Diet Deterioration and Food Retail Structure: Why are Italians Eating Less Fruits and Vegetables?

Alessandro Bonanno[1,2], Francesco Bimbo[1,3], Elena Castellari[4,5] and Paolo Sckokai[5]

[1] Business Economics Group, Department of Social Science.Wageningen University, Wageningen, The Netherlands
[2] Department of Agricultural Economics, Sociology and Education; The Pennsylvania State University, University Park, PA, USA
[3]Dipartimento di Scienze Agrarie, degli Alimenti e dell'Ambiente. Università degli Studi di Foggia. Foggia, Italy
[4]Dipartimento di Economia Agroalimentare.Università Cattolica del Sacro Cuore. Piacenza, Italy
[5]Department of Agricultural and Resource Economics, University of Connecticut, Storrs, USA

Paper prepared for the AAEA invited paper session "**Food Environment, Food Choices and Nutrition Outcomes – An International Perspective"** at the 2015 ASSA annual meeting

Boston, January 2nd -5th 2015

Corresponding Author: Alessandro Bonanno.
Email: alessandro.bonanno@wur.nl & abonanno@psu.edu
Bonanno holds lead authorship; Bimbo, Castellari and Sckokai share equally junior authorship.

Diet Deterioration and Food Retail Structure: Why are Italians Eating Less Fruits and Vegetables?

Abstract

In spite of Italy presenting one of the largest consumption of fruits and vegetables (FV) among EU Countries, the number of adult Italians consuming the recommended daily amounts of FV is declining, especially in the South of the country, were the expansion of the food retail industry has been lagging. In this article we assess whether the food retail structure affects the likelihood of adult Italians consuming five or more daily portions of FV, using 9 years of individual-level data on individuals' lifestyle, including eating habits and perceived access to supermarkets, matched with detailed regional data on the food retail structure. In our analysis we use a Two-Step Instrumental Variable Probit estimator and variables indicating the political climate of the different regions to correct for the potential endogeneity of geographic disparity in retail structures. Results show that increased access and availability of fruits and vegetables affect positively the probability of consuming the daily-recommended amounts of FV. Food retail structure's effect appears less marked for individuals declaring hurdles in accessing supermarkets. While individuals' characteristics play an important role in explaining FV consumption probability among individuals declaring no hurdles in accessing supermarkets, transportation and perceived economic conditions are some of the main determinants for individuals declaring access hurdles.

Keywords: Fruits and vegetables consumption; Food access; Retail structure; Two-Step IV Probit

JEL codes: Q18, L81

Diet Deterioration and Food Retail Structure: Why are Italians Eating Less Fruits and Vegetables?

The inverse relationship between Fruits and Vegetable (FV) consumption and the risk of non-communicable diseases is well established in the nutrition and medical literature. A recent pan-European epidemiological study carried out in the context of the European Prospective Investigation into Cancer and Nutrition (Leenders et al. 2013), shows that an increase of 200 g/day of FV consumption, leads to a 3-6% reduction of mortality risk, largely driven by a reduction in cardiovascular disease mortality, followed by cancer and other non-communicable diseases. However, the share of population consuming the World Health Organization (WHO) recommended daily amount of five portions of FV (or 400g) varies largely among European Countries.[1] In most statistics (e.g. WHO 2008; EUFIC 2012) Italy consistently appears as one of the few European countries where these guidelines are met more often, and, historically, one of the European countries with the largest FV consumption (Reina et al. 2006).

However, the share of individuals meeting the WHO recommended daily amounts of FV consumed in Italy, has declined dramatically in recent years. In the period 2005-2012, statistics produced by the Italian Regional Observatory on Health Care (Rapporto Osservasalute 2013) using Multipurpose Households Survey data from the Italian Institute of Statistics (ISTAT), find a decline of the share of individuals 3 years of age and older eating at least five daily portions of fruits and vegetables falling from 5.3% to 4.7%. Such change is uneven across the Italian territory; while regions in the North of the Country experienced almost no change or even

[1] In recent years, many policies have emerged in Europe to foster healthy eating (Capacci et al. 2012). The tools used include, among the others, aid to consumers informed choice (e.g. public information campaigns, nutritional labeling policies), changes of the market environment (e.g. school fruit schemes) and fiscal tools (e.g. tax/subsidies on food). Among the information policies to enhance FV consumption, one of the most widely adopted scheme is the so-called "5-a-day" program, which encourages consumers to reach the WHO recommended amounts, and whose effectiveness, at least in the case of UK, has been found successful (Capacci and Mazzocchi 2011).

increases in this share (as in the case of Lombardy from 4.9% to 5.8%) southern regions have experienced a steep decline (e.g. Campania form 5.9% to 3.7%; Basilicata, from 3.2% to 1.7%).

Several reasons may be behind these patterns. A first explanation is that the prevalence of Mediterranean diet, rich in fruits and vegetables appears to be declining (Berghöfer et al. 2008) with consumption of whole grains being traded for more animal-based foods and vegetable oils, leading to higher caloric intakes (Belahsen and Rguibi 2006). Bonaccio et al. (2014) highlight also how, following the last economic downturn, adoption of a Mediterranean diet became more dependent on socioeconomic factors. One may conjecture that this may have in turn created disparities in the consumption of FV across areas with different resilience to the economic crisis. This explanation seems consistent with the evidence presented in Kamphuis et al. (2006) literature review, where it is emphasized how several demographic factors, such as household income, marital status, and household size, impact FV consumption, with richer people tending to consume more FV.[2]

Additionally, the geographic disparity in FV consumption in Italy may be due in part to the different availability of fruits and vegetables and to the different structure of the food retailing industry Italian consumers are exposed to. Food retailers' presence in Italy shows an interesting dichotomy: northern areas show larger food stores and a more capillary diffusion than southern ones. In 2013, the selling area per inhabitant was 30% larger in the North-East than in the South. (Castellari and Sckokai 2014).

Several studies analyzing inequalities in FV consumption raise the issue of food access (i.e. availability of food stores with adequate supply of products at accessible prices) as a key factor influencing healthy eating (Giskes et al. 2007). The research assessing the role of perceived vs.

[2] Other studies present an even more complex picture, emphasizing how education levels, social habits and personal values of the households as factors affecting FV consumption (Pollard, Kirk and Cade 2002; Prättälä et al. 2009).

actual food store access in diets (e.g. Caspi et al. 2012; Williams et al. 2012) is growing, albeit with, in some cases, conflicting results (Cummins et al. 2005; Pearce et al. 2008). Food consumption in disadvantaged areas has been extensively analyzed, especially in the U.S. (e.g. Weatherspoon et al. 2012; Hough and Sousa, 2014). Plentiful evidence exists that areas inhabited prevalently by less-privileged individuals are characterized by poorer access to food, either in terms of number of food stores and size (e.g. Alwitt and Donley 1997; Moore and Diez Roux 2006; Powell et al. 2007) or in terms of in-store availability of healthy foods, including FV (Webber, Sobal and Dollahite 2010; Gustafson et al. 2011; Gustafson, Hankins and Jilcott 2012). An increased presence of food stores can be beneficial to consumers on multiple fronts: the expansion of larger food stores may in fact result in lower prices and larger availability of perishable products (Hawkes 2008). Additionally, individuals with limited access to food stores or with access to only small ones may be less able to acquire fruits and vegetables, as they may face higher transportation and search costs, as well as higher prices which may be due to the existing stores' monopolistic position or cost inefficiencies (King, Leibtag and Behl 2004).

With regard to the Italian case, no prior study exists linking the food environment to the consumption of fruits and vegetables. However, as hinted above a relationship may be in place. In Figure 1 we present two maps of Italy. The left panel presents the average selling area per food store in each region, obtained from data provided by Nielsen, while the right panel contains the average share of adult (> 18 years of age) Italians declaring to consume five or more portions of fruits and vegetables daily (FV5share) from the Multipurpose Household Survey (MHS) by ISTAT.[3] Those averages, referring to the period 2003-2012, show northern regions having a medium-high value of both the average store size and FV5share. Differently, regions located in the Center and the South are mostly associated with a medium or medium-low value of both

[3] More details on the data are provided in the "Data and estimation" section.

statistics. Beside few exceptions, regions characterized by larger stores (on average) also register a larger share of people consuming five or more daily portions of FV, suggesting a positive relationship between the two variables.

Using the same data, Figure 2 depicts trends in FV5 share (top panel), average store size in hundreds of square meters (middle panel) and the number of stores per 10,000 inhabitants (bottom panel) in the period 2003-2012, across three geographic macro-areas (North, Center and South and Islands). The share of the population eating daily five or more portions of FV has decreased in all areas, but such decrease is more marked in the South. Even though a common trend exists, with a sharp decline in the year of the economic crisis (2009) and a small recovery in 2010, the successive downtrends appear more marked in the South and the Islands. Different trends can instead be noticed for the average store size across Italian macro-areas. Specifically, northern regions show larger stores compared to the other areas, which have also expanded at a faster rate. In particular, stores in the South and Islands not only are smaller on average, but maintain a small size over time. Interestingly, the geographic rank based on the average store size reflects the same order as FV5share in the three areas. Store density (i.e. number of stores per 10,000 inhabitants, third panel) appears relatively stable for North and Center during the period considered, showing also similar values. Differently, South and Islands experienced a large increase in per-capita stores, bringing this area to have 3.9 stores/10,000 habitants compared to the 3.3 stores/10,000 habitants of the other two areas, and therefore showing, on average, a larger number of smaller stores than the other two areas.

The goal of this article is to assess whether a measurable relationship exists between food retail structure across Italian regions and the probability of adult Italians consuming the daily recommended amounts of FV. Following the discussion highlighted above, we hypothesize that

ease of access and increased assortment and availability create a conducive environment for consuming more FV. To test this hypothesis, we use 9 years of data from a Nielsen database of food stores' location and characteristics in Italy, matched with individual-level data from MHS by the Italian Institute of Statistics. Differently from most existing studies, which use the number of stores in a given area to proxy for food stores availability, we also use other food retail-structure measures such as selling areas, number of in-store scales, and the length of the refrigerated aisles. As food retailers location and features are likely endogenous, we use Newey's minimum chi-squared estimator (Newey 1987), which we will refer to as a Two-Step Instrumental Variable Probit (2SIV) to control for the endogeneity of these variables. Our identification strategy will use results of the regional elections to capture the different tendency of regional governments to support the expansion of retail business and aggregate drivers of food stores location decision.

The Model

In the framework that follows, the choice of consuming a certain number of portions of fruits and vegetables is the outcome of the consumer utility maximization. Consumer *i* in region *l* maximizes her utility, function of income (spent on goods) and leisure (or hours worked), subject to time and budget constraints. The types and features of food outlets available to *i* affects the number of portions of fruits and vegetables through both prices and availability. Let an expression of the FV consumption of consumer *i* in region *l* (FV_{il}) be:

$$FV_{il} = f\left(\mathbf{X}_{il}, \mathbf{RS}_{l}, \mathbf{d}_{l} \mid \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}\right) = f(\mathbf{Z} \mid \boldsymbol{\theta}) + e_{il} \qquad (1)$$

where $\mathbf{X}_{il}$ are consumers and household characteristics; $\mathbf{RS}_{l}$ is a vector of variables capturing the features of the food retail structure available to all consumers in *l*, (e.g number of stores per-

capita, average selling area, availability of fresh food); $\mathbf{d}_l$ are region-specific variables impacting FV consumption; $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\gamma}$ are vectors of parameters conformable to $\mathbf{X}_{il}$, $\mathbf{RS}_l$, and $\mathbf{d}_l$, respectively; and e_{il} is a stochastic term. The vectors $\mathbf{X}_{il}$, $\mathbf{RS}_l$, and $\mathbf{d}_l$ are collapsed in the vector $\mathbf{Z}$ and the first part of the central term in (1) is summarized as $f(\mathbf{Z}|\boldsymbol{\theta})$, $\mathbf{Z}$ being a vector of covariates impacting FV consumption, and $\boldsymbol{\theta}$ a conformable vector of parameters. Equation (1) assumes that food retail structure affects all consumers in region l in the same way, an assumption necessary due to data limitations.[4] For each household, we allow for two levels of fruits and vegetables consumption: the consumption of at least the recommended daily intake (portions) of fruits and vegetables ($FV5$=1) and the consumption of less than the recommended daily amounts ($FV5$=0). The probability of observing $FV5_{il}$ =1 is:

$$\Pr(FV5_{il} = 1|\mathbf{Z}) = \Phi(\mathbf{Z'\theta}); \qquad (2)$$

where $\Phi(.)$ is the standard normal cumulative density function (CDF) and estimates of $\boldsymbol{\theta}$ can be obtained using a probit estimator.

Food retailers' location decision and features may be correlated with unobservable drivers of FV consumption and therefore with the probability that a consumer's diet satisfies the dietetic guidelines. Endogeneity in food retail structure and in food stores features comes from both store location and quality provision are retailers' choice variables. Quality provision has been modeled both using selling area as proxy for it (Ellickson 2006, 2007) or as the case of the provision of services or assortment of specific goods.[5] To resolve the issue of endogeneity in food retail structure, we consider the following reduced form equation:

[4] As the publicly available individual level data on adult Italian consumers only report information on the region of residence, we were forced to use retail structure data at the aggregate, regional, level. However, the features of the food environment in a region may not account for nuances and differences in the local retail environment consumers face. We illustrate the limitations of using regional-level data in the "Discussion" section.

[5] For example, Bonanno and Lopez (2009) show that supermarkets set strategically in-store service levels (e.g. salad bars, deli department and restaurants) to lure-in consumers to their stores and then apply monopolistic prices;

$$RS_{jl} = \mathbf{W}_l^{'}\alpha_j^W + \mathbf{X}_{il}^{'}\alpha_j^X + \mathbf{d}_l^{'}\alpha_j^D + \mathbf{R}_{jl}^{'}\alpha_j^R + \varepsilon_{ijl} = \mathbf{W}'\alpha_j^W + r_{ijl} \qquad (3)$$

where RS_{jl} is the *j*-th food retail structure measure in region *l*, $\mathbf{W}_l$ contains exogenous variables affecting the *j*-th food retail structure measure as well as others'; $\boldsymbol{X}_{il}$ and $\mathbf{d}_l$ are discussed above; the α_j s are conformable vectors of parameters, differing across retail structure measures, and ε_{ijl} is a random term which by assumption satisfies $E(\varepsilon_{ijl}e_{il}) = 0, \forall i, j, l$. The vector $\mathbf{W}$ includes all the exogenous (to both *FV5* and *RS*) variables available to the researcher ($\mathbf{W} = [\mathbf{W}_l, \mathbf{X}_{il}, \mathbf{d}_l]$) while $\mathbf{R}_{jl}$ contains those correlated with equation (1) error so that $r_{ijl} = \mathbf{R}'\alpha_j^R + \varepsilon_{ijl}$ is correlated with e_{il} . Given that ε_{ijl} is assumed to be uncorrelated with e_{il} , r_{ijl} is correlated with e_{il} only through $\mathbf{R}_{jl}$. Consider a vector $\hat{\mathbf{r}}$ whose elements $\hat{r}_{ijl}$ are estimates of r_{ijl} , or the residuals obtained from regressing each *RS* measure on the weakly exogenous vector $\mathbf{W}$. Then one can control for the correlation between *RS* and unobserved drivers of FV including $\hat{\mathbf{r}}$ as explanatory variables in the *FV5* equation. Thus, equation (2) can be rewritten as:

$$\Pr(FV5_{il} = 1 | \mathbf{Z}, \hat{\mathbf{r}}) = \Phi([\mathbf{Z}, \hat{\mathbf{r}}]' \boldsymbol{\eta}^*). \qquad (4)$$

where $\boldsymbol{\eta}^* = [\boldsymbol{\theta}^*, \boldsymbol{\eta}]$. $\boldsymbol{\theta}^*$ and $\boldsymbol{\eta}$ are parameters vector associated with the covariates in equation (1) and $\hat{\mathbf{r}}$, respectively (note that $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}$).[6] Once the vector $\hat{\mathbf{r}}$ is included in the model, unbiased estimates of $\boldsymbol{\theta}^*$ can be obtained using a probit estimator. This approach, also known as the Two-Stage Residual Inclusion (2SRI) method, is similar to Newey's (1987) minimum chi-squared estimator and Rivers and Vuong's (1988) two-stage conditional maximum likelihood estimator,

---

Richards and Hamilton (2006) show retailers competing in product line varieties for apples; and Richards (2007) assesses competition in promotional activities for fresh fruits across stores.

[6] See Wooldridge (2002, pp. 473-474) for a discussion of the relationship between the parameters of equations (2) and (4) in the case of a single endogenous variable.

and in the case of non-linear models (such as probit) performs better than traditional two-stage instrumental variables methods (Terza, Basu, and Rathouz 2008).[7]

Data sources, data manipulation and estimation

The main database we use in the analysis encompasses nine years of individual-level observations from the MHS conducted by the ISTAT in the year 2003, and for the years 2005-2012.[8] This survey contains information on household- and individual-level characteristics (*e.g.* age, gender, level of education, smoking habits, levels of physical activity, time spent watching television, etc…) as well as self-reported frequency of consumption of different food groups including fruit and vegetables. Those individuals declaring to consume at least one portion of fruits, leafy vegetables or other vegetables daily were asked a follow up question on the daily portions consumed. From the recorded amounts of daily portion of fruits and vegetable consumed daily, we coded the indicator FV5, taking the value of one for individuals consuming the daily recommended amounts of fruits and vegetables, 0 otherwise.[9] We focus on adult respondents only, as children may be exposed to voluntary school fruits schemes (or other initiatives) which may affect their FV5 likelihood and which, given the nature of our data, we are unable to control for.

The individual-level MHS data were matched with data on the Italian grocery retail industry from Nielsen. The Nielsen data contain information on store format (hypermarket,

[7] Other applications using similar methods are Bonanno and Li (forthcoming); Stein and Rodney (2012); Lee and Kim (2012); Liu, Lovely, and Ondrich (2010); and Petrin and Train (2010).

[8] The MHS survey data are available since 1993. Questions on the number of portions of fruits and vegetables consumed were added in 2003. MHS data were not collected in 2004.

[9] Modeling directly the declared number of portions of FV consumed was problematic on two fronts. First, the nature of these two questions creates an issue of censoring: individuals who are not asked the follow-up question (number of portions consumed) may still in fact be consuming fruits and vegetables; thus, as no daily amounts were recorded, these were recorded as "false" zeros. Second, data inspection on the daily amount recorded highlighted the presence of positive numbers of daily portions for individuals who declared not to consume FV daily.

supermarket, discount and convenience store), number of establishments, total selling area in square meters, number of scales per store, checkouts and horizontal and vertical meters of refrigerated aisles. The data, originally at the province-level, were aggregated at the regional level[10] so that they can be matched with the MHS individual level database, as the MHS database only reports regional identifiers. Our measures of the regional food retail structure are the average store size (AVSIZE), the average number of stores every 10,000 habitants (PCSTORE), the average number of scales (SCALESQM) and average horizontal and vertical meters of refrigerated aisles for every 100 square meters of selling area (REFRSQM). The average selling area per store (AVSIZE) and the per capita number of stores (PCSTORE) are indirect measurement of the regional food retail structure and their effect should capture how store density (proximity) and store size (assortment) can affect daily consumption of fruit and vegetables. The other two variables (SCALESQM and REFRSQM) are used as proxies for the in-store presence of fresh foods presence. Specifically, with these two variables we aim to capture whether in-store differences on fruit and vegetables presence can play a role on different dietary habits. The MHS also collects information on the declared level of hardship to reach supermarkets. This question allows for four different levels of declared hardship: "no hardship," "some hardship," "considerable hardship," and "I don't know." We use this variable to segment the data in two sub-samples: one including respondents declaring "no hardship" (No Access Issues sample) and another with individuals declaring to have some or considerable hardship to reach supermarkets[11] (Perceived Access Issues sample).

Individual-level control variables that affect the probability of FV5 come from the MHS database and are divided in five groups. 1) availability and use of transportation: number of cars

[10] Where Aosta Valley and Piedmont are considered as an unique regional entity.
[11] Individuals responding "I don't know" were dropped from the sample.

per individual within the household (PCCARS); indicators capturing the absence of public transportation in the municipality of residence (NOPUBTR); if the respondent uses public transportation frequently (every day or more than once a week – PUBTRUSFR) or rarely (monthly or yearly base – PUBTRUSSOME). 2) household composition: household size (HHSIZE), number of children in the household aged 5 years or younger (NCHILD 0-5), number of children aged 6 to 17 years (NCHILD 6-17), age (AGE) of the respondent, age squared (AGE2), and indicators capturing gender (FEMALE) and marital status (NOTMARRIED) of the respondent. 3) education level of the respondents: four indicators for the maximum level of educations achieved by the respondent (College or higher - COLL-HIGHER; some college - SOMECOLL; high school - HSCHOOL; and no education/illiterate - NOEDUC). 4) Income sources and perceived economic status: weather the respondent is not employed (UNEMPL); his/her main source of income comes from retirement (PENSION), income from property and investments (PROP_CAP), or if they are supported by the family (FAMILYSUP); and two indicators capturing whether the economic situation of the household is perceived as scarce (ECONBAD) or absolutely insufficient (ECONVERYBAD). 5) Respondents' habits: variables capturing whether, and where the respondents usually eat lunch away from home (cafeterias or coffee shops – COFFEESH; restaurant - REST, canteen - CANTEEN) number of hours spent in front of the TV (TVHOURS: proxy for inactive time), whether the respondents smoke cigarettes daily (SMOKER), and an indicator for regular practice of physical activity (SPORT). We control for change in food prices using a consumer price index within the food category (CPIFOOD) at the regional level.[12] Last, regional and year fixed-effects were included in the model to further control for unobservables.

[12] The food CPI is collected from the Italian National Institute of Statistics (ISTAT) website: www.istat.it.

Observations showing missing values and outliers were dropped, along with individuals below 18 years of age. The final sample size consists of 254,716 observations, of which 174,741 belong to individuals declaring to have no access issues and the remaining 79,975 who instead declared to have issues in accessing supermarkets. Summary statistics of the variables used in the estimation are presented in Table 1. On average, more than 30% of the sample declares to have some level of difficulty in accessing supermarkets. However, looking at the subsamples we do not observe a substantial difference on the share of individuals consuming five or more portions of FV daily (circa 4.7%) and similar average food retail structure measures. The demographic characteristics of the two sub-samples seem instead to differ. While the sub-sample with No Access Issues seems to have, on average, higher access to private means of transportation (PCCARS), the lack of access to public transportation (NOPUBTR) differs highly between sub-samples with different perceived food access. On average the share of people who own a car but do not have access to public transportation (PCCARS*NOPUBTR) is substantially higher among households perceiving difficulties to reach supermarkets. Moreover, we observe more respondents using public transportation often (PUBTRUSFR) or rarely (PUBTRUSSOME) among those declaring not to have perceived access issues. Furthermore, a larger share of individuals declaring hardships in reaching a supermarket declares to have economic constrains (ECONBAD, ECONVERYBAD) and perceive income from pension (PENSION).

Identification strategy and estimation

We aim to correct for the endogeneity of the food retail structure measures by capturing variation in political climate across the Italian regions. As illustrated elsewhere (Banca d'Italia 2012; Schivardi and Viviano 2012), Italy underwent a process of liberalization of retail trade in

the decade analyzed, which resulted in changes in the retail structure and in the types and numbers of stores allowed to open in each province.[13] Permits to open new stores and to expand existing selling areas were unequal across regions due to the different regional political climates (see AGCM, 2007 for a detailed discussion). Thus, differences in political climate should capture, at least in part, exogenous (to unobserved variations in consumer diets) differences in the structure of the food retail industry.[14]

To this end, we collected data for the results of the regional elections during the sample period from the Italian Ministry of Internal Affairs (*Ministero dell'Interno*). For Sicily, Sardinia and Trentino Alto Adige, this information was collected from regional governments' websites. Specifically, for each election year and region, we collected the share of seats assigned to different groups of parties: right wing (SH_RIGHT), left wing (SH_LEFT), moderate right wing (SH_CENLEFT), moderate left wing (SH_CENRIGHT), "green" (SH_GREEN) parties, as well "federalist" (SH_FEDERALIST), "five star movement[15]" (SH_M5S) and "others" (SH_OTHERS). Moreover, we control for the share of seats assigned to the left and right side coalitions (respectively SH_COALLEFT and SH_COALRIGHT) which is of particular interest since the Italian political environment is usually characterized by numerous parties, and the type of coalitions for each election year and region can capture the nuances of different orientations among the population. We assume that the outcome of a regional election affects *RS* from the year after the election took place; in other words, the results of an election that took place in, say, 2005, will be used as IV for RS for the year 2006 onward, until the next election took place. Additionally, following Bonanno and Li (*forthcoming*) we control for aggregate drivers of store

[13] For a discussion of the economic implication of retail liberalization in Italy see Schivardi and Viviano 2012.
[14] The use of these types of instruments is not new. For examples, Levitt's (1997) identified the impact of police presence on crime rates using election cycles as an instrument for the size of the police force.
[15] "Five Star Movement" (M5S) is an Italian party started on October 2009. M5S is mainly considered "Eurosceptic" and "populist", with the meaning of being a party of citizens.

location. Following the extant entry literature, market size and market growth are two of the most important determinant of retail establishment location, as they capture potential market demand and future profitability (e.g. Bresnahan and Reiss 1991; Ellickson 2006, 2007). As the number of establishments per store is divided by the population, we use average per-capita income by region (INCOME_REG) as a normalized measure of market size and the population growth rate (POP_CHANGERATE) as a proxy for market growth. Furthermore, we also control for population density (POP_DENSITY) to capture the prevalence of more dense location of food stores in more densely populated areas.

As illustrated in the model section we adopt a Two-Stage Residual Inclusion approach. In particular, the estimator of choice is Newey's (1987) minimum chi-squared estimator which, using the terminology of the statistical pa ckage used for data manipulation and estimation, STATA, will be referred to below as a Two-Step IV Probit (2SIV). We use the Amemiya-Lee-Newey (ALN) [16] minimum distance statistics, to test for orthogonality of the over-identifying instruments, distributed as a χ^2 with degrees of freedom equal to the number of overidentifying restrictions, under the null hypothesis of instruments validity. We assess the power of the instruments using an *F*-test on the joint significance of their coefficients in each of the first stage regressions. Following the Staiger and Stock (1997) rule of thumb, a value of the *F*-statistic larger than 10 is considered large enough to dismiss the presence of weak instruments problems. Wald tests for the exogeneity of the RS variables (distributed as a χ^2 with degrees of freedom equal to the number of endogenous variables) were performed on the coefficients of the vector $\hat{\mathbf{r}}$ under the null of exogeneity (i.e. the null hypothesis is that of all the coefficients on $\hat{\mathbf{r}}$ not being statistically different from zero, similar to a Wu-Hausman test).

[16] Lee (1992) showed the equivalence of Newey's estimation method to Amemiya's (1978) Generalized Least Square estimator.

Empirical Results

Empirical results of the estimated parameters of equation (2) and (5) are presented in Table 2. The first column reports maximum likelihood probit estimates for the full sample, while the other three columns report 2SIV probit estimates for the full sample (second column) and for samples segmented according to individual perceived access (No Access Issues vs. Perceived Access Issues). Before illustrating the results, it should be noted that while the pseudo R-squared appear relatively low (circa 0.025) several of the explanatory variables affect the FV5 probability in a statistically significant way. Also, the results of the Wald tests of exogeneity show that, conditionally on the instruments used, the RS variables appear endogenous, although evidence is weaker in the Perceived Access Issue sample: the *p*-values are 0.0021, 0.0171 and 0.1088 for, respectively, Full, No Access Issue and Perceived Access Issue samples. The *p*-values of the ALN tests are all above the cut-off of 0.1 supporting the validity of our identification strategy; last, the instruments seem to have enough explanatory power, as suggested by the *F*-statistic for the joint significance of the instruments' coefficients in the first stage equations exceeding the Staiger and Stock (1997) rule of thumb.

The first result to be highlighted is the difference between the estimated probit and the 2SIV probit RS coefficients. The coefficient for average selling area (AVSIZE) changes from 0.1228 to 0.298 (in both cases statistically significant at the 1% level); that of the per-capita number of stores increases three and a half times, from 0.0563 to 0.2025, with its statistical significance level increasing as well. The coefficient associated with the ratio of refrigerated aisles' meters over selling area increases 38 times, from 0.005 to 0.1930, becoming statistically significant. The least impacted estimate is that of the number of scales per square meter, which,

in spite of increasing in magnitude, remains not statistically significant. Furthermore, the magnitude of the estimated 2SRI coefficients for AVSIZE, PCSTORES and REFRSQM are relatively stable across subsamples. However, the last two lose in statistical significance in the sample of individuals with perceived access issues (REFRSQM showing significance at the 10% level); the coefficients of SCALESQM, being positive and statistically significant at the 10% level in the No Access Issues sample, is negative and not statistically different from zero in the Perceived Access Issues sample. Thus, our empirical results indicate that the food retail structure plays an important role in explaining the probability of consuming five or more daily portions of fruits and vegetables among adult Italians, and that this effect is mostly found among individuals who do not declare to experience hurdles in accessing supermarkets.

The other estimated coefficients show interesting patterns regarding what variables affect the FV5 probability. With respect to variables capturing the availability of transportation, the results indicate that the per-capita number of cars affects the probability of fruits and vegetables consumption positively only for individuals declaring limited access; this effect is magnified for those individuals living in municipalities where public transport is not available. Also while the lack of public transport *per se* seems not to affect FV5 probability, using public transport seems to increase the probability of FV5, particularly in individuals who do not declare access issues.

Demographics and household characteristics have a larger effect among individuals declaring no access issue than for those experiencing hardships in reaching supermarkets: while living in larger household seem to increase the probability of FV5, the presence of children affects it negatively; the probability of FV5 is higher for individuals who are female and not married. Education plays an important role in explaining the probability of FV5, and the estimated coefficients (which do not seem to change largely across samples) grow in magnitude

with higher levels of education, also indicating that individuals with no formal education show lower FV5 probability. Interestingly, while unemployment and being supported by family members mostly does not affect FV5 probability, individuals whose major source of income is capital rents seem to have better diets, particularly in the Perceived Access Issues sample; also being retired is positively related to FV5 probability. Individuals who indicated their economic situation as "bad" or "very bad" show increasingly lower probabilities of consuming FV5, in particular in the Perceived Access Issues sample. Consuming lunch habitually in restaurants or in coffee shops negatively affects the FV5 probability (the former especially for individuals with perceived access issues), where such effect is probably driven by the different variety of food served in the different outlets. As for personal habits, while smoking impacts FV5 likelihood negatively, practicing sport and watching more TV impact it positively. The CPI for food does not play any role. Overall, these results suggest that the economic situation of the household, transportation and education level (which per se, may be correlated with income) seem to be the most important factors explaining FV5 consumption among household in the Perceived Access Sample, while for individuals who do not declare to have access hurdles, there are multiple factors affecting it.

Selected estimated average marginal effects are reported in Table 3.[17] In the first place, the effects of the retail structure variables are dissimilar across samples. An increase in the size of food stores by 100 square meters (or 14% circa) leads to an increase in the probability of consuming FV5 by circa 3% across samples (the effect is 10% lower for individuals declaring to have problems in accessing supermarkets). A marginal increase in the density of food stores' presence (which is a 28.9% increase of this variable) seems to increase FV5 probability among

[17] Marginal effects are obtained using 2SRI estimates with bootstrapped standard errors, since the 2SIV probit routine in STATA does not allow for the calculation of marginal effects.

adult Italians by circa 2%; however, the effect is not statistically different from 0 for individuals declaring access issues. This result suggests that, even if more stores were built, they may not contribute to improve the diets of those who have issues in reaching them; this may also be due to the fact that perceived access seems less related to the actual physical presence of stores and more to income constrains and lack of transportation. The marginal effect of one additional scale per 100 square meter of selling area on FV5 likelihood is only significant at the 10% level among individuals who declared no access issues and the effect is of circa 0.9% (for a unitary change, or 18% circa, of this variable). An additional meter of refrigerated aisle per 100 square meter of selling area (or 16.5% circa) leads to a positive increase in the FV5 probability which vary between 1.85% among individuals with no issues of access, and 1.93 for those who declared to have access issues.

The average effects of the other variables on the probability of adult Italians consuming five or more portions of fruits and vegetables mirror the differences in significance and estimated parameters across samples illustrated above. For example, owning an additional car per-capita in the household in absence of public transportation has twice as high an effect among individuals declaring perceived access issues that in those who do not (for cumulative marginal effects of 0.8% vs. 0.44%). Also, while the presence of children in the households lowers the FV5 probability for individuals declaring no access issues, it plays a much smaller effect on individuals experiencing hardship in accessing supermarkets. However, across the two samples, individuals with college-level or higher education have 1.8% to 2% higher probability of FV5, those with some college 1.61 to 1.41% higher, high school 0.86 to 0.9 higher while having no education from -1.1% to -1.3%. Among individuals with perceived hurdles to accessing supermarkets, those who judge the economic situation of their household as scarce or insufficient

are 0.68% and 1.44 % (circa) less likely to consume fruits and vegetables five or more times daily than those who do not. Last, eating lunch in restaurants or in cafeterias lowers FV5 probability among adult Italians: the latter shows similar effect across individuals with different perceived access (-1.3% vs 1.4% circa), while the former's effect is twice as large for individuals with perceived access issues (-1.31%) than for those with no access issues (-0.65%).

Discussion, conclusions and limitations

The results reported above show that access to food stores, assortment and in-store availability of fresh food show a positive effect on the probability of adult Italians consuming the recommended daily amounts of fruits and vegetables. However, given the dissimilar effects, one may wonder what kind of intervention could foster a diet richer in fruits and vegetables. Given our results, we find that, if the average areas of food stores would double, the increase in probability of FV5 could increase by 22.3% for individuals who declare no access issues and 19.6% for those declaring access issues. Doubling the density of food stores could lead to an increase of 7.6% circa in the probability of consuming the recommended daily portions of fruits and vegetables; however, we have no evidence that this change could lead to an improvement among individuals who have hurdles in accessing supermarkets. Our results indicate also that, if food retailers doubled the number of scales for hundreds of square meters, one will observe a 5.1% circa increase in FV5 probability but only for individuals with no access hurdles. Doubling instead the refrigerated aisles for 100 meters, one would benefit from a 11.4% to 13.2% higher probability of consuming FV5 for individuals declaring to have no access issues and for those declaring to experience hurdles in accessing supermarkets, respectively. Given the structural differences in the Italian food retail across areas highlighted in Figure 2, it seems that current

trends pushing towards larger selling areas, particularly in the North, may lead to an across-the-board benefit in terms of fostering FV consumption. The increased number of (mostly small) stores in the South, may not necessarily foster a higher consumption of FV5, especially among individuals who experience access hurdles. That said, the benefits of small, proximity stores, specialized in the sales of fresh food should not be discarded; as the ratio of the meters of refrigerated aisles over selling areas is the largest in these specialized stores, the existence of these outlets may be beneficial as well. In summary, at least according to our results, the decline in the number of adults consuming the recommended amounts of fruits and vegetables in Italy is not a direct consequence of changes in structural features of the retailing industry (if anything, the opposite). Possibly, other explanations can be found in the still large economic hurdles for Italians struggling to make ends meet following the most recent recession.

The results presented above hold with some caveats. First, because of limitations in the MHS data, our results cannot capture the effects of the local structure of the food retail industry on FV consumption. Thus, since we use region-level food retail structure measures, our results depict *average* impacts of food stores on FV5 likelihood *across* local environments in each region, and are not likely to hold in all Italian provinces. Second, our analysis did not assess in full the trade-off between perceived and actual access. Attempts to include a perceived access indicator as explanatory variable did not produce -statistically significant coefficients (results not shown for brevity, and available upon request) indicating that actual, more than perceived, access is a limiting factor in the adoption of a diet richer in fruits and vegetables. However, as highlighted elsewhere in the literature (e.g. Kamphuis et al. 2006) perceived access is affected by factors such as lack of transport, income, education and *actual* access, therefore, simultaneity may bias our results towards zero. Thus, we attempted to account for the potential endogeneity

of perceived access by estimating a 2SIV probit using lack of public transportation, which we found not to affect FV5 probability, as exclusion restriction. The results (not shown for brevity and available upon request) show again perceived access not playing a role in affecting the FV5 probability.[18] Third, we use FV5 as a proxy of Italian adults' consumption of FV, which per se may be insufficient to capture the overall consumption of fruits and vegetables, let alone the healthiness of the overall diet, which relies on a broader and complex concept of balance among different nutrients.

References

AGCM - Autorità Garante della Concorrenza e del Mercato. 2007. "Qualità della Regolazione e Performance Economiche a Livello Regionale: il Caso della Distribuzione Commerciale in Italia", Studi e Ricerche, Autorità Garante della Concorrenza e del Mercato. http://www.agcm.it/trasp-statistiche/doc_download/2632-sr1.html Accessed 12/01/2014 .

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

Amemiya, T. 1978. "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica* 46(5):1193-1205.

Banca d'Italia. 2012. The Retail Trade Sector and The Food Industry in Italy. Bank of Italy - Occasional papers (Questioni di Economia e Finanza). No 119. Available at http://www.bancaditalia.it/pubblicazioni/econo/quest_ecofin_2/QF_119 Accessed October 23, 2013.

Belahsen, R., and M. Rguibi. 2006. "Population Health and Mediterranean Diet in Southern Mediterranean Countries." *Public Health Nutrition* 9(8):1130-1135.

Berghöfer, A., T. Pischon, T. Reinhold, C.M. Apovian, A.M. Sharma, and S.N. Willich. 2008. "Obesity Prevalence from a European Perspective: a Systematic Review." *BMC Public Health* 8:200-210.

[18] The overall model diagnostics appear similar to those reported in the second column of table 3, that is a low *p*-value of the Wald test for exogeneity, a *p*-value of the ALN statistics above 0.1 and the instruments appearing not to be weak. The estimated parameters also appeared similar to those in the second column of table 2, with the exception of NOPUBTR no longer being included in the model.

Bodor, J.N., D. Rose, T. Farley, C. Swalm, and S.K. Scott. 2008. "Neighbourhood Fruit and Vegetable Availability and Consumption: the Role of Small Food Stores in an Urban Environment." *Public Health Nutrition* 11(4):413-420.

Bonaccio, M., A. Di Castelnuovo, A. Bonanni, S. Costanzo, F. De Lucia, M. Persichillo, F. Zito, M. Donati, G. de Gaetano, L. Iacoviello. 2014 "Decline of the Mediterranean Diet at a Time of Economic Crisis. Results from the Moli-sani Study." *Nutrition, Metabolism and Cardiovascular Diseases* doi: 10.1016/j.numecd.2014.02.014 .

Bonanno, A., and J. Li. Forthcoming. "Food Insecurity and Food Access in US Metro Areas." *Applied Economics Perspectives and Policy* 91(3): in press.

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

Bresnahan, T.F., and P.C. Reiss. 1991. "Entry and Competition in Concentrated Markets." *Journal of Political Economy* 99(5):977–1009.

Capacci, S., and M. Mazzocchi. 2011. "Five-a-day, a Price to pay: An Evaluation of the UK Program Impact Accounting for Market Forces." *Journal of Health Economics* 30(1)87-98.

Capacci, S., M. Mazzocchi, B. Shankar, J. Brambila Macias, W. Verbeke, F. J. Pérez-Cueto, A. Kozioł-Kozakowska, B. Piórecka, A. Niedzwiedzka, D. D'Addesa, A. Saba, A. Turrini, J. Aschemann-Witzel, T. Bech-Larsen, M. Strand, L. Smillie, J. Wills, and W. B. Traill. 2012. "Policies to Promote Healthy Eating in Europe: a Structured Review of Policies and their Effectiveness." *Nutrition reviews* 70(3):188-200.

Caspi, C. E., I. Kawachi, S.V. Subramanian, G. Adamkiewicz, and G. Sorensen. 2012. "The Relationship between Diet and Perceived and Objective Access to Supermarkets among Low-income Housing Residents." *Social science and medicine* 75(7):1254-1262.

Castellari, E. and P. Sckokai. 2014. "La Distribuzione Alimentare al Dettaglio", in Il sistema agro-alimentare dell'Emilia Romagna - Rapporto 2013, (Roberto Fanfani and Renato Pieri editors), Maggioli Editore, Santarcangelo di Romagna, Italy.

Cummins, S., M. Petticrew, C. Higgins, A. Findlay, and L. Sparks. 2005. "Large Scale Food Retailing as an Intervention for Diet and Health: Quasi-Experimental Evaluation of a Natural Experiment" *Journal of Epidemiology and Community Health* 59:1035–1040.

Ellickson, P.B.. 2006. "Quality Competition in Retailing: A Structural Analysis." *International Journal of Industrial Organization* 24(3):521-540.

Ellickson, P.B.. 2007. “Does Sutton Apply to Supermarkets?” *RAND Journal of Economics* 38(1):43-59.

European Food information Council. 2012. Fruit and Vegetable Consumption in Europe – do Europeans get enough? Available at http://www.eufic.org/article/en/expid/Fruit-vegetable-consumption-Europe/ Last accessed 11-11-2014

Giskes, K., F. J. Van Lenthe, J. Brug, J. P. Mackenbach, and G. Turrell. 2007. “Socioeconomic Inequalities in Food Purchasing: the Contribution of Respondent-Perceived and Actual (objectively measured) Price and Availability of Foods.” *Preventive medicine* 45(1):41-48.

Gustafson, A.A., J. Sharkey, C. D. Samuel-Hodge, J. Jones-Smith, M.C. Folds, J. Cai, and A. S. Ammerman. 2011. “Perceived and Objective Measures of the Food Store Environment and the Association with Weight and Diet among Low-income Women in North Carolina.” *Public health nutrition* 14(6):1032-1038.

Gustafson, A., S. Hankins, and S. Jilcott. 2012. “Measures of the Consumer Food Store Environment: a Systematic Review of the Evidence 2000–2011.” *Journal of community health* 37(4):897-911.

Hawkes, C., 2008. Dietary Implications of Supermarket Development: A Global Perspective. *Development Policy Review* 26 (6):657-692.

Hough, G., and M. Sosa. 2014. “Food Choice in Low-Income Populations-a Review.” *Food Quality and Preference*, in press.

Italian Ministry of Health 2013, Rapporto Osservasalute 2013 available at www.osservasalute.it. Last accessed 11-11-2014.

Kamphuis, C.B.M., K. Giskes, G. de Bruijn, W. Wendel-Vos, J. Brug, and F.J. van Lenthe. 2006. “Environmental Determinants of Fruit and Vegetable Consumption Among Adults: a Systematic Review.” *British Journal of Nutrition* 96(4):620–635.

King, R.P., E.S. Leibtag, and A.S. Behl. 2004. “Supermarket Characteristics and Operating Costs in Low-Income Areas.” USDA Economics Research Service, Agricultural Economic Report N. 839.

Lee, L.F. 1992. “Amemiya's Generalized Least Squares and Tests of Overidentification in Simultaneous Equation Models with Qualitative or Limited Dependent Variables.” *Econometric Reviews* 11(3):319-328.

Lee, M.J., and Y.S. Kim. 2012. "Zero-Inflated Endogenous Count in Censored Model: Effects of Informal Family Care on Formal Health Care." *Health Economics* 21(9):1119-1133.

Leenders, M., I. Sluijs, M.M. Ros, H. C. Boshuizen, P. D. Siersema, P. Ferrari, ... and H. B. Bueno-de-Mesquita. 2013. "Fruit and Vegetable Consumption and Mortality. European Prospective Investigation into Cancer and Nutrition." *American Journal of Epidemiology* 178(4):590-602.

Levitt S.P. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review* 87:270-290.

Liu, X., M.E. Lovely, and J. Ondrich. 2010. "The Location Decisions of Foreign Investors in China: Untangling the Effect of Wages Using a Control Function Approach" *Review of Economics and Statistics* 92(1):160-166.

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.

Newey, W.K. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogeneous Explanatory Variables." *Journal of Econometrics* 36(3):231-250.

Pearce, J., R. Hiscock, T. Blakely, and K. Witten. 2008. "The contextual effects of neighbourhood access to supermarkets and convenience stores on individual fruit and vegetable consumption." *Journal of Epidemiology and Community Health* 62(3):198-201.

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

Pollard, J., S. F. Kirk, and J. E. Cade. 2002. "Factors Affecting Food Choice in Relation to Fruit and Vegetable Intake: a Review". *Nutrition research reviews* 15(2):373-387.

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(3):189-195.

Prättälä, R., S. Hakala, A.J.R. Roskam, E. Roos, U. Helmert, J. Klumbiene, H. Van Oyen, E. Regidor, and A.E. Kunst. 2009. "Association between Educational Level and Vegetable Use in Nine European Countries." *Public health nutrition* 12(11):2174-2182.

Reina, G.C., A. Berenguer, and C.A. González. 2006. "Changes in Food Supply in Mediterranean Countries from 1961 to 2001." *Public health nutrition* 9(1):53-60.

Richards, T., and S. Hamilton. 2006. "Rivalry in Price and Variety among Supermarket Retailers." *American Journal of Agricultural Economics* 88(3):710-726.

Richards, T. 2007. "A nested logit model of strategic promotion." *Quantitative Marketing and Economics* 5:63–69.

Rivers, D., and H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39(3):347-366.

Schivardi F., and E. Viviano. 2012. "Entry Barriers in Retail Trade." *The Economic Journal*, 121 (March):145–170.

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

Stein, H., and L. Rodney. 2012. "Do Fertilizer Subsidies Crowd Out Organic Manures? The Case of Malawi." *Agricultural Economics* 43(3):303-14.

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(3):531-543.

Weatherspoon, D.D., J.F. Oehmke, M.A. Coleman, A.S. Dembele, and L.J. Weatherspoon. 2012. "Will Long Term Food Desert Consumers Purchase Fresh Fruits and Vegetables?" *Choices Magazine* 27(3):1-5.

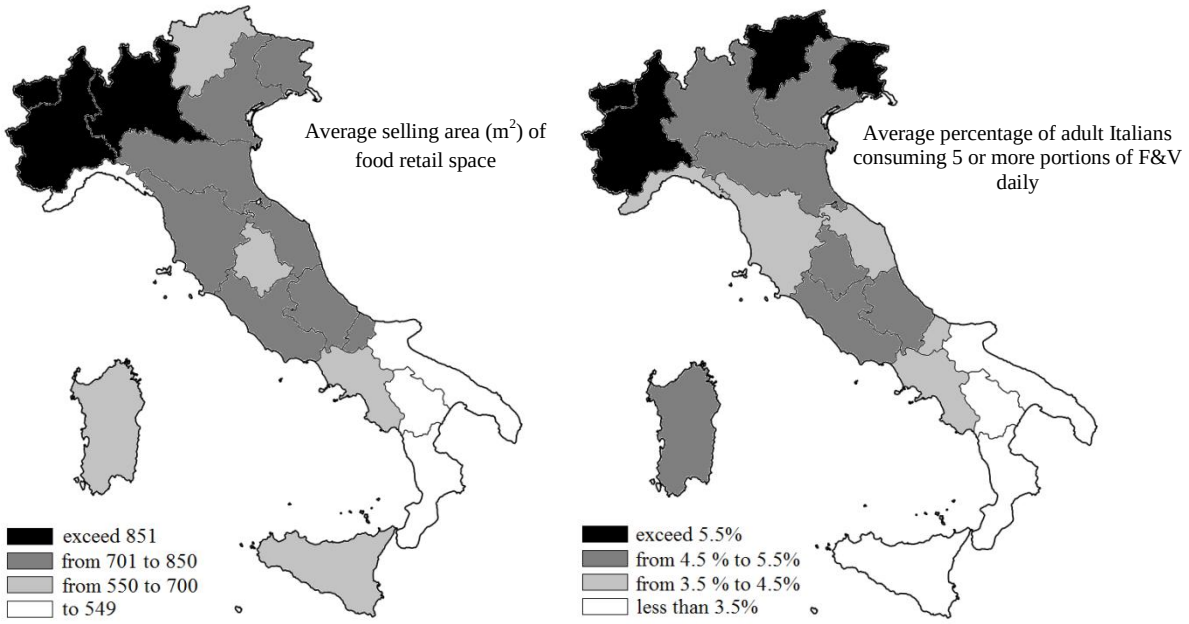
Webber, C. B., J. Sobal, and J. S. Dollahite. 2010. "Shopping for Fruits and Vegetables. Food and Retail Qualities of Importance to Low-Income Households at the Grocery store." *Appetite* 54(2):297-303.

WHO 2008 "The Challenge of Obesity in the WHO European Region and the Strategies for Response" edited by F. Branca, H. Nikogosian and T. Lobstein. 338 pages. Available at http://www.euro.who.int/_data/assets/pdf_file/0010/74746/E90711.pdf. Last access 11-11-2014.

Williams, L. K., L. Thornton, K. Ball, and D. Crawford. 2012. "Is the Objective Food Environment Associated with Perceptions of the Food Environment?" *Public health nutrition 15*(2):291-298.

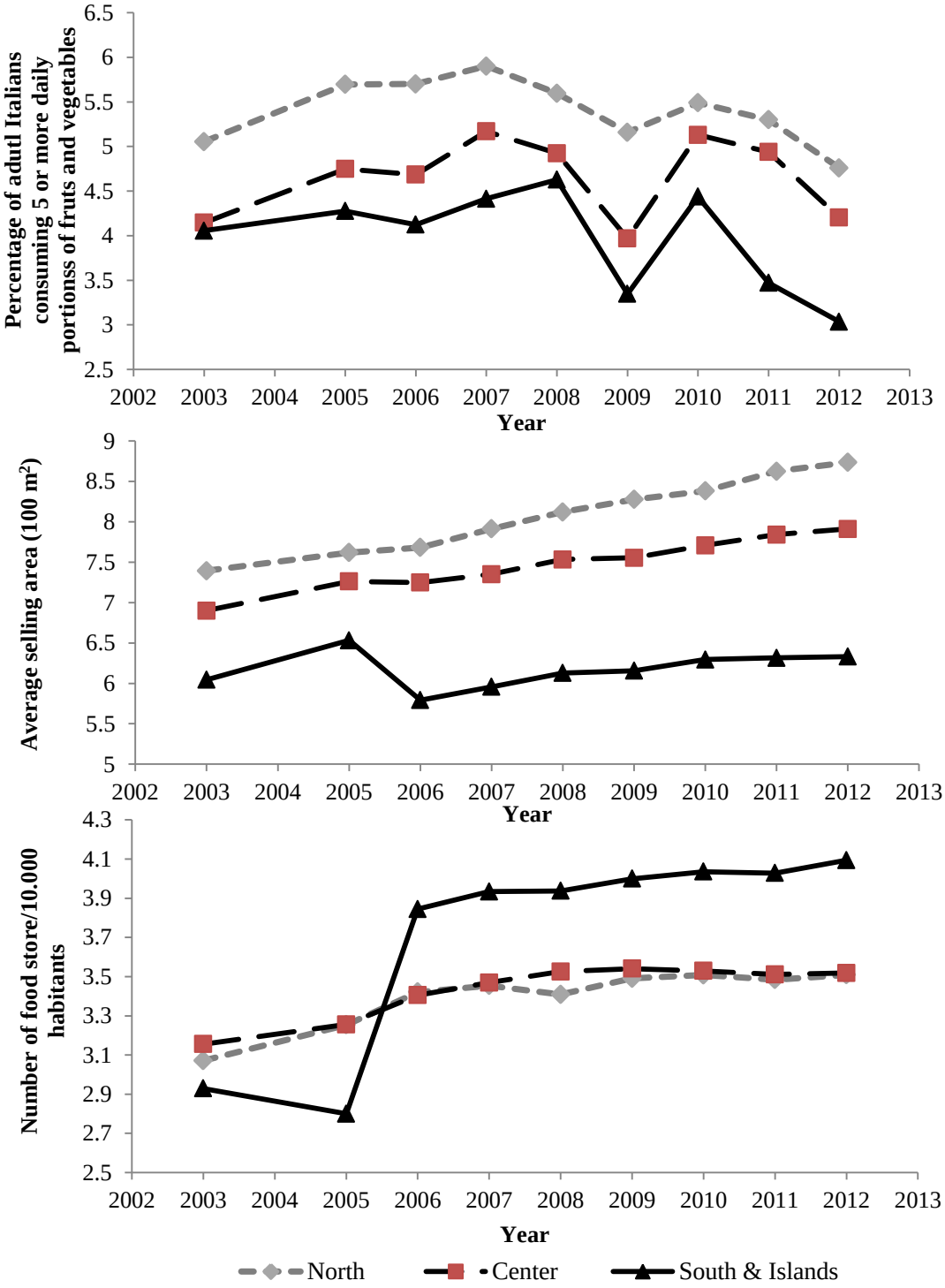
Wooldridge J.M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Figure 1: Average selling area (m^2) of food retail space (left) and percentage of adult Italians consuming more than 5 fruits and vegetables portions daily (right) by region.



Sources: Authors' elaboration from the Italian Institute of Statistics Multipurpose Household Survey and Nielsen database (2003-2012).

Figure 2: Area-specific trends: percentage of adult Italians consuming five or more portions of fruits and vegetables daily (FV5 share, top panel) average food stores selling area (hundreds of square meters - middle panel) and number of food stores per 10,000 inhabitants (bottom panel)



Sources: Authors' elaboration from the Italian Institute of Statistics Multipurpose Household Survey and Nielsen database (2003-2012).

Table 1 Sample Statistics

	Full Sample (N=254,716)		No Access Issues (N=174,741)		Perceived Access (N=79,975)	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std.
FV5	0.047	0.212	0.048	0.214	0.046	0.208
AVSIZE	7.264	1.532	7.263	1.547	7.265	1.499
PCSTORE	3.526	0.822	3.557	0.828	3.458	0.803
SCALESQM	5.721	0.845	5.741	0.850	5.679	0.830
REFRSQM	6.134	0.647	6.148	0.645	6.104	0.651
LOW ACCESS	0.314	0.464				
PCCARS	0.976	0.714	0.996	0.706	0.931	0.731
NOPUBTRAND	0.205	0.404	0.178	0.383	0.263	0.440
PCCARS*NOPUB	0.202	0.518	0.184	0.499	0.243	0.556
PUBTRUSFR	0.091	0.288	0.101	0.301	0.069	0.254
PUBTRUSSOME	0.129	0.335	0.139	0.346	0.107	0.310
HHSIZE	1.729	0.912	1.735	0.913	1.717	0.911
NCHILD(0-5)	0.133	0.403	0.136	0.405	0.128	0.399
NCHILD(6-17)	0.313	0.642	0.317	0.643	0.305	0.641
AGE	50.662	18.197	49.819	17.759	52.506	18.989
FEMALE	0.515	0.500	0.511	0.500	0.523	0.499
NOTMARRIED	0.592	0.492	0.597	0.490	0.579	0.494
COLL-HIGHER	0.084	0.277	0.093	0.290	0.064	0.245
SOMECOLL	0.023	0.149	0.025	0.155	0.018	0.135
HISCHOOL	0.375	0.484	0.391	0.488	0.340	0.474
NOEDUC	0.049	0.215	0.037	0.189	0.074	0.262
UNEMPL	0.061	0.239	0.061	0.240	0.061	0.239
PENSION	0.305	0.460	0.283	0.451	0.353	0.478
PROP_CAP	0.005	0.068	0.005	0.070	0.004	0.064
FAMILYSUP	0.019	0.138	0.019	0.137	0.020	0.141
ECONBAD	0.362	0.481	0.344	0.475	0.401	0.490
ECONVERYBAD	0.056	0.229	0.053	0.223	0.062	0.242
REST	0.035	0.185	0.037	0.188	0.033	0.178
CANTEEN	0.048	0.214	0.050	0.219	0.044	0.204
COFFESHOP	0.025	0.156	0.027	0.162	0.021	0.142
SMOKER	0.228	0.419	0.234	0.424	0.213	0.409
SPORT	0.171	0.376	0.185	0.388	0.139	0.346
TVHOURS	2.917	1.759	2.889	1.730	2.980	1.819
CPIFOOD	1.250	0.095	1.251	0.094	1.249	0.097
Food retail Structure Instruments						
SH_LEFT	0.058	0.034				
SH_CENLEFT	0.406	0.120				
SH_RIGHT	0.001	0.006				
SH_CENRIGHT	0.399	0.107				
SH_FEDERALIST	0.046	0.072				
SH_M5S	0.001	0.007				
SH_GREEN	0.018	0.016				
SH_COALLEFT	0.484	0.132				
SH_COALRIGHT	0.450	0.124				
SH_OTHERS	0.066	0.123				
INCOME_REG	35.115	3.688				
POP_DENSITY	203.336	108.507				
POP_CHANGERA	0.120	0.319				

Sources: Authors' elaboration from the Italian Institute of Statistics, Multipurpose Household Survey and Nielsen database (2003-2012). Food retail structure instruments: author's elaboration on data from the Italian Ministry of Internal Affairs (Ministero dell'Interno) or regional governments' websites (for Sicily, Sardinia and Trentino Alto Adige); regional demographic variables and the CPI for food are from the Italian Institute of Statistics (ISTAT).

Table 2. Estimated Parameters

Variable	Full Sample (*N*=254,716)		No Access Issues (*N*=174,741)	Perceived Access Issues (*N*=79,975)
	PROBIT	2SIV	2SIV	2SIV
AVSIZE	0.1228***	0.2980***	0.3080***	0.2844**
	(0.0270)	(0.0626)	(0.0730)	(0.1230)
PCSTORE	0.0563*	0.2025**	0.1981*	0.2115
	(0.0327)	(0.0846)	(0.1021)	(0.1538)
SCALESQM	0.0371	0.0455	0.0908*	-0.0376
	(0.0276)	(0.0481)	(0.0566)	(0.0912)
FRESHSQM	0.0050	0.1930***	0.1876***	0.2066*
	(0.0235)	(0.0530)	(0.0600)	(0.1115)
PCCARS	0.0178**	0.0178**	0.0130	0.0289*
	(0.0084)	(0.0084)	(0.0101)	(0.0154)
NOPUBTRAN	-0.0126	-0.0137	-0.0153	-0.0082
	(0.0192)	(0.0192)	(0.0252)	(0.0305)
PCCARS*NOP	0.0423***	0.0420***	0.0318*	0.0565**
	(0.0148)	(0.0148)	(0.0192)	(0.0238)
PUBTRUSFR	0.1231***	0.1227***	0.1368***	0.0647**
	(0.0150)	(0.0150)	(0.0173)	(0.0313)
PUBTRUSSOM	0.0447***	0.0451***	0.0509***	0.0282
	(0.0131)	(0.0131)	(0.0153)	(0.0258)
HHSIZE	0.0195***	0.0196***	0.0198**	0.0170
	(0.0065)	(0.0065)	(0.0078)	(0.0117)
NCHILD(0-5)	-0.0616***	-0.0619***	-0.0713***	-0.0403*
	(0.0122)	(0.0122)	(0.0147)	(0.0220)
NCHILD(6-17)	-0.0229***	-0.0231***	-0.0253***	-0.0179
	(0.0074)	(0.0074)	(0.0089)	(0.0136)
AGE	0.0207***	0.0206***	0.0222***	0.0170***
	(0.0018)	(0.0018)	(0.0021)	(0.0031)
AGE2	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
FEMALE	0.1682***	0.1680***	0.1757***	0.1518***
	(0.0100)	(0.0100)	(0.0120)	(0.0181)
NOTMARRIED	0.0345***	0.0347***	0.0430***	0.0163
	(0.0109)	(0.0109)	(0.0132)	(0.0197)
COLL-HIGHER	0.1921***	0.1917***	0.1845***	0.2156***
	(0.0159)	(0.0159)	(0.0185)	(0.0312)
SOMECOLL	0.1625***	0.1615***	0.1643***	0.1513***
	(0.0268)	(0.0268)	(0.0312)	(0.0529)
HISCHOOL	0.0901***	0.0894***	0.0876***	0.0964***
	(0.0102)	(0.0102)	(0.0121)	(0.0188)
NOEDUC	-0.1282***	-0.1283***	-0.1117***	-0.1401***
	(0.0249)	(0.0249)	(0.0330)	(0.0385)
UNEMPL	0.0075	0.0087	-0.0194	0.0672*
	(0.0203)	(0.0203)	(0.0247)	(0.0360)
PENSION	0.0457***	0.0452***	0.0395**	0.0639**
	(0.0148)	(0.0148)	(0.0178)	(0.0269)
PROP_CAP	0.1542***	0.1522***	0.1782***	0.0782
	(0.0571)	(0.0571)	(0.0662)	(0.1142)
FAMILYSUP	0.0125	0.0130	0.0075	0.0315
	(0.0332)	(0.0332)	(0.0403)	(0.0588)
ECONBAD	-0.0355***	-0.0355***	-0.0192*	-0.0731***
	(0.0096)	(0.0096)	(0.0117)	(0.0171)
ECONVERYB	-0.0658***	-0.0663***	-0.0312	-0.1535***
	(0.0211)	(0.0211)	(0.0254)	(0.0381)

REST	-0.0867***	-0.0858***	-0.0658**	-0.1406***
	(0.0256)	(0.0256)	(0.0301)	(0.0490)
CANTEEN	0.0248	0.0259	0.0256	0.0269
	(0.0203)	(0.0203)	(0.0240)	(0.0381)
COFFEESHOP	-0.1414***	-0.1405***	-0.1428***	-0.1388**
	(0.0305)	(0.0305)	(0.0356)	(0.0595)
SMOKER	-0.1514***	-0.1518***	-0.1575***	-0.1412***
	(0.0114)	(0.0115)	(0.0136)	(0.0213)
SPORT	0.1533***	0.1533***	0.1551***	0.1531***
	(0.0115)	(0.0115)	(0.0134)	(0.0223)
TVHOURS	0.0254***	0.0255***	0.0226***	0.0309***
	(0.0025)	(0.0025)	(0.0031)	(0.0044)
CPIFOOD	0.1028	-0.3476	-0.3812	-0.4577
	(0.3493)	(0.5215)	(0.6468)	(0.8937)
Constant	-3.7447***	-6.1214***	-6.4240***	-5.3841***
	(0.4921)	(0.8551)	(0.9934)	(1.6940)
Pseudo R2		0.0237	0.0247	0.0265
P-value Exog		0.0021	0.0171	0.1088
ALN (p-Val)		0.1403	0.1908	0.2643
F-stat inst		$F_{(13,254647)}$	$F_{(13,174627)}$	$F_{(13,79906)}$
	AVSIZE	26307.2	18138.3	8256.62
	PCSTOR	10788.8	7079.02	3688.78
	SCALES	8779.11	6336.73	2487.47
	FRESHM	11942.9	9126.02	2982.57

Note: Regional fixed-effects and year dummies' coefficients excluded for brevity. Asterisks *, **, and *** represent 10%, 5%, and 1% significance levels, respectively. Standard errors appear in parentheses. *Pseudo R2*: Maddala's Pseudo R Squared; *p-value Exog*: p-value of Wald test of exogeneity; ALN test (p-val): p-value of the Amemyia-Lee-Newey minimum distance chi-square statistic; *F-stat inst*.: F-statistic for the test of joint significance of IVs coefficients in first-stage equations.

Table 3 Average Marginal Effects of Selected Variables on FV5 Likelihood

	Full sample		No Access Issues		Perceived Access Issues	
AVSIZE	0.0289	***	0.0304	***	0.0265	**
	(0.0053)		(0.0079)		(0.0111)	
PCSTORE	0.0197	**	0.0195	*	0.0197	
	(0.0082)		(0.0107)		(0.0134)	
SCALESSQM	0.0044		0.0089	*	-0.0035	
	(0.0044)		(0.0055)		(0.0090)	
REFRSQM	0.0188	***	0.0185	***	0.0193	*
	(0.0043)		(0.0056)		(0.0107)	
PCCARS	0.0017	**	0.0013		0.0027	*
	(0.0008)		(0.0009)		(0.0014)	
NOPUBTRAN	-0.0013		-0.0015		-0.0008	
	(0.0016)		(0.0024)		(0.0024)	
PCCARS*NOPUBTRAN	0.0041	***	0.0031	*	0.0053	**
	(0.0012)		(0.0017)		(0.0021)	
PUBTRUSFR	0.0119	***	0.0134	***	0.0060	**
	(0.0015)		(0.0015)		(0.0026)	
PUBTRUSSOME	0.0044	***	0.0050	***	0.0026	
	(0.0014)		(0.0015)		(0.0025)	
NCHILD(0-5)	0.0034	***	0.0042	***	0.0015	
	(0.0010)		(0.0011)		(0.0019)	
NCHILD(6-17)	-0.0060	***	-0.0070	***	-0.0038	**
	(0.0012)		(0.0015)		(0.0018)	
NOTMARRIED	-0.0022	***	-0.0025	***	-0.0017	
	(0.0008)		(0.0009)		(0.0014)	
COLL-HIGHER	0.0186	***	0.0181	***	0.0202	***
	(0.0014)		(0.0019)		(0.0030)	
SOMECOLL	0.0156	***	0.0161	***	0.0141	***
	(0.0027)		(0.0033)		(0.0049)	
HISCHOOL	0.0087	***	0.0086	***	0.0090	***
	(0.0010)		(0.0012)		(0.0018)	
NOEDUC	-0.0124	***	-0.0110	***	-0.0131	***
	(0.0025)		(0.0032)		(0.0036)	
UNEMPL	0.0008		-0.0019		0.0063	*
	(0.0022)		(0.0022)		(0.0034)	
PENSION	0.0044	***	0.0039	**	0.0060	**
	(0.0015)		(0.0018)		(0.0024)	
PROP_CAP	0.0147	***	0.0175	**	0.0073	
	(0.0054)		(0.0072)		(0.0109)	
FAMILYSUP	0.0013		0.0007		0.0029	
	(0.0030)		(0.0037)		(0.0058)	
ECONBAD	-0.0034	***	-0.0019		-0.0068	***
	(0.0009)		(0.0012)		(0.0015)	
ECONVERYBAD	-0.0064	***	-0.0031		-0.0144	***
	(0.0019)		(0.0027)		(0.0036)	
REST	-0.0083	***	-0.0065	**	-0.0131	***
	(0.0025)		(0.0028)		(0.0049)	
CANTEEN	0.0025		0.0025		0.0025	
	(0.0022)		(0.0025)		(0.0037)	
COFFESHOPS	-0.0136	***	-0.0140	***	-0.0130	**
	(0.0025)		(0.0032)		(0.0066)	

Note: Asterisks *, **, and *** represent 10%, 5%, and 1% significance levels, respectively. Standard errors in parentheses, obtained using the delta method.