Food Access and Food Security – An Empirical Analysis

by

Alessandro Bonanno and Jing Li

Alessandro Bonanno is Assistant Professor of Agricultural Economics, Department of Agricultural Economics and Rural Sociology, at the Pennsylvania State University.

Jing Li is a Ph.D. student in Agricultural, Environmental, and Regional Economics, at the Pennsylvania State University.

Corresponding author: Alessandro Bonanno: abonanno@psu.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011

Food Access and Food Security – An Empirical Analysis

Abstract

This paper analyzes the effect of access to different types of food outlets on households' food insecurity levels. Two years (2004 and 2005) of Current Population Survey – Food Security Supplement data are matched with MSA-level data on store counts of Wal-Mart Supercenters, small food stores (small grocery stores and convenience stores), medium and large grocery stores, and convenience stores associated with gas stations. Endogeneity of food stores' location is accounted for to eliminate spurious correlation between households' food security status and food access, using the 2-State Residual Inclusion method (2-SRI). Preliminary results indicate that, before accounting for endogeneity bias, the presence of Wal-Mart supercenters appears to be associated to higher levels of households' food insecurity, while the presence of other food stores is associated with lower levels. After eliminating spurious correlation, only the presence of small food stores appears helping to reduce food insecurity (across measures of food insecurity and data samples) while the presence of gas convenience stores is associated with higher likelihoods of experiencing food insecurity. The presence of Wal-Mart supercenters and that of medium and large grocery stores have little to no impact on the likelihood of a household being food insecure (the first showing only weak evidence of a mitigating effect, the second, instead, showing weak evidence of a magnifying effect).

.

JEL Codes: Q18; L81; P46

Keywords: Food Security, Food Access, 2-Stage Residual Inclusion, Wal-Mart, Convenience Stores, Grocery Stores

Food Access and Food Insecurity – Some Preliminary Findings

1. Introduction

Food insecurity is the outcome of a household being unable to acquire (or being uncertain of having) enough food to meet the needs of all its members (Nord et al., 2010). Due to the most recent economic downturn, recent estimates of households' food insecurity (FI) in the United States have reached worrisome levels. According to USDA estimates, at some point during the year 2009 there were 17.4 million (14.7%) households affected by FI. Of these, 10.6 million (9.0 %) were characterized as Low Food Secure (LFS) households and 6.8 million (5.7 %) as Very Low Food Secure (VLFS) households.[1] Even though these values were substantially unchanged from those of the previous year, such figures are considerably larger than those of one decade ago: in the year 1999 there were 10.1% FI households of which 7.1 % were LFS and 3.0 % VLFS (Nord et al,, 2010).

Although several studies have analyzed the characteristics of food secure households (see for example Rose, Gundersen and Oliveira, 1998; Nord, Andrews and Carlson, 2004; Nord et al, 2010), and many more have analyzed FI in the context of the effectiveness of food assistance programs (e.g. Gundersen and Oliveira, 2001; Jensen, 2002; Borjas, 2004; Kabbani and Kmeid, 2005; Bartfeld and Dunifon, 2006; Gibson-David and Foster, 2006; Yen et al. 2008; Nord and Golla, 2009), few attempts exists aiming to understand the relationship between household's FI status and the surrounding environment.[2]

---

[1] Low Food Secure households have obtained enough food to avoid substantially disruption in their eating patterns or reduced food intake by using a variety of coping strategies, such as eating less varied diets, participating in Federal food assistance programs, or getting emergency food from community food pantries. These households were previously described as "food insecure without hunger." Very Low Food Secure households are affected by disruption of normal eating patterns of one or more household members. Food intake was reduced at times during the year because they had insufficient money or other resources for food. These households were described as "food insecure with hunger."

[2] See for example Bartfeld and Dunifon (2006) analysis of State-level contextual (economic and social) attributes on the likelihood of food security among households with children.

Even though the relationship between food security and the food environment has been widely acknowledged,[3] no empirical analysis has rigorously treated the issue and tried to qualify and quantify the potential impact of the access to different outlets on the likelihood of being food insecure. As causes of food FI are associated with having insufficient income or limited access to other sources of food that do not require expenses (e.g. food pantries, food banks etc...), access to sources where to retrieve food has a pivotal role in ensuring that households have (at least potentially), the ability to be food secure. The presence of food stores can affect FI on two fronts. First, in the presence of limited access, consumers may be charged higher prices as retailers may benefit from monopolistic positions. Second, lack of transportation means may prevent low-income households to reach the most convenient option to them (if such option is available), making it hard to adopt cost-saving strategies (Leibtag and Kaufman, 2003).[4] Not understanding the effect of the food environments on FI may limit the accuracy of policy analysis that would contemplate tools geared to improve food access.[5]

Furthermore, some of the recent structural changes of the food retailing industry (Martinez, 2007) may result in favorable outcomes for low-income households. One of the most important recent changes in food retailing is Wal-Mart's expansion of its Supercenters[6]

---

[3] Several programs exist both at the national and aiming to improve food security through food access. For example, following the passage of the Community Food Security Act in 1996, the USDA launched the Community Food Security Initiative in 1999 to help establishing partnerships between USDA and local communities, (Scott Kantor, 2001).

[4] A vast body of descriptive and/or limited scope analysis shows that areas inhabited by a prevalence of less-privileged individuals are characterized by limited access to large (or "high quality") food stores (see for example Alwitt and Donley, 1997; Ball et. al., 2008; Cotterill and Franklin, 1995; Morland et al., 2002; King et al. 2004; Moore and Diez Roux, 2006; Powell et al., 2007; Zenk et al., 2005).Furthermore, a positive relationship exists between the quality of the food choices that low-income (food stamps recipient) households make and the access to food outlets (Rose and Richards, 2004).

[5] See for example the attempts, in Maryland, to ease tax pressure on grocery stores located in low-income areas (Beherens, 2010).

[6] The company has moved away from its Discount Stores format (carrying a limited number of food products, mostly shelf-stable) to the Supercenter format, which offers fresh produce, meat, bakery, deli and fresh seafood departments, becoming the larger food retailer n the U.S.

format. Such expansion can be beneficial to low-income's households ability to acquire food for two reasons: 1) the company has been found to increase consumers' surplus by offering lower prices and greater product variety (Hausman and Leibtag, 2007) providing relief to low-income individuals giving access to fresh produce at lower prices (Volpe and Lavoie 2008, Basker and Noel, 2009); 2) as Wal-Mart locates its stores preferentially in areas where competition is scant (Jia, 2008; Bonanno, 2010), its expansion could improve food access for low-income households who may have limited access otherwise. However, as the company may preferentially locate its stores in areas characterized by higher percentages of low income individuals[7] isolating casual effect between Wal-Mart Supercenters' presence and likelihood of being food insecure requires controlling for the company's endogenous location decision.

Spurious correlation is an issue that should be accounted for in order to evaluate the impact of any type of food outlets on the households' food security levels, since both consumers and market characteristics affect retailers' location decision. In fact, retailers position themselves *endogenously* into a fringe of low quality stores serving consumers who cannot (due to income constraint) pay for quality, and another tier of high quality stores (Ellickson, 2005; 2007), offering higher prices and a higher level of services (Bonanno and Lopez, 2009). In sum, as demographic characteristics impact location decision and outcomes such as FI, and the likelihood of the presence of unobservables impacting both processes is high, accounting for location endogeneity becomes indispensable.

This analysis aims to investigate whether having access to food retailers of different types, can affect a household's FI status, and whether such effects are larger for low-income households. Household-level data on households' FI status come from the Current Population Survey – Food Security Supplement (CPS-FSS) for the years 2004 and 2005,

---

(Food Marketing Institute, 2007). As of January 31 2011, Wal-Mart operated (in the U.S. alone) 2,747 Supercenters and 803 Discount Stores (Wal-Mart Stores Inc, 2010).

[7] The a rate of conversion of Discount Stores into Supercenters, representing the main strategy followed by the company to expand into food retailing is positively related with higher percentages of population being food stamps' recipients (Bonanno, 2010).

which are matched with food stores data at the MSA level using geography identifiers included in the CPS-FSS. The four types of stores considered are Wal-Mart Supercenters, small sized food stores (convenience stores and small groceries, proxy for easy to access stores, which offer limited assortments), medium-large sized grocery stores (proxy for traditional, full-service food stores, offering broad assortment but potentially hard to reach), and convenience stores attached to gas stations (stores whose location may be hard to reach and that offer only a limited assortment of food products). Our proxies for food access are the MSA-level number of stores of the four types divided by population. We control for food stores' location endogeneity using identification strategies that capture geographic differences in the supply-side determinants of location decision. In the specific case of Wal-Mart Supercenters, the identification strategy uses lagged density of discount stores (as in Basker and Noal, 2009) and the distance from the company's food distribution centers (Bonanno, 2010) as predictors of the density of Supercenters, while for the other food outlets we use geographic variations of cost variables across the different store-types.

Preliminary results show that, after correcting for store location endogeneity, only the presence of small food stores appears to help reducing FI (across measures of food insecurity and samples) while the presence of gas convenience stores is associated with higher likelihoods of experiencing FI. The presence of Wal-Mart supercenters and that of medium and large grocery stores have little to no impact on the likelihood of a household being food insecure (the first showing only weak evidence of a mitigating effect, the second, instead, showing weak evidence of a magnifying effect).

2. An Empirical Model of Food Insecurity and Food Access

The following model is a stylized representation of FI as the outcome of a household`s optimization problem. Household *i* located in area *l* maximizes its utility, which is function of income (spent on goods) and leisure (or hours worked), subject to time and budget constraints. Although the formal derivation is not illustrated here, the interested reader can refer to Jensen (2002) for a thorough discussion. In Jensen's model, (which does not account for the role of the built environment, but considers instead participation in the Food Stamp Program, as a household's decision variable) FI causes disutility due to concerns about having an adequate food supply, and under-consumption of food for some of household members. In the context of this analysis, the household FI status will depend upon both the household's characteristics and the features of the environment they live in. Thus, the FI status of household *i* in area *l*, or FI_{il} will be represented by the following function:

$$FI_{il} = f(X_{il}, FA_l, d_l \mid \beta, \delta, \gamma) + e_{il} = f(Z \mid \theta) + e_{il} \qquad (1)$$

Where X_{il} is a vector of household characteristics, FA_l is a proxy capturing the level of access to food for all households in area *l* (measured by the number of outlets of a given store type N_j divided by the total population in area *l* or N_j / pop_l), d_l is a vector of fixed effects to control for unobservables factors that could impact FI, β, δ, and γ are vectors of parameters conformable to X_{il} , FA_l , and d_l, respectively, and e_{il} is an error term. The first part of the central term in equation (1) can be summarized as $f(Z \mid \theta)$ where *Z* is the vector of all the variables that can influence FI and θ a conformable vector of parameters characterizing the relationship between the covariates in Z and FI_{il}.

Let *h* be a realization of FI_{ij} , i.e. a FI state; the probability that status *h* is observed is:

$$\Pr(FI_{il} = h \mid Z) = \Pr\left(\lambda_h < \sum_k \beta_k X_{kil} + \delta_j FA_{jl} + \sum_s \gamma_s d_s + e_{il} < \lambda_{h+1}\right) \qquad (2)$$

hihc, Considering three possible states for FI (Food Secure, Low Food Secure, and Very Low Food Secure); i.e. $h = \{0,1,2\}$. In this case, the probability of observing a given realization of *h* is:

$$\Pr(FI_{il} = h \mid \boldsymbol{Z}) = L(\lambda_h - \boldsymbol{Z}'\boldsymbol{\theta}) - L(\lambda_{h+1} - \boldsymbol{Z}'\boldsymbol{\theta}); \qquad (3)$$

where $\lambda_0 = -\infty$, $\lambda_3 = +\infty$ and $L(.)$ is the logistic cumulative density function (CDF). If *Z* is uncorrelated with the error terms, the vector of coefficients θ can be estimated be via maximum likelihood, using an ordered logit estimator, and the estimates obtained will be unbiased and consistent.

However, endogeneity bias is likely to be present. [8] From a theoretical standpoint, food retailers play a generic two-stage game (à la Sutton, 1998), where, in the first stage, they decide where to locate (the investment stage) while in a second stage they compete with other incumbent firms in the market (the competition stage). Given the intricate and complex nature of such game, which is dynamic in nature since firms maximize expected profits, there are infinite possible equilibria whose detailed analysis is a daunting task.[9]

Using the simiplifying assumptions that food retail companies have limited ability to choose their store formats[10] (i.e. that the product-type offered by each retailer is given) and that, in a given area and for each establishment type, food retail firms (facing symmetric demand and cost) can be ordered by decreasing profitability (i.e. the most profitable firms enter the market first, as in Berry, 1992),[11] there exists a Nash equilibrium (although not unique), which allows the researcher to treat the observed number of market participants as one of the possible equilibria of a game played by all potential entrants.

Let the observed number of food retail outlets of type *j* in area *l*, N^*_{jl}, be one possible equilibrium outcome of the location game discussed above. As, in general terms, larger markets could support a larger equilibrium number of firms, and market size impacts differently the equilibrium of different types of firms (Sutton, 1998), market size could become a limiting factor for firms investing heavily in fixed cost, which applies to most of

---

[8] Traditional models of firm's location consider firms' location to occurr along a continuum of possible outcomes. Some industry will be mostly supply-oriented, (e.g., coal mines), others demand oriented (e.g. family doctor practices), while others will p[ay attention to both supply and demadn factoers, such as retailers. See Kilkenny & Thisse (1999) for a review.
[9] See Jia (2008) for a formal analysis of a dynamic location game by retail firms.
[10] Strictly speaking, retail firms present different formats (making them differentiated products). Including the format-type decision in the game, will complicate the analysis further (see for example Mazzeo, 2002; or Seim, 2006).
[11] This assumption is consistent with those of seminal models of firms' entry (e.g. Bresnahan and Reiss, 1991; Berry, 1992)

traditional food stores (Ellickson, 2006).[12] Assuming that, in area *l*, market size is a proportion of the population (pop_l)[13], let's consider the following reduced form equation representing the equilibrium measure of store density for store-type *j* in area *l*:

$$\frac{N^*_{jl}}{Pop_l} = g^D(\mathbf{X}_{jl}, \mathbf{X}_{-jl}; \alpha^D_j) + g^C(C_{jl}, K_{jl}; \alpha^C_j) + \varepsilon_{jl} \qquad (4)$$

where the $g^D(.)$ and $g^C(.)$ are functions representing, respectively, the role of demand and cost factors on the equilibrium store density levels; $\boldsymbol{X}_{jl}$ and $\boldsymbol{X}_{-jl}$ are vectors of demand characteristics in area *l* (both for the *j*-th store type and for that of other store-types)[14]; C_{jl} and K_{jl} are vectors of format- and market- specific cost variables (variable and fixed cost, respectively), the α^D_j and α^C_j are conformable vectors of parameters qualifying the relationship between demand and cost factors (respectively) and store density, and ε_{jl} is an idiosyncratic error term.

Since $FA_{jl} \equiv \frac{N^*_{jl}}{Pop_l}$, unless one controls for all demand-side factors impacting both FI_{il} and FA_{jl}, the likelihood that, in equation (2) FA_{jl} is correlated with e_{il} is high. However, if the cost variables C_{jl} and K_{jl} are truly exogenous, *and* if the source of correlation of FA_{jl} and e_{il} is exclusively due to the presence of the term $g^D(\mathbf{X}_{jl}, \mathbf{X}_{-jl}; \alpha^D_j)$, that is, $E(\varepsilon_{jl}e_{il}) = 0, \forall i, j, l$ the presence of spurious correlation in equation (2) can be easily resolved.

Rewriting equation (4) as

[12] Such considerations apply to most industries whose firms commit to a specific location. Asplund and Sandin (1999) point out in their analyses of Swedish regional markets for driving schools, as profits per capita decrease in market size, capacity will tend to impose a limit to the possibility of observing a higher number of equilibrium firms.

[13] As market size, representing the potential demand for the goods offered by the firm *j* is a function of market characteristics such proportion is not a constant but depends on other factors such as income and consumers' heterogeneity which, in equation (4) below are represented by the vectors $\boldsymbol{X}_j$ and $\boldsymbol{X}_{-j}$. See Asplund and Sandin (1999) for more details.

[14] Demand characteristics across store types enter equation (4) because in the second stage of the game firms are likely to compete with those of other formats.

$$\frac{N^*_{jl}}{Pop_l} = g^C(C_{jl}, K_{jl}; \alpha^C_j) + r_{jl} \quad (5)$$

where the term $r_{jl} = g^D(\mathbf{X}_{jl}, \mathbf{X}_{-jl}; \alpha^D_j) + \varepsilon_{jl}$ is by construction, correlated with the errors in equation (2). It is easy to show that, if one introduced r_{jl} (or an unbiased estimate of it) in equation (2) the source of correlation between FA_{jl} and e_{il} will be accounted for in the model and the new resulting error term uncorrelated with FA_{jl}. Thus, let $\hat{r}_{jl}$ be the errors obtained from a first-stage linear regression of the ***j***-th FA indicator on relevant cost variables:

$$\frac{N^*_{jl}}{Pop_l} = (C_{jl}, K_{jl})'\hat{\alpha}^C_j + \hat{r}_{jl} \quad (6)$$

The model in equation (2) can then be rewritten as follows:

$$\Pr(FI_{il} = h \mid Z) = \Pr\left(\lambda^{2SRI}_h < \sum_k \beta^{2SRI}_k X_{kil} + \delta^{2SRI}_j FA_{jl} + \delta^r_j \hat{r}_{jl} + \sum_s \gamma^{2SRI}_s d_s + e^{2SRI}_{il} < \lambda^{2SRI}_{h+1}\right) \quad (7)$$

which, under the assumptions in (3),can be estimated via Ordered Logit.

The approach illustrated above, the 2-Stage Residual Inclusion (2SRI) method, is superior to classical 2-stage instrumental variable methods in non-linear models such Ordered logit. Terza, Basu, and Rathouz (2008) show that, while classical "2-stage" approaches can produce inconsistent estimates, the 2SRI method produces unbiased and consistent estimates for a broad family of non-linear estimators. Application of such method can be found in several areas such as policy analysis (Alvarez and Glasgow, 1999), health economics (see for example Terza, Basu, and Rathouz; 2008) and marketing (Petrin and Train, 2010).

3. Data and Estiamtion

3.1. Data Sources and Variable Definitions

The database used in the estimation is obtained combining different data sources. Data on households' FI status and their characteristics come from two years of individual-level observations of the Current Population Survey -Food Security Supplement (CPS-FSS) of the U.S. Census Bureau and Bureau of Labor Statistics, December 2004 and 2005.[15]

The CPS-FSS reports different measured of household FI. The survey respondents are asked a series of eighteen questions related to the availability of food in their household, including limitations in food consumptions and the number of meals skipped, distinguishing for disruptions in eating habits for both adults and children in the households. In base of the household's responses to these questions (or to a subset of it) "raw" FI indicators are constructed and then manipulated to obtain Rasch-based scores which are then coded to obtain discrete FI indicators.[16] The indicators used in this analysis are the households' "Food security summary status, 12-month" and "Food security summary status, 30-day," referred to as FI-12m and FI-30d, respectively. The categories chosen as statuses of FI are Food Secure (FI = 0), which includes High and Marginal Food Security statues (FS); Low Food Security (LFS; FI = 1), and Very Low Food Security (VLFS; FI = 2).

While the public access files of the CPS-FSS do not disclose the exact location of the individuals' surveyed, most observations have state and MSA-code identifiers attached to them, which allows the CPS-FSS data to be matched with other, market level databases. Only observations presenting valid entries of both FI indicators and geographic indicators are retained in the database.

Data on traditional food retailers' location were obtained from the County Business Pattern (CBP) of the U.S. Census Bureau/Bureau of Labor Statistics (BLS). The industries considered are NAICS 445110: Grocery Stores; NAICS 445120, Convenience Stores, and

---

[15] The choice of using the years 2004 and 2005 was made on two grounds. First, as the CPS-FSS had to be matched to MSA-level data coming from other sources, and MSA definitions changed across years *and* across databases, the data for the years 2004 and 2005 allowed for a perfect match with other MSA-level data. Second, as data on Wal-Mart supercenters location is only available until January 2006, through T. J. Homes Store location database (Holmes, 2010), no subsequent years were used.

[16] The illustration of how the Food Security indicators are obtained from the original eighteen measures collected is not immediate and it is left out for brevity. See Nord (2002) for more details.

NAICS 447110, Convenience stores with Gas Station. Data on Wal-Mart Supercenters' stores number and location are obtained from T. J. Holmes database (Holmes, 2010). County-level CBP data and the Wal-Mart data were aggregated to the MSA level to match the geographic indicators of the CPS-FSS.[17]

Using the information on the number of employees contained in the CBP, grocery stores' establishments with less than 50 employees are combined with the number of establishments belonging to NAICS 445120 to obtain a proxy of the number of small (proximity / low assortment) food stores; the food access measure SMALL is then obtained divided this number by total population (in tens of thousands), obtained from the Population Estimates Program (PEP). A proxy for access to supermarkets and other traditional food outlets defined GROC, is obtained dividing the number of NAICS 445120 establishment with 50 or more employees by MSA-level population in hundreds of thousands (PEP); the variable GSCNV, a proxy for access to outlets characterized by limited accessibility (as for cars are necessary) and assortments is obtained dividing the number of NAICS 447110 establishments by population in hundreds of thousands. Lastly the variable WMSC is obtained dividing the aggregated, MSA-level number of Wal-Mart Supercenters by population in millions.

Household-level variables from the CPS-FSS survey are used to control for household's characteristic: age of the household head (AGE), number of children in the household (CHILD), highest education level in household (three binary variables indicating, respectively High-School, HIGHSC; Some College, SOMCOL and Bachelor degree or more, COLMOR), gender of the household head (GEND: 1=Male, 0=Female), and a series of binary variables accounting for race of the household head (Black, Asian and Hispanic, respectively), home ownership (HMOWN), single-head household (SINGLEH), unitary household (SINGLUN), and for the presence of any non-citizens (NOCITIZ), unemployed (UNEMPL)

[17] CBP data at the MSA –level could not be directly used due to discrepancies in some of the classifications across the two databases.

and disabled (DISABL) individuals in the household. Interval regression on the 16-level household income bracket indicators in the CPS, on a series of demographic predictors and variables assessing type of employment, is used to obtain a proxy for the income level of the household adjusted by household size. Such proxy (INC_PR) was dividing the predicted value from the interval regression by household size.[18] Lastly, state-level fixed effects are obtained using the state identifiers in the CPS-FSS and are included in the model to control for households' unobserved heterogeneity.

Households showing invalid entries of the demographic variables illustrated above and of the 16-level household income indicators are also dropped from the database. The data points used in the estimation consisted of 36,887 observations (18,356 for the year 2004 and 18,531 for 2005). From this database, which will be referred to as the Full sample, a subsample including only households whose income is below the 185% of the current poverty threshold, referred to as the low income (Low-Inc) sample. The number of observations for this sample is 7,487 (3,817 for the year 2004 and 3,670 for 2005).

Summary statistics for the FI indicators and the FA variables for the different years and different subsamples are reported in Table 1. The values show that, as expected, the percentages of FI households are much larger in the low-income sample than in the full sample. In particular, 27.41 % of the households in the Low-Inc sample experienced FI in the 12-months prior to the survey (18.45 % LFS and 8.96 % VLFS) vs. 9.51 in the full sample

---

[18] The variables used in the regression to predict household income are the demographic indicators described in the main text, indicators for the head's employment (civil employment, part-time and full-time employment, number of hours worked), and different size of the areas where they live in. State-level fixed effects are also used. Borrowing from Cameron (1988) and Cameron and Huppert (1989), we model the probability of household income being between two thresholds represented by two consecutive levels of the 16-brackets household income indicator in the CPS, can be represented by the difference of two standard normal CDFs (after appropriate standardization of the intervals), and that the relationship between income and the covariates is linear. The model is estimated using Maximum Likelihood assuming log-normality of the errors to ensure only positive income values. The predicted values where then converted in levels and divided by number of individuals in the household, obtaining the variable INC_PR.

(6.48 % LFS and 3.03 % VLFS) while more modest percentages are recorded for Fi occurrences in the 30 days prior to the survey. The FA variables show a picture consistent with the evidence that Wal-Mart tends to locate its food stores preferentially in areas where there may be a higher concentration of low-income individuals, since the average number of WMSC in the full sample is 10% lower than in the Low-Inc sample (2.65 vs. 3.04, respectively). The sample averages for the other food stores' density are instead relatively similar across full and low-income samples, although the presence of small grocery and convenience stores (SMALL) is 10% lower for the Low-Inc subsample. Lastly, a list of all the household-level variables and summary statistics for the two samples are illustrated in the top half of Table 2.

3.2. Identification Strategy

To implement the 2SRI method, one needs to find viable exogenous variables to be used in the first stage regressions. The rationale behind the choice of what variables (i.e. the identification strategy) to use for each different FA measure is discussed below. Although such variables are referred to as "instruments" the reader should be aware that the 2-SRI method adopted here differs from standard IV methods.

Our strategy to account for the endogeneity of WMSCs[19] uses two facts that are based on the company's unique store location strategy. First, as the company's expansion into food retailing capitalizes on converting its mass merchandize Discount Stores (DSs) into supercenters (see Bonanno 2010), the lagged number of DSs is used as instrument for WMSCs as it represents a good predictor of SCs density (approach similar to that used by Basker and Noal, 2009). Furthermore, as Holmes (2011) shows, another major driver of the

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

company's location decision is the proximity of a distribution center, allowing the company to capitalize from economies of density, consistently with the Hub-and-spoke logistic system of the company (Walton and Huey, 1992). Thus, a weighted average of the inverse of distance form food distribution centers (see Bonanno, 2010, for more details), whose location is available in Holmes (2010) database, is used as additional instrument for the density of Wal-Mart Supercenters.

Our identification strategy for the non-Wal-Mart measures of food access, is based on the simple notion that, given the size of a market (i.e. the potential demand), food retail establishments would locate preferentially in areas where pre-existing infrastructures provide ease of transportation and implementation of logistics structure, where the price of land may be lower, and where the prices of operation specific (and other location-specific) costs are smaller. To this end, we use historical and current information on infrastructure to capture exogenous (to current changes in food security) variation in store-density across retailers, as well as store-type specific sources of costs. To account for pre-existing infrastructures which could facilitate transportation and logistics operations we use the state-level miles of federal highways in 1950 (U.S. Department of Transportation, Federal Highway Administration, 1950); as small convenience and grocery stores tend to locate mostly in urban areas, the length of 1950 highways in Urban areas is used as instrument for SMALL, while as larger operations (as well as gas stations) may benefit from the capillarity of the highway system, the length of 1950 highways in secondary areas are used as instruments for GROG and GSCNV. The proxy for land prices is the MSA-level number of vacant housing units, from the 2000 Census, (U.S., Bureau of Census, 2000) divided by square miles of land, from the Census Gazetteer of counties (U.S. Bureau of Census, 2001). Population density (thousands of individuals by square mile), is also accounted for in the first-stage regressions of the three non-Wal-Mart FA variables.

Store-type specific cost shifters are energy prices collected from the U.S. Department of Energy, Energy Information Administration. State-specific wholesale (refiner) gasoline price ($/gal), is used for GSCNV, annual state-level price of electricity for commercial use (KW/h) is used for both SMALL and GROC, while as larger stores necessitate of more frequent delivery of goods and they may operate their own fleet, the "On-highway" price of diesel (all types) in $/gal is used as instrument for GROC. Lastly, as some states set very low (as low as zero) minimum corporate tax rates to attract small businesses, the minimum corporate tax rate for the lowest net income level is used as additional instrument for SMALL; similarly, as large companies will be discouraged to operate in areas where the business tax rates for large operations are higher, the corporate tax rate for the highest net income level is used as additional instrument for GROC. Both variables come from the Tax Foundation of the U.S. Bureau of Census. A list of all the instruments used and some summary statistics are illustrated in the bottom half of Table 2.

3.3. Tests and estimation

One advantage of the 2-SRI method is that the significance of the estimated coefficient associated with the residuals from the first-stage regression indicates whether endogeneity was present in the original model, following the same rationale of the classical Hausman (1978) test used in linear models. However, no formal method to determine the validity of the instruments used for each of the FA variables in the first stage regressions exists. In order to have an indication of the validity of the instruments, a linear version of the model was estimated via IV methods (Generalized Method of Moments – GMM) using the Rash scores as dependent variables and the orthogonality condition (of the instruments to the error terms) evaluated via Hansen's (1982) *J*-tests, while Staiger and Stock (1997), rule of thumb (the value of an *F*-statistic of a test for the joint significance of the instruments exceeding 10) is used to establish whether the model is affected by "weak instrument" problem.

In sum, two samples of data were used: one including all the households with valid FI entry and geographic indicators, and one including only households below 185% of the poverty level. Discrete 12-month and 30-day food insecurity indicators were used as dependent variables. As the correlation between the FA variables is large and most of the instruments used to correct for their endogeneity are at the state-level, estimating a model where the FA variables were used simultaneously was not feasible due to problems of multicollinearity. Thus, 16 models (combining each of the FI indicators with the FA variables) where estimated first via Ordered Logit, and, after testing for the presence of endogeneity bias, the models were re-estimated via 2-SRI method using the residuals of the first-stage regressions which adopted the set of exogenous variables illustrated above.[20] All data manipulation and estimation were performed in STATA v. 11.

4. Empirical Results

4.1 OL and 2SRI-OL estimates

The results of ordered logit estimation of equation (3) for the full sample, not accounting for endogeneity of the FA variables are presented in Table 3. Generally speaking, the use of different FA variables does not impact the overall performance of the model (the pseudo R-squared show approximately the same values of 0.13, and the likelihood ratio tests for the joint significance of the coefficients shows similar values across models).

The estimated coefficient for WMSC is positive (0.0150) and significant at the10% level, while the presence of GSCNV seem not to have an impact on the likelihood of being FI (its coefficient, -0.0241, is not statistically significant). The relationship between FI and the

[20] The model was re-estimated using binary FI indicators as dependent variables, combining households showing LFS and VLFS status. The results of these models, estimated via logit and 2-SRI/logit were virtually identical to those which will be illustrated in the main text and therefore excluded. Furthermore, the use of a more flexible estimator, the generalized ordered logit (Williams, 2006), which relaxes the proportional odds assumption of the Ordered Logit, was attempted. As convergence was in many cases impossible to achieve, the partial results obtained are excluded from the manuscript.

density of proximity and grocery stores (SMALL and GROG) is negative and statistically significant (the estimated coefficients are, respectively, -0.1255 and -0.0137, both significant at the 1%). If such results were unbiased they would indicate that a higher concentration of WMSCs would increase the likelihood of being FI while that of traditional food stores would decrease it. In sum, in spite the price decreasing effects due to the company's presence and its strategic location in low-income area, the company's presence would increase the likelihood of being FI, while the presence of traditional outlets (grocery and convenience stores) helps reducing the likelihood of a household being FI. Similar patterns of results are obtained in the low-income sample (not reported in table form for brevity): the estimated coefficient for WMSC is positive (0.0188) and significant at the 10% level, the coefficients for SMALL and GROC are negative and statistically significant at the 1% level (the estimated coefficients are, respectively, -0.1619 and -0.0165), while that of GSCNV is, again, negative and not statistically significant (-0.0299).

Before assessing in detail the endogeneity bias of the results, it should be noted that the estimated coefficients for the different demographics used in the model are mostly consistent with the characteristics of FI households highlighted in other studies (see for example Daponte and Stephens, 2004; Bartfeld and Dunifon, 2006; Nord et al. 2004, Nord et al. 2010). Furthermore, sign, magnitude and overall significance of the estimates is largely unaffected by the use of different food access measures. The factors affecting the likelihood of being FI in a negative and statistically significant way are: age of the household head, household head being male, higher levels of education in the household (in particular the COLMOR and SOMCOL dummies), the predicted average income of a household member,[21] and house ownership of a home. Factors showing instead a positive relationship with the likelihood of

[21] Models including a household aggregate income indicators, as well as a full set of dummies capturing all the income brackets were also estimated. In both cases the Pseudo R-squared increased but the significance of most of the household-level variables decreased largely. The results from these different specifications of the model are excluded for brevity but are available upon requests.

being FI are single head households, Asian and Black ethnicities of the head, living in a single-unit household, number of children in the household, as well as the presence of unemployed and non-citizens in the household.

Table 4 reports the OL/2-SRI estimates of the equation (7), which accounts for endogeneity bias, along with the results of the tests for the validity of the instruments performed on the GMM results using Rasch scores (full-sample only).[22] The results of Hansen's *J* test show that the hortogonality condition is only satisfied in one case, as three out o for of the *p*-values are below the 0.1 customary rejection thresholds, (the *p*-values are, respectively, 0.058 for the model including WMSC, 0.5421 for SMALL, 0.0011 for GSCNV, and 0.0228 for GROC). The *F*-statistics for the joint significance of the instruments are large enough to ensure that the instruments are not weak (in all cases the values exceed the rule of thumb of 10 set by Staiger and Stock, 1997). In sum, the results of the 2SRI models discussed below could still be biased in the same direction as the OL ones.

The values presented in Table 4 show that the residuals from the first-stage regressions of the FA variables are statistically significant for three out of four variables considered; in detail, the estimates of the FA coefficients for WMSC, SMALL and GSCNV in table 3 were all likely to be biased; in spite of the coefficient for the residual of GROC not being statistically significant, the *p*-value associated with the Sargan *C* was 0.0001 which suggests enedogeneity bias. The coefficient of WMSC switches sign, becoming negative (-0.0428) but not statistically significant, while the negative coefficient for SMALL is statistically significant and show a magnitude 90% larger than the OL estimate (-0.2378). Interestingly, the sign of both GSCNV and GROG coefficients switch from negative to positive, although the second is not statistically significant. In particular, a higher concentration of gas stations

[22] The endogeneity of the FA variables was detected using Rasch scores as dependent variables, estimated via GMM, by means of *C* statistics, obtained as difference of two Sargan statistics (Hayashi, 2000, pg. 232). The results of these tests (as well as detailed results of the GMM regressions across samples and FI indicators) are omitted for brevity, unless specifically needed.

with convenience stores shows a positive association with the likelihood of a household showing positive values of FI-12m, the estimated coefficient being 0.2470 and statistically significant that the 1% level. These results suggest that store proximity, more than assortment, may be a key component in improving the likelihood of being FS, as a larger presence of stores which are usually characterized by ease of access (SMALL) seems to reduce the likelihood of being FI, while that of stores which are harder to reach (GSCNV) seem to increase such likelihood. It should also be noted that the inclusion of the instruments for the first stage regressions does not affect the performance of the model (the values of the Pseudo R^2 are unchanged) and the behavior of the estimated parameters associated with the households' characteristics, resemble closely those in table 3.

Before illustrating the actual impact of FA on the likelihood of being FI -- i.e. the marginal effects of the FA variables in the likelihood of experiencing LFS and VLFS -- the detailed results of equation (7) estimated using WMSC as FA variable across the different subsamples (Full and Low-inc) and FI indicators (FI-12m and FI-30d), which are presented in table 5 are discussed. The estimated parameters show that, in the case of WMSC, a higher concentration of the company's stores is negatively, although weakly, associated with FI, with a statistically significant coefficient (at the 10% level) only for FI-30d indicator and the low-income household's sample. Also, the estimated coefficients arte larger in the Low-Inc sample, although the lack of significance of both the coefficients obtained for the Full sample (the estimated coefficients are -0.0239, F1-12m/Full; -0.0348 FI-12m/Low-Inc; -0.0204, FI-30d/Full; all of them not statistically significant, and -0.0490, FI-30d/Low-inc). These results suggest that the presence of a higher density of the company's stores could be beneficial to reduce the risk of experiencing FI among low-income individuals, however such effect is weak and only limited to recent occurrences, indicating that the expected combined effects of low prices and strategic location in underserved areas may be mitigated by the need for transportation to reach the stores.

Some changes in estimated coefficients of the demographic variables indicate that the demographic profile of FI households can be different across FI definitions and samples. In particular, the role of most demographic indicators is weakened among low-income household, in particular, age of the household head, some of the ethnic profile, secondary education, being a single unit household and the presence of non-citizens in the household. It should also be noted that the coefficients of the income indicator lose statistical significance and in one occurrence show perverse sign while being statistically significant at the 10% level (only occurrence among the 16 estimated models is WMSC model using the FI-12m indicator as dependent variable).

4.2 Marginal Effects

Table 6 presents a summary of the OL/2SRI estimates of the FA variables across samples and measures of FI, along with the respective marginal effects. The estimated marginal effects associated with WMSC indicate that, if the number of supercenters per 1,000,000 individuals increases by one unit (corresponding to approximately a 35% increases in number of stores) on average, among low-income households, the likelihood of being food insecure during the 30-day period prior to the survey is reduced by -0.52% (the marginal effects for the other FI indicators and samples are not discussed since they are not statistically significant). The increase of 1 store per millions of people results in a decrease of 0.26% in the likelihood to experience LFS; and an additional - 0.26 % of that of experiencing VLFS. In other words, if the availability of Wal-Mart supercenters was increased by 1 additional unit for (approximately) 170,000 people, low-income household would have been 1% less likely to be FI in the month prior to the survey.

The presence of small (proximity) stores seems to have a significant negative effect on the likelihood of being FI across measure and samples. The coefficients vary from -0.1956, (FI-12m/Low-Inc sample), and -0.2378 (FI-12m, Full sample). The marginal effects associated with this variable indicate that, for a 1-unit increase in small food outlets per 10,000 individuals (corresponding to doubling their numbers, in the full sample and a 110% increase in the Low-Inc sample) household would have experienced (on average) a 1.8% increase in the likelihood of being FS during the last 12-months (1.14% of the likelihood of being LFS and an additional -0.66% of that of being VLFS), effect which increases to -3.66 % for the Low-Inc sample (-2.11% of LFS and an additional -1.55% of that of being VLFS). Considering instead FI-30d, the marginal increases in the likelihood of being FS obtained for the full and low-income samples are 0.79 % and 2.41% respectively. In sum, these results indicate that increasing the number of proximity stores, helps providing access to food to low-income individuals who would otherwise be underserved, as these households may benefit from the advantages coming from easy to reach locations.

The estimated coefficients for GSCNV show that the presence of this outlet plays a worsening role in determining the likelihood of being FI, effect which is particularly marked among low-income households. The marginal effects associated with GSCNV indicate that, for a 1-unit increase in gas stations with convenience stores per 100,000 individuals (i.e., approximately, a 55% increase) would have caused a decrease in the likelihood of being food secure during the past year equal to 1.87% decrease (the likelihood of experiencing LFS and VLFS are 1.18% and an additional 0.69%, respectively) which amounts to -5.34% among low-income households (the likelihood of experiencing LFS goes up by 3.08 %, while an additional 2.25% is recorder for that of being VLFS). A similar effect, although more modest in magnitude is also obtained on the likelihood that households were FS in the 30 day period prior to the survey: the marginal effects would have been that of an average 1.08% decreased in the likelihood of being FS (+0.56% of the likelihood of being LFS and an additional +0.53% of that of being VLFS), effect which increases to +2.99 % for the Low-Inc sample (for an increase of 1.5% the likelihood of being LFS and an additional 1.50% of that of being VLFS). These results indicate that as the location of these stores may not be convenient, and that as higher price and lower quality could be associated with these outlets, their presence may cause an increase in the likelihood of being FI. In other words, an increase in the presence of this outlet could create both direct (prices and travel) and indirect (necessity to go to other outlets to complete their food basket) costs on households, leading to a decrease in the likelihood of being FS. Lastly, the effect of GROC on the likelihood of being FI, is statistically significant only in one instance, for FI-12m, low-income sample, showing a positive coefficient of 0.0871, associated with a marginal decreasing effect of the likelihood of being FS of 1.63%, for an approximate increase of the number of these stores by 19%. A combination of higher prices and perhaps access which requires a means of transportation may be at the source of this effect.

In sum, focusing the attention on low-income households only, larger presence of stores which are easily accessible (small grocery and convenience stores) and, in a much more modest measure, of stores offering a higher variety of low priced foods (for which lower prices offset travel costs), are the only two alternatives likely to mitigate the risk of being FI. Higher concentrations of stores that offer limited assortment, and hard to access location (gas-convenience stores) is instead associated with higher likelihoods of experiencing food insecurity, with some modest magnifying effect also coming from traditional large retailers (grocery stores).

5 Concluding remarks

The preliminary results presented in this paper show that the presence of different food stores has a role in impacting the likelihood of households' food security status. In particular, as one considers different income levels, only the presence of proximity stores (small grocery and convenience stores) and, at a much more limited extent low-priced alternatives (Wal-Mart supercenters) seem to play an effective role to alleviate this issue. In contrast, the presence of hard food outlets, which may offer limited assortments, hard to reach location (and perhaps higher prices) may jeopardize food security. We find also limited evidence of medium and large grocery stores to have a weak impact on worsening the likelihood of being food secure.

Future efforts will focus on refining our identification strategy for the FA variables, trying to include more variations in the instruments used as to avoid issues of multicollinearity and include more than one FA variable in the model simultaneously.

REFERENCES

Alvarez, R.M., and G. Glasgow. 1999. "Two-Stage Estimation of Nonrecursive Choice models." *Political Analysis* 8:147–165.

Alwitt, L.F., and T.D. Donley. 1997. "Retail Stores in Poor Urban Neighbourhoods." *The Journal of Consumer Affairs* 31:139–163.

Asplund, M. And S. Rickard,.1999. "The Number of Firms and Production Capacity in Relation to Market Size." *Journal of Industrial Economics* 47(1): 69-85.

Ball K., A. Timperio, and D., Crawford. 2008. "Neighborhood Socio Economic Inequalities in Food Access and Affordability." *Health and Place* (15):578–585.

Bartfeld, J., and R. Dunifon, R. 2006. "State-level predictors of Food Insecurity among Households with Children." *Journal of Policy Analysis and Management* 25(4):921-942.

Basker, E. and M. D. Noel. 2009. "The Evolving Food Chain: Competitive Effects of Walmart's Entry into the Supermarket Industry." *Journal of Economics and Management Strategy*.

Beherens, A. 2010. "Maryland's Grocery Store Tax Credit Bill Could Improve Food Deserts'" Livable futures. March 25th, 2010. Accessed April 15th 2010. Available at http://www.livablefutureblog.com/2010/03/marylands-grocery-store-tax-bill-could-improve-food-deserts/

Berry, S.T. 1992. "Estimation of a Model of Entry in the Airline Industry." *Econometrica* 60: 889–917.

Bonanno, A. 2010. "An Empirical Investigation of Wal-Mart's Expansion into Food Retailing." *Agribusiness: An International Journal* 26(2):220–242.

Bonanno, A. and R. A. Lopez. 2009. "Competition effects of Supermarket Services." *American Journal of Agricultural Economics* 91(3): 555–568.

Borjas, G. 2004. "Food Insecurity and Public Assistance." *Journal of Public Econo*mics 88 (7):1421–1423.

Bresnahan, T.F., and P.C. Reiss. 1991. "Entry and competition in concentrated markets." *Journal of Political Economy* (99):977–1009.

Cameron, T. A. 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." *Journal of Environmental Economics and Management*, 15(4):335-379

Cameron, T. A. and D. D. Huppert. 1989. "OLS versus ML Estimation of Non-Market Resource Values with Payment Card Interval Data." *Journal of Environmental Economics and Management*, 17 (3):230-246

Cotterill, R.W., and A. W. Franklin. 1995. "The urban grocery store gap." Food Marketing Policy Issue Paper 8. Food Marketing Policy Center, University of Connecticut.

Daponte, B. O., and M. Stephens. 2004. "The Relationship Between Food Assistance, the Value of Food Acquired, and Household Food Supply." Joint Center for Poverty Research Working Paper 351. Chicago: The Center.

Ellickson, P. B. 2005. "Does Sutton Apply to Supermarkets?" *RAND Journal of Economics*. 38(1): 43-59.

Ellickson, P. B. 2007. "Quality Competition in Retailing: A Structural Analysis" *International Journal of Industrial Organization* 24:521– 540.

Food Marketing Institute. 2007. Top U.S. Supermarket and Grocery Chains. Available at http://www.fmi.org/docs/facts_figs/faq/top_retailers.pdf. Accessed 12/07/2007

Gibson-David, C.M., and E.M. Foster. 2006. "A Cautionary Tale: Using Propensity Scores To Estimate the Effect of Food Stamps on Food Insecurity," *Social Service Review* 80: 93-126.

Goetz, S. J., and H. Swaminathan. 2006. "Wal-Mart and Family Poverty in US Counties." *Social Science Quarterly* 83:211-225.

Gundersen, C. and V. Oliveira. 2001. "The Food Stamp and Food Insufficiency." *American Journal of Agricultural Economics* 83(4): 875-887.

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

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4):1029-54.

Hayashi F. 2000, *Econometrics*, Princeton, NJ: Princeton University Press.

Hausman, J.A., 1978. "Specification Tests in Econometrics." *Econometrica* 46:1251–1271.

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

Holmes, T. J. 2011. "The Diffusion of Wal-Mart and Economies of Density." *Econometrica*, 79: 253–302.

Jensen, H. H. 2002. "Food Insecurity and the Food Stamp Program." *American Journal of Agricultural Economics* 84(5): 1215-1228.

Jia, P. 2008. "What Happens when Wal-Mart Comes to Town: an Empirical Analysis of the Discount retail industry." *Econometrica* 76(6): 1263-1316.

Kabbani, N.S., and M.Y. Kmeid. 2005. "The Role of Food Assistance in Helping Food Insecure Households Escape Hunger," *Review of Agricultural Economics* 27: 439-445.

Kilkenny, M. and Thisse, J. 1999. "The Economics of Location: A Selective Survey." *Computers and Operations Research*, 26(14), 1369-1394.

King, R.P, E.S. Leibtag, and A.S. Behl. 2004. "Supermarket Characteristics and Operating Costs in Low-Income Areas." USDA-ERS Agricultural Economic Report N. 839.

Leibtag E. L. and P. R. Kaufman. 2003 "Exploring Food Purchase Behavior of Low-Income Households How Do They Economize?" ERS-USDA. Agriculture Information Bulletin 747-07 www.ers.usda.gov/publications/aib747/aib74707.pdf

Martinez S.W. 2007. "The U.S. Food Marketing System: Recent Developments 1997-2006" USDA/ERS Economic Research Service Report Number 42.

Mazzeo, M.J. 2002. "Product choice and oligopoly market structure." *The RAND Journal of Economics* 33:221–242.

Moore, L.V., and A. V. Diez Roux. 2006. "Associations of Neighborhood Characteristics with the Location and Type of Food Stores." *American Journal of Public Health* 96 (2): 325–331

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

Nord, M. 2002. "A 30-Day Food Security Scale for CPS Food Security Supplement data" E-FAN No. (02015) 23 pp, August 2002. Available on-line at http://www.ers.usda.gov/publications/efan02015/efan02015.pdf accessed 1/10/2010.

Nord, M., and A. M. Golla. 2009. "Does SNAP Decrease Food Insecurity? Untangling the Self-Selection Effect" USDA-ERS Report N.85 October 2009. Available on-line at http://www.ers.usda.gov/Publications/ERR85/ERR85.pdf. accessed 1/10/2010.

Nord, M., M. Andrews, and S. Carlson. 2004. "Household food security in the United States, 2003." Food and Rural Economics Division, ERS-USDA. Food Assistance and Nutrition Research Report No. 42.

Nord, M., A. Coleman-Jensen, M. Andrews, and S. Carlson. 2010. "Household Food Security in the United States, 2009." Economic Research Report No. (ERR-108) 68 pp. Available at http://www.ers.usda.gov/Publications/err108/

Petrin. A.. and K. Train. 2010. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research* 47(1): 3-13.

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

Rose, D., G. Gundersen, and V. Oliveira. 1998. "Socio-economic Determinants of Food Insecurity in the United States: Evidence from the SIPP and CSFII datasets. Food and Rural Economics Division, ERS-USDA Technical Bulletin No.1869.

Rose, D., and R. Richards. 2004. "Food Store Access and Household Fruit and Vegetable Use among Participants in the US Food Stamp Program." *Public Health Nutrition* 7(8): 1081-8.

Scott Kantor, L., 2001. "Community Food Security Programs Improve Food Access." *Food Review*, 24 (1): 20-26.

Sutton J. 1998. *Technology and Market Structure: Theory and History*: Cambridge, Massachussts, MIT Press 676 pp

Staiger, D., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557-586.

The Tax Foundation. Corporate Income Tax Rates 2000-2011. Available at http://www.taxfoundation.org/taxdata/show/230.html Accessed 06/01/2010.

Terza, J. V., A. Basu, and P. J. Rathouz. 2008. "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics* 27: 531-543.

U.S. Bureau of Census. Population Estimates Program: 2004, 2005

________ 2001. Gazetteer of counties: Year 2000 Available at http://www.census.gov/geo/www/gazetteer/places2k.html Accessed 04/01/2006

U.S. Bureau of Census and U. S. Bureau of Labor Statistics, Current Population Survey -Food Security Supplement 2004, 2005 http://www.bls.gov/cps/.

________ County Business Pattern 2004, 2005.

U.S. Department of Energy. Energy Information Administration. Refiner Gasoline Prices by Grade and Sales Type – Regular

__________ Current and Historical Monthly Retail Sales, Revenues and Average Revenue per Kilowatthour by State and by Sector (Form EIA-826) - commercial prices.

__________ Weekly retail gasoline and diesel prices: Diesel - all types. http://www.eia.doe.gov/dnav/pet/pet_pri_gnd_a_epd2d_pte_dpgal_a.htm

U.S. Department of Transportation, Federal Highway Administration, Highway Statistics. 1950. http://www.fhwa.dot.gov/policy/ohpi/hss/index.cfm

Volpe. R. J. III. and N. Lavoie, 2008. "The Effect of Wal-Mart Supercenters on Grocery Prices in New England, Richard." *Review of Agricultural Economics* 30 (1):4-26.

Walde, P. E. 2007 "Measuring the effect of Food Stamps on Food Insecurity and Hunger: research and Policy Considerations." *Journal of Nutrition* 137(2): 307–310.

Wal-Mart Stores Inc. 2011. 2010 Annual Report. Retrieved March 29, 2011, from http://walmartstores.com/sites/annualreport/2010/

Walton, S., and J. Huey. 1992. *Sam Walton, Made in America: My Story*. New York City: Doubleday Publisher.

Williams, R. 2006. "Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables." *The Stata Journal*, 6, 58–82.

Yen, S. T., M. Andrews, Z. Chen, and D. B. Eastwood. 2008. "Food Stamp Program Participation and Food Insecurity: An Instrumental Variables Approach," *American Journal of Agricultural Economics* 90(1):117-132.

Zenk, S., Schulz, A.J., Israel, B.A., James, S.A., Bao, S., Wilson, M.L. 2005. "Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit." *American Journal of Public Health* 95(4): 660–667.

Table 1. FI Indicators and Food Access Variables: Descriptive Statistics across samples.

Variable	Full Sample				Low-Income Sample			
		% FS	% LFS	%VLFS		% FS	% LFS	%VLFS
FI-12m		90.49	6.48	3.03		72.60	18.45	8.96
FI-30d		95.65	2.36	2.00		87.39	6.86	5.74
	Mean	St. Err	Min	Max	Mean	St. Err	Min	Max
WMSC	2.65	4.93	0.00	30.06	3.04	5.23	0.00	30.06
SMALL	0.97	0.98	0.01	4.86	0.90	0.94	0.01	4.86
GSCNV	1.83	2.39	0.04	17.08	1.86	2.46	0.04	17.08
GROC	5.16	7.44	0.04	54.49	5.12	8.69	0.04	54.49

Legend and data sources:

FI-12m: 12-month general-scale FI status (0=FS; 1=LFS; 2=VLFS). Source: CPS-FSS

FI-30d : 30-day general-scale FI status: (0=FS; 1=LFS; 2=VLFS). Source: CPS-FSS

WMSC: Number of WM Supercenters/ 1,000,000 population. Source: CBP / PEP

GSCNV: Number of NAICS 447110 Stores /100,000 population. Source: Holmes(2010) Database / PEP

SMALL: Number of NAICS 445120+NAICS 445110 stores <50 employees /10,000 Population. Source: CBP / PEP

GROC: Number of NAICS 445110 stores >=50 employees /100,000 population. Source: CBP / PEP

Table 2. Demographic Control Variables, Instruments and Sources

Variable	Description	Full Sample		LowInc Sample	
		Mean	St.dev.	Mean	St.dev.
Demographic variables – Continuous					
AGE	Age of Household Head	43.07	12.38	38.95	13.16
INC_PR	Estimated Average Income of a Household Member	26.79	15.34	15.14	10.49
CHILD	Number of children in household<18 year	0.72	1.06	1.17	1.33
Control Variables - Discrete		Frequency of 1			
GEND	Head is Male		57.13		47.87
HIGHSC	Max Educational Attainment: High school		24.84		36.56
SOMCOL	Max Educational Attainment: Some College		29.87		29.40
COLMOR	Max Educational Attainment College Degree or Higher		37.82		12.91
ASIAN	Race is Asian Household Head		3.55		3.35
BLACK	Race is Black Household Head		10.62		18.29
HISP	Hispanic Ethnicity Household Head		10.71		25.94
HMOWN	Own living quarters (for household)		70.15		41.82
SINGLEH	Single Head Household		16.38		32.31
SINGLUN	Single Unit household		29.28		28.50
NOCITIZ	Non-citizen in household		12.27		24.39
UNEMPL	Unemployed in household		5.88		10.24
DISABL	Disabled in the household		0.49		0.80
FA Instruments					
distFDC	Inverse of distance from WM food distribution centers	9.53	14.93	9.12	15.09
Ndslag	Lagged density of WM DSs / 1,000,000 population	3.18	4.35	2.96	4.24
pop07_sqmi	Population density (.000/square mile)	16.88	45.27	13.19	40.28
Vacden	Number of vacant units/total square miles	18.01	17.65	16.90	17.68
Fedhw50sec	Length of Federal Aid highway - secondary system (1950)	9.39	6.43	9.75	6.31
Fedhw50urb	Length of Federal Aid highway - urban system (1950)	0.47	0.38	0.47	0.37
P_gas	State-level refiner gasoline price ($/gal)	1.59	0.20	1.50	0.21
P_diesel	Area-level diesel price (On-Highway) All Types ($/gal)	2.15	0.31	2.14	0.31
P_elect	State-level retail electricty price, commercial use (c/kWh)	8.56	2.21	8.50	2.19
tax_corp	Corporate tax rate for the highest net income level	0.06	0.03	0.06	0.03
tax_min	Corporate tax rate for the lowest net income level	0.07	0.03	0.06	0.03

Note: all the demographic variables come from the CPS-FSS

Table 3. Ordered Logit estimates of equation (3), FI-12m; Full sample

FA Variable	WMSC		SMALL		GSCNV		GROC	
FA	0.0150	*	-0.1255	***	-0.0241		-0.0137	***
	(0.0080)		(0.0330)		(0.0178)		(0.0052)	
AGE	-0.0043	**	-0.0042	**	-0.0043	**	-0.0043	**
	(0.0018)		(0.0018)		(0.0018)		(0.0018)	
GEND	-0.3481	***	-0.3469	***	-0.3467	***	-0.3463	***
	(0.0410)		(0.0410)		(0.0410)		(0.0410)	
HIGHSC	-0.2241	***	-0.2268)	***	-0.2236	***	-0.2244	***
	(0.0625)		(0.0626)		(0.0625)		(0.0625)	
SOMCOL	-0.2197	***	-0.2252	***	-0.2205	***	-0.2227	***
	(0.0651)		(0.0652)		(0.0651)		(0.0651)	
COLMOR	-1.0495	***	-1.0594	***	-1.0519	***	-1.0558	***
	(0.0882)		(0.0883)		(0.0882)		(0.0882)	
HISP	-0.2922	**	-0.2837	**	-0.2912	**	-0.2890	**
	(0.1287)		(0.1288)		(0.1287)		(0.1287)	
ASIAN	0.3280	***	0.3448	***	0.3316	***	0.3363	***
	(0.0543)		(0.0546)		(0.0544)		(0.0545)	
BLACK	0.3656	**	0.3715	**	0.3764	**	0.3750	**
	(0.1850)		(0.1850)		(0.1849)		(0.1849)	
HMOWN	-0.8953	***	-0.8978	***	-0.8941	***	-0.8946	***
	(0.0443)		(0.0443)		(0.0443)		(0.0443)	
INC_PR	-0.0193	***	-0.0185	***	-0.0192	***	-0.0189	***
	(0.0029)		(0.0029)		(0.0029)		(0.0029)	
SINGLEH	0.6876	***	0.6928	***	0.6884	***	0.6906	***
	(0.0526)		(0.0527)		(0.0527)		(0.0527)	
SINGLUN	0.5506	***	0.5442	***	0.5494	***	0.5479	***
	(0.0582)		(0.0582)		(0.0582)		(0.0582)	
CHILD	0.2029	***	0.2052	***	0.2034	***	0.2043	***
	(0.0211)		(0.0211)		(0.0211)		(0.0211)	
NOCITIZ	0.1563	***	0.1893	***	0.1620	***	0.1712	***
	(0.0571)		(0.0577)		(0.0572)		(0.0574)	
UNEMPL	0.5263	***	0.5254	***	0.5259	***	0.5256	***
	(0.0639)		(0.0639)		(0.0638)		(0.0639)	
DISABL	0.3303		0.3215		0.3304		0.3262	
	(0.2164)		(0.2167)		(0.2165)		(0.2166)	
δ_1	1.3288	***	1.2731	***	1.2782	***	1.2763	***
	(0.1706)		(0.1704)		(0.1714)		(0.1706)	
δ_2	2.6466	***	2.5916	***	2.5959	***	2.5943	***
	(0.1724)		(0.1722)		(0.1732)		(0.1723)	
Pseudo-R^2	0.1298		0.1302		0.1298		0.1299	
L-ratio$\chi^2_{(65)}$	3,577.07		3,588.00		3,575.39		3,580.33	
p-values	0.0000		0.0000		0.0000		0.0000	

*Note: *, **, and *** represent 10, 5 and 1% significance levels* – Standard errors in parenthesis. State–level fixed effects coefficients omitted for brevity.

Table 4. Ordered Logit/2SRI estimates FI-12m, full samples

FAVariable	WMSC		SMALL		GSCNV		GROC	
FA	-0.0239		-0.2378	***	0.2470	***	0.0309	
	(0.0170)		(0.0741)		(0.0876)		(0.0367)	
FA-RES	0.0489	***	0.1502	*	-0.2663	***	-0.0453	
	(0.0187)		(0.0954)		(0.0842)		(0.0365)	
AGE	-0.0042	**	-0.0041	**	-0.0041	**	-0.0042	**
	(0.0018)		(0.0018)		(0.0018)		(0.0018)	
GEND	-0.3524	***	-0.3523	***	-0.3520	***	-0.3501	***
	(0.0410)		(0.0410)		(0.0410)		(0.0410)	
HIGHSC	-0.2077	***	-0.2071	***	-0.2073	***	-0.2060	***
	(0.0632)		(0.0633)		(0.0633)		(0.0633)	
SOMCOL	-0.2039	***	-0.2038	***	-0.2025	***	-0.2027	***
	(0.0659)		(0.0659)		(0.0659)		(0.0659)	
COLMOR	-1.0420	***	-1.0458	***	-1.0418	***	-1.0421	***
	(0.0886)		(0.0887)		(0.0886)		(0.0886)	
HISP	0.1366	**	0.1585	**	0.1509	**	0.1408	**
	(0.0636)		(0.0639)		(0.0639)		(0.0637)	
ASIAN	-0.2521	*	-0.2363	*	-0.2381	*	-0.2443	*
	(0.1302)		(0.1303)		(0.1303)		(0.1303)	
BLACK	0.3557	***	0.3703	***	0.3635	***	0.3590	***
	(0.0556)		(0.0558)		(0.0557)		(0.0557)	
HMOWN	-0.8928	***	-0.8953	***	-0.8957	***	-0.8932	***
	(0.0443)		(0.0444)		(0.0444)		(0.0444)	
INC_PR	-0.0180	***	-0.0171	***	-0.0175	***	-0.0178	***
	(0.0029)		(0.0029)		(0.0029)		(0.0029)	
SINGLEH	0.6927	***	0.6956	***	0.6934	***	0.6931	***
	(0.0527)		(0.0527)		(0.0527)		(0.0527)	
SINGLUN	0.5505	***	0.5439	***	0.5452	***	0.5471	***
	(0.0582)		(0.0582)		(0.0582)		(0.0582)	
CHILD	0.2055	***	0.2077	***	0.2068	***	0.2063	***
	(0.0211)		(0.0211)		(0.0211)		(0.0211)	
NOCITIZ	0.1185	*	0.1457	**	0.1374	**	0.1299	**
	(0.0609)		(0.0613)		(0.0613)		(0.0611)	
UNEMPL	0.5317	***	0.5336	***	0.5332	***	0.5321	***
	(0.0640)		(0.0640)		(0.0640)		(0.0640)	
DISABL	0.3352	***	0.3360	***	0.3413	***	0.3383	***
	(0.2164)		(0.2166)		(0.2166)		(0.2165)	
δ_1	1.2849	***	1.2498	***	1.7468	***	1.4086	***
	(0.1754)		(0.1780)		(0.2196)		(0.1866)	
δ_2	2.6030	***	2.5684	***	3.0652	***	2.7267	***
	(0.1771)		(0.1797)		(0.2211)		(0.1883)	
Pseudo-R^2	0.1301		0.1303		0.1301		0.1300	
L-ratio$\chi^2_{(66)}$	3584.33		3592.37		3585.46		3582.52	
p-values	0.0000		0.0000		0.0000		0.0000	
Hansen J	$\chi^2_{(1)}$=1.234		$\chi^2_{(4)}$=3.09		$\chi^2_{(3)}$=6.66		$\chi^2_{(5)}$=13.1	
p-values	*p*=0.058		*p*=0.5421		*p*=0.0011		*p*=0.0228	
F-stat(weak)	853.73		8689.78		235.04		581.57	

*Note: *, **, and *** represent 10, 5 and 1% significance levels* – Standard errors in parenthesis. State–level fixed effects coefficients are omitted.

Table 5. Ordered Logit/2SRI: FI-12m and FI-30d; Full and Low-Inc samples – Wal-Mart SCs

Indicator; Sample(size)	FI-12m Full(36,887)		FI-12m LowInc(7,487)		FI-30d Full(36,887)		FI-30d LowInc(7,487)	
WMSC	-0.0239		-0.0348		-0.0204		-0.0490	*
	(0.0170)		(0.0235)		(0.0232)		(0.0299)	
WM-RES	0.0489	***	0.0652	**	0.0388		0.0644	*
	(0.0187)		(0.0259)		(0.0258)		(0.0339)	
AGE	-0.0042	**	-0.0020		0.0004		0.0018	
	(0.0018)		(0.0024)		(0.0025)		(0.0032)	
GEND	-0.3524	***	-0.2178	***	-0.3732	***	-0.2130	***
	(0.0410)		(0.0611)		(0.0579)		(0.0821)	
HIGHSC	-0.2077	***	-0.1341	*	-0.0976		-0.0352	
	(0.0632)		(0.0759)		(0.0887)		(0.1046)	
SOMCOL	-0.2039	***	-0.0370		-0.0498		0.0793	
	(0.0659)		(0.0831)		(0.0921)		(0.1128)	
COLMOR	-1.0420	***	-0.7731	***	-0.8094	***	-0.4464	**
	(0.0886)		(0.1336)		(0.1259)		(0.1791)	
HISP	0.1366	**	-0.0970		0.0160		-0.2289	**
	(0.0636)		(0.0853)		(0.0895)		(0.1162)	
ASIAN	-0.2521	*	-0.2868	*	-0.3487	*	-0.2931	
	(0.1302)		(0.1719)		(0.1950)		(0.2369)	
BLACK	0.3557	***	0.1604	**	0.1696	**	-0.0182	
	(0.0556)		(0.0765)		(0.0771)		(0.1022)	
HMOWN	-0.8928	***	-0.5143	***	-0.8807	***	-0.5087	***
	(0.0443)		(0.0625)		(0.0628)		(0.0855)	
INC_PR	-0.0180	***	0.0094	*	-0.0243	***	0.0046	
	(0.0029)		(0.0050)		(0.0041)		(0.0065)	
SINGLEH	0.6927	***	0.4357	***	0.6945	***	0.4075	***
	(0.0527)		(0.0752)		(0.0750)		(0.1026)	
SINGLUN	0.5505	***	-0.0026		0.7140	***	(0.1436	
	(0.0582)		(0.0898)		(0.0825)		(0.1201)	
CHILD	0.2055	***	0.1189	***	0.1330	***	0.0487	
	(0.0211)		(0.0265)		(0.0297)		(0.0364)	
NOCITIZ	0.1185	*	0.0930		-0.0718		-0.0817	
	(0.0609)		(0.0807)		(0.0880)		(0.1113)	
UNEMPL	0.5317	***	0.4547	***	0.6521	***	0.5244	***
	(0.0640)		(0.0831)		(0.0828)		(0.1059)	
DISABL	0.3352	***	0.3559	***	0.2296	***	0.1337	***
	(0.2164)		(0.2688)		(0.3024)		(0.3732)	
δ_1	1.2849	***	1.2901	***	2.0368	***	1.9024	***
	(0.1754)		(0.2615)		(0.2363)		(0.3399)	
δ_2	2.6030	***	2.6928	***	2.8733	***	2.7835	***
	(0.1771)		(0.2637)		(0.2378)		(0.3417)	
Pseudo-R^2	0.1301		0.0391		0.1073		0.0338	
L-ratio$\chi^2_{(27)}$	3584.33		441.67		1655.63		235.35	
p-values	0.0000		0.0000		0.0000		0.0000	
Hansen J	$\chi^2_{(1)}$=1.23		$\chi^2_{(1)}$=2.69		$\chi^2_{(1)}$=4.63		$\chi^2_{(1)}$=1.72	
p-values	*p*=0.058		*p*=0.1011		*p*=0.0315		*p*=0.1776	
F-stat(weak)	853.73		137.74		853.43		157.74	

*Note: *, **, and *** represent 10, 5 and 1% significance levels* – Standard errors in parenthesis. State–level fixed effects coefficients are omitted.

Table 6. Estimated Parameters, Marginal Effects of the food access variables on the likelihood of being Food Insecure: FI-12m and FI-30d; Full and Low-Inc samples

Indicator Sample(size)		FI-12m Full(36,887)		FI-12m LowInc(7,487)		FI-30d Full(36,887)		FI-30d LowInc(7,487)	
Estimated Coefficients									
	WMSC	-0.0239		-0.0348		-0.0204		-0.0490	*
		(0.0170		(0.0235)		(0.0232)		(0.0299)	
	SMALL	-0.2378	***	-0.1956	*	-0.2026	**	-0.2267	*
		(0.0741)		(0.1012)		(0.1031)		(0.1362)	
	GSCNV	0.2470	***	0.2851	**	0.2773	**	0.2812	*
		(0.0876)		(0.1248)		(0.1241)		(0.1666)	
	GROC	0.0309		0.0871	*	0.0543		0.0661	
		(0.0367)		(0.0517)		(0.0514)		(0.0691)	
Marginal Effects									
FI=0	WMSC	0.0018		0.0065		0.0008		0.0052	*
(FS)		(0.0013)		(0.0044)		(0.0009)		(0.0032)	
	SMALL	0.0180	***	0.0366	*	0.0079	**	0.0241	*
		(0.0056)		(0.0189)		(0.0040)		(0.0145)	
	GSCNV	-0.0187	***	-0.0534	**	-0.0108	**	-0.0299	*
		(0.0066)		(0.0233)		(0.0049)		(0.0177)	
	GROC	-0.0023		-0.0163	*	-0.0021		-0.0070	
		(0.0028)		(0.0097)		(0.0020)		(0.0073)	
FI=1	WMSC	-0.0011		-0.0038		-0.0004		-0.0026	*
(LFS)		(0.0008)		(0.0025)		(0.0005)		(0.0016)	
	SMALL	-0.0114	***	-0.0211	*	-0.0041	*	-0.0120	*
		(0.0035)		(0.0109)		(0.0021)		(0.0072)	
	GSCNV	0.0118	***	0.0308	**	0.0056	**	0.0150	*
		(0.0042)		(0.0135)		(0.0025)		(0.0089)	
	GROC	0.0015		0.0094	*	0.0011		0.0035	
		(0.0018)		(0.0056)		(0.0010)		(0.0037)	
FI=2	WMSC	-0.0007		-0.0028		-0.0004		-0.0026	*
(VLFS)		(0.0005)		(0.0019)		(0.0004)		(0.0016)	
	SMALL	-0.0066	***	-0.0155	*	-0.0039	*	-0.0121	*
		(0.0021)		(0.0080)		(0.0020)		(0.0073)	
	GSCNV	0.0069	***	0.0225	**	0.0053	**	0.0150	*
		(0.0024)		(0.0099)		(0.0024)		(0.0089)	
	GROC	0.0009		0.0069	*	0.0010		0.0035	
		(0.0010)		(0.0041)		(0.0010)		(0.0037)	

*Note: *, **, and *** represent 10, 5 and 1% significance levels* – Standard errors in parenthesis. Marginal Effects are calculated at the sample average of the variables.