### Food Policy 61 (2016) 163-175

Contents lists available at ScienceDirect

**Food Policy** 

journal homepage: www.elsevier.com/locate/foodpol

# Food environment in the United States as a complex economic system



Department of Agricultural Economics, Texas A&M University, USA

### ARTICLE INFO

Article history Received 9 September 2013 Received in revised form 3 March 2016 Accepted 28 March 2016 Available online 6 April 2016

JEL classification: C40 D83 D85 I18

Keywords: Food environment Obesity Poverty Food insecurity Food assistance Directed acyclic graphs

# 1. Background information

# The food environment is complex. According to the Economic Research Service of the United States Department of Agriculture, food choices, health and well-being as well as community characteristics and various factors affecting these as a whole are defined as the food environment (USDA, 2014). It includes all physical and social aspects that influence what, where, how, and when we consume our food as well as government policies influencing food production, food prices, food taxes, and food marketing. The food environment also is affected by other economic factors such as income, unemployment, poverty, food insecurity, and food assistance programs. Nutrition and obesity also are contributory factors to the food environment (Capps, 2009).

Food environment factors such as proximity to stores/restaurants, food prices, food and nutrition assistance programs, and community characteristics interact to influence food choices and diet quality (USDA, 2014). More specifically, indicators of food choices such as access to and proximity of a grocery store, number

# ABSTRACT

The food environment in the United States is complex. Sixteen socio-economic-demographic variables from various public data sources are studied with a machine learning algorithm to ascertain the causality structure associated with the food environment in the United States. High levels of unemployment and poverty are direct causes of high levels of food insecurity, while low income causes high levels of food insecurity via increased levels of poverty. Unemployment is a common cause for both increased levels of food insecurity and poverty. We find that food insecurity and participation in Supplemental Nutrition Assistance Program (SNAP) are related, yet no direct causality is observed. Contrary to past studies which find that SNAP participation decreased the occurrences of poverty, in contemporaneous time, we find that poverty and SNAP participation are related through several back-door paths, via food insecurity, unemployment, race and food taxes. Obesity and SNAP participation are indirectly related via several back-door paths, namely, race income, poverty and food insecurity and unemployment. Also, food insecurity and obesity are related by several back-door paths. Low income, high food taxes, and race (being Black and non-Hispanic) are direct causes of obesity. The complex causality structure in the US food environment reveals that policy variables cannot be treated independently of their rich causal structure. Government agencies responsible for designing policies for food assistance, poverty alleviation, combating food insecurity and obesity need to consider the interrelationships among these variables.

© 2016 Elsevier Ltd. All rights reserved.

of food stores and restaurants, expenditures on food away from home, participation in food and nutrition assistance programs, food prices, food taxes and availability of local foods are important factors contributing to the food environment. As far as the health and well-being of a community's food environment is concerned, food insecurity, presence of food deserts, adult and childhood obesity and physical activity levels are of great concern (USDA, 2014). Additionally, there may be other characteristics such as demographic composition, income, poverty status, and availability of recreation and fitness centers that may impact the food environment. In addition to the aforementioned factors affecting the food environment, other factors such as macroeconomic shocks (unemployment, interest rates, inflation, mortgage crises, divorce or separation, disability), asset availability and liquidity, food prices and government support programs for agricultural commodities may influence the food environment of communities (Dharmasena et al., 2014a).

Research is beginning to emerge documenting the complexity and interaction of factors affecting the food environment. However, given the complexity of the interaction among the aforementioned variables, more research is necessary to identify causal relationships among these factors, thereby providing the paths for effective policy interventions (USDA, 2014). Even though we find several studies in the extant literature addressing issues





OOD POLICY

<sup>\*</sup> Corresponding author at: Department of Agricultural Economics, TAMU 2124, Texas A&M University, College Station, TX 77843-2124, USA.

E-mail addresses: sdharmasena@tamu.edu (S. Dharmasena), d-bessler@tamu. edu (D.A. Bessler), ocapps@tamu.edu (O. Capps Jr.).

related to food insecurity, food deserts, food assistance, health and other factors of food environment, these studies have considered only few variables at a time in piecemeal fashion, hence not giving the holistic picture of the complex economics of the food environment (see for example, Wolf and Colditz, 1998; Casey et al., 2001, 2006; Jyoti et al., 2005; Dubois et al., 2006; Yen et al., 2008; Finkelstein et al., 2009; Nord and Golla, 2009; Dixon, 2010; Cawley and Meyerhoefer, 2012; Gundersen et al., 2011a; Tiehen et al., 2012). As a result, in the absence of complete picture pertaining to factors affecting the food environment, the design of appropriate food and nutrition policies can be adversely affected.

We discuss several studies addressing different components or variables of the food environment in the United States, centering attention on four food and nutrition policies. They are food security (or insecurity), food assistance (Supplemental Nutrition Assistance Program, SNAP), poverty and obesity. The articles are chosen with regard to addressing various food environment-related variables that are affecting or affected by aforementioned four policy variables. This selected literature provides a representation of food environment-related issues albeit not an exhaustive treatment.

#### 1.1. Food insecurity and food assistance programs

Coleman-Jensen et al. (2012) reported that in 2011, 14.9% of U.S. households were food insecure, and 5.4% households had very low food security. Furthermore, 57% of food-insecure households reported that they received assistance from one or more of the three largest Federal food and nutrition assistance programs (the Supplemental Nutritional Assistance Program, SNAP; the National School Lunch Program, NSLP; and the Supplemental Nutrition Assistance Program for Women, Infants and Children, WIC). Also, they found that food insecurity varied among households with different economic and demographic characteristics, such as the prevalence of high food insecurity with households with low income, all households with children and specifically those with children under six years of age, households with children headed by a single woman, and Black and Hispanic households.

Wilde and Nord (2005) showed that higher incomes reduced food insecurity as expected; however they did not find clear evidence to support the notion that households' participation in food assistance programs decreased food insecurity. Gundersen et al. (2011a,b) reported that information beyond poverty status is important for understanding food insecurity. Other factors possibly determining food insecurity are unemployment rate, income, race and ethnic origin. Gundersen and Gruber (2001) and Leete and Bania (2010) found that households with no liquid assets were substantially more likely to be food insecure compared to those with liquid assets.

There are numerous consequences of food insecurity in the United States. A large body of literature has focused on correlational relationships rather than on causal relationships of food insecurity in attempting to explain various outcomes (Gundersen et al., 2011a). Gundersen et al. (2011a, p. 289) provided an extensive review of correlational studies relating food insecurity to health outcomes. Furthermore, Gundersen et al. (2011a) stated that, while food insecurity may cause health problems, the host of potential unobserved variables between food insecurity and poor health could be a problem trying to model causality effects of food insecurity and health outcomes. For example, food insecurity itself may not be directly contributing to worse health outcomes; however, other factors associated with households such as low income may be a contributory factor to worse health outcomes along with food insecurity. Also Gundersen et al. (2011a) explained possible effects of SNAP and NSLP in dealing with food insecurity in the United States. Economists have suggested that participation in the SNAP is likely to be endogenous, hence regressing food insecurity

on SNAP, not controlling for unobserved variables, potentially would lead to spurious empirical results. Identification of proper instruments for SNAP participation has been challenging; however, Gundersen and Oliveira (2001) were able to take care of this endogeneity issue, identifying variables that control the selection into SNAP and food insecurity. The participation in the NSLP in alleviating food insecurity faces similar problems associated with endogeneity of selection into and participation in this program. However, after controlling for selection and measurement error problems, Gundersen et al. (2012) found evidence that the NSLP substantially affects reducing food insecurity. Furthermore, acting on the premise that SNAP decreases food insecurity, Gundersen et al. (2009) found that increases in participation in SNAP decreased the relative well-being of program participants.

Yen et al. (2008) estimated the relationship between SNAP participation and household food insecurity to find that participation in SNAP reduces the severity of food insecurity. Nord and Golla (2009) measured U.S. household food insecurity before and after participation in the SNAP to discover that food insecurity declines with household's participation in the SNAP (households self-select into SNAP when they are severely food insecure). Although all income-poor households did not participate in food assistance programs (self-selection issues). D. Ribar and Hamrick (2003) and D.C. Ribar and Hamrick (2003) found that income-poverty had a direct negative relationship with food sufficiency. For the same reason, Bartfeld and Dunifon (2006) concluded that it was difficult to document the relationship between food assistance programs and their impact on food security, although a decline in food insecurity was observed when such programs are in place.

Jensen (2002) discussed similar effects and relationships between income and food security (insecurity) as well as between food expenditure of low-income households and its relation to poverty. Furthermore, Jensen (2002) discussed household participation in U.S government food assistance programs (like SNAP) and its relationship to food insecurity. Another extensive body of literature, represented by Gundersen and Kreider (2008), Bhattacharya et al. (2004), Bitler et al. (2005), Borjas (2004), Furness et al. (2004), Gundersen (2008), Gundersen et al. (2003), Laraia et al. (2006), D. Ribar and Hamrick (2003), D.C. Ribar and Hamrick (2003), Van Hook and Balistreri (2006), found that participation in food assistance programs decreased food insecurity, particularly in adults, but not in children as found by Bhattacharya et al. (2004). Olson et al. (2004) investigated factors contributing to food insecurity especially centering attention on rural communities of the United States.

#### 1.2. Food assistance programs and poverty

Tiehen et al. (2012) found on average a 4.4% decline in the prevalence of poverty as a result of SNAP, while the decline in depth and severity of poverty was on average 10.3% and 13.2% respectively. They concluded that the SNAP significantly improved the welfare of low-income households. Additionally, Bitler and Haider (2011) explained the linkage between low income and food assistance and the role of food assistance in alleviating problems associated with such low incomes. Case et al. (2002) explained the relationship of household income to health, more specifically with respect to children's health, finding that children from lower-income households with chronic conditions had worse health outcomes than did those from higher-income households.

# 1.3. Obesity and food insecurity

Finkelstein et al. (2005) discussed economic causes and consequences of the obesity epidemic in the United States. According to them, reduction of body energy expenditure (as a result of technological advancements in the workplace and lack of participation in exercise activities), intake of calorie dense foods, rising real income and easy accessibility to cheap calorie dense food, were contributory factors for obesity epidemic in the United States. Furthermore, consequences of the obesity epidemic were found to be increasing obesity-related medical costs/expenditures, absenteeism and low productivity in the workplace, and loss of quality of life-related issues. Gundersen and Kreider (2009) found that food security had statistically significant positive impact on children's health. Additionally, Casey et al. (2001, 2006), Dubois et al. (2006) and Jyoti et al. (2005) found evidence to support a positive relationship between food insecurity and (childhood) obesity.

### 1.4. Obesity and food assistance programs

Cawley and Meyerhoefer (2012), Dixon (2010), Finkelstein et al. (2009) and Wolf and Colditz (1998) found that participation in food assistance programs led to increase in obesity rates in the United States. Additionally, Ver Ploeg and Ralston (2008) found that the SNAP may have contributed to high rates of obesity among below-poverty households or individuals in the United States.

Chen et al. (2012) investigated the effect of distance to fast-food restaurants on obesity to find that decreasing access to fast-food lowers BMI by a statistically significant, but economically small, amount. These findings complemented findings of similar studies which investigated relationship between fast-food consumption and obesity in the United States (Dunn, 2010; Currie et al., 2004).

Another aspect of the food environment is the distance food travels before being purchased by consumer (distance to supermarket and/or farmer's market) and its relation to prices consumers pay (or are willing to pay) for different foods as well as the poverty status of consumers. Grebitus et al. (2013) found that the average willingness-to-pay for apples and wine decreased related to distance traveled. Broda et al. (2009) showed that poor consumers paid less for food (or goods in general) they consumed, compared to what their richer counterparts paid for similar food (or goods).

All aforementioned studies, centering attention to food insecurity, food assistance, obesity and poverty, considered only a limited number of variables (affecting or affected by) in a piecemeal fashion; hence, a holistic picture of the "food environment" as a complex economic system has not yet been put forward. Consequently, public policies generated on the basis of limited information may be suboptimal, hence potentially misleading. To improve policy-making, it is imperative to develop a more complete understanding of the food environment in the United States. We recognize that Dharmasena et al. (2014a) and Dharmasena et al. (2014b) have investigated similar questions using similar procedures in the past in understanding food environment variables in the United States. These are unpublished conference papers. However, to best of our knowledge, work investigated in this manuscript is the first *published work* in the literature to study causality structures of food environment factors modeled through artificial intelligence and directed acyclic graphs (DAGs), recognizing its complexity. Given this background, the specific objectives of this study are: (1) to model the food environment in the United States as a complex economic system using causality structures developed through DAGs, and (2) to discuss policy analysis as informed by graphical causal structures gleaned from objective 1.

Using data from various public sources pertaining to 16 socioeconomic-demographic variables, and graphical causal structures based on artificial intelligence and DAGs, we estimated causality relationships among variables determining the food environment of the United States. In particular, we are interested in uncovering causality structures pertaining to four important nutrition and food policy variables in the United States. They are policies related to obesity, poverty, food insecurity and participation in food assistance programs. We find that being unemployed and/or poor increased the incidence of food insecurity. Furthermore, being low-income, unemployed, having high food taxes, being Hispanic and non-White increased the occurrence of being poor. Increased participation in food assistance programs (SNAP) was observed for those who are unemployed, non-Hispanic, and face high food taxes. Being low income, Black, non-Hispanic and having high food taxes increased the incidence of being obese. Furthermore, low income caused SNAP participation to go up via increased unemployment and low income also caused an increase in food insecurity via high incidence of poverty.

Participation in SNAP and poverty are related vis-à-vis two back-door paths, i.e. via unemployment and food taxes. Obesity and expenditure food away from home are related in a back-door path via income. Our analysis of this complex causality structure reveals that policy variables cannot be dealt with independently. Government agencies responsible for designing policies for food assistance, poverty, food insecurity and obesity need to consider the complex interrelationships among these policy variables.

This paper is organized as follows. Section 2 presents the analytical framework followed by Section 3, dealing with data and variables used in this study. Section 4 is devoted to estimation followed by Section 5 concerning empirical results and discussion. Concluding remarks, implications and suggestions for future work follow in Section 6. The methodology behind graphical causal structures is discussed in Appendix.

#### 2. Analytical framework<sup>1</sup>

Causal relationships come from recent efforts in computer science (Pearl, 1995, 2000; Spirtes et al., 2000). One reason for studying causal models, represented here as  $X \rightarrow Y$ , is to predict the consequences of changing the effect variable (Y) by changing the cause variable (X). Manipulating Y by way of manipulating X is at the heart of causation. Hausman (1998, page 7) writes: "Causation seems connected to intervention and manipulation: one can use causes to 'wiggle' their effects." Prediction of such 'wiggling' through intervention in food and nutrition policy variables is what gives us hope for improving the appropriate policy design to address food insecurity, poverty, obesity and food assistance in the United States. A manipulation-based definition of causation is generally more in line with a philosophical definition of causality than other definitions that focus exclusively on predictability. For example, Bunge (1959) argues that causality requires a productive or genetic principle that describes "how something comes into being." X causes Y if X is productive of Y. Definitions that focus on prediction alone, without distinguishing between intervention and subsequent realization, may mistakenly label cause variables that are associated only through an omitted variable. For example, Granger-type causality (Granger, 1980) focuses solely on prediction, without considering intervention and manipulation. Here we follow work of Spirtes et al. (2000) and Pearl (2000), in representing causal structures in graphical form and using separation notions to help identify causal flows with observational data.

Three types of directed acyclic graphs can be identified as follows. They are "causal chains", "causal forks" and "inverted causal forks". Let us assume three variables, *A*, *B*, and *C* to be related as follows,  $A \rightarrow B \rightarrow C$ . This relation is a causal chain. It is read as "A causes *B* and *B* causes *C*" (or *A* causes *C* via *B*). If  $\rho$  represents correlation between variables, then for a causal chain,  $\rho(A, C) \neq 0$ ;  $\rho(A, C|B) = 0$ . That is to say, unconditional correlation between *A* and *C* is not equal to zero (*A* and *C* are related). However, the conditional correlation between *A* and *C* is zero, meaning that additional information about *B* makes *A* 

<sup>&</sup>lt;sup>1</sup> This discussion follows Bessler (2002).

and C independent. Below in our results section we find this causal chain: income  $\rightarrow$  poverty  $\rightarrow$  food insecurity. So income and food insecurity are unconditionally related (their correlation is not zero). However, if we condition on the middle variable (poverty), income and food insecurity are unrelated, as poverty captures the income effect on food insecurity.

If *X*, *Y*, and *Z*, are related as follows,  $Y \leftarrow X \rightarrow Z$ , then it would give rise to a causal fork. This relation is read as "*X* causes *Y*" and "*X* causes *Z*" or "*Y* and *Z* are caused by *X*, meaning that *X* is a "common cause". Again, given  $\rho$  represents correlation between variables, then for a causal fork,  $\rho(Y,Z) \neq 0$ ;  $\rho(Y,Z|X) = 0$ . This means that conditioning on a common cause (i.e. *X*) screens off the association between joint effects (joint effects are *Y* and *Z*). As we explain below in the results section, unemployment, food insecurity and participation in food assistance programs (i.e. SNAP) are related to depict a causal fork: food insecurity  $\leftarrow$  unemployment  $\rightarrow$  SNAP. As such, food insecurity and SNAP are related (their correlation is not zero), however, conditioning on unemployment makes food insecurity and SNAP not related. This is because unemployment now explains variation in SNAP and food insecurity, thereby becoming the common cause.

If *P*, *Q*, and *R*, are related as follows, it is called an inverted causal fork,  $P \rightarrow R \leftarrow Q$ . Thus would result in the following unconditional and conditional correlations  $\rho(P,Q) = 0$ ;  $\rho(P,Q|R) \neq 0$ ; This means that *P* and *Q* are not related, however conditioning on the extra information about *R* makes *P* and *Q*, related. In other words, conditioning on a collider (*R* in this case, the common effect) opens up a path between its causes. As explained below in the results section, food insecurity, unemployment and poverty are related as follows to make an inverted causal fork: unemployment  $\rightarrow$  food insecurity  $\leftarrow$  poverty. Here, unemployment and poverty are not related, however conditioning on food insecurity makes unemployment and poverty related. That is to say, food insecurity is explained both by unemployment and poverty, making unemployment and poverty related, once we know food insecurity<sup>2</sup>.

## 3. Data

Data from various public sources (such as USDA Food Environment Atlas, United States Census Bureau, Center for Disease Control and Prevention, United States Bureau of Labor Statistics, USDA Food and Nutrition Service; see Table 1 below for more information about data sources) pertaining to the 48 contiguous states of the United States is used in our analysis. This database contains a host of broad categories of variables affecting the food environment of the United States, and they are categorized into several major groups (see Table 1, column 3). These groups are given as follows: health, socio-economic character, food assistance, food prices, food taxs, availability of food stores, food away from home expenditures, and food insecurity. The variables we employed are (see Table 1, column 1), percentage of obese adults, poverty rate, median household income, participation rate in food assistance programs (SNAP), soda prices, milk prices, soda taxs, number of grocery stores, per capita expenditures at fast-food restaurants, per capita expenditures at full-service restaurants, percentage of Whites, percentage of Blacks, percentage of Hispanics, percentage of Asians, unemployment rate and percentage of food insecure households. In this study, we use state level data<sup>3</sup> of 48 observations per variable for 16 food environment variables. Table 1 provides the food environment variables used in this study along with a brief description. The geographic level at which the variables are measured represent calendar years 2008–2010 for all 16 variables used in this study, hence all data variables are contemporaneous and will capture this effect. Table 2 shows summary statistics of each variable.

During the period of this study (2008-2010), on average, the percentage of obese adults in the lower 48 contiguous states of the United States was about 27%. The percentage of households under poverty was about 14% on average in the same time period. Food insecure households, based on food insecurity without hunger (diets lacked variety and quality, however, food intake was not reduced) on average per state was just under 14%. The annual average unemployment rate across the 48 contiguous states was between 7% and 8%. In 2008-2010, Whites comprised about 82% of the population of the United States, while Blacks and Hispanics were each 11% and 10%, respectively. On average, 3% of the population was Asian. Average household income was \$50,056, and the average SNAP participation rate as a percentage of eligible persons was about 74%. The CPI for soda and milk was 1.00 and 0.98 respectively, while the average soda tax was about 3%. The average number of grocery stores per state was 1326. The average per capita fast-food expenditure was about \$389 per year per state, while the average per capita expenditure on full-service restaurants was about \$463 per year per state.

#### 4. Finding graph structures

As indicated above, to model the food environment factors in the United States as a complex economic system using causality structures developed through a machine learning algorithm, initially we develop the correlation matrix of variables (or variance-covariance matrix of variables under consideration). Table 3 shows the lower-diagonal correlation matrix of the aforementioned variables (16 variables considered in this study; see the data section for a list of the variables).

The machine learning algorithm labeled the Greedy Equivalence Search (GES) (Chickering, 2002) is one of several offerings under the umbrella TETRAD IV (Glymour et al. (2014)) offered by research workers in the Department of Philosophy at Carnegie Mellon University. The GES algorithm is a search algorithm that involves two phases. In the first phase the algorithm attempts to add edges (connections between variables) from equivalent classes (explained below) starting from a DAG representation with no edges (a DAG with no edges implies that all variables are independent of all the other variables). The GES algorithm proceeds in stepwise fashion attempting to added edges, scoring each graph using the Bayesian Information Criterion (BIC) metric. After scoring comparisons among all possible equivalence classes, the one equivalence class that increases the score most is chosen for the next step. This procedure is repeatedly conducted until no further replacement increases the score. Once a local maximum is reached in the first phase, the second phase begins by deleting a single edge and comparing the scores of DAG in equivalence classes repeatedly until a local maximum is reached. When the algorithm reaches a local maximum, it obtains the optimal solution. The algorithm is described in Chickering (2002). The GES algorithm (Chickering, 2002; Glymour et al., 2014) is run several iterations to find the optimum causal graph structure based on minimizing the Bayesian Information Criteria (BIC) for each graph. In this analysis, the minimum BIC for the entire graphical structure for the best model was -112.8093. Also, we looked at the significance levels of edges connecting variables (see Table 4) in deciding the best graphical structure; a 95% significance level is considered in this study (p-value = 0.05) to reject (fail to reject) the null hypothesis that edges are statistically different from zero.

<sup>&</sup>lt;sup>2</sup> As will be demonstrated below, the relationships between unemployment and poverty is a bit more complicated than just explained in an inverted fork.

<sup>&</sup>lt;sup>3</sup> The Food Environment Atlas has county-level data for some variables; however some of these data do not show enough variability at the county level. That is, some county level data are simply repetitions of state level mean (average) values. As a result, we did not use county level data for our work. Therefore, we are not able to capture intra-state variation as pointed out by a reviewer.

# Table 1 Names, category and description of variables used in the study. Source: Please refer to the "Data Sources for Variable" column.

Variable name	Data source for variable	Variable category based on USDA Food Environment Atlas Classification	Variable description
Obese adults (%)	Centers for Disease Control and Prevention (CDC), USDA Food Environment Atlas	Health	Percentage of obese adults of non-institutionalized adult population, aged 20 and older per state
Poverty (%)	United States Census Bureau, USDA Food Environment Atlas	Socio-economic character	Percentage of poor households per state based on poverty threshold calculated using money income, size of the family and ages of family members
Household income (\$)	United States Census Bureau, USDA Food Environment Atlas	Socio-economic character	Inflation adjusted real household income measured at the median per state
SNAP participation rate (%)	USDA Food and Nutrition Service	Food assistance	State level SNAP participation rate as a percentage of all eligible persons
Soda prices (index)	United States Bureau of Labor Statistics, USDA	Food price	Annual Consumer Price Index for carbonated beverages
	Food Environment Atlas		(base 1982–1984 = 100)
Milk prices (index)	United States Bureau of Labor Statistics, USDA Food Environment Atlas	Food price	Annual Consumer Price Index for milk (base 1997 = 100)
Soda taxes (%)	USDA Food Environment Atlas	Food tax	Soda sales tax at retail stores
Grocery stores (#)	USDA Food Environment Atlas	Availability of food stores	Number of grocery stores per state
Fast-food expenditures (\$/person/state/year)	Consumer Expenditure Survey (CES) by United States Bureau of Labor Statistics	Food away-from-home expenditure	Per Capita Fast-food Restaurant Expenditure per state per year
Full-service restaurant expenditures (\$/person/state/year)	Consumer Expenditure Survey (CES) by United States Bureau of Labor Statistics	Food away-from-home expenditure	Per Capita Full Service Restaurant Expenditure per state per year
White (%)	United States Census Bureau, USDA Food Environment Atlas	Socio-economic character	Percentage of Whites
Black (%)	United States Census Bureau, USDA Food Environment Atlas	Socio-economic character	Percentage of Black
Hispanic (%)	United States Census Bureau, USDA Food Environment Atlas	Socio-economic character	Percentage of Hispanic population
Asian (%)	United States Census Bureau, USDA Food Environment Atlas	Socio-economic character	Percentage of Asian population
Unemployment (%)	United States Bureau of Labor Statistics	Socio-economic character	Ratio of the number of individuals who are unemployed to the number of individuals within the labor force
Food insecure households (%)	Current Population Survey of the USDA, and USDA Food Environment Atlas	Food insecurity	Percentage of food insecure households per state; prevalence of both low or very low food security is used (food insecurity without hunger) (diets lacked variety and quality, however, food intake is not reduced)

Note: All data used are available for calendar years 2008–2010.

#### Table 2

Summary	v statistics	of the	variables	across 4	48 States	of the	United	States	used in	n the	e analy	sis. S	Source:	Calculated	by	authors
---------	--------------	--------	-----------	----------	-----------	--------	--------	--------	---------	-------	---------	--------	---------	------------	----	---------

Variable name	Mean	Standard deviation	Minimum	Maximum
Obese adults (%)	27.24	3.10	19.70	34.27
Poverty (%)	13.90	2.93	8.33	21.67
Household income (\$)	50,056	7,910	37,191	96,536
SNAP participation (%)	73.92	10.24	53.00	98.00
Soda prices (CPI Index)	1.00	0.06	0.92	1.20
Milk prices (CPI Index)	0.98	0.12	0.72	1.18
Soda taxes (%)	0.03	0.02	0.00	0.07
Grocery stores (#)	1326	1786	101	9560
Fast-food expenditures (\$/person)	388.96	162.67	131.02	765.63
Full-service restaurant expenditures (\$/person)	463.31	194.59	159.00	872.38
White (%)	82.06	9.52	60.25	95.63
Black (%)	10.98	9.70	0.42	37.21
Hispanic (%)	10.80	10.19	1.22	46.44
Asian (%)	3.03	2.53	0.64	13.55
Unemployment (%)	7.53	1.75	3.80	11.47
Food insecure households (%)	13.86	2.51	7.10	19.40

Note: These are averages over the three year period, 2008-2010.

The BIC approximation from Schwarz is expressed in Eq. (1) below.

$$S(\mathcal{G}, \boldsymbol{D}) = \ln p(\boldsymbol{D}|\hat{\boldsymbol{\theta}}, \mathcal{G}^{h}) - \frac{d}{2}\ln m$$
(1)

where  $\hat{\theta}$  is the maximum-likelihood estimate of the unknown parameters, *d* is the number of free parameters (not equal to zero) of graph  $\mathcal{G}$ , and *m* is the number of observations in data, **D**. The  $S(\mathcal{G}, \mathbf{D})$  function offers a tradeoff between fit given by  $\ln p(D|\hat{\theta}, \mathcal{G}^h)$  and parsimony is given by  $-\frac{d}{2} \ln m$ .

Equivalence class, as described in Chickering (2002) is explained as follows. If two DAGs, G and G' are equivalent, it is said that these two DAGs are in same equivalence class. To satisfy this equivalence class, DAGs must be *distributionally equivalent* as well as *independence equivalent*. Two DAGs are *distributionally equivalent* if the corresponding graphs have same probability distribution structure (or same skeleton and V structure), and DAGs are *independence equivalent* if the independence constraints in the two DAGs are identical. GES algorithm only searches for acyclic graphs by definition (Chickering, 2002, page 510; Verma and Pearl, 1991, Theorem 1).

The TETRAD IV (and the GES algorithm) rests on three assumptions:

- The Causal Sufficiency Condition there is no variable omitted from the study which causes two or more of the variables included in the study.
- (2) The Causal Markov Condition the probability distribution generating our observed data on *n* variables can be written as the product of *n* conditional distributions, where we need only condition on the parents of each variable, not grandparent, aunts, uncles etc.
- (3) The Faithfulness Condition if a zero correlation (or partial correlation) between variables is noted in the data this correlation arises because there is no edge between these variables in the "true" model and not because of cancellations of deeper parameters connecting these variable in the "true" model.

To clarify the identification of the aforementioned causal flows, we model causal structures (directed acyclic graphs or DAGs) among each variable. The GES-Algorithm (Spirtes et al., 2000), is applied to achieve such identification. Applications of these algorithms have become prevalent in recent years following the suggestion of Swanson and Granger (1997), as well as an early

application by Bessler and Akleman (1998) and a more recent application of Rettenmaier and Wang (2013).

#### 5. Empirical results and discussion

In this section, we provide initially a narrative on unconditional correlations (simple correlations) of the 16 variables considered in this study, in particular as it relates to four important nutrition and food policy related variables, namely obesity, poverty, food insecurity and food assistance programs in the United States. Second, we discuss the graphical directed acyclic structure constructed using all 16 variables. Third, we select several subsets of DAGs and compare the DAG-generated causal structures with similar relation-ships observed in the extant literature.

The unconditional correlation matrix of 16 variables considered is presented in Table 3. The prevalence of adult obesity is positively correlated with poverty, race (Black households), unemployment status, and food insecurity status, participation in SNAP, milk prices and soda taxes, while it is negatively correlated with household income, soda prices, expenditures food away from home (fast food restaurants and full-service restaurants), number of grocery stores, race (White households), and Hispanic and Asian racial groups. Poverty is positively correlated with prevalence of adult obesity, participation in SNAP, soda taxes, number of grocery stores, Black and Hispanic racial groups, unemployment status and food insecure households. It is negatively correlated with household income, expenditures at full-service and fast-food restaurants, soda prices, milk prices, and White and Asian racial groups. Participation in food assistance programs, such as SNAP is positively correlated with poverty, obesity, soda prices, the White racial group, unemployment and household food insecurity. SNAP is negatively correlated with Black, Hispanic and Asian households, Fullservice and fast-food restaurants, number of grocery stores, soda taxes and milk prices, and household income. Household food insecurity is positively related to adult obesity, poverty, participation in SNAP, milk prices, soda taxes, number of grocery stores, Black and Hispanic households, and unemployment. It is negatively related to household income, expenditure on full-service and fast-food restaurants, soda prices, and White and Asian racial groups.

Fig. 1 shows the directed acyclic graph developed for 16 food environment variables in the United States considered in this study. The numerical value between each directed edge represents the value of the partial derivative between the variables, when the variable where the arrow head is directed (effect variable) is regressed on the variable where the arrow is coming from (cause

	boo	nsecurity																	
	Jnemployment F	i																_	.4 1
	Asian I																	.32	-0.08
	panic A																2 1	33	4
	s His															2	7 0.5	3 0.2	1 0.2
	Black														1	-0.1	0.0	0.3	0.3
	Whites													1	-0.92	-0.12	-0.34	-0.34	-0.30
	Full serv	exps										1		-0.22	0.16	0.24	0.57	0.21	-0.15
	Fast-food	expenditures									1	0.86		-0.19	0.15	0.22	0.46	0.16	-0.12
	Grocery									1	0.24	0.38		-0.33	0.19	0.39	0.65	0.30	0.16
S.	Soda	taxes							1	0.26	0.18	0.21		-0.12	0.10	0.11	0.25	0.22	0.18
a by author	Milk	prices						1	0.12	0.06	-0.06	0.13		-0.40	0.40	-0.08	0.16	-0.05	0.04
:: calculate	Soda	prices					1	-0.14	-0.20	-0.08	0.10	0.12		0.20	-0.38	0.23	0.26	-0.11	-0.07
y. source	SNAP					1	0.02	-0.02	-0.28	-0.19	-0.17	-0.13		0.21	-0.08	-0.46	-0.36	0.12	0.06
nne arn	Income				1	-0.35	0.13	0.23	0.13	0.20	0.47	0.59		-0.06	-0.08	0.22	0.66	-0.09	-0.54
considered	Poverty			1	-0.84	0.19	-0.13	-0.08	0.05	0.07	-0.31	-0.37		-0.33	0.38	0.09	-0.32	0.35	0.74
n variables i	Obesity		1	0.67	-0.68	0.27	-0.35	0.08	0.09	-0.08	-0.17	-0.31		-0.36	0.50	-0.36	-0.40	0.13	0.43
רסוופומנוטו ווומנווג טו נווב אזגניני			Obesity	Poverty	Income	SNAP	Soda prices	Milk prices	Soda taxes	Grocery	Fast-food expenditures	Full-service restaurant	expenditures	White	Blacks	Hispanic	Asian	Unemployment	Food insecurity

Table

variable). The 0.0000 value represent the *p*-value associated with the estimated partial derivative. Table 4 shows the partial derivative values, standard errors associated with the partial derivative value, and associated *t*-statistic and *p*-value with respect to each directed edge under consideration. In other words, these numbers show the underlying structural model with respect to 16 food environment variables considered in this study. Alternative graphical structures were investigated based on minimizing the Bayesian Information Criteria (BIC). The directed acyclic graph shown in Fig. 1 is associated with the minimum BIC score of -112.8093. For brevity, all alternative graphical structures and associated BIC scores are not reported here, however they are available from authors upon request. As depicted in Fig. 1, the causality structures among these variables are complicated. These causality structures associated with nutrition and food policy variables (obesity, poverty, food insecurity and food assistance) reveal valuable information that can be used in prudent policy making.

Obesity, food insecurity and SNAP variables are strictly endogenous, meaning in graph-theoretic language, all arrows are pointed toward these variables and no arrows are emanating from them. According to the DAG (Fig. 2) and estimates of partials and associated significance levels (we use *p*-value 0.05 as significance level in this study) shown in Table 4, low household income, being race Black and not being Hispanic have resulted in increased levels of obesity in the United States. Being obese does not increase the levels of poverty, nor does being poor not increase incidence of obesity directly. However, having low household income has contributed to have more incidences of poverty and obesity. Therefore, in graph-theoretic language, income is identified as a common cause for both obesity and poverty.

High levels of unemployment, being non-Hispanic, and low levels of food tax (soda tax) have increased the incidence of SNAP participants in the United States. Poverty is neither strictly endogenous nor exogenous, meaning it is weakly endogenous (or weakly exogenous); arrows come into the poverty variable, as well as, they emanate from it.

Low levels of income, high levels of unemployment, high food taxes (soda taxes), being Hispanic and not being White have increased the incidences of being poor in the United States. Being poor does not increase the participation in SNAP directly, however poverty and SNAP participation are related via income and unemployment (back-door paths). Income is a common cause of both poverty and unemployment, meaning higher income levels would lead to reduced levels of unemployment and poverty. Additionally, small positive change in unemployment will lead to positive change in SNAP participation and food insecurity. That is to say, in graph-theoretic language, unemployment is a common cause of SNAP participation, food insecurity and poverty. Furthermore, high incidence of poverty causes to have increased levels of food insecurity.

Although there is no direct causality between obesity and food insecurity, there are two back-door paths connecting obesity to food insecurity, via income, unemployment and poverty. Income, poverty, and food insecurity form a causal chain, i.e. income  $\rightarrow$ poverty  $\rightarrow$  food insecurity. As a result, if one wishes to find out the effects of income on food insecurity, conditioning on poverty, i.e. including poverty as an explanatory variable in regression analvsis, blocks the path from income to food insecurity. Similar logic is applied if one wishes to measure the effect of income on SNAP participation; conditioning on unemployment blocks the path from income to SNAP participation, since income, unemployment and SNAP forms a causal chain (income  $\rightarrow$  unemployment  $\rightarrow$  SNAP). Likewise, according to the sub-graph (see Fig. 2) which shows the relationship between obesity, income, and food insecurity, i.e. obesity  $\leftarrow$  income  $\rightarrow$  poverty  $\rightarrow$  food insecurity, if one wants to find out the effects of obesity on food insecurity, one has to

#### Table 4

Parameter estimates (partial value) for each edge and their associated significance. *Source:* calculated by authors

Edge		Partial value	Std error	t-stat	p- value
Asian	Grocery	1 0212	0 1 3 0 7	7.81	0.0000
Asian	Income	0.7263	0.1507	6 30	0.0000
Asian	Soda tax	0.3273	0.1431	2 29	0.0267
Asian	Unemployment	0.9568	0.2203	4 34	0.0001
Blacks	SNAP	-0.2436	0.1288	-1.89	0.0647
Blacks	Milk price	0.4231	0.1311	3.23	0.0023
Blacks	Obesity	0.4144	0.082	5.06	0.0000
Blacks	Soda price	-1.2624	0.3228	-3.91	0.0003
Blacks	Unemployment	1.1221	0.4121	2.72	0.0091
Full serv exp	Fast food exp	0.8778	0.0733	11.97	0.0000
Grocerv	Full serv exp	0.2137	0.1202	1.78	0.0818
Hispanic	SNAP	-0.5462	0.125	-4.37	0.0001
Hispanic	Obesity	-0.1917	0.0835	-2.29	0.0263
Hispanic	Poverty	0.2222	0.0445	4.99	0.0000
Hispanic	Full serv exp	0.562	0.115	4.89	0.0000
Income	Grocery	-0.4452	0.1274	-3.49	0.0010
Income	Milk price	0.2624	0.1311	2.00	0.0511
Income	Obesity	-0.6291	0.0833	-7.55	0.0000
Income	Poverty	-0.9101	0.0441	-20.65	0.0000
Income	Unemployment	-0.5747	0.1664	-3.45	0.0012
Milk price	Fast food exp	-0.1674	0.0733	-2.28	0.0270
Poverty	Food insecurity	0.6536	0.1004	6.51	0.0000
Soda price	Grocery	-0.2852	0.0986	-2.89	0.0058
Soda price	Soda tax	-0.2851	0.1431	-1.99	0.0521
Soda tax	SNAP	-0.2814	0.1218	-2.31	0.0253
Soda tax	Obesity	0.1538	0.0823	1.87	0.0679
Soda tax	Poverty	0.0908	0.0436	2.08	0.0426
Unemployment	SNAP	0.3855	0.133	2.90	0.0057
Unemployment	Food insecurity	0.2477	0.1004	2.47	0.0173
Unemployment	Full serv exp	0.1978	0.1183	1.67	0.1011
Unemployment	Poverty	0.096	0.0471	2.04	0.0473
Whites	Income	0.1838	0.1152	1.59	0.1175
Whites	Poverty	-0.3166	0.0449	-7.05	0.0000
Whites	Soda price	-0.9566	0.3228	-2.96	0.0048
Whites	Unemployment	0.9792	0.4429	2.21	0.0319

Note: p-value = 0.05 for rejecting the null hypothesis that edges are statistically not different from zero. Significant p-values are presented in bold font.

condition only on income (again conditioning on poverty blocks the path). Similarly, if one wants to find out the effects of obesity on SNAP participation, one should only condition on income, because conditioning on unemployment will block the path from obesity to SNAP participation, i.e. causality is obesity  $\leftarrow$  income  $\rightarrow$  unemployment  $\rightarrow$  SNAP.

Being Hispanic is a common cause for both obesity (negative) and poverty (positive), i.e. obesity  $\leftarrow$  Hispanic  $\rightarrow$  food insecurity. Also, the following relationships between obesity, Blacks, unemployment status, poverty, food insecurity, and SNAP participation exists, obesity  $\leftarrow$  Blacks  $\rightarrow$  SNAP, obesity  $\leftarrow$  Blacks  $\rightarrow$  unemployment  $\rightarrow$  food insecurity, obesity  $\leftarrow$  Blacks  $\rightarrow$  unemployment  $\rightarrow$  poverty.

In the following section, we compare and contrast the DAGgenerated sub-graphical model associated with major policy variables (obesity, poverty, food insecurity and food assistance) with that of similar work in the extant literature concerning the same policy variables. Fig. 2 shows the DAG-generated aforementioned sub-graphical model. Fig. 3 depicts the similar relationships extracted from this literature. A reviewer questioned the validity of comparisons of results coming from aggregated data versus micro-level data. In principle, we agree with the reviewer's comment. However, we do not have individual level data for all variables considered here. Therefore, we fall back on aggregate data, using a representative household as our focus. Our data are aggregated and represent information for a representative individual/ household in the aggregate. For the variables used in our study, agent's behavior at the representative aggregate level is considered a good proxy for more micro-level behavior; hence comparisons of our results with those of extant literature are considered to be legitimate to the best of our knowledge. To support this contention, we find plausible causal linkages at the aggregate level (macro level) that either confirm or elaborate on results found with micro level data (for example, see the discussion below with respect to comparing our findings with those of Gundersen et al. (2011a)).

Let us compare our work with that of Gundersen et al. (2011a). According to them, food insecurity is directly caused by poverty, income, race and unemployment (see Fig. 3); being poor, low income, being Black and unemployed increased the incidences of food insecurity. However, they did not consider other policy variables such as obesity and food assistance (SNAP) in their analysis. Our results find a similar causal structure to that of Gundersen et al. (2011a) with respect to unemployment, and poverty (they are direct causes of food insecurity): however we find income (low income) to cause food insecurity via the causal chain, poverty and unemployment, i.e. income  $\rightarrow$  poverty  $\rightarrow$  food insecurity and income  $\rightarrow$  unemployment  $\rightarrow$  food insecurity. In other words, if one wants to find the effects of income on food insecurity, having poverty in the model (conditioning on poverty) blocks the path from income to food insecurity, thereby not identifying the appropriate effect of income on food insecurity. Similarly, if one wants to find the effect of income on food insecurity, having unemployment in the model blocks the effect of income on food insecurity. In Gundersen et al. (2011a,b) work, they found that race variables directly cause food insecurity. However, our findings show that race (White, Black, Asian or Hispanic) variables cause food insecurity via poverty, income and unemployment. For example, being Hispanic causes food insecurity via poverty. That is to say, being Hispanic increased the incidence of being poor and then being poor increased the incidence of being food insecure. Also, being White causes food insecurity via either poverty (being White decreased the occurrence of being poor, hence less food insecure) or via income and poverty (being White increased the occurrence of being high income, hence less poor and less food insecure) or via income and unemployment (being White increased the incidence of being high income, hence less unemployed and therefore more food secure). Furthermore, being Asian caused food insecurity via income and unemployment (being Asian increased the occurrence of being high income, hence less unemployed and more food secure) or via income and poverty (being Asian increased the incidence of being high income, hence less poor and more food secure). Also, being Black caused food insecurity via unemployment (being Black increased the occurrence of being unemployed, which increased the food insecurity among them).

Additionally, we shed light on causality structure involving obesity and food assistance (SNAP). Unlike Nord and Golla (2009) and Yen et al. (2008) who found direct causal relationship of food insecurity with SNAP participation, i.e. SNAP  $\rightarrow$  Food insecurity (increased SNAP participation decreased food insecurity), we find in the contemporaneous time that there is no direct causality between SANP participation and food insecurity, however they are related through unemployment and food taxes and poverty in a back-door path.

Tiehen et al. (2012) found SNAP  $\rightarrow$  poverty (a negative relationship or participation in SNAP reduces poverty). In contemporaneous time, however we find that poverty and SNAP are related through unemployment, food taxes and race (Blacks and Hispanics) in several back-door paths (i.e. no direct causality relationships).

Cawley and Meyerhoefer (2012), Meyerhoefer and Yang (2011), Dixon (2010), Finkelstein et al. (2009) and Wolf and Colditz (1998) found direct causality relation of SNAP participation to obesity, i.e. SNAP  $\rightarrow$  obesity (see Fig. 3). In other words, participation in SNAP increased the incidence of obesity levels. In our work, we find that



Fig. 1. The directed acyclic graph associated with the sixteen food environment variables. Note: Numbers printed on edges represent the partial effect of one variable on the other (parameter estimate shown in Table 4). *P*-value associated with each parameter estimate is shown next to the variable name.



Fig. 2. A subset of directed acyclic graph associated with the sixteen food environment variables (policy variables) constructed using information in Fig. 1.

obesity and SNAP participation are related, but nonetheless we find no direct causality. We find that obesity and SNAP participation are related in four back-door paths, via unemployment, poverty and income; via food taxes; via race (Blacks and Hispanics). Therefore, if one wants to find the effect of SNAP participation on obesity, one must condition only on income (see from Fig. 2 that conditioning on income captures the effect of SNAP on obesity via unemployment, food insecurity, poverty and race). Casey et al. (2001, 2006) and Jyoti et al. (2005) found a direct link between household being food insecure (especially children) to high levels of obesity. In our work, we do not find a direct causality relationship of food insecurity and obesity. We find that food insecurity and obesity are related via several back-door paths. They are: (1) causality via poverty and income, (2) causality via poverty, unemployment and income, (3) causality via unemployment and income, (4) causality via unemployment, SNAP, food Gundersen, Brown, Engelhand, and Waxman (2011a)



Cawley and Meyerhoefer (2012); Dixon (2010); Finkelstein et al.,(2009); Wolf and Coldtiz (1998)



Nord and Golla (2009); Yen et al., (2008)



Tiehen, Jolliffe, Gundersen (2012)







Fig. 3. Relationships of key policy variables from the extant literature.

tax, (5) causality via unemployment, SNAP, Blacks, (6) causality via unemployment, SNAP and Hispanic, and (7) causality via poverty, White and income, to name a few.

#### 6. Conclusions and implications

Research is beginning to emerge documenting the complexity and interaction of factors affecting the food environment in the United States. However, given the complexity of the interaction among a myriad of variables, more research is necessary to identify and measure the underlying causal relationships among these factors, hence providing the paths for effective policy interventions. Several studies in the extant literature have addressed the issues related to food insecurity, food deserts, food assistance, and health and other factors of food environment. However, these studies have considered only a limited number of variables in a piecemeal fashion; hence, a holistic picture of the "food environment" as a complex economic system has not yet emerged. Public policies generated on the basis of limited information may be suboptimal, hence potentially misleading. To improve policy-making, it is necessary to develop a more complete understanding of the food environment. We make use of a novel modeling approach, causality structures modeled through artificial intelligence and directed acyclic graphs (DAGs) (Bayesian probability networks), recognizing the complexity of the food environment.

Using data from various public sources (such as USDA Food Environment Atlas, United States Census Bureau, Centers for Disease Control and Prevention, United States Bureau of Labor Statistics, USDA Food and Nutrition Service) with regard to 16 socio-economic-demographic variables, and graphical causal structures based on artificial intelligence and DAGs, we estimated causality relationships among variables determining the food environment of the United States. In particular, we centered attention on uncovering causality structures pertaining to four important nutrition and food policy variables in the United States, namely, obesity, poverty, food insecurity and food assistance programs.

We find that food insecurity, obesity and SNAP participation (food assistance) to be endogenous, meaning that they are caused by several of other variables. Poverty is weakly endogenous, meaning that there are variables that cause poverty and poverty causes food insecurity in a positive relationship. Unemployment is found to be a significant positive cause of food insecurity, poverty and SNAP participation. However, income causes food insecurity via poverty (meaning low income would make one poor hence food insecure). The importance of knowing this causality relationship is that if one wants to find the effect of income on food insecurity, conditioning on poverty blocks the path from income to food insecurity, thereby not identifying the effect of income on food insecurity. That is to say, if one were to run a regression analysis of food insecurity on income, having poverty variable on the right-hand side of the model will give rise to statistically insignificant parameter estimate of the marginal effect of income on food insecurity. The measure of poverty blocks the path between the measure of income and the measure of food insecurity. Also, as explained above, in our model we find unemployment to be a common cause for both food insecurity and poverty (meaning low income causes one to be food insecure and also poor). Therefore, if one wants to find the effect of poverty on food insecurity, conditioning on unemployment would make poverty and food insecurity conditionally orthogonal. Again. econometrically, if one were to regress food insecurity on poverty, having unemployment on the right-hand side of the regression will block the information flow from poverty to food insecurity, hence giving rise to statistically insignificant parameter estimate. Such econometric modeling is naturally the next step in investigating quantitative effects as a result of regressing one variable on the other, once the causality pattern among variables is identified through directed acyclic graph. Following Pearl (2000) (see Pearl, 2000, Chapter 5 pages 133–170), such modeling and subsequent policy simulation can be carried out with Structural Econometric Modeling (SEM).

Additionally, we were able to shed light on the causal structure linking obesity and food assistance (SNAP). Food insecurity is related to SNAP participation via a back-door path, unemployment. However, if one wants to find out the effect of food insecurity on SNAP participation, conditioning on unemployment would make food insecurity and SNAP conditionally orthogonal. Unlike past studies which found that SNAP participation decreased incidence of food insecurity, we find that in the contemporaneous time, food insecurity and SNAP participation is related in a complex relationship via several back-door paths. Again, contrary to past studies which found that SNAP participation decreased occurrences of poverty, in contemporaneous time, we find that poverty and SNAP participation is related via several complex back-door paths, i.e. via unemployment, food taxes, race (Hispanic and Black status) and obesity and income. In this light, if one wants to measure the effect of poverty on SNAP participation, conditioning on food insecurity opens up the path between poverty and unemployment, however if one conditions on unemployment, it would block the information flow from poverty to SNAP participation.

The current literature finds a direct causality relation between SNAP participation and obesity. However, we find that obesity and SNAP participation are related, but in an indirect relationship. We find that obesity and SNAP participation are related in four back-door paths, via unemployment, poverty and income; via food tax; via race (Blacks and Hispanics) and via unemployment, food insecurity, poverty and income. Although past studies found a direct causality relationship with respect to food insecurity and obesity, we find that food insecurity and obesity are related via several back-door paths namely; via poverty and income; via poverty, income and unemployment; via unemployment and income; via unemployment, SNAP, food taxes; via unemployment, SNAP, Blacks; via unemployment, SNAP and Hispanics, and via poverty, White and income.

It is apparent from the complex causality structure that policy variables cannot be dealt with independently. Government agencies responsible for designing policies for food assistance (such as SNAP), need to pay attention to several variables such as prevalence of food insecurity, poverty, income, obesity, food taxes, unemployment, and evidence of several racial groups being affected by poverty, low income, unemployment and obesity. Policy makers centering attention to curb the obesity epidemic in the United States, according to our model, need to not only design appropriate tax policies (if at all), but also policies that help consumers overcome poverty, unemployment, food insecurity, low income, and to provide appropriate food assistance programs. Since obesity also is related to poverty, policies directed to combat poverty should pay attention to obesity reduction measures as well. A similar argument applies to policies directed to curb food insecurity, since food insecurity is related (indirectly) to obesity. Some racial groups are more vulnerable (Blacks and Hispanics) because they typically are poor, food insecure, obese, and have low incomes.

It is evident that interrelationships and causality structures among obesity, poverty, food insecurity, food assistance and associated variables are very complex. Using a richer set of variables along with novel modeling techniques, we shed light on complex causality relationships intertwining important policy variables. The importance of understanding these causality relationships is that, if one wants to model factors affecting a given policy variable, causality structure provides the information about the variables analysts should condition on. Conditioning on the wrong variable may block an important path or opens another back-door path connecting variables. We deal with contemporaneous causality structures of 16 variables affecting the food environment of the United States. Our work stimulates future research to where one may want investigate the dynamic behavior of these variables in determining the causality structure through alternative modeling approaches. As well, the complex system studied here can be augmented further by studying even more variables, potentially uncovering other latent paths between and among variables. Furthermore, as suggested by Pearl (2000) (also explained above), once the underlying causality structure is identified, Structural Equation Modeling (SEM) of variables affecting the food environment complex using econometric methods can be suggested as a natural next step of this research.

#### **Funding sources**

We wish to acknowledge partial funding provided by Texas A&M AgriLife Research for the initial stages of this project.

# Acknowledgements

Authors wish to thank three anonymous reviewers along with the editor of this journal, Dr. Satoru Shimokawa for providing valuable comments which increased the quality of the paper. Special thank goes to Dr. Judea Pearl, Cognitive Systems Lab, Computer Science Department of University of California at Los Angeles (UCLA) for reading initial drafts of this manuscript and providing important comments and direction.

Also, Senarath Dharmasena would like to thank several undergraduate research assistants of the Department of Agricultural Economics at Texas A&M University and Dr. Jessica Todd of United States Department of Agriculture – Economic Research Service for providing assistance with data. Senarath Dharmasena also would like to extend special thanks to Dr. David A. Bessler, Regents Professor at Texas A&M University, for introducing him to the exciting world of causality modeling, decade and half ago, then as his mentor and currently colleague at the Department of Agricultural Economics at Texas A&M University. Last, but not least, Senarath Dharmasena extends special thanks and love for his dear wife Inoka Chandrasena for keeping him afloat during the writing of this paper.

Authors would like to note that the views expressed in this article do not represent those of their institutions but solely of their own.

## Appendix

### A.1. Directed graphs

Essentially, a directed graph is an illustration using arrows and vertices to represent the causal flow among a set of variables, whose values are measured in non-time sequence (e.g., cross section data). A graph is an ordered triple  $\langle V, M, E \rangle$  where V is a non-empty set of vertices (variables), M is a non-empty set of marks (symbols attached to the end of undirected edges) and **E** is a set of ordered pairs. Each member of *E* is called an edge. Vertices connected by an edge are said to be adjacent. If we have a set of vertices {A, B, C, D} the undirected graph contains only undirected edges (e.g. A - B). A directed graph contains only directed edges (e.g.  $C \rightarrow D$ ). A directed acyclic graph is a directed graph that contains no directed cyclic paths. An acyclic graph has no path that leads away from a variable only to return to that same variable. (The path  $A \rightarrow B \rightarrow C \rightarrow A$  is labeled "cyclic" as here we move from A to B, but then return to A by way of C.) Only acyclic graphs are used in our analysis. Given very small finite time intervals, directed cyclic paths cannot exist. That is to say, given two variables, A and *B*, and their causality relationship as  $A \rightarrow B$ , this truly means that  $A_t \rightarrow B_t$  (A at time t causes B at time t). For very small finite time intervals, we rule out  $B_t \rightarrow A_t$  and interpret the reverse causality as  $B_t \rightarrow A_{t+1}$  (meaning that B at time t cases A at time t + 1). As a result it is true that  $A_{t+1} \rightarrow B_{t+1}$ .

It is helpful and valid to use terms from genealogy when referring to variables and their position in a causal structure as, for example, parents, grandparents, children, grandchildren, ancestors or descendents, etc. So in the path  $A \rightarrow B \rightarrow C \leftarrow D$ , the variables A, Band D are ancestors of variable C. As such, variable C is a descendent of variables A, B and D. The variable A is a parent of variable B. Variable C has two parents, B and D, and one grandparent, variable A.

Directed acyclic graphs are pictures (illustrations) for representing conditional independence as given by the recursive decomposition:

$$Pr(v_1, v_2, v_3, \dots v_n) = \prod_{i=1}^n Pr(v_i | pa_i)$$
(2)

where *Pr* is the probability of vertices (variables)  $v_1$ ,  $v_2$ ,  $v_3$ , ...  $v_n$ , the notation  $pa_i$  represents the realization of some subset of the variables that precede (come before in a causal sense)  $v_i$  in order ( $v_1$ ,  $v_2$ ,  $v_3$ , ...  $v_n$ ) and the symbol  $\Pi$  represents the product operation, with index of operation denoted below (start) and above

(finish) the symbol. Pearl (1995) proposed d-separation as a graphical characterization of conditional independence. That is, d-separation characterizes the conditional independence relations given by the above product. If we formulate a directed acyclic graph in which the variables corresponding to pa<sub>i</sub> are represented as the parents (direct causes) of  $v_i$ , then the independencies implied by the product given above can be read off the graph using the notion of d-separation as defined in Pearl (1995):

**Definition.** Let *X*, *Y* and *Z* be three disjoint subsets of vertices [variables] in a directed acylic graph G, and let p be any path between a vertex [variable] in *X* and a vertex [variable] in *Y*, where by 'path' we mean any succession of edges, regardless of their directions. *Z* is said to block p if there is a vertex w on p satisfying one of the following: (i) w has converging arrows along p, and neither w nor any of its descendants are on *Z* or (ii) w does not have converging arrows along p, and w is in *Z*. Furthermore, *Z* is said to d-separate *X* from *Y* on graph G, written  $(X \perp Y \mid Z)_G$ , if and only if *Z* blocks every path from a vertex [variable] in *X* to a vertex [variable] in *Y*.

Geiger et al. (1990) demonstrated that there is a one-to-one correspondence between the set of conditional independencies,  $X \perp Y \mid Z$ , implied by the above factorization and the set of triples, *X*, *Y*, *Z*, that satisfy the d-separation criterion in graph G. If G is a directed acyclic graph with vertex set V, and *X*, *Y* and *Z* are in V, then G linearly implies the correlation between *X* and *Y* conditional on *Z* is zero if and only if *X* and *Y* are d-separated given *Z*.

## References

- Bartfeld, J., Dunifon, R., 2006. State-level predictors of food insecurity among households with children. J. Policy Anal. Manage. 25 (4), 921–942.
- Bessler, D.A., 2002. On world poverty: Its causes and effects. Paper delivered to the Food and Agricultural Organization (FAO) of the United Nations. Rome. http:// agecon2.tamu.edu/people/faculty/bessler-david/WebPage/Bessler%20Poverty. pdf (internet access January 1, 2016).
- Bessler, D.A., Akleman, D.G., 1998. Farm prices, retail prices, and directed graphs: results for pork and beef. Am. J. Agric. Econ. 80 (5), 1144.
- Bhattacharya, J., Currie, J., Haider, S., 2004. Poverty, food insecurity, and nutritional outcomes in children and adults. J. Health Econ. 23 (4), 839–862.
- Bitler, M., Haider, S.J., 2011. An economic view of food deserts in the United States. J. Policy Anal. Manage. 30 (1), 153–176.
- Bitler, M., Gundersen, C., Marquis, G., 2005. Are WIC non-recipients at less nutritional risk than recipients? An application of the food security measure. Rev. Agric. Econ. 27 (3), 433–438.
- Borjas, G., 2004. Food insecurity and public assistance. J. Public Econ. 88 (7–8), 1421–1443.
- Broda, C., Leibtag, E., Weinstein, D.E., 2009. The role of prices in measuring the poor's living standards. J. Econ. Perspect. 23 (2), 77–97.
- Bunge, M., 1959. Causality and Modern Science. Harvard University Press, Cambridge.
- Capps Jr., O., 2009. Issues indigenous to consumer economics and food marketing: opportunities for research contributions. J. Agric. Appl. Econ. 41 (2), 315–322.
- Case, A., Lubotsky, D., Paxson, C., 2002. Economic status and health in childhood: the origins of the gradient. Am. Econ. Rev. 92 (5), 1308–1334.
- Casey, P., Simpson, P., Gossett, J., Bogle, M., Champagne, C., Connell, C., Harsha, D., McCabe-Sellers, B., Robbins, J., 2006. The association of child and household food insecurity with childhood weight status. Pediatrics 118, 1406–1413.
- Casey, P., Szeto, K., Lensing, S., Bogle, M., Weber, J., 2001. Children in foodinsufficient, low-income families: prevalence, health and nutrition status. Arch. Pediatr. Adolesc. Med. 155, 508–514.
- Cawley, J., Meyerhoefer, C.D., 2012. The medical cost of obesity: an instrumental variables approach. J. Health Econ. 31, 219–230.
- Chen, S.E., Florax, R.J., Snyder, S.D., 2012. Obesity and fast food in urban markets: a new approach using geo-referenced micro data. Health Econ. http://dx.doi.org/ 10.1002/hec.2863.
- Chickering, D.M., 2002. Optimal structure identification with greedy search. J. Mach. Learn. Res. 3, 507–554.
- Coleman-Jensen, A., Nord, M., Andrews, M., Carlson, S., 2012. Household Food Security in the United States in 2011. Economic Research Report No. 141. USDA Economic Research Service.
- Currie, S., Grossman, M., Saffer, H., 2004. The effect of fast food restaurants on obesity and weight gain. Am. Econ. J.: Econ. Policy 2, 32–63.
- Dharmasena, S., Ishdorj, A., Capps, O., Bessler, D.A., 2014a. Dynamics of macroeconomic shocks on food assistance programs. Selected paper presented at the Southern Agricultural Economics Association (SAEA) Annual Meetings, Dallas, Texas.

- Dharmasena, S., Bessler, D.A., Todd, J., Capps Jr., O., 2014b. Dynamics of the food environment in the United States. Selected paper presented at the Agricultural and Applied Economics Association (AAEA) Annual Meetings, Minneapolis, MN.
- Dixon, J.B., 2010. The effect of obesity on health outcomes. Mol. Cell. Endocrinol. 316 (2), 104–108.
- Dubois, L., Farmer, A., Girard, M., Porcherie, M., 2006. Family food insufficiency is related to overweight among preschoolers. Soc. Sci. Med. 63, 1503–1516.
- Dunn, R.A., 2010. The effect of fast-food availability on obesity: an analysis by gender, race, and residential location. Am. J. Agric. Econ. 92, 1149–1164.
- Finkelstein, E.A., Ruhm, C.J., Kosa, K.A., 2005. Economic causes and consequences of obesity. Annu. Rev. Public Health 26, 239–257.
- Finkelstein, E.A., Trogdon, J.G., Cohen, J.W., Dietz, W., 2009. Annual medical spending attributable to obesity: payer-and service-specific estimates. Health Aff. 28 (5), w822–w831.
- Furness, B., Simon, P., Wold, C., Asarian-Andersen, J., 2004. Prevalence and predictors of food insecurity among low-income households in los angeles county. Public Health Nutr. 7 (6), 791–794.
- Geiger, D., Verma, T., Pearl, J., 1990. Identifying independencies in bayesian networks. Networks 20, 507–534.
- Glymour, C., Scheines, R., Spirtes, P., Ramsey, J., 2014. The TETRAD Project: Causal Models and Statistical Data. Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA (project page: http://www.phil.cmu.edu/tetrad/ current.html).
- Granger, C.W.J., 1980. Testing for Causality: a personal viewpoint. J. Econ. Dyn. Control 2, 329–352.
- Grebitus, C., Lusk, J.L., Nayga, R.M., 2013. Effect of distance of transportation on willingness to pay for food. Ecol. Econ. 88, 67–75.
- Gundersen, C., 2008. Measuring the extent, depth and severity of food insecurity: an application to American Indians in the United States. J. Popul. Econ. 21 (1), 191–215.
- Gundersen, C., Kreider, B., 2008. Food stamp and food insecurity: What can be learned in the presence of neoclassical measurement error? J. Hum. Resour. 43 (2), 352–382.
- Gundersen, C., Kreider, B., 2009. Bounding the effects of food insecurity on children's health outcomes. J. Health Econ. 28, 971–983.
- Gundersen, C., Gruber, J., 2001. The dynamic determinants of food insufficiency. In: Andrews, Margaret, Prell, Mark (Eds.), Second Food Security Measurement and Research Conference, Volume II: Papers. USDA, ERS Food Assistance and Nutrition Research Report 11-2, pp. 91–109.
- Gundersen, C., Oliveira, V., 2001. The food stamp program and food insufficiency. Am. J. Agric. Econ. 84 (3), 875–887.
- Gundersen, C., Kreider, B., Pepper, J., 2011a. The economics of food security in the United States. Appl. Econ. Perspect. Policy 33 (3), 281–303.
- Gundersen, C., Kreider, B., Pepper, J., 2012. The impact of the national school lunch program on child health: a nonparametric bounds analysis. J. Econom. 166 (1), 79–91.
- Gundersen, C., Jolliffe, D., Tiehen, L., 2009. The challenges of program evaluation: when increasing program participation decreases the relative well-being of participants. Food Policy 34, 367–376.
- Gundersen, C., Brown, J., Engelhard, E., Waxman, E., 2011b. Map the Meal Gap. Feeding America, Technical Brief.
- Gundersen, C., Weinreb, L., Wehler, C., Hosmer, D., 2003. Homelessness and Foodinsecurity. J. Hous. Econ. 12 (3), 250–272.
- Hausman, D., 1998. Causal Asymmetries. Cambridge University Press, Cambridge, UK.
- Jensen, H.H., 2002. Food insecurity and the food stamp program. Am. J. Agric. Econ. 84 (5), 1215–1228.
- Jyoti, D., Frongillo, E., Jones, S., 2005. Food insecurity affects school children's academic performance, weight gain, and social skills. J. Nutr. 135, 2831– 2839.
- Laraia, B., Siega-Riz, A., Gundersen, C., Dole, N., 2006. Psychological factors and socioeconomic indicators are associated with household food insecurity among pregnant women. J. Nutr. 136 (1), 177–182.
- Leete, L., Bania, N., 2010. The effect of income shocks on food insufficiency. Rev. Econ. Household 8, 505–526.
- Meyerhoefer, C., Yang, M., 2011. The relationship between food assistance and health: a review of the literature and empirical strategies for identifying program effects. Appl. Econ. Perspect. Policy 33 (3), 304–344.
- Nord, M., Golla, A.M., 2009. Does SNAP decrease food insecurity? Untangling the Self-Selection Effect. Economic Research Report No. 85. U.S. Department of Agriculture, Economic Research Service.
- Olson, C.M., Anderson, K., Kiss, E., Lawerence, F.C., Seiling, S.B., 2004. Factors protecting against and contributing to food insecurity among rural families. Family Econ. Nutr. Rev. 16 (1), 12–20.
- Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82 (4), 669–710. Pearl, J., 2000. Causality: Models, Reasoning and Inference. Cambridge University Press, New York, NY, USA.
- Rettenmaier, A.J., Wang, Z., 2013. What determines health: a causal analysis using county level data. Eur. J. Health Econ. 14, 821–834.
- Ribar, D., Hamrick, K., 2003. Dynamics of Poverty and Food Sufficiency. Food Assistance and Nutrition Research Report 33. U.S. Department of Agriculture, Economic Research Service, Washington D.C.
- Ribar, D.C., Hamrick, K.S., 2003. Dynamics of Poverty and Food Sufficiency. Food Assistance and Nutrition Research Report No. 36. U.S. Department of Agriculture, Economic Research Service.

Spirtes, P., Glymour, C.N., Scheines, R., 2000. Causation, Prediction, and Search. The MIT Press.

- Swanson, N.R., Granger, C.W.J., 1997. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. J. Am. Stat. Assoc. 92 (437), 357–367.
- Tiehen, L., Jolloffe, D., Gundersen, C., 2012. Alleviating Poverty in the United States: The Critical Role of SNAP Benefits. Economic Research Report No. 132. U.S. Department of Agriculture, Economic Research Service.
- USDA, 2014. Food Environment Atlas http://www.ers.usda.gov/data-products/foodenvironment-atlas/about-the-atlas.aspx (internet access November 29, 2014).
- Van Hook, J., Balistreri, K., 2006. Ineligible parents, eligible children: food stamp recipient, allotments, and food insecurity among children of immigrants. Soc. Sci. Res. 35 (1), 228–251.
- Verma, T., Pearl, J., 1991. Equivalence and Synthesis of Causal Models. Technical Report, R-150. Department of Computer Science, University of California at Los Angeles.
- Ver Ploeg, M., Ralston, K., 2008. Food Stamps and Obesity: What Do We Know? Economic Information Bulletin No 34. U.S. Department of Agriculture, Economic Research Service. Washington, D.C.
- Wilde, P., Nord, M., 2005. The effect of food stamps in food security: a panel data approach. Rev. Agric. Econ. 27 (3), 425–432.
- Wolf, A.M., Colditz, G.A., 1998. Current estimates of the economic cost of obesity in the United States. Obes. Res. 6 (2), 97–106.
- Yen, S.T., Andrews, M., Chen, Z., Eastwood, D.B., 2008. Food stamp program participation and food insecurity: an instrumental variables approach. Am. J. Agric. Econ. 90 (1), 117–132.