**A Complex Model of Consumer Food Acquisitions: Applying Machine Learning and Directed Acyclic Graphs to the National Household Food Acquisition and Purchase Survey (FoodAPS)**

Mark C. Senia
Department of Agricultural Economics
Texas A&M University
mcsenia@tamu.edu

Senarath Dharmasena
Department of Agricultural Economics
Texas A&M University
sdharmasena@tamu.edu

Jessica E. Todd
Economic Research Service (ERS)
United States Department of Agriculture (USDA)
jtodd@ers.usda.gov

***Selected Paper prepared for presentation at the Southern Agricultural Economics Association's 2018 Annual Meeting, Jacksonville, Florida, February, 3-6 2018***

**Funding**: This research received funding from the United States Department of Agriculture Co-Operative research grant award No. 58-4000-5-0042.

## Abstract

Complex causal relationships among a large set of variables that affect the U.S. households' food acquisition and purchase decisions were estimated using machine learning algorithms and directed acyclic graphs. Asians and Hispanics live in an environment with high concentrations of fast- and non-fast food restaurants. Obesity is less prevalent among Asians. Being Hispanic makes one to be more food insecure. Those with higher incomes are food secure and obesity is less prevalent among them. Being Black positively causes to be a SNAP participant and food insecure. Obesity is positively caused by fair/poor health and diet status.



Keywords: Directed Acyclic Graphs, National Household Food Acquisition and Purchase Survey, Food Environment, Food Insecurity, Obesity

*JEL Classification: C40, D83, I18*

The interaction between the local food environment, an individual's dietary pattern, food prices, health outcomes, and policy variables is complex and rarely considered in its entirety. Many studies examine the interactions of these factors in piecemeal fashion but usually not together as a complex system. Some studies examine the link between individual and household characteristics and the local food environment (Powell, Chaloupka, and Bao, 2007). Others examine the results between individual characteristics and an individual's dietary pattern (Darmon and Drewnowski, 2008). Some even examine the link between the local food environment and an individual's dietary pattern (Moore et al., 2008) or an individual's health outcomes (Chen, Jaenicke, and Volpe, 2016). More research is needed to examine the complex interactions among all these factors, which will help shape correct policy decisions.

This research is important because the food environment and an individual's interactions within the food environment are critically important in determining an individual's dietary quality and risk of negative health outcomes such as obesity. As explained by Finkelstein, Ruhm, and Kosa (2005) there are numerous economic causes and consequences of obesity among adults and children in the United States. Obesity is a major risk factor for diabetes, cardiovascular disease, cancer, sleep apnea, nonalcoholic fatty liver disease, osteoarthritis, and other problems (Ahima and Lazar, 2013). According to Dharmasena and Capps (2012), two-thirds of adults in the United States are either overweight or obese. Ogden *et al*., (2014) shows that childhood obesity has more than doubled in children and quadrupled in adolescents in the past 30 years in the United States. This rate has slowed, as there have been no significant changes in the obesity prevalence in youth or adults between 2003-2004 and 2011-2012. On a typical day, 30% of children report consuming fast food (Bowman et al., 2004). It is important to study the acquisitions of food away from home as this account for 32% of total calories consumed

(Guthrie, Lin, and Frazao, 2004). Thus, policies that want to reduce the health problems associated with obesity need a full picture of the interactions among all variables.

The contribution of this research is to use the individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables to estimate a graphical causality structure using the National Household Food Acquisition and Purchase Survey (FoodAPS) (USDA, ERS, 2017). This is in contrast to studies that consider these variables in a fragmented approach using *a-priori* endogenous/exogenous relationships and not as a complex system that shows many possible interactions among variables considered. Then we will calculate the parameter estimates underlying the structural relationships built from the causality structure developed through machine learning methods and directed acyclic graphs. Finally, we will make comparisons of the causality effects from the estimated directed acyclic graph and parameter estimates to current research in the field.

The estimation of a graphical causal structure is done using directed acyclic graphs (DAGs). The DAG is generated using two machine learning algorithms: Greedy Equivalence Search (GES) and Linear non-Gaussian Orientation Fixed Structure Rule Three (LOFS R3) (Chickering, 2002; Ramsey, Sanchez-Romero, and Glymour, 2014) First, the GES algorithms is run on the data to build a graphical causal structure. Then, the LOFS R3 algorithm is run on the resulting structure to orient any edges that were not oriented by the GES algorithm. The DAG is generated under assumptions made by imposing *a priori* knowledge on the structure.

Our main findings can be briefly summarized as follows: Asian individuals live in an environment with high concentrations of fast food and non-fast food restaurants. Hispanic individuals live in areas with a higher concentration of fast food restaurants and food stores.

Obesity is less prevalent among Asian individuals. Hispanic individuals are more likely to report a fair or poor diet. Those with higher incomes are less likely to report low food security and obesity is less prevalent among individuals in high-income groups. For two products (oils and vegetables), the quantity moves the price. However, no price directly affects the product that it represents in the current time. In regards to the paths between poverty, race and food insecurity, we find a number of paths. We find that Hispanic individuals are more likely to be food insecure. There is also a direct path between the percent of poverty level and food insecurity and a path between college education and food insecurity. We find a causal chain from those who are classified as Black to supplemental nutrition assistance program (SNAP) participation to low food security. A similar casual chain also exists for Hispanic individuals to low food security via SNAP participation.

The remainder of this paper proceeds as follows. In the Literature Review section, we discuss the existing literature on diet, health, food assistance programs, and directed acyclic graphs. In the Theoretical Background section, we discuss the theory of directed acyclic graphs. In the Data section, we give a detailed description of the data and the variables. In the Estimation Procedure, we discuss the algorithms, *a priori* knowledge, and parameter estimates underlying the structural relationships. In the Results section, we discuss the results and compare to the current literature. Finally, in the Conclusions section we conclude and discuss limitations.

**Literature Review**

In the existing literature, the individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, prices, health outcomes, and policy variables are usually not considered together at the same time in a complex system. Here we provide a brief review of literature, but a more thorough comparison of current literature with our findings is presented in the results section.

***Diet and Health***

As diet could possibly be linked to obesity, this is a popular area of research. Beatty, Lin, and Smith (2014) look at changes in the distribution of dietary quality among adults in the United States over the period 1989–2008. They find improvements for both low-income and higher-income individuals alike with 63% of the improvements attributed to changes in food formulation and demographics. Stewart et al. (2011) finds that low-income households spend too high a share of their budget on high-calorie, high-fat food and not a high enough share on fruits and vegetables. The authors argue that reallocating their budget to more fruit and vegetables would allow these households to meet government dietary guidelines.

A section of literature also argues that individual characteristics are associated with an individual's dietary pattern. Darmon and Drewnowski (2008) find that diets of whole grains, lean meats, fish, low-fat dairy products, and fresh vegetables and fruit are more likely to be consumed by those with higher income and more education. In contrast, the consumption of refined grains and added fats has been associated with lower income and education. Dubowitz et al. (2008) find that a higher income and education is associated with a higher level of fruit and vegetable consumption. Zagorsky and Smith (2017) find that middle-class individuals eat more fast food than the poor or wealthy and that those in the poorest income quintile eat fast food much less often than those in higher quintiles.

Some literature argues that the local food environment is affected by household characteristics. Kwate (2008) argues that Black neighborhoods will have a higher share of fast food restaurants. The author explores a number of possible explanations for this outcome ranging from racial separation to income inequality.  Powell, Chaloupka, and Bao (2007) also find associations between individual characteristics and the local food environment. These authors find that nationally, Black neighborhoods have a lower availability of restaurants compared to White and Hispanic neighborhoods, but Black neighborhoods in urban areas have a higher proportion of fast-food restaurants relative to full service. The authors conclude that these results may explain racial differences in the obesity rates.

***Diet and Food Assistance Programs***

The food consumption behavior of households that participate in the Supplemental Nutrition Assistance Program (SNAP), which provides low-income households benefits with which to purchase food for at-home consumption, has been extensively studied. Liu, Kasteridis, and Yen (2013) find no evidence that SNAP participation promotes consumption of more food away from home (which tends to be less healthy). Wilde and Ranney (2000) find that among SNAP households who conduct major grocery shopping trips only once per month (42% of all SNAP households), calorie intake drops by the fourth week of the SNAP benefit month. Results on the relationship between SNAP participation and food security are subject to problems of selection bias and endogeneity (Gundersen and Oliveira 2001; Jensen 2002). Gundersen and Oliveira (2001) show that once one controls for adverse selection, SNAP recipients have the same probability of food insufficiency as non-recipients. Further research suggests that SNAP participation may reduce the percentage of food insecure households by 4.6% to 10.6% (Malbi et al., 2013). One estimate of the reduction in the likelihood of food insecurity due to SNAP participation is as high as 30% (Ratcliffe, McKernan, and Zhang, 2011).

Policymakers are interested in studies about possible incentives to improve the diet and health of residents, especially low-income households. Of particular importance is improving the diet of residents living in food deserts. Generally, these are areas where residents have limited access to food stores. More specific definitions vary and may include the number of stores in an area or the types of stores available to residents (see Walker, Keane, and Burke (2010) for a review of food desert literature). Andrews, Bhatta, and Ploeg (2013) suggest that economic incentives should be considered as an alternative to store development in food desert communities. This would include options such as allowing SNAP households to use a portion of their benefits to fund transportation to locations with more economical shopping options.

Lin, Yen, Dong, and Smallwood (2010) look at a number of ways to increase consumption of fruit, vegetables, and milk. The authors find that a 10% price subsidy would reduce consumption deficiencies by 4%–7% at an estimated cost of $734 million a year. Studies intended to promote policies need to fully understand the interactions among health, dietary patterns, individual characteristics, and the local food environment.

***Directed Acyclic Graphs***

Directed acyclic graphs (DAGs) are a way to graphically illustrate the causal structure among a set of variables. The structure of these graphs is found with machine learning techniques and this allows inference when the interactions among variable are too complex or there are too many for human minds to comprehend (also sort out interactions). DAGs have been used in studies ranging from mapping the integration of brain networks (Ramsey, Hanson, and Glymour, 2011; Smith et al, 2011; Mumford and Ramsey, 2014, Ramsey, Sanchez-Romero, and Glymour, 2014) to studying the relationship money and prices (Bessler and Lee, 2002) to modeling vehicle collision with pedestrians (Davis, 2003). Few studies on consumer food demand or the food environment exist that use directed acyclic graphs. Wang and Bessler (2006) use DAGs to analyze U.S. meat consumption for beef, chicken, turkey, and pork, focusing on price quantity endogeneity in food products. Some meat products show evidence of endogeneity while others do not. Lai and Bessler (2015) use a directed acyclic graph to examine causal relationships among retail prices, manufacturer prices, and number of packages sold for carbonated soft drink. The results show that the retail price leads to manufacturer price and quantity sold. This means that the retail store has more pricing power than the soft drink manufacturers.

The most similar paper to our research is Dharmasena, Bessler, and Capps (2016). The authors use directed acyclic graphs to model the food environment in the United States. They used state level aggregated data to model the food environment in the United States in contrast to the individual level data used in this research. The results indicate that food insecurity and participation in SNAP are related but do not seem to have a direct causal link. The authors also find that poverty and SNAP participation are related indirectly through their links to food insecurity, unemployment, race, and food taxes.

***FoodAPS***

The National Household Food Acquisition and Purchase Survey (FoodAPS) is still quite recent, but there have been a few studies using the data. Taylor and Villas-Boas (2016) use the FoodAPS data with a multinomial mixed logit model to estimate food store choices as a function of type and household attributes. They find that households are willing to pay between $12 and $17 per week in distance traveled for superstores, supermarkets, and fast food, while they are willing to pay significantly less for the remaining outlets. They conclude that policymakers should consider incentivizing the building of the outlets of which there is a higher willingness to pay among consumers.

Smith et al. (2016) use the FoodAPS data to examine spending patterns over the SNAP benefit month. The authors find evidence of short-run impatience and fungibility of income in SNAP participants. Wilde, Llobrera, and Ploeg (2014) use a random sample of census block groups from the FoodAPS data and find that census blocks with high poverty have a closer proximity to a supermarket than other blocks. Basu, Wimer, and Seligman (2016) use FoodAPS to examine the association between cost of living and nutrition among low-income individuals, finding that low-income individuals have worse nutrition in counties with a high cost of living.

**Theoretical Background**

Recent literature has focused more on inferring causal relationships from observational data in the absence of controlled experiments (Pearl, 2009; Spirtes et al., 2000). These methods rely on algorithms that allow causal inferences to arise without explicitly formed hypotheses. The causal structures that arise from these algorithms can be represented in graphical form as a DAG (directed acyclic graph). The following discussion is drawn from a number of sources (Pearl, 1995; Spirtes et al., 2000; Pearl, 2009).

A graph is an ordered triple {**V**, **M**, **E**} where **V** is a non-empty set of variables (or nodes), **M** is a non-empty set of symbols attached to the ends of undirected edges, and **E** is a set of ordered pairs (edges). More simply, a graph is a diagram containing a number of nodes (which represent variables) and arrows that depict relationships among the variables. A directed graph contains only directed edges (X →Y). Lines without arrows (X − Y) are undirected edges and are used to indicate correlations with unknown directionality. A DAG is a graph that does not include any cycles that start and end at the same node. For example, $X_t \rightarrow Y_t \rightarrow Z_t \rightarrow X_t$ represents a cyclic graph because it starts at the variable X and ends on the same variable X in the contemporaneous time.

Figure 1 is an example of a directed acyclic graph. The boxes represent variables and the arrows depict the directionality of causal relationships. The variables included and the ordering of the variables is motivated by the extant literature. For example, individuals and household characteristics are usually assumed to be strictly exogenous. This is illustrated by having arrows directed outward of the variables and not having any arrows directed into the variables. Figure 1 is also the model of the interactions among a set of variables extracted from the FoodAPS data, which will be modeled using DAG methods discussed later. Given the complexity, not all possible interactions are illustrated in this figure.

Nodes (variables) are sometimes referred to with the terminology of parents and children or ancestors and descendants. A parent of a node is any other node with an arrow into that node. An ancestor is any node that appears earlier than a node in a chain of nodes. A child of a node is any other node with an arrow into it from that node. A descendant is a node that occurs later in a chain.

A DAG may also be referred to as a Bayesian network when its joint probability density function can be written as a product of the individual conditional density functions:

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^{n} \mathrm{P}(X_i|pa_i) \tag{1}$$

where $P$ is the probability of the variables $X_1, X_2, \dots, X_N$ and $pa_i$ is the set of variables that precede $X_i$ (the parents). Thus, for any set of random variables, the probability of any member of a joint distribution can be calculated from conditional probabilities using the chain rule (given an ordering of $\boldsymbol{X}$) as follows:

$$P(X_1, X_2, \dots, X_N) = P(X_1)P(X_2|X_1)P(X_3|X_2, X_1) \cdots P(X_N|X_1, X_2, \dots, X_{N-1}) \tag{2}$$

Pearl (1986) proposes that *d*-separation is a graphical version of this conditional independence. Pearl (1995, p. 671) defines *d*-separation as follows:

Let X, Y and Z be three disjoint subsets of nodes in a directed acyclic graph G, and let p be any path between a node in X and a node in Y, where by 'path' we mean any succession of arcs, regardless of their directions. Then Z is said to block *p* if there is a node *w* on *p* satisfying one of the following two conditions: (i) w has converging arrows along p, and neither w nor any of its descendants are in Z, or, (ii) w does not have converging arrows along p, and w is in Z. Further, Z is said to d-separate X from Y, in G, written $(X \perp Y|Z)_G$, if and only if Z blocks every path from a node in X to a node in Y.

The above is a method used to read the conditional independencies of equation (1) directly off a graph. Geiger, Verma, and Pearl (1990) show that there is a one-to-one correspondence between the conditional independencies of equation (1) and the set of triples {X, Y, Z} that satisfies the *d*-separation defining in the graph G.

Further understanding of this concept can be gained by examining the three types of structures possible in a DAG: causal chains, causal forks, and colliders (inverted causal forks). A casual chain implies the variables *X*, *Y*, and *Z* are related as $X \rightarrow Y \rightarrow Z$. This implies a causal ordering in the variables such that *X* causes *Y* and *Y* causes *Z* (or *X* causes *Z* via *Y*). Figure 1 contains the causal chain: *Local Food Environment* → *Dietary Pattern* → *Health Outcomes*. This means that the variables that determine the local food environment affect the individuals' dietary patterns, which intern affects health outcomes. This also means that knowing information with regards to dietary pattern blocks the information flow from local food environment variables to health outcomes, since dietary pattern gets the information from local food environment variables.

A casual fork implies the variables *X*, *Y*, and *Z* are related as $X \leftarrow Y \rightarrow Z$. This implies that *Y* is a common cause of *X* and *Z*. Figure 1 does not contain any causal forks. Assume we introduce a new node to the figure such that *Local Food Environment* ← *Household Characteristics* → *Nutrition Assistance Participation*. This means that the household characteristics are a common cause for both local food environment variables and participation in nutrition assistance programs. Furthermore, nutrition assistance programs and local food environment variables may be related in the absence of a common cause, household characteristics, however extra information with regards to household characteristics that causes both nutrition assistance programs and local food environment may separate the information flow from later two.

A collider implies the variables $X$, $Y$, and $Z$ are related as $X \rightarrow Y \leftarrow Z$. This means that $Y$ is caused by $X$ and $Z$. Figure 1 contains the collider: *Local Food Environment* → *Dietary Pattern* ← *Prices*.  Local food environment variables and prices may be not related, however, extra information with regards to a collider, dietary pattern (or a common effect), may open up a causality path between local food environment and prices (or joint causes).

Three assumptions are needed to find a DAG. *Causal Sufficiency* means there are no omitted variables that can cause any of the included variables (Spirtes et al., 2000, p. 45).The *Causal Markov condition*, which relies on *d*-separations, states that any node is independent of its non-descendants conditional on its parents (Spirtes et al., 2000, p. 53). The *Causal Faithfulness condition* implies that any zero correlation observed between two variables is entailed by the application of the Causal Markov condition to the graph (Spirtes et al., 2000, p. 56).

**Data**

The National Household Food Acquisition and Purchase Survey (FoodAPS) is a nationally representative panel of 4,826 U.S. households containing information about each household's food purchases and acquisitions (USDA, ERS, 2017). Details were collected about foods purchased or acquired for consumption at home and away from home, including foods purchased using SNAP or other food assistance benefits. The survey is unique in that it oversamples SNAP households and low-income households not participating in SNAP in relation to higher income households. The publicly available version of the FoodAPS dataset was used in this analysis. This data have been modified to prevent disclosing the identity of any respondent, which is standard practice in the provision of public-use files (more information about this procedure can be found in the FoodAPS user guide [USDA, ERS, 2016]).

FoodAPS was fielded between April 2012 and January 2013 and collected information on all food acquisitions and purchases at home and away from home by all members of the household over a seven-day period. Households had to scan barcodes, save receipts, and record other information in food journals. Information obtained from the household includes the quantities and expenditures for all at home and away from home foods and beverages purchased or acquired by all household members, eating occasions by household members. Further information was collected about household characteristics (e.g., income, program participation, food security, health status, etc.) and household access to food (e.g., location of purchase and distance to food stores and restaurants) (USDA, ERS, 2017). The USDA added information about nutrient content of purchased food and the local retail environmental based on scanned barcodes of products and household locations.

Information is available at the individual level for the 14,317 individuals who participated in FoodAPS. As obesity is a variable of interest, we restrict the sample to only those where the information is available on body mass index (BMI). This is given for 13,336 individuals in the sample. The FoodAPS survey only has BMI information for individuals 2 years old (24 months) or older. This is due to a committee on childhood obesity that concluded that for children under 2 years of age, BMI values are not helpful (Barlow, 2007). Further, we drop a few cases that are missing racial or ethnic status as this characteristic would be difficult to impute. This leaves a final sample of 13,329 individuals.

Table 1 gives descriptions and summary statistics for all variables used in the DAG. All means discussed in this section are weighted using the Taylor series linearization method from appendix D of the FoodAPS user's guide (USDA, ERS, 2016, p. 55). The first section presents individual or household characteristics. Household characteristics include the size of the household and the household's average monthly income as a percent of the poverty guideline. The average monthly income is the sum of average imputed income for each member of the household. This income is then used to find the percent of the poverty guideline given the household's characteristics. The imputation procedure and calculation as percent of the poverty guideline were given by the FoodAPS survey. The average household has 3.29 members and an income that is 366% of the poverty guideline. For reference, the 2012 poverty guidelines give the poverty guideline for a family of four as $23,050 (HHS, 2012). This roughly translates into a weighted household income for a family of four in our sample of $84,363.

Next, a number of individual characteristics are presented in the table. All of these individual characteristics are indicator variables except for age. The average age of individuals in the sample is around 40 years. Around 52% of the sample is female. In regards to racial identification, 74% are White, 13% are Black, 4% are Asian, 1% are American Indian, and the rest identify as another race. Around 16% of the sample claim Hispanic ethnicity.

A few education and employment indicators are also included in the table. Close to 48% of the sample is currently employed. This may seem low, but a number of individuals under 18 are included in this sample. These individuals be included in the unemployed category and the less than high school degree category as they are currently in school. This sample contains 30% with less than a high school degree or currently enrolled in school, 22% with a high school degree, 24% with some college completed or Associate's degree, and 24% with a college degree.

The next section of the table presents the local food environment characteristics. For the FoodAPS households, Todd and Scharadin (2016) find that 87% visited large grocery stores and supermarkets, and 85% visited restaurants and other eating-places at least once. Given in the table are the number of fast food restaurants, non-fast food restaurants, and SNAP-authorized supermarkets and superstores within 5 miles of the household. Within five miles of the household, there is an average of 61 fast food restaurants, 281 non-fast food restaurants, and 22 supermarkets and superstores. Around 33% of the households live in a rural Census tract.

FoodAPS provides this data for a wide number of ranges, but the five-mile distance was chosen for inclusion. This should cover the range of travel dictated by much of the research in the area. A distance of 5 miles should cover around 80% of visits to sit-down restaurants and fast-food outlets with an average distance between the food establishments and homes of 2.6 miles (Liu, Han, and Cohen, 2015). Further, only considering supermarkets and supercenters for the choice of food at home store density may seem overly restrictive as it excludes convenience stores and small grocery stores. However, supermarkets and supercenters are the dominant store of choice for most U.S. households. According to Ver Ploeg et al. (2015), around 44% of households do their primary grocery shopping at supercenters, while another 45% do their primary shopping at supermarkets. Household on average travel 3.8 miles to their primary food store of choice.

Next in table 1, we present prices faced by the household. These prices are calculated for six USDA main food categories: dairy, fruit, grains, proteins, vegetables, and oils. Prices are calculated at the household level as the same prices likely apply to all individuals in the household. Most items in FoodAPS are matched to a USDA food category code and a USDA food code which provides the nutrients and food pattern equivalent values contained in each food (Bowman et al., 2017). The total grams of each food obtained by the household are summed together for each of the six food groups and calories. Unit values (proxy for prices) are calculated by taking the total expenditure in a category and dividing this by the total weight (grams) purchased. Missing prices are imputed using an auxiliary regression of quantity purchased on household income, household size, and location variables.[1] This approach is not without precedent and is standard procedure used in the price imputation literature (Capps, et al, 1994; Alviola and Capps, 2010, Kyureghian, Nayga and Capps, 2011, and Dharmasena and Capps, 2014).

The use of unit values is also likely to create bias in the form of measurement error. It is possible the aggregates are endogenous to the choice of quality. We utilize the procedure described by Cox and Wohlgenant (1986) to correct for endogeneity in prices. For this procedure, we regress the difference between the unit price and the mean unit price for each category on a number of household demographics.

$$p_i^u - \bar{p}_i^u = \sum_j \beta_{ij} D_{ij} + v_i \tag{3}$$

For equation (3), $p_i^u$ is the unit price for the commodity $i$, $\bar{p}_i^u$ is the mean unit price for commodity *i* across all households, $\beta_{ij}$ is a set of coefficients to be estimated, $D_{ij}$ is a vector of characteristics for household $j$, and $v_i$ is the error term. The demographics used in the regression

[1] For some items with missing quantities, the USDA will attempt to impute grams if sufficient product information is available. We use this imputed value when available.

are the characteristics of each household in the sample. We include income, household size, and dummy variable indicating the region of the country. In order to get the quality-adjusted price, we used the estimated coefficients from equation (3) and then calculate the following,

$$\hat{p}_i = p_i^u - \sum_j \hat{\beta}_{ij} D_{ij}, \tag{4}$$

where $\hat{p}_i$ are the prices to be used in the estimation in place of the observed unit prices.

The prices presented in table 1 are the quality-adjusted prices in U.S. Dollars per gram at the household level. The quality adjusted prices range from \$0.0069 per gram for dairy to \$0.0275 per gram for oils. Fruits and vegetables has similar quality-adjusted prices at \$0.0083 and \$0.0091 per gram respectively. The weighted means of the quality-adjusted prices have the highest standard error for fruit and for oils while the grains and vegetables are the lowest. Negative quality-adjusted prices are interpreted as those individuals needing to be compensated in order to purchase that product.

Next in table 1 are the amounts acquired per day for six USDA main food categories (dairy, fruit, grains, proteins, vegetables, and oils) and calories as a percent of the recommended intake per day for each individual. To determine an individual's calorie requirement per day, we use the estimates of daily calorie requirements by age, gender, and activity level provided by the HHS and USDA (2015, p. 77-78, and presented and presented in table 2). Details of the HHS and USDA assumptions are given in the notes for table 2. Each individual is assigned a daily calorie recommendation based on gender and age from the moderately active column.

To estimate the total amounts acquired in away-from-home food consumption we total the calories and contributions to the six food categories for each event. Each food away from home event lists the household members present and if any guests were present. First, we find the household share of the meal as the percent of people present at the event that are members of the household. For example, if five people are present at the meal but only four household members are among the participants, then the household's share of the meals calories and food group contributions is 80%.

Once we have the household's share of the meal, we need to find each individual's share of calories and food group contributions. The individual share is that individual's daily calorie recommendation as a percent of the total calorie recommendation for all individual in the household present at the food away from home event. For example, if an individual's daily recommendation is 2000 calories and the household's daily recommendation is 10,000 calories for those at the food away from home even, then the individual's calorie share of the event is 20% of the household share. The individual shares are used to calculate the portion of the food event's calories and contributions to the six food categories go to each individual. Then we sum across all food away from home events over the week and divide by seven to get the daily estimate of calories acquired and of the six food groups acquired.

A similar process is followed to estimate calories and food group contributions from food at home.[2] For food at home, the household share of the food at home acquisitions is assumed to be 100%. The individual share is that individual's daily calorie recommendation as a percent of the total calorie recommendation across all individual in the household. For example, if an individual's daily recommendation is 2,000 calories and the household's daily recommendation is 10,000 calories, then the individual's calorie share is 20% of the household's total food-at-home acquisitions. This individual share is then used to calculate the calories and contributions to the six food categories that go to that individual.

[2] Food at home is usually not consumed at the time it is acquired but rather eaten over time. Since FoodAPS tracks acquisitions over a seven-day period, the share for food at home represents a maximum possible intake given the foods acquired over that period.

To convert from the contributions to the food categories to the percent of the daily recommendations for the six food categories we use the Healthy U.S.-Style Pattern presented in table 3 (HHS and USDA, 2015, p. 80-82). This gives the recommended levels of consumption across a number of food groups. The totals found using the process above are divided by the daily recommendations to give the percent acquired each day as a percent of the recommended amount. For our sample, individuals on average acquire the following percent of the daily recommendations: 52% for dairy, 35% for fruit, 88% for grain, 66% for proteins, 43% for vegetables, 91% for oils, and 90% for calories. These figures may seem low but it is important to recall that FoodAPS collected acquisitions information, not consumption, and like all survey data, there is likely to be some under reporting of food acquisitions.

Next in table 1 we present health measures for the individual and household. Low food security is based on the USDA's 30-day Adult Food Security Scale. Seventeen percent of households are identified as having low food security. Next is a measure of the household's own assessment of the health of their overall diet. This variable indicates that 22% of households placed themselves in the fair or poor category (two lowest of five possible responses). Then a variable indicates if the household believes they eat too few fruits and vegetables. This indicates that 69% of households believe they need to eat more fruits and vegetables. Next, we have an indicator of the individual's belief that their health is fair or poor (two lowest of five possible responses). In the sample, 15% of individual placed themselves in these categories. Then, we present and indicator if the individual is obese based on reported BMI. For adults age 18 and older, obese is determined by BMI ranges 30.0 and above. For children, obese is determined by a BMI percentile at or above the 95th percentile. For this sample, 28% of individuals are considered obese.

Finally, at the end of table 1 we include two program participation variables: whether anyone in the household is receiving SNAP benefits and whether anyone is receiving WIC benefits. For our sample, 16% of individuals are living in households where at least one member receives SNAP benefits and 6% of individuals are living in a household where at least one member receives WIC benefits. For the U.S. in 2012, there were 46,609,000 individuals participating in SNAP (Gray, 2014, p. 12) and a total population of 313,998,379 (U.S. Census Bureau, Population Division, 2017). This would mean that 14.8% of individuals in the U.S. participated in SNAP for the same year as our sample. Thus, the weighted average of 16% from our sample is close to the U.S. participation rate of 14.8%.

## Estimation Procedure

### ***Finding a Graphical Causal Structure with GES***

The Greedy Equivalence Search (GES) algorithm is used to find a graphical causal structure by searching over Markov equivalence classes (Meek, 1997; Chickering, 2002). The algorithm will assign all graphs in the same equivalence classes the same score. Two graphs are equivalent (in the same equivalence class) if the DAGs are *distributionally equivalent* and *independence equivalent*. Two graphs are distributionally equivalent if under the Markov condition (the probability structure of a graph can be written with the probabilities of the variables conditionals just on the variables' parents), the graphs share the same joint probability distribution. For three variables, this reduces the search space from 25 possible DAG structures to a search over to 11 equivalence classes (Kwon and Bessler, 2011, p. 95). One example, the graphs $A \rightarrow B \rightarrow C$ and $A \leftarrow B \leftarrow C$ will have the same joint probability structure.[3] Two DAGs

[3] The joint probability for $A \rightarrow B \rightarrow C$ is $P(A,B,C) = P(A)*P(B|A)*P(C|B)$. The joint probability for $A \leftarrow B \leftarrow C$ is $P(C,B,A) = P(C)*P(B|C)*P(A|B)$. Bayes' theorem is applied to this joint probability and the result is $P(C,B,A)=P(C)*P(C|B)*P(B)/P(C)*P(B|A)*P(A)/P(B)$. This simplifies to $P(C,B,A)= P(A)*P(B|A)*P(C|B)$. Thus, $P(A,B,C) = P(C,B,A)$ and they are distributionally equivalent.

are independence equivalent if the independence constraints in the two DAGs are identical.[4] Further, it is assumed that the true causal model is acyclic and there are no common hidden causes existing between variables. The variables are assumed to have direct causal influence on other variables in a linear manner with each variable having a Gaussian distribution. Given the above assumptions, the GES algorithm follows inclusion optimality, that the result is the most parsimonious model that contains the true model (Chickering and Meek, 2002).

The GES algorithm works by first doing a forward search and then doing a backward search. The search begins with an empty graph of all the variables. In the forward search, GES begins adding edges between nodes that increase the score. This continues until no additional edge increases the score. After this search, the algorithm compares across all equivalence classes and chooses the class with the highest score. The forward search will find the equivalence class that includes the true DAG for the data. Edges are oriented according to the orientation rules described in Sprirtes et al (2000) and Meek (1995). Appendix D provides more details on these orientation rules. In the backward search, the algorithm removes edges until no single edge removal increases the score. Once no additional edge removal increases the score, the algorithm stops. The DAG found from the second step will be the one that best represents the data.

The scoring algorithm is important for the function of the GES algorithm. The Bayesian Information Criterion (BIC) is the most commonly used criteria to score the graph. The BIC is a measure of the marginal likelihood of the data given the graph structure and is defined as,

$$BIC = 2lnP\left(D \middle| \hat{\theta}, G\right) - c \cdot k \cdot \ln(n) \tag{5}$$

Where $P$ represents the probability, $D$ is the data, $\hat{\theta}$ is the maximum likelihood estimate, $G$ represents the structure of the DAG, $c$ is a penalty parameter, $k$ is the number of parameters, and

[4] For the graph $A \rightarrow B \rightarrow C$, the independence constraint is $A \perp C|B$. For the graph $A \leftarrow B \leftarrow C$, the independence constraint is $A \perp C|B$. The graphs share the same independence constraint and hence are independence equivalent.

$n$ is the number of observations. The BIC is used to balance between fit and parsimony. The penalty parameter can be used to speed up searches and reduce the number of false positive results by producing a sparser graph. Some authors suggest a discount penalty that increases as the number of edges in the true graph increases (Ramsey, 2010).

***Further Orienting Causal Edges with LOFS***

Even though the GES algorithm is highly effective at orienting edges, after running the algorithm some edges may still be un-oriented (e.g., $A - B$). The Linear non-Gaussian Orientation Fixed Structure (LOFS) algorithms can be run on the GES results to orient any un-oriented edges. The LOFS works by considering higher moments of the data. A brief outline is presented here and more detail is available in other sources (Ramsey, Hanson, and Glymour 2011; Mumford and Ramsey, 2014). Three types of LOFS algorithms will be discussed: Rule 1 (*R1*), Rule 2 (*R2*), and Rule 3 (*R3*). Ramsey, Sanchez-Romero, and Glymour (2014) discuss these three rules and other alternatives. The algorithms differ from the GES in that they rely on assumptions of non-normality of the variables. Orientations are made to maximize the non-normality of variables. A scoring method commonly used for these algorithms is the Anderson-Darling statistic for normality (Anderson and Darling, 1952).[5]

The *R1* algorithm works by adding a single directed edge, testing the residuals of all possible models from adding that edge, and then selecting the model with the most non-normal residual. For example, consider two variables *A* and *B* that are adjacent in a graph. The variable *A* is regressed on the empty set and on variable *B*. The Anderson-Darling statistic is calculated for the residuals in both regressions. If the regression of *A* on *B* has a higher Anderson-Darling

[5] The Anderson-Darling test statistic for the null hypothesis that the data follow a normal distribution is $A^2 = -n - S$. Given the ordered data $Y_i$, a number of observation n, and the normal cumulative distribution function $\Phi$, then

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} [ln(\Phi(Y_i)) + ln(1 - \Phi(Y_{n+1-i}))].$$

statistic, then $B$ must be a parent of $A$ (i.e., $B \rightarrow A$) (Ramsey, Sanchez-Romero, and Glymour, 2014).

For the *R2* algorithm, consider an undirected edge between variables $A$ and $B$ with $P_A$ being the candidate parents of $A$ excluding $B$ and $P_B$ being the candidate parents of $B$ excluding $A$. First *R2* checks if the Anderson-Darling statistic of $A$ conditional on $B$ and $P_A$ is greater than then Anderson-Darling statistic of $B$ conditional on $A$ and $P_B$. Then *R2* checks if the Anderson-Darling statistic of $B$ conditional on $P_B$ is greater than then Anderson-Darling statistic of $A$ conditional on $P_A$. If both conditions are satisfied then $A \rightarrow B$. If both conditions are reversed then $B \rightarrow A$ (Ramsey, Hanson, and Glymour 2011).

The *R3* algorithm checks whether the Anderson-Darling statistic of residuals from the regression of $A$ on $B$ plus the Anderson-Darling statistic of variable $B$ is greater than the Anderson-Darling statistic of residuals from the regression of $B$ on $A$ plus the Anderson-Darling statistic of variable $A$. If this is true, then *R3* orients the edge as $B \rightarrow A$. If the reverse relationship were true, then the edge would be oriented as $A \rightarrow B$ (Ramsey, Sanchez-Romero, and Glymour, 2014).

***Imposing a Priori Knowledge***

In addition to the assumption inherent to the algorithms described earlier, a number of constraints can be further imposed on the search to speed estimation and produce a sparser graph. This is also important because sometimes the data cannot distinguish between two graphs. For example, the graphs $A \rightarrow B \rightarrow C$ and $A \leftarrow B \leftarrow C$ will have the same joint probability structure. The data cannot distinguish between the two graphs. We need to impose a priori information about the direction that the arrows flow. This background knowledge can be imposed in three forms.

First, edges between two variables can be forbidden.  This means that the algorithm will not be allowed to connect two variables in any direction no matter what the data may imply. Second, edges between two variables can be required. This means that the algorithm will be required to connect two variables in some direction no matter what the data may imply. Third, temporal tiers may be imposed. Edges from a later tier are forbidden in earlier knowledge tiers. These knowledge tiers provide an ordering to the variables and are helpful if we know one variable or group of variables must precede another variable or group of variables.

For our data, we use four knowledge tiers and some forbidden edges. Figure 1 provides a depiction of our four knowledge tiers. Descriptions of the variables in each category can be found in table 1.  In the first tier, we place individual and household characteristics. These are the only variables that are required to be fully exogenous in our model. For this tier, we also forbid edges within the tier. We are more interested in how these variables affect other tiers rather than the interactions among these characteristics. This also has the effect of speeding the estimation process. In the second tier, we place characteristics of the local food environment. Recall that variables in the first tier are not required to connect to variables in the second tier, they are only given the opportunity. In the third tier, we place the individual's dietary pattern and prices. These are placed in the same tier because other authors have found evidence that prices and quantities are predetermined using DAGs (Wang and Bessler, 2006). In the fourth tier, we place the health outcomes of the individual. The policy variables used in this study are not placed in any tier and are allowed to be endogenous or exogenous as determined by the algorithm. However, the policy variables are not allowed to cause the characteristics in tier 1 (forbidden edges into tier 1 variables). For example, SNAP participation is restricted not to cause an individual's characteristics (e.g., race, gender, or age).

***Estimating Structural Models from Graphical Structures***

After finding a DAG, one may find a corresponding structural relationship among variables that represents the graph. A short overview of estimating structural relationships is presented. A more in depth discussion on estimating these relationships can be found in Bollen (1989). The structural relationships consist of the DAG found from the algorithmic search (the variables and directed edges) and new nodes representing each error term. All edges in the graph must be directed in order to estimate the structural relationships. The causal structure of a structural relationship is indicated using directed edges from the DAG. For example, the directed edge in $A \rightarrow B$ indicates that *A* is the right hand side variable and *B* is the left hand side variable (i.e. $B = \beta A + \epsilon$). Bi-directed edges, such as $A \leftrightarrow B$, represent that the error terms between variable *A* and *B* are correlated. When constructing the structural relationship from a DAG, we assume that it is linear (for simplicity) with Gaussian errors (since the GES algorithm requires that the underlying data generating process to be Gaussian). A multiple linear regression is used to estimate coefficients and residual variances. The number of partial effects to be estimated is equal to the number of edges in the graph.

When estimating the parameters, two criteria are important for consideration to identify parameters (Pearl, 2009). The first to consider is the *back-door criteria*. Suppose we are interested in the causality relationship between variables *X* and *Y* and have the graph of figure 2. The variables *Z* satisfy the *back-door criteria* if: (1) no variables in *Z* are descendants of *X* (meaning no arrow head is pointing at *Z* from *X*) and (2) *Z* blocks every path between *X* and *Y* that contains an arrow into *X*. In order to block the information flow from *X* to *Y* via *back-door path*, run a regression of *Y* on *X* and *Z* (i.e. $Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon$). The conditioning on *Z* will block information flow via the *back-door* and provide and unbiased and consistent estimate of $\delta Y/\delta X$.

The second is the *front-door criteria*. Suppose we are interested in the causality relationship between variable $X$ and $Y$ and have the graph shown in figure 3. The set of variables $W$ meet the *front-door criteria* if: (1) $W$ intercepts all paths directed from $X$ to $Y$, (2) there are no unblocked *back-door paths* from $X$ to $W$, and (3) all *back-door paths* from $W$ to $Y$ are blocked by $X$. The method to block the *front-door path* works in two steps. First, regress $Y$ on $W$ and $X$ (i.e., $Y = \beta_0 + \beta_1 W + \beta_2 X + \epsilon$) to get an estimate of $\delta Y/\delta W$. Second, regress $W$ on $X$ (i.e., $W = \beta_0 + \beta_1 X + \epsilon$) to get an estimate of $\delta W/\delta X$. The unbiased and consistent estimate of $\delta Y/\delta X$ can be found by multiplying $\delta Y/\delta W$ by $\delta W/\delta X$.

**Results**

The TETRAD V software developed by Glymour et al. (2016) is used to estimate the DAG and estimate the parameters for the structural relationships. The variables in table 1 are used to estimate the structure of the graph for individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, food prices, health outcomes, and policy variables given the knowledge discussed earlier. After running the GES algorithm, two edges remained undirected among the quality-adjusted price variables: *Pqa_dairy – Pqa_protein* and *Pqa_grain – Pqa_protein*. The R3 LOFS algorithm is run on the graphical structure given by GES and orients the edges as *Pqa_protein* ← *Pqa_dairy* and *Pqa_protein* → *Pqa_grain*. The final graphical structure is given in figure 4. This figure shows the direction of causality among the variables. The partial values along with the direction for each edge can be found in table 4.

The DAG in Figure 4 shows how complicated the relationships are among the variables. It can be helpful to examine the *Markov blankets* of some variables. The Markov blanket of a node (variable) is the set of its parents, its children, and any parents of its children. This will render the variable conditionally independent from the rest of the graph. In essence, the Markov blanket of the node is the most important knowledge in predicting the behavior of a node. Figure 5 provides the Markov blanket for SNAP participation. Figure 6 provides the Markov blanket for WIC participation. Figure 7 provides the Markov blanket for obesity. Each of the figures also includes the partial effects from table 4 for quick reference. SNAP participation includes the most variables in its Markov blanket while obesity includes the least. WIC participation and SNAP participation appear in each other's Markov blanket indicating a strong dependence between the two. These causality results from Markov blanketing are discussed below, while making comparisons to results from the extant literature, where available.

*Link between Individual/Household Characteristics and the Local Food Environment*

Kwate (2008) argues that those classified as Black neighborhoods will have a higher share of fast food restaurants. In contrast, Powell, Chaloupka, and Bao (2007) find that predominantly Black neighborhoods have fewer full-service and fast food restaurants. The authors also find that there are significantly fewer restaurants available in predominantly Hispanic neighborhoods and that middle-income neighborhood have more restaurants than low and high-income neighborhoods.

Figure 4 does not show any direct path between those who are classified as Black individuals and restaurant density. However, the figure does show a path between Black and rural variables. Table 4 indicates that being Black decreases the occurrences that the individual live in a rural area of the United States. The variable Black is related to non-fast food restaurant density and supermarket and superstore density via the rural variable, ($Black \underset{(-)}{\longrightarrow} Rural$ $\underset{(+)}{\longrightarrow} nonfast\ food\ restaurant\ density$ and $Black \underset{(-)}{\longrightarrow} Rural$ $\underset{(-)}{\longrightarrow} supermarket\ and\ superstore\ density$). According to this relationship and the sign on the associated partial values with respect to these variables shown in table 4, (which shows the *front-door criterion* discussed above), being Black decreased the incidence of living in a rural area, which in turn increased incidence of their living in an area with fast-food restaurants. Also, being Black increased the incidence of them living in an area with high superstore/supermarket density. As shown in figure 4 and table 4, Hispanic variable has the same two causal chains, ($Hiapanic$ $\underset{(-)}{\longrightarrow} Rural \underset{(+)}{\longrightarrow} nonfast\ food\ restaurant\ density$ and $Hispanic \underset{(-)}{\longrightarrow} Rural$ $\underset{(-)}{\longrightarrow} supermarket\ and\ superstore\ density$. This means that being Hispanic decreased the incidence that they live in a rural area which in turn increased the incidence of them living in an area with fast food restaurants and also increased the incidence of them living in an area with high superstore/supermarket density. These results are confirmed though two direct causality relationships that Hispanic variable has with fast food density and superstore/supermarket density. They are shown in table 4 as follows: $Hispanic \underset{(+)}{\longrightarrow} fastfood\ density$ and $Hispanic$ $\underset{(+)}{\longrightarrow} supermarket\ and\ superstore\ density$.

That is to say that being Hispanic increases the incidence of one living in an area with more fast food density as well as more supermarket/superstore density. This information and causality relationship is important in estimating the effects of Black and Hispanic racial/ethnic groups on supermarket/superstore density and presence of non-fast food outlet density. Including the Rural variable in a regression of supermarket/superstore density on Black or Hispanic variables will block the information flow from Black or Hispanic variables to superstore/supermarket density, due to the causal chain relationship between these variables. Similar explanation is true for a regression of non-fast food density variable on Black or Hispanic variables, again due to causal chain relationship between these variables, via variable Rural. Being Asian increases the incidence of them living in environments with high concentrations of fast food and non-fast food restaurants (as shown in figure 4 and table 4; $Asian \underset{(+)}{\longrightarrow} fastfood\ restaurant\ density$ and $Asian \underset{(+)}{\longrightarrow} nonfastfood\ restaurant\ density$). Being White decreases the incidence of them living in areas with a lower density of fast food restaurants (i.e.$White \underset{(-)}{\longrightarrow} fastfood\ restaurant\ density$). The graph also indicates a connection between household size and restaurant density and store density. Table 4 indicates that being a larger household has more incidences of them living in areas with a higher density of fast food restaurants, lower density of non-fast food restaurants, and a larger selection of food stores. No paths are found between income and the local food environment variables.

*Link between Individual/Household Characteristics and the Individual's Dietary Pattern*

Darmon and Drewnowski (2008) find that diets of whole grains, lean meats, fish, low-fat dairy products, and fresh vegetables and fruit are more likely to be consumed by groups of higher income and education. In contrast, the consumption of refined grains and added fats has been associated with groups of lower income and education. Little evidence indicates that income or education affects total energy intakes. Dubowitz et al. (2008) find that a higher income and education is associated with a higher level of fruit and vegetable consumption.

Figure 4 does show a number of paths from the individual and household characteristics to the dietary pattern variables. Table 4 shows a small positive effect between those with higher education and income with fruit consumption. Being Black causes individuals to consume less vegetables and dairy (i.e. $Black \underset{(-)}{\longrightarrow} dairy\ consumption$ and $Black$

$\underset{(-)}{\longrightarrow} Vegetable\ consumption$), while being White has high a incidence of consumption of dairy (i.e. $White \underset{(+)}{\longrightarrow} Dairy\ consumption$. Being female causes one to consume less grain products ($Female \underset{(-)}{\longrightarrow} Grain\ producuts\ consumption$. Being a college graduate increased the incidence of consuming fruits ($College\ Graduate \underset{(+)}{\longrightarrow} Fruit\ Consumption$) while being employed decreased the incidence of consumption of fruits ($Employed \underset{(-)}{\longrightarrow} Fruit\ Consumption$. Being a large household decreased the incidence of consumption of proteins and vegetables ($Large\ Household\ Size \underset{(-)}{\longrightarrow} Protein\ Consumption, Large\ Household\ Size$ $\underset{(-)}{\longrightarrow} Vegetable\ Consumption$). Our DAG does not have any other direct connections between the individual characteristics and the dietary pattern, but there are a few causal chains through the food environment variables. For example, a causal chain runs from White to fast food restaurant density to fruit consumption. The link between white and fast food density is strong and negative, and the link between fast food density and fruit consumption is small and positive (i.e. $White \underset{(-)}{\longrightarrow} Fast\ Food\ Density \underset{(+)}{\longrightarrow} Fruit\ Consumption$). This would mean that, being White have less incidence of living in an area with more fast food restaurants, also will consume a smaller amount of the recommended daily fruit requirement. If one wants to find the effect of being White on consumption of daily fruit requirement, fruit consumption must only be regressed on White variable, since conditioning on fast food density variable would block the information flow from White to fruit consumption.

*Link between Individual Household Characteristics and Health Outcomes*

This is a more interesting area of research as some dimensions of socioeconomic status cause health, some are caused by health, and some are mutually determined with health (Cutler. Lleras-Muney, and Vogl, 2011). Smith (2007) generally finds that individuals with lower socioeconomic status have much worse health outcomes and that the primary culprit appears to be education and not an individual's financial resources. Williams et al. (2010) find that race, socioeconomic status, and gender all matter for health separately and in combination and that that racial disparities in health persist at every level of socioeconomic status.

Figure 4 shows a number of paths between individual characteristics and health outcomes. Table 4 shows that the age variable has a positive link with fair or poor health status and a negative link with too few fruits and vegetables consumed. This means that being older increased the incidences of being in fair/poor health status and decreased the incidences of consuming too few fruits and vegetables (i.e. $Age \underset{(+)}{\longrightarrow} FairPoor\ Health\ Status, Age \underset{(-)}{\longrightarrow} too\ few\ Fruits\ and\ Vegetables$). Being Asians have a lower incidence of being obese ($Asian \underset{(-)}{\longrightarrow} Obese$). Being a college graduate decreased the incidences of reporting fair/poor diet, fair/poor health and low food security ($College\ Graduate \underset{(-)}{\longrightarrow} FairPoor\ Diet, ge\ Graduate \underset{(-)}{\longrightarrow} FairPoor\ Health\ Status\ , College\ Graduate \underset{(-)}{\longrightarrow} Low\ Food\ Security$). Being Hispanic increased the incidences of reporting fair/poor diet and high food insecurity (low food security) (i.e.$Hispanic \underset{(+)}{\longrightarrow} FairPoor\ Health\ Status, Hispanic \underset{(+)}{\longrightarrow} Low\ Food\ Security$). High levels of income increased the incidences of being less food insecure and also being less obese ($Income \underset{(+\_}{\longrightarrow} Food\ Security, Income \underset{(-)}{\longrightarrow} Obesity$). The graph also indicates causal chains thorough other groups of variables. For example, $College\ Graduate \underset{(+)}{\longrightarrow} Percent\ of\ recommended\ fruit\ perday \underset{(-)}{\longrightarrow} Low\ food\ security$. This causal chain can be explained as follows. Being a college graduate increased the incidences of consuming daily recommended amounts of fruits which in turn decreased the food insecurity among them. Again, if one were to investigate the effect of being a college graduate on that individual being food (in)secure, one should not condition on the percent of recommended fruit per day variable, since it blocks the information flow from college graduate variable to low food security variable. Also, a causal chain exists from individual characteristics through dietary pattern to health outcomes ($Black \underset{(-)}{\longrightarrow} percent\ of\ recommended\ vegetables\ per\ day$ $\underset{(+)}{\longrightarrow} percent\ of\ recommended\ fruits\ per\ day \underset{(-)}{\longrightarrow} low\ food\ security$). This causal chain can be explained as follows. Being Black decreased the incidences of consuming daily recommended percentage of fruits and vegetables, which in turn made them food insecure. Again, if one wants to find out the effect of being Black on food (in)security, one must take caution on conditioning on variables such as percent of daily recommended fruits and vegetables, because these variables may block the information flow from variable, Black to variable, low food security. This is true for any causal chain relationship, if there are no *back-door* and *front-door* paths from the cause to the effect. Thus, research may need to be careful conditioning on variables that might block the paths from individual characteristics to health outcomes.

*Link between the Local Food Environment and the Individual's Dietary Pattern*

Moore et al. (2008) finds that individuals with no supermarkets near their home are much less likely to have a healthy diet. Those with a low store density are less likely to have a healthy diet than those with a higher density. Morland, Wing, and Roux (2002) find that individuals that live in areas with a higher presence of supermarkets consume more fruits and vegetables. Timperio et al. (2008) find that for children, a higher density of fast food outlets is associates with a lower likelihood of consuming the recommended amount of fruit.

Figure 4 shows a few edges that connect the local food environment to the dietary pattern. The parameter estimates from table 4 indicate that an individual living in an area with higher density of fast food outlets lead to consuming less dairy and more fruit, though the effects are small ($Density\ of\ Fast\ Food\ Outlets \underset{(-)}{\longrightarrow} Dairy\ Consumption,$

$Density\ of\ Fast\ Food\ Outlets \underset{(+)}{\longrightarrow} Fruit\ Consumption$). In addition, an individual living in an area with higher concentration of food stores have a low incidence of consuming too few fruits and vegetables ($Super\ market\ and\ super\ stores \underset{(+)}{\longrightarrow} Fruits\ and\ Vegetables$), but this effect does not show up directly in the dietary pattern. The rural indicator does not have any paths to any dietary pattern variables. That we find few edges between the local food environment and the dietary pattern variables indicates the effect of the food environment on diet is limited. This certainly is in contrast to the literature that finds strong associations between the two.

*Link between the Local Food Environment and Health Outcomes*

Chen, Jaenicke, and Volpe (2016) find that food environment factors are associated with obesity status even after controlling for individual and household factors. Garasky, Morton, and Greder (2006) find that the local food environment has a large impact on household food insecurity. Mehta and Chang (2008) find that a high fast food restaurant density is associated with a higher BMI and a higher full service restaurant density is associated with a lower BMI. Reitzel et al. (2014) find that a high density of fast food restaurants is positively associated with BMI but only for individuals with lower incomes.

Figure 4 shows only one direct path from the local food environment to health outcomes. Individuals living in an area with higher concentration of superstores/supermarkets have a high incidence of consuming more fruits and vegetables. There are no direct paths from the density of fast food and non-fast food restaurants to the obesity or food insecurity measure. Thus for this sample, the local food environment does not appear to have much influence on health outcomes directly. However, there is a causal chain from local food environment to health outcomes via the individual's dietary pattern ($Fast\ Food\ Restaurant\ Density \underset{(+)}{\longrightarrow} Percent\ of\ Recommended\ Fruits\ Per\ Day \underset{(-)}{\longrightarrow} Low\ Food\ Security$). This relationship means that, individuals living in areas with high fast food restaurant density have an increased incidence of consuming percent of recommended fruits per day, which in turn decreased the low food security (increased food security) among them.

*Link between the Individual's Dietary Pattern and Health Outcomes*

Bradlee et al. (2009) find that the intakes of dairy, grains and total fruits and vegetables are inversely associated with obesity among adolescents. Wolongevicz et al. (2010) find that women with a lower diet quality are more likely to become obese than those with a higher quality diet. Wosje et al. (2010) find that a diet high in dark green and deep-yellow vegetables was associated with a lower fat mass for children.

Figure 4 shows that the DAG makes a connection between fruit consumption and dietary outcomes. As seen in table 4, being an individual with consumption of higher amounts of fruits in their diets have low incidence of claiming a fair/poor diet, and low food security ($Percent\ of\ Fruits\ in\ Diets \underset{(-)}{\longrightarrow} FairPoor\ Diet, Percent\ of\ Fruits\ in\ Diets \underset{(-)}{\longrightarrow} Low\ Food\ Security$). That only the fruit consumption seems to affect the health variables means that fruit consumption as a percent of the daily recommended amount may be a good proxy for food (in)security. Many of the edges stay within the dietary pattern group and do not connect to any of the health outcomes.

*Link between Prices and Dietary Patterns or Health Outcomes*

Wang and Bessler (2006) use DAGs to find that for some meat products in the United States, prices and quantities purchased are contemporaneous. Beydoun *et al*. (2011) find that a higher price index of fruits and vegetables is associated with a higher BMI. Powell and Han

(2011) find that fast food and food at home prices are not associated with any broad food consumption categories.

Our price variables have mixed results in the DAG in figure 4. Individual and household characteristics, local food environment, and the policy variables do not have any effect on the prices. This may be due to the procedure used to correct for quality described earlier. Only one price affects any part of the dietary pattern; the price of diary positively leads to the consumption of oils. For two items, the item's quantity causes the item's price: $Percent\ of\ Recommended\ Oils\ Per\ Day \underset{(+)}{\longrightarrow} Quality\ Adjusted\ Price\ of Oils$, $Percent\ of\ Recommended\ Vegetables\ Per\ Day \underset{(-)}{\longrightarrow} Quality\ Adjusted\ Price\ of\ Vegetables$. For these two items, the quantity is determined before the prices. The quantity of dairy has a link to the price of oils ($Percent\ of\ recommended\ Dairy\ Per\ Day \underset{(+)}{\longrightarrow} Quality\ Adjusted\ Price\ of\ Oils$). However, no price directly affects the product that it represents.

*Links with SNAP Participation, WIC Participation, Food Insecurity, and Obesity*

Finally, we concentrate our discussion of results for some popular variables of policy interest, such as participation in the SNAP and WIC programs, U.S. food insecurity, and U.S. obesity epidemic, to other research. We compare results from other authors to our figures 5, 6, and 7 for direct effects and we refer back to figure 4 for any causal chains. Dharmasena, Bessler, and Capps (2016) use of DAGs reveals obesity, food insecurity, and SNAP participation in the United States to be strictly endogenous (i.e. caused by other variables). Our results find obesity to be endogenous but that food insecurity and SNAP participation are not strictly endogenous. Also similar to Dharmasena, Bessler, and Capps (2016), we find no direct causality between obesity and food insecurity. However, we find a *back-door* path between obesity and food insecurity via income ($Obesity$

$\underset{(-)}{\leftarrow} Household\ Income\ as\ Percent\ of\ Household\ Poverty\ Guideline$

$\underset{(-)}{\rightarrow} Low\ Food\ Insecurity \underset{(+)}{\rightarrow} FairPoor\ Health\ Status \underset{(+)}{\rightarrow} Obesity$). This means that low-income households have increased incidence of being obese (this is direct causality) as well as being low-income causes to have high incidence of being food insecure, which in turn leads to being in fair/poor health status, which ultimately leads to high rates of incidence of obesity among these households. If one wants to find out the effect of income (Household Income as Percent of Household Poverty Guideline variable) on obesity, the *back-door* path leading to obesity via low food insecurity must be blocked (by including both income and low food insecurity variables in the right-hand side of the regression).

Gundersen *et al*. (2014) modeled food insecurity with income, race, and unemployment at a state level. They find that the unemployed and those in poverty are more likely to be food insecure. No link was found between those who are classified as Black population and food insecurity but a negative association was found with a state's Hispanic population. Nord et al. (2010) find that households headed by a Black, Hispanic, or less educated individuals are all more likely to be food insecure.

Figure 4 (and table 4) shows paths between poverty, race and food insecurity. We find that being Hispanic increases the incidences of being food insecure. There is also a direct path between the percent of poverty level and food insecurity and a path between college education and food insecurity. Having an income above the poverty level and being college educated makes an individual more food secure. We find a causal chain from Black to SNAP participation to low food security. Those who are classified as Black have high incidence of being a SNAP participant, which in turn leads to a higher level of food insecurity (that is, $Black$

$\underset{(+)}{\longrightarrow} SNAP\ Participation \underset{(+)}{\longrightarrow} Low\ Food\ Insecurity$). Also, those who are classified as Black has high incidence of being a SNAP participant, which in turn results in fair/poor health status, and ultimately being obese ($Black \underset{(+)}{\longrightarrow} SNAP\ Participation \underset{(+)}{\longrightarrow} FairPoor\ Health\ Satus \underset{(+)}{\longrightarrow} Obesity$). Being Asian has a low incidence of being a SNAP participant, which would in turn reduce the level of food insecurity ($Asian \underset{(-)}{\longrightarrow} SNAP\ Participation \underset{(-)}{\longrightarrow} Food\ Insecurity$). These results mean that if SNAP participation is included in a model with this race variable and food insecurity, the path between race and food insecurity will be blocked by the inclusion of SNAP participation (since this is a causal chain). This supports the results from Dharmasena, Bessler, and Capps (2016).

Gundersen and Ziliak (2015) report that food insecure children are more likely to report being in fair or poor health and that food insecurity is generally negatively associated with health. Our DAG shows similar results. There is a direct and positive path from low food security to fair or poor health status. There is also an indirect path from food insecurity to fair or poor health status via fair or poor diet status. Thus, including diet status in a model of food insecurity and health status might block the path between food insecurity and health status (again due to the effect of a causal chain).

**Conclusions**

The objective of this research was to use the individual and household characteristics, characteristics of the local food environment, the individual's dietary pattern, food prices, health outcomes, and policy variables to estimate a complex causality structure using the National Household Food Acquisition and Purchase Survey (FoodAPS). This was in contrast to studies that consider these variables in a fragmented approach and with which we made comparison in the results section.

To accomplish this, we estimated a graphical causality structure by way of a directed acyclic graph (DAG). The DAG is generated using two algorithms: GES and LOFS R3. First, the GES algorithms are run on the data. Then, the LOFS R3 algorithm is run on the resulting structure to orient any edges that were not oriented by the GES algorithm. The DAG is generated under assumptions made by imposing *a priori* knowledge on the structure. From this DAG we are able to construct structural relationships and estimate partial effects for the edges, which allowed for the comparisons to other research.

We found a number of interesting results for the relationship between individual characteristics and the local food environment. Asian individuals live in areas with higher concentrations of fast food and non-fast food restaurants. White individuals live in areas with few fast food restaurants. Hispanic individuals live in areas with a higher concentration of fast food restaurants and food stores.

Also, we find some interesting results between individual characteristics and health outcomes. Being Asian has a high incidence of being less obese. Being a college graduate has a low incidence of being under fair/poor diet, low incidence of being under fair/poor health status, and low incidence of being food insecure. Being Hispanic makes an individual reporting being in fair or poor diet and then also reporting food insecure. Being high-income has a low incidence of being food insecure which in turn makes them less obese.

Our price variables have mixed results in the DAG. Individual and household characteristics, local food environment, and the policy variables do not have any causal effect on the prices. This is likely due to the procedure used to correct for quality described earlier. Only one price affects any part of the dietary pattern; the price of diary positively causes the consumption of oils. Some quantities affect price in the DAG, such as the quantity of vegetables affecting the price of vegetables. The quantity of oils also affects the price of oils. However, no price directly affects its corresponding product.

In regards to the paths between poverty, race and food insecurity, we find a number of paths. We find that being Hispanic makes one to be more food insecure. There is also a direct positive causality path between the percent of poverty level and food insecurity and a negative causality path between college education and food insecurity. We find a positive causal chain from Black to SNAP participation to low food security. A similar casual chain also exists for Hispanic individuals to low food security via SNAP participation. These results mean that if SNAP participation is included in a model with these race variables and food insecurity, the path between race and food insecurity will be blocked by the inclusion of SNAP participation. This is similar to results from Dharmasena, Bessler, and Capps (2016). Obesity is positively caused by fair/poor health status and fair/poor diet status.

A number of directions for future research are immediately apparent. Moving beyond assuming a Normal distribution and the absence of latent variables is ripe for future research. Use of an algorithm in developing the DAG that allows for latent variables such as the FCI (Fast Causal Inference) may give further insight into the interactions of variables (Spirtes, Meek, and Richardson, 1999). Similarly, an algorithm that allows non-Gaussian errors such as LiNGAM may be helpful (Shimizu et al., 2006). In addition, we may need to consider the choice of variables in the model. For example, perceptions of the local food environment may be more important in influencing the consumption of fruits and vegetables than actual store density measures (Lucan and Mitra, 2012).

.

***References***


Ahima, R.S., and M.A. Lazar. 2013. "The Health Risk of Obesity-Better Metrics Imperative." *Science* 23:856-858.

Alviola, P.A., and O.Capps, Jr. 2010. "Household demand analysis of organic and conventional fluid milk in the United States based on the 2004 Nielsen Homescan panel." *Agribusiness* 26(3):369-388.

Anderson, T. W. and D.A. Darling. 1952. "Asymptotic Theory of Certain "Goodness-of-Fit" Criteria Based on Stochastic Processes". *Annals of Mathematical Statistics* 23:193–212.

Andrews, M., R. Bhatta, and M.V. Ploeg. 2013. "An Alternative to Developing Stores in Food Deserts: Can Changes in SNAP Benefits Make a Difference?" *Appl. Econ. Perp. And Policy* 35(1): 150-170.

Barlow, S.E. 2007. "Expert Committee Recommendations Regarding the Prevention, Assessment, and Treatment of Child and Adolescent Overweight and Obesity: Summary Report." Pediatrics 120 (Supplement 4) S164-S192.

Basu, S., C. Wimer, and H. Seligman. 2016. "Moderation of the Relation of County-Level Cost of Living to Nutrition by the Supplemental Nutrition Assistance Program." *American Journal of Public Health* 106(11):2064-2070.

Beatty, T.K.M., B.H. Lin, and T.A. Smith. 2014. "Is Diet Quality Improving? Distributional Changes in the United States, 1989-2008." *Amer. J. Agr. Econ.* 96(3): 769-789.

Bessler, D.A. and S. Lee. 2002. "Money and Prices: U.S. Data 1869-1914 (A Study with Directed Graphs)." *Empirical Economics* 27:427-446.

Beydoun, M.A., L.M. Powell, X. Chen, and Y. Wang. 2011. "Food Prices Are Associated with Dietary Quality, Fast Food Consumption, and BodyMass Index among U.S. Children and Adolescents." *The Journal of Nutrition* 141:304–311.

Bollen, K. A. 1989. *Structural Equations with Latent Variables.* New York: John Wiley & Sons.

Bradlee, M.L., M.R. Singer, M.M. Quresh, and L.L. Moore. 2009. "Food Group Intake and Central Obesity among Children and Adolescents in the Third National Health and Nutrition Examination Survey (NHANES III)." *Public Health Nutrition* 13(6):797–805.

Bowman S.A., J.C. Clemens, J.E. Friday, K.L. Lynch, and A.J. Moshfegh. 2017. *Food Patterns Equivalents Database 2013-14: Methodology and User Guide [Online]*. Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland. May. Available at: http://www.ars.usda.gov/nea/bhnrc/fsrg

Capps, Jr. O., R. Tsai, R. Kirby, and G. Williams. 1994. "A comparison of demand for meat products in the Pacific Rim region." *Journal of Agricultural and Applied Economics* 19(1): 210-224.

Chen, D., E.C. Jaenicke, and R.J. Volpe. 2016. "Food Environments and Obesity: Household Diet Expenditure versus Food Deserts." *American Journal of Public Health* 106(5):881-888.

Chickering, D.M. 2002. "Optimal Structure Identification with Greedy Search." *Journal of Machine Learning Research* 3:507-554.

Chickering, D.M. and C. Meek. 2002. "Finding Optimal Bayesian Networks." *Proceeding of the Eighteenth Conference on Uncertainty in Artificial Intelligence.* pp. 94-102.

Cox, T., and M. Wohlgenant. 1986. "Prices and Quality effects in cross-section demand analysis." *Am. J. Agric. Econ*. 68:908-919.

Cutler, D.M., A. Lleras-Muney, and T. Vogl. 2011. "Socioeconomic Status and Health: Dimensions and Mechanisms." In S. Glied and P.C. Smith, ed. *Oxford Handbook of Health Economics*. Oxford, U.K.: Oxford University Press, pp. 124-168.

Darmon, N. and A. Drewnowski. 2008. "Does social class predict diet quality?" *Am J Clin Nutr* 87:1107–1117.

Davis, G.A. 2003. "Bayesian Reconstruction of Traffic Accidents." *Law, Prob. & Risk* 2:69-89.

Dharmasena, S., D.A. Bessler., and O. Capps Jr. 2016. "Food Environment in the United States as a Complex Economic System." *Food Policy* 61: 163-175.

Dharmasena, S., and O. Capps, Jr. 2012. "Intended and Unintended Consequences of a Proposed National Tax on Sugar-Sweetened Beverages to Combat the U.S. Obesity Problem." *Health Economics* 21(6):669-694

Dharmasena, S. and O.Capps Jr. 2014. "Unraveling demand for dairy-alternative beverages in the United States: the case of soymilk." *Agricultural and Resource Economics Review* 43(1):140-157

Dubowitz, T., M. Heron, C.E. Bird, N. Lurie, B.K. Finch, R. Basurto-Da'vila, L. Hale, and J.J. Escarce. 2008. "Neighborhood Socioeconomic Status and Fruit and Vegetable Intake among Whites, Blacks, and Mexican Americans in the United States." *Am J Clin Nutr* 87:1883–91.

Finkelstein, E.A., C.J. Ruhm, and K.M. Kosa. 2005. "Economic Causes and Consequences of Obesity." *Annual Rev. of Pub. Health.* 26:239-257.

Garasky, S., L.W. Morton, and K.A. Greder. 2006. "The Effects of the Local Food Environment and Social Support on Rural Food Insecurity." *Journal of Hunger & Environmental Nutrition* 1(1):83-103.

Geiger, D., T. Verma, and J. Pearl. 1990. "Identifying Independencies in Bayesian Networks." *Networks* 20:507-534.

Glymour, C., R. Scheines, P. Spirtes, and J. Ramsey. 2016. The TETRAD Project: Causal Models and Statistical Data. Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, http://www.phil.cmu.edu/tetrad/current.html.

Gray, K.F. 2014. *Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2013.* Washington DC: U.S. Department of Agriculture, Food and Nutrition Service, Supplemental Nutrition Assistance Program Report No. SNAP-14-CHAR, December.

Gundersen, C., and V. Oliveira. 2001. "The Food Stamp Program and Food Insufficiency." *Amer. J. Agr. Econ.* 83(4): 875-887.

Gundersen, C. and J.P. Ziliak. 2015. "Food Insecurity and Health Outcomes." *Health Affairs* 34(11):1830-1839.

Gundersen, C., B. Kreider, and J. Pepper. 2011. “The Economics of Food Security in the United States.” *Appl. Econ. Perspect.* 33(3):281-303.
Gundersen, C., A. Dewey, A.S. Crumbaugh, M. Kato, and E. Engelhard. 2014. *Map the Meal Gap 2016: Technical Brief*, Feeding America.
Guthrie, J.F., B.H. Lin, and E. Frazao. 2002. “Role of Food Prepared Away From Home in the American Diet, 1977-78 versus 1994-96: Changes and Consequences.” *J. of Nutrition Education and Behavior* 34(3): 140-150.
Jensen, H. H. 2002. “Food Insecurity and the Food Stamp Program.” *American Journal of Agricultural Economics* 84 (5): 1215–1228.
Kwate, N.O.A. 2008. “Fried Chicken and Fresh Apples: Racial Segregation as a Fundamental Cause of Fast Food Density in Black Neighborhoods.” *Health & Place* 14:32-44.
Kwon, D.H. and D.A. Bessler. 2011. “Graphical Methods, Inductive Causal Inference, and Econometrics: A Literature Review.” *Computational Economics* 38(1):85-106.
Kyureghian, G., Capps Jr, O. and Nayga Jr, R.M., 2011. “A Missing Variable Imputation Methodology with an Empirical Application.” *Advances in Econometrics,* 27: 313-337.
Lai, P. and D.A. Bessler. 2015. “Price Discovery between Carbonated Soft Drink Manufacturers and Retailers: A Disaggregate Analysis with PC and LiNGAM Algorithms.” *Journal of Applied Economics* 18(1):173-197.
Lin, B.H., S.T. Yen, D. Dong, and D.M. Smallwood. 2010.”Economic Incentives For Dietary Improvements Among Food Stamp Recipients.” *Contem. Econ. Policy* 28(4): 524-536.
Liu, J.L., B. Han, and D.A. Cohen. 2015. “Beyond Neighborhood Food Environments: Distance Traveled to Food Establishments in 5 US Cities, 2009–2011.” *Prev Chronic Dis* 12(6).
Liu, M., P. Kasteridis, and S.T. Yen. 2013. “Breakfast, Lunch, and Dinner Expenditures Away From Home in the United States.” *Food Policy* 38:156-164.
Lucan, S.C and N. Mitra. 2012. “Perceptions of the Food Environment are associated with Fast-Food (Not Fruit-And-Vegetable) Consumption: Findings from Multilevel Models.” *International Journal of Public Health* 57:599–608.
Mabli, J., J. Ohls, L. Dragoset, L. Castner, and B. Santos. 2013. *Measuring the Effect of Supplemental Nutrition Assistance Program (SNAP) Participation on Food Security.* Prepared by Mathematica Policy Research for the U.S. Department of Agriculture, Food and Nutrition Service, August.
Meek, C. 1995. “Causal Inference and Causal Explanation with Background Knowledge.” In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95),* Morgan Kaufmann Publishers, Inc. pp. 403-411
Meek, C. 1997. “Graphical Structures: Selecting Causal and Statistical Models.” PhD thesis, Carnegie Mellon University.
Mehta, N.K. and V.W. Chang. 2008. “Weight Status and Restaurant Availability: A Multilevel Analysis.” *Am J Prev Med* 34(2):127-133.

Moore, L.V., A.V.D. Roux, J.A. Nettleton, and D.R. Jacobs. 2008. “Associations of the Local Food Environment with Diet Quality—A Comparison of Assessments based on Surveys and Geographic Information Systems: The Multi-Ethnic Study of Atherosclerosis.” *American Journal of Epidemiology* 167(8):917-924.

Morland, K., S. Wing, and A.D. Roux. 2002. “The Contextual Effect of the Local Food Environment on Residents’ Diets: The Atherosclerosis Risk in Communities Study.” *American Journal of Public Health* 92(11):1761-1767.

Mumford, J.A. and J.D. Ramsey. 2014. “Bayesian Networks for fMRI: A Primer.” *NeuroImage* 86:571-582.

Nord, M, A. Coleman-Jensen, M. Andrews, and S. Carlson, 2010. *Household Food Security in the United States, 2009*. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No. 108.

Ogden CL, Carroll MD, Kit BK, Flegal KM. 2014. ”Prevalence of Childhood and Adult Obesity in the United States, 2011-2012.” *Journal of the American Medical Association*, 311(8): 806-814.

Pearl, J. 1986. “Fusion, Propagation, and Structuring in Belief Networks” *Artificial Intelligence* 29:241-288.

Pearl, J. 1995. “Causal Diagrams for Empirical Research.” *Biometrika* 82(4):668-710

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference,* 2nd ed. New York: Cambridge University Press.

Powell, L.M. and E. Han. 2011. “The Costs of Food at Home and Away From Home and Consumption Patterns among U.S. Adolescents.” *Journal of Adolescent Health* 48:20–26.

Powell, L.M., F.J. Chaloupka, and Y. Bao. 2007. “Community, State, and Other Environmental Issues The Availability of Fast-Food and Full-Service Restaurants in the United States: Associations with Neighborhood Characteristics.” *American Journal of Preventive Medicine* 33(4S):S240–S245.

Ramsey, J. 2010. “Bootstrapping the PC and CPC Algorithms to Improve Search Accuracy.” Department of Philosophy. Technical Report No. CMU-PHIL-187, Carnegie Mellon University.

Ramsey, J.D., S.J. Hanson, and C. Glymour. 2011. “Multi-Subject Search Correctly Identifies Causal Connections and Most Causal Directions in the DCM Models of the Smith et al. Simulation Study.” *NeuroImage* 58:838-848.

Ramsey, J.D., R. Sanchez-Romero, and C. Glymour. 2014. “Non-Gaussian Methods and High-Pass Filters in the Estimation of Effective Connections.” *NeuroImage* 84:986-1006.

Ratcliffe, C., S. McKernan, and S. Zhang. 2011. “How Much Does the Supplemental Nutrition Assistance Program Reduce Food Insecurity?” *American Journal of Agricultural Economics* 93(4):1082–1098.

Reitzel, L.R., S.D. Regan, N. Nguyen, E.K. Cromley, L.L. Strong, D.W. Wetter, and L.H. McNeill. 2014. “Density and Proximity of Fast Food Restaurants and Body Mass Index among African Americans.” *American Journal of Public Health* 104(1):110-116.

Shimizu, S., P. Hoyer, A. Hyvarinen, and A. Kerminen. 2006. “A Linear Non-Gaussian Acyclic Model for Causal Discovery.” *Journal of Machine Learning Research* 7:2003-2030.

Smith, J.P. 2007. “The Impact of Socioeconomic Status on Health over the Life-Course.” *Journal of Human Resources* 42(4):739-764.

Smith, T.A., J.P. Berning, X. Yang, G. Colson, and J.H. Dorfman. 2016. “The Effects of Benefit Timing and Income Fungibility on Food Purchasing Decisions among Supplemental Nutrition Assistance Program Households.” *Am J Agric Econ* 98(2):564-580.

Smith, S.M., K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich. 2011. “Network Modelling Methods for FMRI.” *NeuroImage* 54(2):875-891.

Spirtes, P., C.N. Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search*, Cambridge, MA: MIT Press.

Spirtes, P., Meek, C. and Richardson, T. 1999. “An Algorithm for Causal Inference in the Presence of Latent Variables and Selection Bias.” In *Computation, Causation, and Discovery* 211–252. AAAI Press, Menlo Park, CA.

Stewart, H., J. Hyman, E. Frazao, J.C. Buzby, and A. Carlson. 2011. “Can Low-income Americans Afford to Satisfy MyPryamid Fruit and Vegetable Guidelines?” *J. Nutrition Education and Behavior* 43(3): 173-179.

Taylor, R., and S.B. Villas-Boas. 2016. “A Discrete Choice Analysis of the National Household Food Acquisition and Purchase Survey (FoodAPS).” *Amer. J. Agr. Econ.* 98(2): 513-532.

Timperio, A., K. Ball, R. Roberts, K. Campbell, N. Andrianopoulos, and D. Crawford. 2008. “Children's Fruit and Vegetable Intake: Associations with the Neighbourhood Food Environment.” *Preventive Medicine* 46:331–335.

Todd, J., and B. Scharadin. 2016 *Where Households Get Food in a Typical Week: Findings From USDA’s FoodAPS,* Washington DC: U.S. Department of Agriculture, Economic Research Service, EIB-156, July.

U.S. Census Bureau, Population Division. 2017. Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016. Available at https://factfinder.census.gov/bkmk/table/1.0/en/PEP/2016/PEPANNRES/0100000US|0100000US.04000|0200000US1|0200000US2|0200000US3|0200000US4.

U.S. Department of Agriculture, Economic Research Service (USDA, ERS). 2016. *National Household Food Acquisition and Purchase Survey (FoodAPS): User’s Guide to Survey Design, Data Collection, and Overview of Datasets*. Washington DC: November.

U.S. Department of Agriculture, Economic Research Service (USDA, ERS). 2017. *National Household Food Acquisition and Purchase Survey (FoodAPS).* http://www.ers.usda.gov/data-products/FoodAPS-national-household-food-acquisition-and-purchase-survey.aspx

U.S. Department of Health & Human Services (HHS). 2012. “2012 HHS Poverty Guidelines.” *Federal Register* (77)17:4034-4035

U.S. Department of Health and Human Services and U.S. Department of Agriculture (HHS and USDA). 2015. *2015–2020 Dietary Guidelines for Americans.* 8th Edition. Available at http://health.gov/dietaryguidelines/2015/guidelines/.

Ver Ploeg, M., L. Mancino, J. E. Todd, D.M. Clay, and B. Scharadin. 2015. *Where Do Americans Usually Shop for Food and How Do They Travel To Get There? Initial Findings From the National Household Food Acquisition and Purchase Survey*, Washington DC: U.S. Department of Agriculture, Economic Research Service, EIB-138, March.

Walker, R.E., C.R. Keane, and J.G. Burke. 2010. "Disparities and Access to Healthy Food in the United States: A Review of Food Deserts Literature." *Health & Place* 16(5):876-884.

Wang, Z. and D.A. Bessler. 2006. "Price and Quantity Endogeneity in Demand Analysis: Evidence from Directed Acyclic Graphs." *Agricultural Economics* 34:87-95.

Wilde, P., and C.K. Ranney. 2000. "The Monthly Food Stamp Cycle: Shopping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework." *Amer. J. Agr. Econ.* 82: 200-213.

Wilde, P., J. Llobrera, and M.V. Ploeg. 2014. "Population Density, Poverty, and Food Retail Access in the United States: An Empirical Approach." *International Food and Agribusiness Management Review* 17(A):171-186.

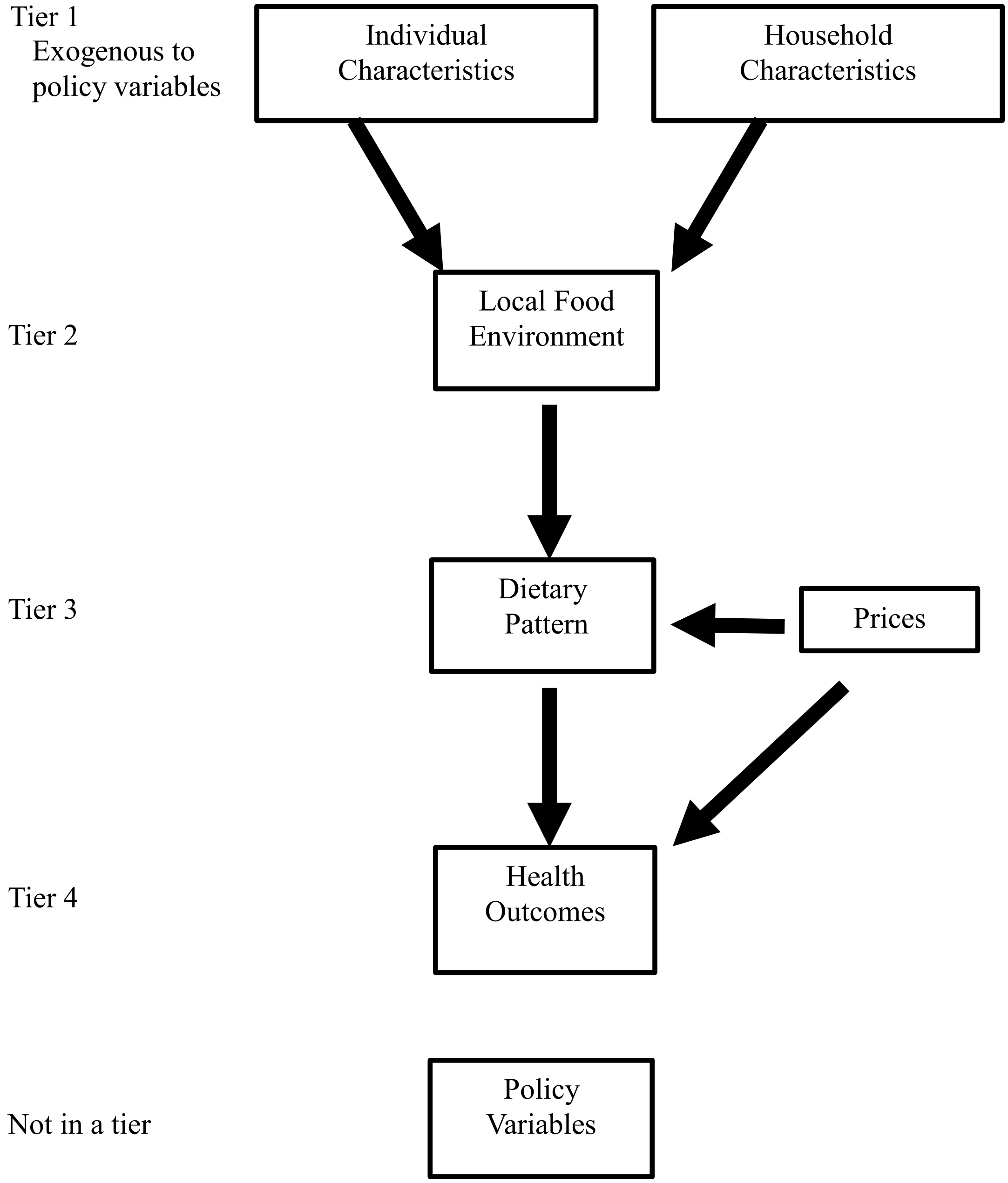
Williams, D.R., S.A. Mohammed, J. Leavell, and C. Collins. 2010. "Race, Socioeconomic Status, and Health: Complexities, Ongoing Challenges, and Research Opportunities." *Ann. N.Y. Acad. Sci* 1186:69-101.

Wolongevicz, D.M., L. Zhu, M.J. Pencina, R.W. Kimokoti, P.K. Newby, R.B. D'Agostino, and B.E. Millen. 2010. "Diet Quality and Obesity in Women: The Framingham Nutrition Studies." *British Journal of Nutrition* 103:1223–1229.

Wosje, K.S., P.R. Khoury, R.P. Claytor, K.A. Copeland, R.W. Hornung, S.R. Daniels, and H.J. Kalkwarf. 2010. "Dietary Patterns Associated with Fat and Bone Mass in Young Children." *Am J Clin Nutr* 92:294–303.

Zagorsky, J.L. and P.K. Smith. 2017. "The Association between Socioeconomic Status and Adult Fast-Food Consumption in the U.S." *Economics and Human Biology,* Available online https://doi.org/10.1016/j.ehb.2017.04.004.

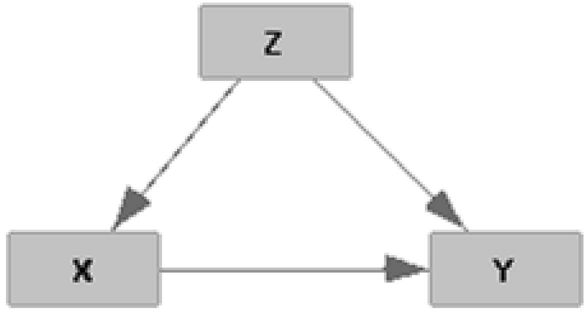
**Figure 1. Example Directed Acyclic Graph (DAG) with Imposed Knowledge**




*Note:* This is an example of a DAG. The figure also demonstrates knowledge imposed during estimation. Arrows indicate the direction of causality. Not all possible paths are included for simplicity. Variable in a given tier can only affect variables in higher number tiers. The policy variables are not in a tier and can be endogenous or exogenous to any tier except that they are not allowed to cause variable in tier 1.

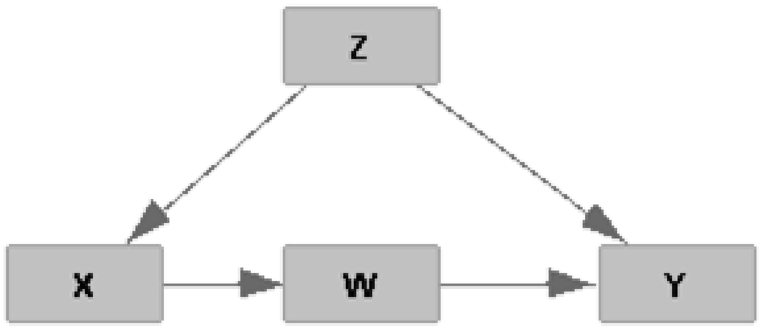
*Source*: Produced by authors.

**Figure 2. Representation of Back Door Criteria**



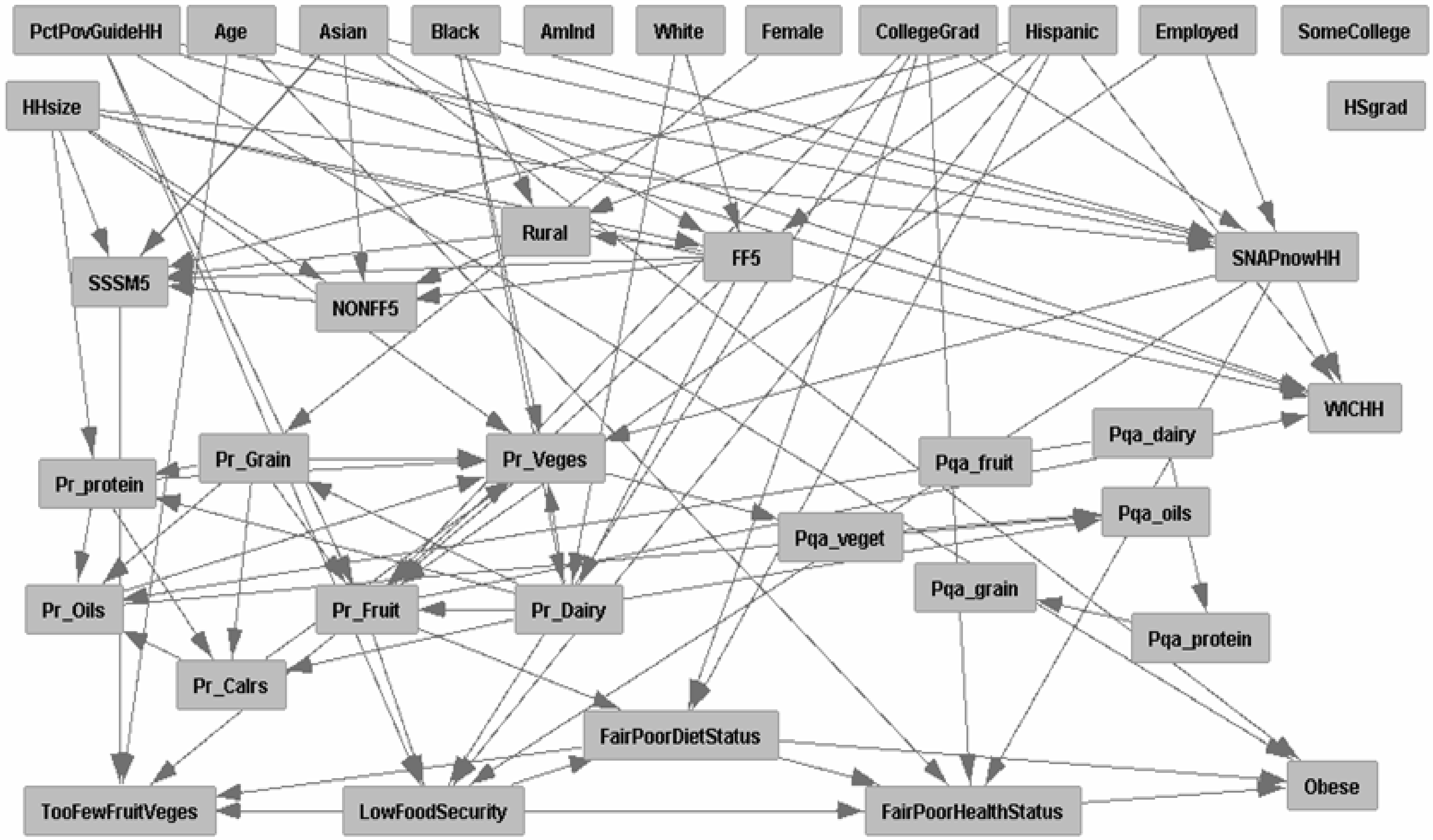

*Source*: Produced by authors.

**Figure 3. Representation of Front Door Criteria**




*Source*: Produced by authors.

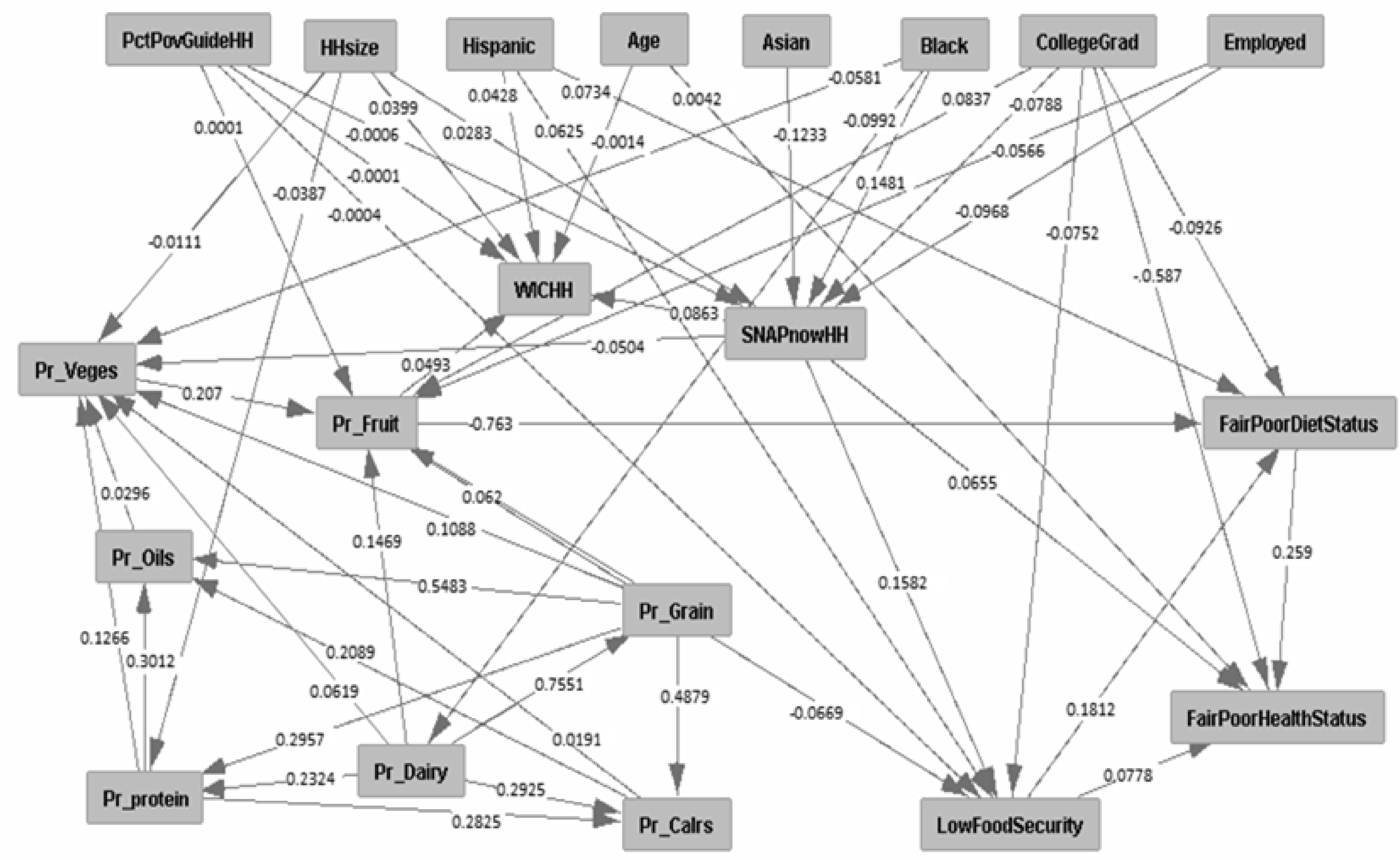
**Figure 4. Directed Acyclic Graph Associated with the Variables of Interest after Running GES Algorithm and R3 Algorithm**




*Note:* The graph contains 88 edges between the 37 nodes. The max number of edges into a node is eight and the max number of edges out of a node is seven. The maximum total number of edges into and out of a node is 11.

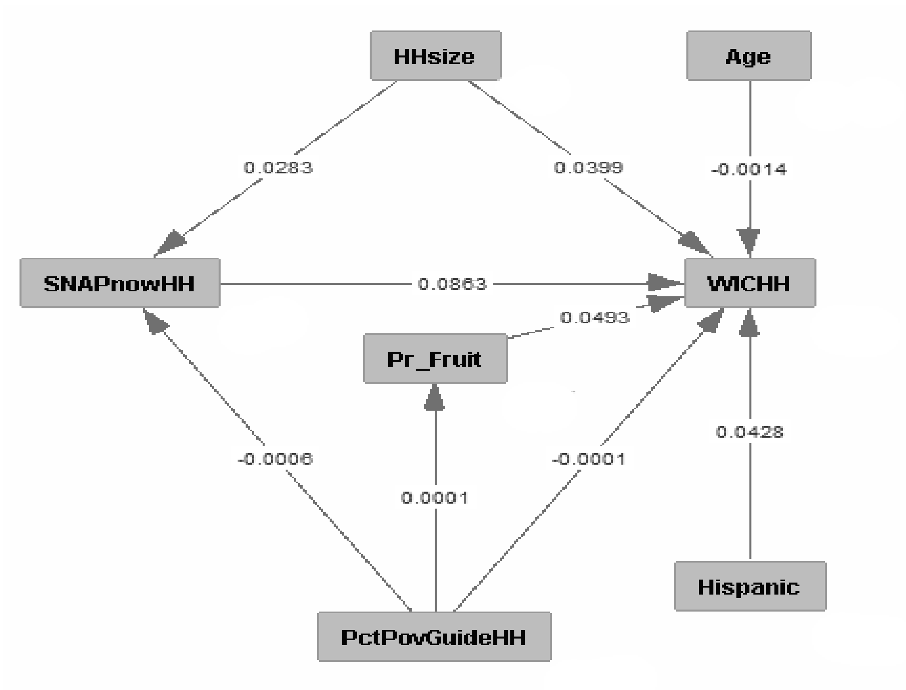
*Source*: Produced by authors.

**Figure 5. Markov Blanket for SNAP Participation with Partial Effects**



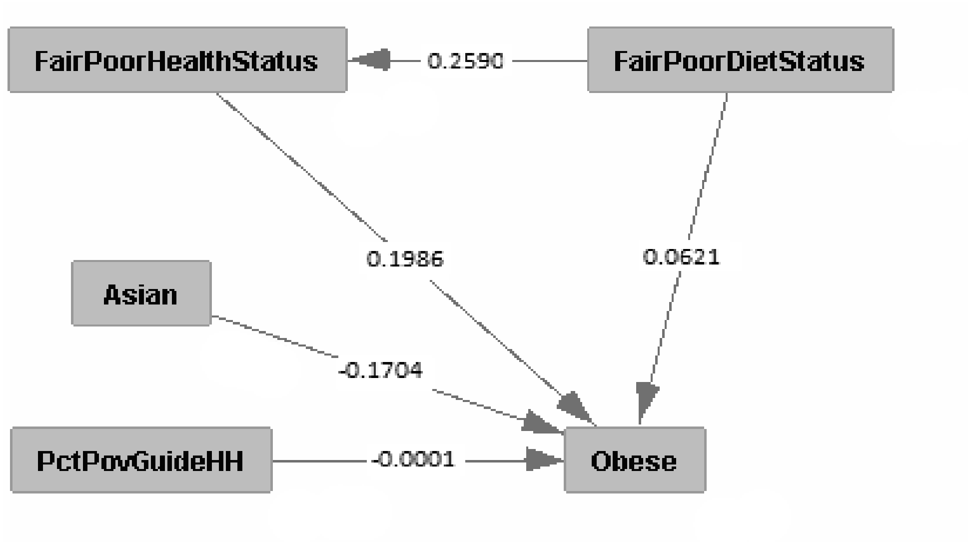

*Source*: Produced by authors.

**Figure 6. Markov Blanket for WIC Participation with Partial Effects**




*Source*: Produced by authors.

**Figure 7. Markov Blanket for Obesity with Partial Effects**




*Source*: Produced by authors.

**Table 1. Descriptions of Variables and Summary Statistics**

| Variable Name | Variable Description | Mean | Std. Error of Mean | Min | Max |
|---|---|---|---|---|---|
| *Household or Individual Characteristics* | | | | | |
| HHsize | Number of people at residence, excluding guests | 3.2854 | 0.0711 | 1.0000 | 14.0000 |
| PctPovGuideHH | Household average (monthly) income as sum of average imputed income per member as percent of household poverty guideline | 366.0040 | 15.6906 | 0.0000 | 2755.5953 |
| Female | Indicates if individual is female | 0.5195 | 0.0045 | | |
| Age | Approximate midpoint of individual's age group | 39.6783 | 0.5423 | 1.0000 | 85.0000 |
| White | Individual in White racial category (base is other race) | 0.7375 | 0.0220 | | |
| Black | Individual in Black racial category (base is other race) | 0.1304 | 0.0183 | | |
| AmInd | Individual in American Indian or Alaskan Native racial category (base is other race) | 0.0059 | 0.0015 | | |
| Asian | Individual in Asian or Native Hawaiian or Other Pacific Islander racial category (base is other race) | 0.0431 | 0.0064 | | |
| Hispanic | indicates individual Hispanic (base in non-Hispanic) | 0.1636 | 0.0270 | | |
| Employed | Individual is currently employed (base in not employed) | 0.4840 | 0.0086 | | |
| HSgrad | Individual is high school grad (base is less HS or currently in school) | 0.2180 | 0.0101 | | |
| SomeCollege | Individual completed some college (base is less HS or currently in school) | 0.2388 | 0.0091 | | |
| CollegeGrad | Individual is college grad (base is less HS or currently in school) | 0.2429 | 0.0166 | | |
| *Local Food Environment Characteristics* | | | | | |
| FF5 | Number of fast food restaurants within 5 mi of household | 61.3541 | 7.2144 | 0.0000 | 429.0000 |
| NONFF5 | Number of non-fast food restaurants within 5 mi of household | 280.5988 | 45.6687 | 0.0000 | 3639.0000 |
| SSSM5 | Number of snap-authorized supermarkets and superstores within 5 miles of household | 21.7820 | 3.9756 | 0.0000 | 383.0000 |
| Rural | Indicates household lives in census rural area | 0.3336 | 0.0337 | | |

*(continued)*

**Table 1. *(continued)***

| Variable Name | Variable Description | Mean | Std. Error of Mean | Min | Max |
|---|---|---|---|---|---|
| *Prices* | | | | | |
| Pqa_dairy | Quality adjusted price of dairy in $/gram | 0.0069 | 0.0011 | -0.0047 | 1.0059 |
| Pqa_protein | Quality adjusted price of protein in $/gram | 0.0243 | 0.0010 | -0.0204 | 3.1173 |
| Pqa_grain | Quality adjusted price of grain in $/gram | 0.1042 | 0.0003 | -0.0043 | 0.3431 |
| Pqa_fruit | Quality adjusted price of fruit in $/gram | 0.0083 | 0.0016 | -0.0064 | 1.7725 |
| Pqa_veget | Quality adjusted price of vegetables in $/gram | 0.0091 | 0.0003 | -0.0061 | 0.4162 |
| Pqa_oils | Quality adjusted price of oils in $/gram | 0.0275 | 0.0014 | -0.0045 | 0.8591 |
| *Quantities acquired* | | | | | |
| Pr_dairy | Percent recommended amount of diary per day | 0.5150 | 0.0176 | 0.0000 | 14.2052 |
| Pr_fruit | Percent recommended amount of fruit per day | 0.3511 | 0.0144 | 0.0000 | 4.8440 |
| Pr_grain | Percent recommended amount of grain per day | 0.8844 | 0.0195 | 0.0000 | 21.8428 |
| Pr_protein | Percent recommended amount of protein per day | 0.6591 | 0.0205 | 0.0000 | 20.5850 |
| Pr_veges | Percent recommended amount of vegetables per day | 0.4277 | 0.0107 | 0.0000 | 5.5386 |
| Pr_calrs | Percent recommended amount of calories per day | 0.9030 | 0.0330 | 0.0000 | 83.3027 |
| Pr_oils | Percent recommended amount of oils per day | 0.9145 | 0.0265 | 0.0000 | 39.2867 |
| *Health Measures* | | | | | |
| LowFoodSecurity | Household level 30-day measure of food security, Indicator indicates low food security | 0.1659 | 0.0104 | | |
| FairPoorDietStatus | Household level own assessment of health of diet, Indicator indicates rating of fair or low diet | 0.2224 | 0.0112 | | |
| TooFewFruitVeges | Household level assessment if enough produce is consumed, Indicator indicates belief consumes too few fruits/vegetables | 0.6897 | 0.0148 | | |
| FairPoorHealthStatus | Respondent's rating of individual's general health is fair or poor | 0.1489 | 0.0070 | | |
| Obese | Indicator variable for obesity. For adults, obese determined by BMI ranges 30.0 and above. For children, obese is determined by ranges of BMI percentile at or above the 95th percentile. | 0.2816 | 0.0098 | | |
| *Policy Variables* | | | | | |
| SNAPnowHH | Indicator if anyone in household is receiving SNAP benefits | 0.1599 | 0.0107 | | |
| WICHH | Indicator if anyone in household receiving WIC benefits | 0.0638 | 0.0059 | | |

*Note:* This table contains summary statistics for all variables used in the DAGs. Means are weighted using the Taylor series linearization method from appendix D of the FoodAPS user's guide (USDA, ERS, 2016, p. 55). The variable names are the same as those in DAG figures. The variables with unreported minimums and maximums are indicator variables.*Source*: Calculated by authors.

**Table 2. Recommended Calories per Day**

| | Males | | | Females | | |
|---|---|---|---|---|---|---|
| Age | Sedentary | Mod. Active | Active | Sedentary | Mod. Active | Active |
| 2 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| 3 | 1,000 | 1,400 | 1,400 | 1,000 | 1,200 | 1,400 |
| 4 | 1,200 | 1,400 | 1,600 | 1,200 | 1,400 | 1,400 |
| 5 | 1,200 | 1,400 | 1,600 | 1,200 | 1,400 | 1,600 |
| 6 | 1,400 | 1,600 | 1,800 | 1,200 | 1,400 | 1,600 |
| 7 | 1,400 | 1,600 | 1,800 | 1,200 | 1,600 | 1,800 |
| 8 | 1,400 | 1,600 | 2,000 | 1,400 | 1,600 | 1,800 |
| 9 | 1,600 | 1,800 | 2,000 | 1,400 | 1,600 | 1,800 |
| 10 | 1,600 | 1,800 | 2,200 | 1,400 | 1,800 | 2,000 |
| 11 | 1,800 | 2,000 | 2,200 | 1,600 | 1,800 | 2,000 |
| 12 | 1,800 | 2,200 | 2,400 | 1,600 | 2,000 | 2,200 |
| 13 | 2,000 | 2,200 | 2,600 | 1,600 | 2,000 | 2,200 |
| 14 | 2,000 | 2,400 | 2,800 | 1,800 | 2,000 | 2,400 |
| 15 | 2,200 | 2,600 | 3,000 | 1,800 | 2,000 | 2,400 |
| 16 | 2,400 | 2,800 | 3,200 | 1,800 | 2,000 | 2,400 |
| 17 | 2,400 | 2,800 | 3,200 | 1,800 | 2,000 | 2,400 |
| 18 | 2,400 | 2,800 | 3,200 | 1,800 | 2,000 | 2,400 |
| 19-20 | 2,600 | 2,800 | 3,000 | 2,000 | 2,200 | 2,400 |
| 21-25 | 2,400 | 2,800 | 3,000 | 2,000 | 2,200 | 2,400 |
| 26-30 | 2,400 | 2,600 | 3,000 | 1,800 | 2,000 | 2,400 |
| 31-35 | 2,400 | 2,600 | 3,000 | 1,800 | 2,000 | 2,200 |
| 36-40 | 2,400 | 2,600 | 2,800 | 1,800 | 2,000 | 2,200 |
| 41-45 | 2,200 | 2,600 | 2,800 | 1,800 | 2,000 | 2,200 |
| 46-50 | 2,200 | 2,400 | 2,800 | 1,800 | 2,000 | 2,200 |
| 51-55 | 2,200 | 2,400 | 2,800 | 1,600 | 1,800 | 2,200 |
| 56-60 | 2,200 | 2,400 | 2,600 | 1,600 | 1,800 | 2,200 |
| 61-65 | 2,000 | 2,400 | 2,600 | 1,600 | 1,800 | 2,000 |
| 66-70 | 2,000 | 2,200 | 2,600 | 1,600 | 1,800 | 2,000 |
| 71-75 | 2,000 | 2,200 | 2,600 | 1,600 | 1,800 | 2,000 |
| 76 & up | 2,000 | 2,200 | 2,400 | 1,600 | 1,800 | 2,000 |

*Note:* Estimates based on Estimated Energy Requirements (EER) equations, using reference heights (average) and reference weights (healthy) for each age-sex group. For children and adolescents, reference height and weight vary. For adults, the reference man is 5 feet 10 inches tall and weighs 154 pounds. The reference woman is 5 feet 4 inches tall and weighs 126 pounds. Source: *2015-2020 Dietary Guidelines for Americans*, p. 77-78.

**Table 3. Recommended Dietary Pattern by Calorie Requirement**

| Calories | 1,000 | 1,200 | 1,400 | 1,600 | 1,800 | 2,000 | 2,200 | 2,400 | 2,600 | 2,800 | 3,000 | 3,200 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Food Group* | | | | | | | | | | | | |
| Vegetables (c-eq/day) | 1 | 1.5 | 1.5 | 2 | 2.5 | 2.5 | 3 | 3 | 3.5 | 3.5 | 4 | 4 |
| Fruits (c-eq/day) | 1 | 1 | 1.5 | 1.5 | 1.5 | 2 | 2 | 2 | 2 | 2.5 | 2.5 | 2.5 |
| Grains (oz-eq/day) | 3 | 4 | 5 | 5 | 6 | 6 | 7 | 8 | 9 | 10 | 10 | 10 |
| Dairy (c-eq/day) | 2 | 2.5 | 2.5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| Protein (oz-eq/day) | 2 | 3 | 4 | 5 | 5 | 5.5 | 6 | 6.5 | 6.5 | 7 | 7 | 7 |
| Oils | 15 | 17 | 17 | 22 | 24 | 27 | 29 | 31 | 34 | 36 | 44 | 51 |

*Note:* The Healthy U.S.-Style Pattern is based on the types and proportions of foods Americans typically consume, but in nutrient-dense forms and appropriate amounts. It is designed to meet nutrient needs while not exceeding calorie requirements and while staying within limits for overconsumed dietary components.
Source: *2015-2020 Dietary Guidelines for Americans*, p. 80-82.

**Table 4. Parameter Estimates (Partial Values) for each Edge and Associated Significance**

| Edge | | Partial | | | |
|---|---|---|---|---|---|
| From | To | Value | Std. Error | t-stat | p-value |
| Age | FairPoorHealthStatus | 0.0042 | 0.0001 | 28.1328 | 0.0000 |
| Age | TooFewFruitVeges | -0.0014 | 0.0002 | -8.1234 | 0.0000 |
| Age | WICHH | -0.0014 | 0.0001 | -9.7968 | 0.0000 |
| Asian | FF5 | 91.3143 | 3.1597 | 28.8997 | 0.0000 |
| Asian | NONFF5 | 234.753 | 11.0775 | 21.192 | 0.0000 |
| Asian | Obese | -0.1704 | 0.0195 | -8.757 | 0.0000 |
| Asian | SNAPnowHH | -0.1233 | 0.0192 | -6.4236 | 0.0000 |
| Asian | SSSM5 | -8.131 | 1.0382 | -7.8319 | 0.0000 |
| Black | Pr_Dairy | -0.0992 | 0.0153 | -6.4799 | 0.0000 |
| Black | Pr_Veges | -0.0581 | 0.008 | -7.2804 | 0.0000 |
| Black | Rural | -0.1449 | 0.0094 | -15.472 | 0.0000 |
| Black | SNAPnowHH | 0.1481 | 0.0106 | 13.9842 | 0.0000 |
| CollegeGrad | FairPoorDietStatus | -0.0926 | 0.0111 | -8.3485 | 0.0000 |
| CollegeGrad | FairPoorHealthStatus | -0.0587 | 0.0094 | -6.2261 | 0.0000 |
| CollegeGrad | LowFoodSecurity | -0.0752 | 0.0115 | -6.5408 | 0.0000 |
| CollegeGrad | Pr_Fruit | 0.0837 | 0.0094 | 8.9368 | 0.0000 |
| CollegeGrad | SNAPnowHH | -0.0788 | 0.0119 | -6.6141 | 0.0000 |
| Employed | Pr_Fruit | -0.0566 | 0.0064 | -8.7944 | 0.0000 |
| Employed | SNAPnowHH | -0.0968 | 0.008 | -12.066 | 0.0000 |
| FairPoorDietStatus | FairPoorHealthStatus | 0.259 | 0.0072 | 36.0712 | 0.0000 |
| FairPoorDietStatus | Obese | 0.0621 | 0.0093 | 6.6774 | 0.0000 |
| FairPoorDietStatus | TooFewFruitVeges | 0.1197 | 0.0087 | 13.7721 | 0.0000 |
| FairPoorHealthStatus | Obese | 0.1986 | 0.0105 | 18.8831 | 0.0000 |
| Female | Pr_Grain | -0.1713 | 0.0135 | -12.708 | 0.0000 |
| FF5 | NONFF5 | 5.8746 | 0.0329 | 178.792 | 0.0000 |
| FF5 | Pr_Dairy | -0.0005 | 0.0001 | -8.7007 | 0.0000 |
| FF5 | Pr_Fruit | 0.0004 | 0.0000 | 9.6697 | 0.0000 |
| FF5 | Rural | -0.0026 | 0.0000 | -56.883 | 0.0000 |
| FF5 | SSSM5 | -0.1524 | 0.0056 | -27.288 | 0.0000 |
| HHsize | FF5 | 2.3362 | 0.3138 | 7.4449 | 0.0000 |
| HHsize | NONFF5 | -12.769 | 1.1098 | -11.505 | 0.0000 |
| HHsize | Pr_Protein | -0.0387 | 0.0028 | -13.696 | 0.0000 |
| HHsize | Pr_Veges | -0.0111 | 0.0015 | -7.5314 | 0.0000 |
| HHsize | SNAPnowHH | 0.0283 | 0.002 | 14.3775 | 0.0000 |
| HHsize | SSSM5 | 0.7353 | 0.1029 | 7.144 | 0.0000 |
| HHsize | WICHH | 0.0399 | 0.0016 | 25.3534 | 0.0000 |
| Hispanic | FairPoorDietStatus | 0.0734 | 0.0088 | 8.3868 | 0.0000 |

*(continued)*

**Table 4. *(continued)***

| Edge | | Partial | | | |
|---|---|---|---|---|---|
| From | To | Value | Std. Error | t-stat | p-value |
| Hispanic | FF5 | 61.2195 | 1.4495 | 42.2342 | 0.0000 |
| Hispanic | LowFoodSecurity | 0.0625 | 0.0087 | 7.1603 | 0.0000 |
| Hispanic | Rural | -0.1241 | 0.0083 | -14.992 | 0.0000 |
| Hispanic | SSSM5 | 7.6242 | 0.5052 | 15.0926 | 0.0000 |
| Hispanic | WICHH | 0.0428 | 0.0066 | 6.4883 | 0.0000 |
| LowFoodSecurity | FairPoorDietStatus | 0.1812 | 0.0083 | 21.7309 | 0.0000 |
| LowFoodSecurity | FairPoorHealthStatus | 0.0778 | 0.0072 | 10.7815 | 0.0000 |
| LowFoodSecurity | TooFewFruitVeges | 0.0765 | 0.0085 | 8.9979 | 0.0000 |
| NONFF5 | SSSM5 | 0.0894 | 0.0008 | 113.942 | 0.0000 |
| PctPovGuideHH | LowFoodSecurity | -0.0004 | 0.0000 | -21.077 | 0.0000 |
| PctPovGuideHH | Obese | -0.0001 | 0.0000 | -7.061 | 0.0000 |
| PctPovGuideHH | Pr_Fruit | 0.0001 | 0.0000 | 7.3627 | 0.0000 |
| PctPovGuideHH | SNAPnowHH | -0.0006 | 0.0000 | -36.001 | 0.0000 |
| PctPovGuideHH | WICHH | -0.0001 | 0.0000 | -8.2187 | 0.0000 |
| Pqa_dairy | Pqa_protein | 1.7695 | 0.0167 | 106.008 | 0.0000 |
| Pqa_dairy | Pr_Oils | 3.1553 | 0.4354 | 7.2469 | 0.0000 |
| Pqa_protein | Pqa_grain | 0.0257 | 0.002 | 13.1601 | 0.0000 |
| Pqa_veget | Pqa_oils | 0.8572 | 0.0323 | 26.579 | 0.0000 |
| Pr_Calrs | Pr_Oils | 0.2089 | 0.0095 | 21.9885 | 0.0000 |
| Pr_Calrs | Pr_Veges | 0.0191 | 0.0027 | 7.0174 | 0.0000 |
| Pr_Dairy | Pqa_oils | 0.0084 | 0.0007 | 11.4947 | 0.0000 |
| Pr_Dairy | Pr_Calrs | 0.2925 | 0.0204 | 14.3622 | 0.0000 |
| Pr_Dairy | Pr_Fruit | 0.1469 | 0.0068 | 21.655 | 0.0000 |
| Pr_Dairy | Pr_Grain | 0.7551 | 0.0134 | 56.3933 | 0.0000 |
| Pr_Dairy | Pr_Protein | 0.2324 | 0.0122 | 18.9823 | 0.0000 |
| Pr_Dairy | Pr_Veges | 0.0619 | 0.0064 | 9.7174 | 0.0000 |
| Pr_Fruit | FairPoorDietStatus | -0.0763 | 0.0097 | -7.8721 | 0.0000 |
| Pr_Fruit | LowFoodSecurity | -0.0669 | 0.0096 | -6.9373 | 0.0000 |
| Pr_Fruit | TooFewFruitVeges | -0.0909 | 0.0098 | -9.311 | 0.0000 |
| Pr_Fruit | WICHH | 0.0493 | 0.0071 | 6.9252 | 0.0000 |
| Pr_Grain | Pr_Calrs | 0.4879 | 0.0124 | 39.401 | 0.0000 |
| Pr_Grain | Pr_Fruit | 0.062 | 0.0042 | 14.7555 | 0.0000 |
| Pr_Grain | Pr_Protein | 0.2957 | 0.0071 | 41.6817 | 0.0000 |
| Pr_Grain | Pr_Oils | 0.5483 | 0.0139 | 39.5445 | 0.0000 |
| Pr_Grain | Pr_Veges | 0.1088 | 0.0042 | 25.6402 | 0.0000 |
| Pr_Protein | Pr_Calrs | 0.2825 | 0.0141 | 19.9916 | 0.0000 |
| Pr_Protein | Pr_Oils | 0.3012 | 0.0157 | 19.2007 | 0.0000 |
| Pr_Protein | Pr_Veges | 0.1266 | 0.0045 | 28.1305 | 0.0000 |

*(continued)*

**Table 4. *(continued)***

| Edge | | Partial | | | |
|---|---|---|---|---|---|
| From | To | Value | Std. Error | t-stat | p-value |
| Pr_Oils | Pqa_oils | 0.0051 | 0.0003 | 19.3807 | 0.0000 |
| Pr_Oils | Pr_Veges | 0.0296 | 0.0024 | 12.2563 | 0.0000 |
| Pr_Veges | Pqa_veget | -0.0019 | 0.0002 | -7.8099 | 0.0000 |
| Pr_Veges | Pr_Fruit | 0.207 | 0.0089 | 23.2745 | 0.0000 |
| Rural | NONFF5 | 123.179 | 5.6173 | 21.9284 | 0.0000 |
| Rural | SSSM5 | -7.4742 | 0.5199 | -14.376 | 0.0000 |
| SNAPnowHH | FairPoorHealthStatus | 0.0655 | 0.0068 | 9.6373 | 0.0000 |
| SNAPnowHH | LowFoodSecurity | 0.1582 | 0.0084 | 18.9293 | 0.0000 |
| SNAPnowHH | Pr_Veges | -0.0504 | 0.006 | -8.4163 | 0.0000 |
| SNAPnowHH | WICHH | 0.0863 | 0.0062 | 13.8887 | 0.0000 |
| SSSM5 | TooFewFruitVeges | -0.0006 | 0.0001 | -7.4703 | 0.0000 |
| White | FF5 | -22.468 | 1.353 | -16.606 | 0.0000 |
| White | Pr_Dairy | 0.0985 | 0.0121 | 8.1646 | 0.0000 |

*Source*: Calculated by authors.