899) are statistically significant, while the latter (−1.592 and –0.831) are not, suggesting a negative correlation between restaurants and the incidence of fruits and vegetables consumption among the adult population. The coefficients for limited service restaurants become more negative after endogeneity is controlled for, although it is not statistically significant, suggesting that the density of these establishments is negatively correlated with our crude proxy for healthy eating and that, because fruits and vegetables consumption is likely negatively related to obesity incidence, our results are consistent with previous studies (Chou, Grossman, and Saffer, 2004; Chen, Florax, and Snyder, 2009; Currie et al. 2009); however, this relationship is likely to be non-causal. Instead, once endogeneity is controlled for the sign of the full service restaurant coefficient becomes positive, although significant at the 10% level only in the model without SNAP-Ed.

This result seems to support Anderson and Matsa's (2011) argument that excess calories from meals at restaurants could be offset by reducing consumption on other occasions, or in this case, by increasing the frequency of consumption of fruits and vegetables. However, not all individuals may make such calorie-offsetting decisions: the fact that the coefficient loses statistical significance when our proxy for nutrition education is controlled for, may indicate that only a portion of the population (perhaps consumers who are more educated from a nutritional stand point) would make better dietary choices in the presence of more restaurants, and that once education is controlled for only a correlation persists.

The effect of WMSCs on FV5 incidence is negative and statistically significant across models and estimation methods, with the GMM coefficients showing a lower magnitude than the OLS coefficients, while the coefficients are eight to twelve percent larger when SNAP-Ed expenditure is accounted for. This supports Courtemanche and Carden's (2011) contention that the real income-increasing effect of Wal-Mart' lower prices does not translate into consumption of healthier foods; as the negative relationship between the incidence of adults consuming fruits and vegetables five or more times a day and the company's presence becomes more marked controlling for nutrition education, it may suggest that the less educated consumers are those engaging the most in unhealthy eating practices once they are exposed to more varieties at lower prices. Our results indicate that an increase of one WMSC per 100,000 individuals (a 200% increase in store numbers, the sample average of which is 0.49) would reduce FV5 incidence by 0.78 to 0.9 percentage points.

The coefficients associated with SNAP-Ed expenditures are positive and significant using both OLS and GMM (1.821 and 1.371, respectively), indicating that an increase in per-capita expenditure in SNAP-Ed of $1/year would raise FV5 incidence by approximately 1.4 to 1.8

percentage points. In other words, investing about $0.55/year per person in SNAP-Ed could increase FV5 by up to one percent. As the average real per-capita SNAP-Ed in our sample is circa 0.27 $/year, such a 1% increase would require increasing SNAP-Ed expenditures by 200 to 260%. The reader should keep in mind that since this variable is obtained dividing SNAP-Ed expenditure by the total population, these increases assume that the campaign reaches the entire population of a state. The estimated coefficients of the other variables in the model show similar magnitude and significance across specifications. Physical activity, per capita income and education are positively related to healthy eating (the first being significant only in the OLS estimates), and so is living in areas characterized by warmer climate (significant at the 10% in the model estimated via GMM) and higher price levels. The percent of population that is black or belongs to minority ethnic groups is negatively related to FV5 incidence, while percent of families without children, average age, and the percent of female population have no impact.

Table 3 presents results for equation 2: OLS estimates in the first column and GMM in the second; the third and fourth columns represent the respective long-run parameters. Overall, the direction and magnitude of most of the estimated coefficients are comparable to those presented in Table 2, with improved goodness of fit (0.611 and 0.589 for the models estimated via OLS and GMM, respectively).[27] The coefficients associated with $FV5_{it-1}$ are positive and significant, i.e., we find that the values of state-level of FV5 persist over time. The OLS and GMM coefficients are, respectively, 0.3601 and 0.3092, indicating that the impact of previous eating habits could be overestimated by circa 18 % if one does not account for the endogeneity of food store density. The behavior of the estimated FA coefficients is similar to that of equation (1).

Discussing only the statistically significant GMM estimates, the estimated long-run effects show that increasing the number of F&V stores by one unit per 100,000 people could, in the long-run, raise FV5 incidence by 1.23 percentage points while an increase of one WMSC per 100,000 individuals would lower it by 1.53 percentage points. In the long-run, an increase of $1/year per person in SNAP-Ed expenditure will increase FV5 incidence by 1.79 (GMM estimates) to 2.32 (OLS) percentage points. Thus, in the long-run, FV5 incidence could increase by one percentage point if SNAP-Ed expenditures increased by 0.43 to 0.56 $/person per year.

[27] Also, in this case, the instruments for food store locations satisfy the orthogonality condition (*p*-value of Hansens's *J*-test is 0.5099) and show sufficient explanatory power. The performance of the first-stage regressions is very similar to that illustrated for equation (1) in footnote 26, and is not presented for the sake of brevity.

Table 3. Equation 2 – Estimated coefficients: incidence of consuming five servings of fruits and vegetables per day and long-run parameters

	Estimated Coefficients		**Long-run Parameters**	
	OLS	*GMM*	*OLS*	*GMM*
Lag FV5	0.3601**	0.3092*		
	(0.1629)	(0.1837)		
Grocery	4.4194**	0.3054	6.9067 *	0.4421
	-1.9860	-5.5938	(3.6443)	(8.0404)
FV stores	0.8619***	0.8503**	1.3469 **	1.2310 *
	(0.3232)	(0.4026)	(0.5703)	(0.7036)
Lim Serv. Res	-1.3502	-3.4589	-2.1102	-5.0073
	(2.1693)	(4.0061)	(3.3798)	(5.9355)
Full Serv. Res	-2.5601***	0.7393	-4.0010 **	1.0702
	(0.7157)	(1.8877)	(1.6657)	(2.6267)
WMSCs	-1.0746***	-1.0590***	-1.6794 ***	-1.5331 ***
	(0.2104)	(0.2272)	(0.4911)	(0.5325)
SNAP-Ed	1.4848***	1.2347**	2.3205 **	1.7875 *
	(0.4710)	(0.5726)	(0.9399)	(0.9653)
Phys Act	0.0031	0.0124	0.0049	0.0180
	(0.0794)	(0.0800)	(0.1243)	(0.1167)
No Child	0.0612	0.0228	0.0956	0.0330
	(0.0919)	(0.0883)	(0.1546)	(0.1311)
> College	0.1573*	0.1300	0.2458 **	0.1881
	(0.0920)	(0.0950)	(0.1088)	(0.1210)
Income	0.1942*	0.2772	0.3035 **	0.4012 *
	(0.1061)	(0.2058)	(0.1536)	(0.2505)
% Oth. Race	-0.2740***	-0.2586***	-0.4282 ***	-0.3744 ***
	(0.0837)	(0.0945)	(0.1292)	(0.1148)
% Black	-0.0624	-0.0671	-0.0975 *	-0.0971 *
	(0.0434)	(0.0524)	(0.0555)	(0.0579)
% Fem	(0.1912)	0.1336	-0.2988	0.1935
	(0.3596)	(0.5558)	(0.6026)	(0.7745)
Age	-0.1759	-0.2388	-0.2748	-0.3457
	(0.2450)	(0.3137)	(0.4021)	(0.4433)
Temp	0.0376	0.0511	11.5955	13.0750
	(0.0320)	(0.0527)	(8.4214)	(8.4790)
CPI	7.4196	9.0318*	0.0588	0.0740
	(4.5355)	(4.7364)	(0.0514)	(0.0699)
Constant	12.9670	-3.7785		
	(19.5750)	(28.9290)		
R-squared	0.6108	0.5893		
Hansen's J [$\chi^2_{(4)}$]		3.2693		
p-value J		0.5099		

Note. *, **, and *** represent 10, 5 and 1% significance levels, respectively. Standard errors in parenthesis. Regional fixed-effects and time dummy coefficients excluded for brevity.

Table 4 presents the results of the obesity equations (3) and (4). The first two columns contain OLS and GMM estimated parameters of equation (3) while the third and fourth columns report the GMM estimates of equation (4) and the calculated long-run parameters. The goodness of fit of the models is comparable (the R-squared are 0.8709 and 0.8423, for equation 3, OLS and GMM respectively, and 0.8567 for equation 4, GMM). While no relationship seems to emerge between FV5 and adult obesity incidence in the OLS estimates (the coefficient being positive, small and not statistically significant), once endogeneity of eating habits is accounted for,[28] the results show that a one percentage point increase in FV5 incidence lowers adult obesity incidence by approximately −0.24 % (in the short run), while a long-run marginal decrease can reach -0.44%.

Most of the estimated coefficients for the other variables show similar magnitude, sign and significance across specifications and behave as expected. The percent of population holding a college degree or higher shows negative and significant coefficients as do having no children; the signs associated with average age, smoking and belonging to minorities are negative but not statistically significant. Also, the percent of population married and the percent of black population do not show a statistically significant effect. The coefficients of two variables have signs counter to expectation: physical activity and per capita income, both of them positive, although the latter not statistically significant: as these results do not account for the indirect effect on adult obesity incidence though their impact on healthy eating, the full effect should be considered instead, which is discussed below.

Table 5 presents the cumulative impact of SNAP-Ed expenditure, selected FA and demographic variables on adult obesity obtained combining the estimated parameters of equations 1 and 3, with those of equations 2 and 4 (GMM results). As they are obtained from parameters of separate equations, we are unable to provide standard errors associated with them, a caveat that readers should keep in mind. Increasing SNAP-Ed expenditure by $1/year per individual (that is, under the caveat that the policy would have to reach the entire population) could reduce the incidence of adult obesity by at least 0.3 percentage points. Such an increase translates into a long-run 0.8% decrease in the rate of adult obesity. An increase of one F&V store per 100,000 individuals reduces adult obesity incidence by approximately −0.2 percentage points in the short-run and −0.54% in the long run. The sample average of F&V stores density being about 1, we can conclude that 1) doubling the density of F&V stores reduces adult obesity incidence by half a percentage point, or that 2) increasing the density of F&V by 180% would lower obesity by one percent. The presence of WMSCs results in a short-run obesity-increasing effect ranging from 0.2 to 0.25, and a long-run marginal increase of 2/3 of a percentage point.

---

[28] The excluded variables from the *FV5* equations, used as instruments in the ObI equation, satisfy the orthogonality condition (*p*-values of Hansen *J* are 0.6290 and 0.3092, for equation (3) and (4), respectively). The R-squared of the regressions are 0.5556 and 0.5884; the coefficients of WMSC instrument and CPI were statistically significant in the first stage regression (the second at the 10 % level in equation (4)), while the other variables showed less explanatory power, although SNAP-Ed's coefficient approaches the 10% significant levels in both first stage regressions. In spite of the different performances across models, the parameters of the over-identifying instruments are jointly significant at the 1% level in both cases.

Table 4. Equations 3 and 4 – Estimated coefficients: adult obesity incidence and long-run parameters.

		Eq (3)	**Eq (4)**	**LR parameters**
	OLS	GMM	GMM	Eq. 4 (GMM)
Lag ObI			0.4634***	
			(0.0856)	
FV5	0.0055	-0.2338*	-0.2371*	-0.4419 *
	(0.0255)	(0.1394)	(0.1267)	(0.2309)
Phys Act	0.1006***	0.1228***	0.0619*	0.1154 *
	(0.0241)	(0.0376)	(0.0346)	(0.0684)
No Child	-0.1615***	-0.1864***	-0.1316***	-0.2452 ***
	(0.0386)	(0.0510)	(0.0494)	(0.0950)
> College	-0.3427***	-0.3227***	-0.1901***	-0.3542 ***
	(0.0309)	(0.0353)	(0.0412)	(0.0649)
% Smokers	-0.0116	-0.1134	(0.0881)	-0.1642
	(0.0381)	(0.0746)	(0.0755)	(0.1346)
Income	0.0104	0.0939	0.0474	0.0884
	(0.0494)	(0.0692)	(0.0683)	(0.1242)
% Oth. Race	-0.0084	-0.0877	0.0713	0.1328
	(0.0392)	(0.0646)	(0.0743)	(0.1509)
% Black	0.07654***	0.05276**	-0.0116	-0.0217
	(0.0217)	(0.0259)	(0.0268)	(0.0513)
% Fem	0.5401***	0.5185***	0.6510***	1.2131 ***
	(0.1637)	(0.1747)	(0.1740)	(0.3945)
Age	-0.0272	0.0265	-0.1162	-0.2165
	(0.1161)	(0.1373)	(0.1337)	(0.2541)
% Married	0.0281	-0.0284	0.0149	0.0277
	(0.0292)	(0.0447)	(0.0430)	(0.0817)
Constant	8.5778	19.2020	-1.6358	
	(9.5569)	(12.1710)	(13.7270)	
R-squared	0.8709	0.8423	0.8507	
Hansen's *J*	$\chi^2_{(3)}$= 1.7357;	*p*-val = 0.6290	$\chi^2_{(4)}$=4.7935; *p*-val = 0.3091	

Note. *, **, and *** represent 10, 5 and 1% significance levels, respectively. Standard errors in parentheses. Regional fixed-effects and time dummy coefficients excluded for brevity

The cumulative marginal effects of the socio-demographic variables show that increases in income, education and the level of physical activity all have an effect on containing obesity, which supports previous findings. However, the marginal effect of physical activity calculated from the coefficients of equations not including the dependent variable lag, shows perverse (positive) sign; this points to the importance of including dynamics in the model. The results indicate that, as income increases, consumers will consume more overall, but proportionately more fruit and vegetables. Looking at the results of the models accounting for system dynamics, the marginal effect of an increase of $1,000 per individual results in a 0.15 (short-run) to 0.26

(long-run) points decrease in adult obesity. Comparing this effect to that of SNAP-Ed, an increase of 0.33 $ per capita in SNAP-Ed among the entire population would (approximately) have the same effect on obesity as that of increasing consumers' income by $1,000. The cumulative effect of an increase in regular physical activity within the population is to reduce the adult obesity rate, suggesting that the effect of physical activity on obesity may not simply be due to the burning of more calories but also to an improvement in eating habits, because of the adoption of healthier life styles. Last, even if the percent of population holding at least a college degree did not significantly affect healthy eating, its direct effect on adult obesity prevails, leading to a short-run marginal effect of −0.14 and a long-run effect of −0.26.

Table 5. Calculated marginal impact of SNAP-Ed, selected FA and demographic variables on adult obesity incidence (GMM results only).

	Eq. 1 and 3	Eq. 2 and 4(SR)	Eq. 2 and 4 (LR)
SNAP-Ed	-0.3205	-0.2928	-0.7898
FV stores	-0.2093	-0.2016	-0.5439
WMSCs	0.2096	0.2511	0.6774
Income	-0.0125	-0.1534	-0.2585
Phys Act	0.1021	-0.1893	-0.3516
> College	-0.3888	-0.1442	-0.2559

Discussion, Implications for the Agribusiness Sector, and Limitations

The results illustrated above have a series of policy implications. In the first place, at the aggregate-level, we find no evidence of a negative causal relationship between the density of food-service establishments and the state-level incidence of adult healthy eating (similar to Collins and Baker, 2009, who find no "Granger causality" on obesity incidence using nationwide data), suggesting that policies aiming to restrict access to these outlets may have little impact on improving healthy diets. This result could be seen as an average (aggregate) outcome of consumers' eating habits, which is consistent with Anderson and Matsa's (2011) findings, i.e., that consumers who eat at restaurants more often could offset calories by reducing consumption on other occasions. Another result which may have policy implications is the detrimental effect of WMSCs on eating habits. Consistent with Courtemanche and Carden (2011) we find that the price-decreasing effects of the company may induce consumers to increase consumption overall, but not necessarily that of healthier foods. However, the company has announced (January 2011) a five-year plan to reduce the price of produce and the sodium, trans fat and added sugars content in several food produces under their private brands, as well as pledging to push major suppliers to follow their example (The New York Times, 2011). Policy makers should monitor closely whether this initiative impacts eating habits which, in light of the magnitude of our findings, could have a large impact on obesity rates.

Similarly, our results indicate that expenditures on nutrition education programs can improve eating habits and, indirectly, curb the incidence of adult obesity. However, increases in nutrition education efforts would have to be substantial. Using the combined long-run effects of SNAP-Ed on obesity, our results indicate that quadrupling average expenditure on nutrition education (see below for some additional caveats on the interpretation of this result) could reduce adult obesity by 0.8%; the feasibility of such a large spending increase as a policy tool is unlikely. However, as the presence of FV stores appears to be a catalyst of healthy eating, the use of local subsidies or zoning laws that enhance the presence of these outlets may be combined with larger nutrition education expenditures and have a synergic effect.

Our findings are relevant for the agribusiness industry on several fronts. In the first place, they advise against large-scale policies to regulate the structure of the food environment; although the results and the data used do not allow for evaluating the effectiveness of such interventions in specific contexts, they indicate that, at the aggregate level, their validity can be questioned. Second, an implication of our results is that agribusiness firms could considerably benefit from investing in nutrition education campaigns (or support already existing ones), on two fronts: 1) in general, they could benefit (indirectly) in terms of their public image as they will contribute to a more widespread adoption of healthy eating habits; 2) more specifically, fruit and vegetables producers and processors could directly benefit from an expanded consumer base which could result from an effective nutrition education campaign.

Our study is, however, not without limitations. First, we use the incidence of adults who consume more than five servings of fruits and vegetables daily as proxy for eating habits among the adult population; that is, as mentioned above, a rough and incomplete proxy for “healthy eating;” future research should consider using more refined measures of the overall quality of individuals’ diets. Second, given the aggregate nature of the data used, we cannot rule out that part of the population could be affected negatively by the presence of outlets such as fast-food restaurants. While some consumers may show marked preference for healthier alternatives, they may still opt for healthy diets regardless of the food environment, others, more susceptible to the surrounding environment, may be affected negatively by the presence of fast foods. These two effects may cancel each other out, resulting in a zero aggregate effect: as such, our results, which depict average aggregate effects, cannot and should not be used to draw inferences on how an individual’s fruit and vegetables consumption habits are affected by the presence of a particular food outlet. Third, although our statistical evidence supports the validity of our identification strategy, our findings are conditional on the choice of instruments for the food density measures. Additional research should explore an alternative identification strategy to shed more light on whether the relationship between the food environment and diet (and its outcomes, such as obesity) is causal or not, and whether different identification strategies produce different results. Fourth, as we estimate the two equations separately, we are unable to provide standard errors for the indirect and combined marginal effects of some of the explanatory variables on adult obesity incidence. Because of the small number of observations and the aggregate nature of the data used in this analysis, the limited variation in the data did not allow us to use successfully simultaneous equation modeling. Future research using more detailed databases and more sophisticated empirical approach could overcome such limitation.

Lastly, two remarks on our use of the per capita real expenditure on SNAP-Ed are warranted. First, it could be argued that because SNAP-Ed reaches only a small portion of the population, i.e., SNAP participants, it may not be worthwhile to investigate the possibility of increasing

SNAP-Ed funding as a policy tool. The reader should however keep in mind that, because of the population it reaches, SNAP-Ed actually targets individuals in greatest need of education from a nutritional standpoint and those who may therefore derive the greatest benefit from it. As a result, the FV5-increasing effect of SNAP-Ed may come largely from the fact that more disadvantaged consumers engage in "healthy eating"; as such, an increase in SNAP-Ed expenditure could be an effective policy tools to improve eating habits among low income households. However, given the size of the increase in SNAP-Ed expenditures needed to achieve a sizeable reduction in obesity, expanding the reach of the policy to target more population may, in practice, be infeasible.[29] The infeasibility of such approach is even more evident if one considers that we normalized federal SNAP-Ed expenditure by population, i.e., all of the estimated impacts of this variable refer to a hypothetical scenario where the funds were spent to educate the entire state population and not only SNAP recipients.

Second, our results do depict causal effects if and only if the measure of nutrition education used, per-capita SNAP-Ed, is in fact exogenous. To benefit from SNAP-Ed, states are required to match federal funds; therefore states in which diets are worse are also those where there may be more interest in investing in nutrition education. From an empirical standpoint, this means that one should control for enough factors associated with both diets and policymakers' decisions to invest in nutrition education, to guarantee that the estimated parameters of equations (1) and (2) are unbiased (McGeary, 2009). The risk of endogeneity bias is, however, limited as we control for consumers' heterogeneity through aggregate demographic indicators, and exogenous determinants of diets (such as temperature).[30]

Concluding Remarks

This paper has examined the simultaneous impact of different food outlets' density and expenditures in nutrition education programs on the incidence of healthy eating among the U.S. adult population and, indirectly, on adult obesity incidence, using a state-level panel data set and controlling for different sources of biases in the estimates. We find that even after controlling for omitted variable, endogeneity bias and the lagged fruit and vegetables consumption incidence, the density of fruits and vegetable stores is associated with higher shares of adults consuming fruits and vegetables regularly, and therefore lower obesity rates. The presence of Wal-Mart Supercenters, in contrast, across specifications, is associated with lower percentages of individuals consuming fruit and vegetables regularly and, as a consequence, with higher levels of obesity. The effect of SNAP-Ed expenditures is consistently that of more healthy eating and, consequently, reduced obesity.

Our results advise against large-scale policies to regulate the structure of the food environment as they indicate instead that investing in nutrition education could be (at the aggregate level), a more suitable tool to improve healthy eating incidence among the adult population. As a result,

[29] Also, as an anonymous reviewer pointed out, SNAP-Ed may also capture different attitudes towards SNAP within the states. Addressing this question is beyond the scope of our current study but worth exploring in future research.

[30] McGeary (2009) also points out that both funding levels and diets could change in response to local economic conditions. As we control for variables related to economic conditions, e.g., real income and CPI (related to inflationary trends) such a bias should not be an issue here.

while all agribusiness firms could benefit from investing in nutrition education campaigns, as they may gain in terms of image, fruit and vegetables producers and processors would benefit directly from it as they may experience an expanded consumer base.

Future research could use more refined data and econometric analysis to separate the direct and indirect impact of the food outlets on obesity, accounting for the trade-off between having access to more food outlets vs. that of having access to healthier ones. Alternatively, different types of food outlets (for example farmers' markets) or more refined measures of healthy eating (including for example whole grains, low fat dairy, etc.) could be examined.

References

Anderson, M., and D. A. Matsa. 2011. Are restaurants really supersizing America? *American Economic Journal: Applied Economics* 3(1):152–88.

Arsenault, J. E. 2010. Can nutrition labeling affect obesity? *Choices Magazine* 25(3): 27-30.

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(4): 977–1009.

Bonanno, A. 2010. An empirical investigation of Wal-Mart's expansion into food retailing. *Agribusiness: An International Journal* 26(2): 220–243.

Bowman, S. A., and B. T. Vinyard. 2004. Fast food consumption of U.S. adults: Impact on energy and nutrient intakes and overweight status. *Journal of the American College of Nutrition* 23(2):163-168.

Center for Disease Control and Prevention. 1996-2005. *Behavioral Risk Factor Sur*veillance System Survey Data. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Chen, S. E., R. J.G.M. Florax, and S. D. Snyder. 2009. *Obesity, Fast Food, and Grocery Stores: Evidence from Geo-referenced Micro Data.* National Poverty Center. Available at http://www.npc.umich.edu/news/event s/food-access/index.php

Chou, S., M. Grossmam, and H. Saffer. 2004. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics* 23: 565–587.

Collins R., and G.A. Baker. 2009. Of junk food and junk science. *International Food and Agri-business Management Review* 12(3) 111-126

Courtemanche, C., and A. Carden. 2011. Supersizing supercenters? The impact of Wal-Mart supercenters on Body Mass Index and obesity. *Journal of Urban Economics* 69 (2):165-181.

Cummins, S., M. Petticrew, C. Higgins, A. Findlay, and L. Sparks. Large scale food retailing as an Intervention for diet and health: quasi-experimental evaluation of a natural Experiment. *Journal of Epidemiology Community and Health* 59(12): 1035–1040.

Currie J., S. DellaVigna, E. Moretti, and V. Pathania. 2009. The effect of fast food restaurants on obesity. NBER Working .Paper No. 14721 http://www.nber.org/papers/w14721 (accessed September 12, 2011).

Cutler, D. M., E. L. Glaeser, and J. M. Shapiro. 2003. Why have Americans become more obese? *Journal of Economic Perspectives* 17 (3):93-118.

Drewnowski, A., and S.E. Specter. 2004. Poverty and obesity: the role of energy density and energy costs. *American Journal of Clinical Nutrition* 79 (1): 6-16.

Dunn, R. A. 2008. Obesity and the availability of fast-Food: an instrumental variables approach. Available at the SSRN eLibrary Paper No. 989363.

Earth System Research Laboratory. Average mean temperature index by month: 1971-2000 Available at: http://www.esrl.noaa.gov/psd/data/usclimate/tmp.state.112000.climo

Egger, G., and B. Swinburn. 1997. An "ecological" approach to the obesity pandemic. *BMJ* 1997, 315:477.

Elbel, B., R. Kersh, V. L. Brescoll, and B. L. Dixon. 2009. Calorie labeling and food choices: a first look at the effects on low-income people in New York City. *Health Affairs* 28(6): 1110-21.

Flegal, K. M., M. D. Carroll, C. L. Ogden, and L. R. Curtin. 2010. Prevalence and trends in obesity among US adults, 1999-2008 *Journal of the American Medical Association* 303(3): 235-241.

Goetz, S. J., and H. Swaminathan. 2006. Wal-Mart and family poverty in US counties. *Social Science Quarterly* 83(2006): 211-225.

Greene W. H. 2003. *Econometric Analysis*. 4th Edition. New Jersey, Prentice Hall.

Hansen, L. P. 1982. Large sample properties of Generalized Method of Moments estimators. *Econometrica* 50(4): 1029-54.

Holmes, T. J. 2010. Opening dates of Wal-Mart stores and Supercenters, 1962-Jan 31, 2006. http://www.econ.umn.edu/~holmes/data/WalMart/index.html (accessed June 4, 2011).

Kuchler, F., Tegene, A., and J.M. Harris. 2005. Taxing snacking foods: manipulating diet quality or financing information programs? *Review of Agricultural Economics* 27(1): 4-20.

Lakdawalla, D. and T. Philipson. 2002. The growth of obesity and technological change: A theoretical and empirical examination. NBER Working paper Series N. 8946. http://www.nber.org/papers/w8946 (accessed August 3, 2011).

Lin, B.-H. and E. Frazão. 1997. Nutritional quality of foods at and away from home. *Food Review* 20(1): 33-40.

Lin, B. and R. M. Morrison. 2002. Higher fruit consumption linked with lower Body Mass Index. *Food Review* 25(1): 28–32.

Mc Geary, Kerry A. 2009. The impact of State-Level Nutrition-Education Program funding on BMI: Evidence from the Behavioral Risk Factor Surveillance System, NBER working Paper w15001. http://www.nber.org/papers/w15001 (accessed August 3, 2011).

Mela, D. J. 1999. Food choice and intake: the human factor. *Proceedings of the Nutrition Society* 58(3): 513-521.

Maddock, J. 2004. The relationship between obesity and the prevalence of fast food restaurants: state-level analysis. *American Journal of Health Promotion* 19(2):137-143.

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

Morland, K., Diez Roux, A., and S. Wing. 2006. Supermarkets, other food stores, and obesity. The Atherosclerosis Risk in Communities Study. *American Journal of Preventive Medicine* 30(4): 333-339.

Philipson, T. J., and R. A. Posner. 1999. The long-run growth in obesity as a function of technological change. NBER Working Paper No. 7423.

Powell, L. M., Auld, C., Chaloupka, F. J., O'Malley, P. M. and L. D. Johnston. 2007.Access to fast food and food prices: relationship with fruit and vegetable consumption and overweight among adolescents. *Advances in Health Economics and Health Services Research* 17(1): 23–48.

Rose, D., and R. Richards. 2004. Food store access and household fruit and vegetable use among participants in the U.S. Food Stamp Program. *Public Health Nutrition* 7(8): 1081-1088.

Sahyoun, N. R., Pratt, C. A., and A. Anderson. 2004. Evaluation of nutrition education interventions for older adults: a proposed framework. *Journal of the American Dietetic Association* 104(1): 58-69.

Schroeter, C., L. House, and A. Lorence. 2007. Fruit and vegetable consumption among college students in Arkansas and Florida: Food culture vs. health knowledge. *International Food and Agribusiness Management Review* 10(3): 63-89

Staiger, D., and J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65(3): 557-586.

Stolberg, S.G. 2011. Wal-Mart shifts strategy to promote healthy foods. *The New York Times.* January 20.

Todd, J., and C. Zhen. 2010. Can taxes on calorically sweetened beverages reduce obesity? *Choices Magazine*, 25(3): 31-37.

U.S. Bureau of Census. 1998–2006. County business patterns database. http://www.census.gov/epcd/cbp/download/cbpdownload.html

U.S. Bureau of Census. 1998–2006. Population estimates program. http://www.census.gov/popest/estimates.php

U.S. Bureau of Census. 2001. Gazetteer of counties: Year 2000. http://www.census.gov/geo/www/gazetteer/places2k.html.

U. S. Department of Agriculture, Food and Nutrition Service. 1998-2006. Federal Outlays in Supplemental Assistance Nutrition Program.

U. S. Department of Agriculture, Food and Nutrition Service. 2010. SNAP-Ed factsheet. http://www.nal.usda.gov/snap/SNAP-EdFactsheet2010.pdf

U.S. Department of Transportation. Federal Highway Administration. 1952. Highway Statistics.

U.S. Tax Foundation. Corporate Income Tax Rates: 2000-2011. http://www.taxfoundation.org/taxdata/show/230.html. (accessed 06/01/2011).

Variyam, J.N., Blaylock, J., Lin, B.-H., Ralston, K., and D. Smallwood. 1999. Mother's nutrition knowledge and children's dietary intake. *American Journal of Agricultural Economics* 81(5): 373-384.

Wal-Mart Inc. 1998-2006. Shareholders' Annual Reports. www.walmartstores.com

White, Martin. 2007. Food access and obesity. *Obesity Reviews* 8(1): 99-107.

Wooldridge J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Appendix 1: A Brief Illustration of the Geometric Lag Model

The following illustration of the geometric lag model follows the discussion in Greene (2003), chapter 19. Note that Greene discusses two versions of the geometric distributed lag model: the model used here is similar to the "partial adjustment" model.

Let the covariates impacting eating habits be *X*, and *j* be their subscript. The process generating eating habits in area *i* at time *t* ($t = 1,\dots, T$) is $FV5_{it} = \gamma_0 + \sum_{j=1}^{J}\sum_{k=0}^{\infty}\lambda^k \gamma_j X_{ji(t-k)} + e_{FVit}; \quad 0 < \lambda < 1$

where λ^k is a weight given to the past values of the explanatory variables, and e_{FVti} contains an autoregressive component or $e_{FVit} = \lambda e_{FVit-1} + \varepsilon_{FVit}$, while ε_{FVit} is a mean 0 normally distributed disturbance. Subtracting $\lambda FV5_{it-1}$ from both sides of the equation and noting that

$\lambda FV5_{it-1} = \lambda\gamma_0 + \lambda\sum_{j=1}^{J}\sum_{k=0}^{T}\lambda^k \gamma_j X_{ji(t-1-k)} + \lambda e_{FVit-1}$ one has

$FV5_{it} - \lambda FV5_{it-1} = \gamma_0 - \lambda\gamma_0 + \sum_{j=1}^{J}\gamma_j X_{jit} + e_{FVit} - \lambda e_{FVit-1}$. Defining $\gamma_0^{'} = \gamma_0 - \lambda\gamma_0$ gives

$FV5_{it} = \gamma_0^{'} + \lambda FV5_{it-1} + \sum_{j=1}^{J}\gamma_j X_{jit} + \varepsilon_{FVit}$ which resembles equation (2).

Appendix 2. Moment Conditions

Following the notation in the "Empirical Methods" section, let's define vectors containing al the exogenous variables in equation 1 and 3, respectively as

$X_{FV5} = [NEd, OH, SD, WMSC, REG, T, En]$ (where $WMSC$ is the per-capita instrumented number of Wal-Mart Supercenters) and $X_{OBI} = [OH^{Ob}, SD^{Ob}, REG, T]$.
Let the exogenous variables used to capture supply-side variation in the determinant of food stores location in equation 1 and 3 be Z_{FA}. The moment condition to be satisfied in equation 1 and 3, respectively, are $E([X_{FV5}, Z_{FA}]'\varepsilon_{FV5}) = 0$ and $E([LFV5, X_{FV5}, Z_{FA}]'\varepsilon_{FV5}) = 0$, where $LFV5$ is the instrumented lag of FV5, while the moment condition that need to hold in equation 2 and 4, respectively, are

$E([X_{ObI}, NEd, WMSC, En]'\varepsilon_{ObI}) = 0$ and $E([X_{ObI}, LObI, LFV5, NEd, WMSC, En]'\varepsilon_{ObI}) = 0$