HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

WP 18/09

Family Ties and Children Obesity in Italy

Federico Crudu; Laura Neri and Silvia Tiezzi

April 2018

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Family Ties and Children Obesity in Italy

March 2018

Abstract

This paper estimates the influence of overweight family members on weight outcomes of Italian children aged 6 to 14 years. We use a new dataset matching the 2012 cross sections of the Italian Multipurpose Household Survey and the Household Budget Survey. Endogenous peer groups within the family are accounted for using a set of instrumental variables. We find evidence of a strong, positive effect of both overweight adults and peer children in the family on children weight outcomes. The impact of overweight peer children in the household is larger than the impact of adults. These findings can help identifying the main factors driving the rise in Italian children obesity in the past few decades.

Key words: children obesity; family ties; IV probit; heteroskedasticity.

JEL classification: I12

1 Introduction

In the past three decades, children overweight and obesity prevalence has risen substantially in most high-income countries (Lobstein et al., 2015). Obesity rates are low in Italy compared to most OECD countries, but the picture is different for children. The OECD reports that the rate of overweight children in Italy is one of the highest in OECD and non-OECD countries (OECD, 2015). According to the fourth wave of the Italian Surveillance System OKkio alla Salute, in 2014 the rate of overweight and obese primary school children in Italy was, respectively, 20.9% and 9.8% (Lauria et al., 2016).[1] Even though obesity prevalence has slightly decreased between 2008 and 2014, a children obesity divide exists with southern regions displaying higher prevalence compared to northern regions (Figure 1).

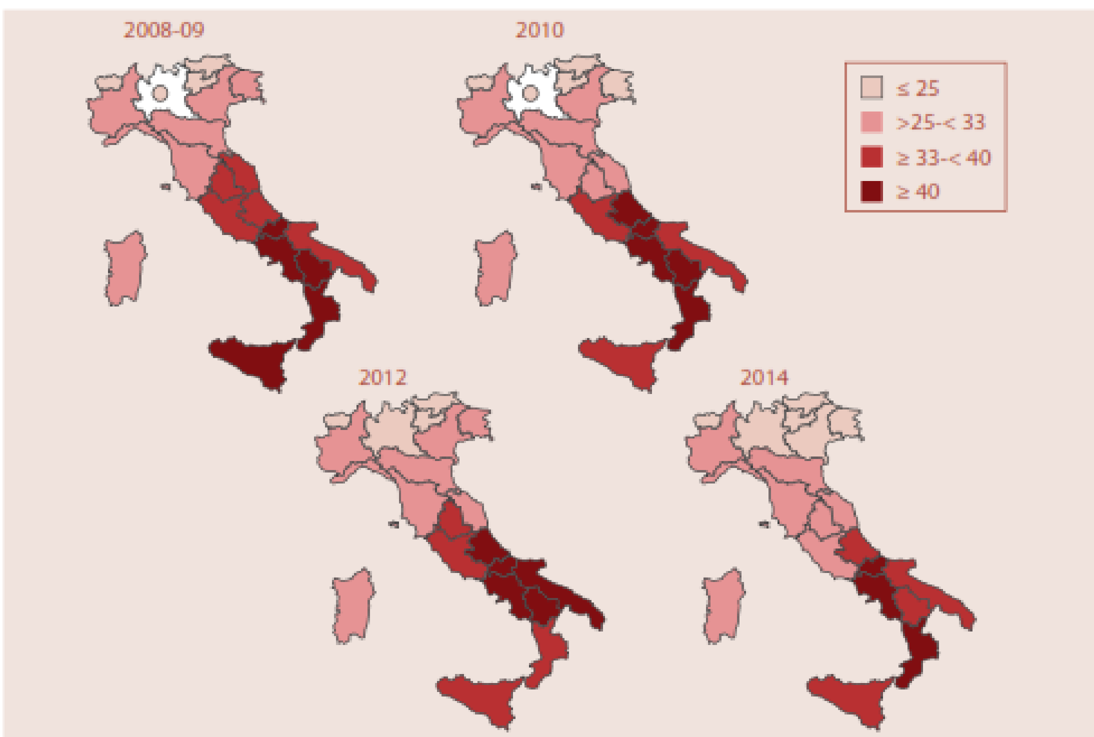

Figure 1: Prevalence of overweight and obese children aged 8-9 years in Italy, 2008-2014. Source: Istituto Superiore di Sanità (ISS), 2016.

[1]The Surveillance System OKkio alla Salute (http://www.epicentro.iss.it/okkioallasalute/) monitors overweight and obesity of Italian children in primary schools (6-11 years of age). The System, promoted and financed by the Italian Ministry of Health, was started in 2007 and participates into the World Health Organization (WHO) European Childhood Obesity Surveillance Initiative (COSI).

Besides representing a threat for children's health and a cost to society, childhood obesity has documented consequences for adult life: effects on self-esteem, body image and confidence, and lower wages as several habits take shape in early life and persist into adulthood (Schwartz et al., 2011). The existing literature on children obesity in Italy mainly belongs the medical sector. Binkin et al. (2010) estimated the prevalence of overweight and obese third-grade children by geographic area using the 2008 wave of the nationally represented nutrition survey OKkio alla Salute. In addition to explaining the high level of childhood obesity in the overall population, higher than that of most Western countries, they produced evidence of substantial geographic differences, with the prevalence of obesity twice as high in the South compared to the North. More recently, Bracale et al. (2013), using data from a special survey conducted in 2008, evaluated the prevalence of childhood overweight and obesity in a sample of school-age children (6-11 years of age) living in Milan and examined socio-cultural, parental and lifestyle factors associated with children's Body Mass Index (BMI) that might affect the risk of obesity. Only moderate levels of above average weight and obesity were found among Milan children. This study confirmed the relevance of socio-cultural aspects as factors affecting weight outcomes in a school-age children population. Despite the potential role for the presence of other overweight and obese family members in shaping children's weight outcomes, no previous research that we are aware of has assessed social interaction within the family as a determinant of childhood obesity. The purpose of this paper is to examine the impact of overweight or obese members of the family on Italian children's weight outcomes. We use a dataset of Italian observations for the year 2012 on children and adolescents aged 6 to 14 years. The dataset is the result of statistical matching of the 2012 cross-section from two distinct surveys that share a set of variables and are representative of the same population. SM allows us to exploit information on consumption expenditures that is included in one of the two surveys but not in the other. After accounting for a large set of controls and using instrumental variables estimation, our data support the hypothesis of a powerful, positive and significant effect of both *obese peer children* and of *obese adults* in the family on Italian children's obesity. Our paper makes three contributions to the literature on social interaction and obesity. First, it is the first study investigating social interaction within the family as a determinant of childhood obesity. Specifically, our main purpose is to investigate the causal impact of the presence of other overweight/obese children or adults in the family on the obese child under consideration. Second, it is the first study on social interaction and childhood obesity in Italy. Several stylized facts make Italian children obesity especially interesting. Firstly, the issue of childhood obesity is paradoxically more prominent in Italy, where the Mediterranean diet is prevalent, than in other countries. Secondly, even though maternal employment is usually associated with higher children weight outcomes, this is hardly the case in Italy where the labor market participation of mothers is much lower than in other European countries (Brilli et al., 2016), while children obesity is higher. Lastly, family ties are culturally strong in Italy thus making social interaction within the family a particularly interesting case to explore. Third, this paper represents a novel contribution to a very scant literature on peer effects and children obesity. To our knowledge, only three papers deal with peer effects in children (Nie et al., 2015; Asirvatham et al., 2014; Gwozdz et al., 2015). Of these, only Gwozdz et al. (2015) deal with European data. This is surprising given the recognition that children consumption decisions are affected by those of their peers (Dishion & Tipsord, 2011) and the recent finding that peer effects are more pronounced in children compared to adolescents (Nie et al., 2015). Moreover, several habits take shape in early life and persist into adulthood (Schwartz et al., 2011). The remainder of the paper proceeds as follows. Section 2 summarizes the literature. Section 3 describes the data and the statistical matching. Section 4 discusses the identification strategy. Section 5 presents the estimation methods and the main results. Section 6 concludes.

2 Literature

The main recognized cause of the rise in children/adolescents obesity is an imbalance between calories intake and calories expenditure. The factors driving this imbalance have been studied by a large literature. One strand of literature has addressed the relationship between maternal employment and children obesity in many developed countries (Cawley & Liu, 2012; Champion et al., 2012; Fertig et al., 2009; Gaina et al., 2009; García et al., 2006; Greve, 2011; Gwozdz et al., 2013; Liu et al., 2009; Morrill, 2011, to cite only a few). Overall, these studies find empirical evidence of a positive relationship between maternal employment and childhood obesity. Maternal employment affects children weight outcomes through a number of channels. For example, Cawley & Liu (2012) find that employed women spend significantly less time eating and playing with their children and are more likely to purchase prepared foods. Fertig et al. (2009) find that maternal employment is related to children BMI through the average number of meals consumed in a day, through reading/talking/listening to the music and through TV watching. A related factor is the increasing use of non-parental child-care, as a key mechanism affecting children weight outcomes (Hubbard, 2008; Herbst & Tekin, 2011). The growing use of non-parental care may play a crucial role in shaping children habits through quality of foods offered and the level of physical activity. Herbst & Tekin (2011) find that center-based care is associated with large and stable increases in BMI throughout the BMI distribution, while the impact of other non-parental arrangements appears to be concentrated at the tails of the distribution. Hubbard (2008) also finds that using non-parental child-care (informal care from a relative, informal care from a baby-sitter, and center care) increases the likelihood of obesity. Although a central research focus has been understanding the mechanisms through which more calories are ingested and less calories are expended, a recent strand of literature, initiated by Christakis & Fowler (2007) has emerged in the field of health economics that addresses the influence of social interactions, particularly of peers, on adolescents' health status. Christakis & Fowler (2007) conducted a study to determine whether obesity might also spread from person to person. Their starting point was that people embedded in social networks are influenced by the behaviors of those around them such that weight gain in one person might influence weight gain in others. They found that, among pairs of adult siblings, if one sibling became obese, the chance that the other would become obese increased by 40% and that if one spouse became obese, the likelihood that the other spouse would also become obese increased by 37%. A subsequent study by Fowler & Christakis (2008) produced evidence of person-to-person spread of obesity in adolescents. A recent review by Powell et al. (2015) has identified social contagion, i.e. the phenomenon whereby the network in which people are embedded influences their weight over time, as one of the social processes explaining the role of social networks in the development of adult overweight and obesity. The general finding in this literature is that weight-related behaviors of adolescents (Fowler & Christakis, 2008; Mora & Gil, 2013; Trogdon et al., 2008) are affected by peer contacts. Less is known about peer effects and childhood obesity. As reported by Nie et al. (2015) most of the empirical literature on peer effects and obesity refers to adolescents or adults and uses US data. We are aware of only three studies, besides ours, analyzing children. Asirvatham et al. (2014) study peer effects in elementary schools using measured obesity prevalence for cohorts within schools using a panel dataset at the grade level from Arkansas public schools. They found that changes in the obesity prevalence at the oldest grade are associated with changes in obesity prevalence at younger grades and that the magnitude of the effect is greater in kindergarten to fourth-grade schools than in kindergarten to sixth-grade schools. Nie et al. (2015) analyze peer effects on obesity in a sample of 3 to 18 years old children and adolescents in China. Peer effects are found to be stronger in rural areas, among females and among individuals in the upper end of the BMI distribution. Finally, Gwozdz et al. (2015) analyze peer effects on childhood obesity using a panel data of children aged two to nine from eight European countries. They show that peer effects are larger in Spain, Italy, and Cyprus – compared to the other European countries in the sample. As far as we know, no study has yet analyzed the impact on children obesity of the obesity status of other members of their own family. The specific purpose of this paper is to extend the recent literature on the influence of peers on children obesity by investigating the effect on Italian obese children aged 6 to 14 years of other obese children in their family. In addition, we also investigate the role of obese adult members of the family on children's weight outcomes.

3 Data and Matching

Studies of peer effects and childhood obesity usually include information on economic characteristics of the household, such as income in addition to personal and socio-demographic information as low-income individuals, for example, are usually found more likely to be obese than high-income ones (Trogdon et al., 2008). In addition, the relationship between income and weight can vary by gender, race-ethnicity and age.[2] For example, the risk of obesity is higher for women and children in low-income or low-socioeconomic groups compared to men. Due to the lack of a single Italian dataset containing both individual weight outcome and socio-economic variables, we have matched two datasets, the 2012 cross-section of the Multipurpose Survey on Households: Aspects of Daily Life (MSH), and the 2012 cross-section of The Household Budget Survey (HBS) including information on households' consumption expenditures.

The MSH and HBS are official sample surveys, conducted by the Italian National Statistical Institute (ISTAT). The MSH for the year 2012 is a large nationally representative sample survey covering 19,330 households and 46,463 family members, including children aged 6 to 14 years. The questionnaire, administered by paper and pencil, contains 3 blocks of questions: a general questionnaire on individual characteristics of the first 6 members of the household; a family questionnaire collecting information about the household habits and lifestyles; a diary on health and nutritional information of each member of the household. For children and adolescents aged 6-17 a binary indicator for whether the child is overweight or obese is also included. The identification of a child as being overweight or obese is based on BMI threshold values for children aged 6 to 17 developed by Cole et al. (2000) and adopted by the International Obesity Task Force (IOTF). MSH, does not contain information on income or expenditures that could potentially be important covariates in our empirical model. The 2012 cross-section of the HBS provides instead information on monthly consumption expenditures of 22,933 Italian households. ISTAT uses a weekly diary to collect data on frequently purchased items and a face-to-face interview to collect

[2] Food Research and Action Center: http://frac.org/obesity-health/relationship-poverty-obesity.

data on large and durable expenditures. Current expenditures are classified in about 200 elementary goods and services. The survey also includes detailed information on the household structure and socio-demographic characteristics (such as regional location, household size, gender, age, education and employment condition of each household member). For both surveys, annual samples are independently drawn according to a two-stage design.[3] Apart from having a large set of variables in common, the two surveys share many characteristics such as the same target population, sampling method, geographic frame, data collection procedure. These common characteristics allow us to use statistical matching (SM henceforth) as an ideal method for combining information on household members' quality of life and children weight outcomes with information on households' consumption expenditures.

3.1 Statistical matching

SM aims at integrating two or more datasets drawn from the same target population (Rodgers, 1984; Cohen, 1991; Radner et al., 1980). In the basic framework there are two data sources (generally data from two different surveys), $\mathcal{A}$ and $\mathcal{B}$. $\mathcal{A}$ contains vector-valued variables (X, Y), whereas $\mathcal{B}$ contains vector-valued variables (X, Z) such that X is shared by both sources. SM uses the X variables common to both surveys as a bridge to create records containing (X, Y, Z) which can then be used to investigate the relationship between Y and Z (D'Orazio et al., 2006). In practice, matching procedures impute the target variables from a donor to a recipient survey. Our purpose was to integrate households' total consumption expenditures (*totexp*) from the HBS (denoted survey $\mathcal{A}$) into the MSH dataset (survey $\mathcal{B}$), as shown in Table 1.

The first step was to identify the matching variables X. Since $\mathcal{A}$ and $\mathcal{B}$ are representative samples of the same population, the common variables are expected to share the same

[3]Details on the sampling procedure used to collect data in both surveys can be found for the MSH survey in: ISTAT (2012) Indagine Multiscopo sulle Famiglie, aspetti della Vita quotidiana, Anno 2012. For the HBS survey in ISTAT (2012) File Standard-Indagine sui Consumi delle Famiglie-Manuale d'uso, anno 2012, Downloadable at $http://www.istat.it/it/archivio/4021$.

HBS Data Set ($\mathcal{A}$, the donor)	MSH Data Set ($\mathcal{B}$, the recipient)	Resulting Data Set
$(X, totexp)$		
	$(X, obechild)$	$(X, obechild, totexp_{imputed})$

Table 1: Statistical Matching Scheme. HBS Data Set (donor) contains vector-valued variables $(X, totexp)$, Multipurpose Survey Data Set contains vector-valued variables $(X, obechild)$, X is shared by both datasets. Statistical matching uses the matching variables X, common to both surveys, as a bridge to create records containing $(X, obechild, totexp_{imputed})$.

marginal/joint distribution. This check was performed using the Cramer's V association measures. Potentially, all the variables identified and chosen according to this check could be used in the SM. In fact, just the most relevant ones have been identified and selected according to a linear model for predicting the logarithm of the target variable "household total expenditure" (*ltotexp*). Table 2 shows the set of matching variables used to predict/impute the target variable. The listed common variables explain 70% of the total variability of the target variable. In addition to these, we include a number of interaction terms in the specification.

NO	Household living in the North West
NE	Household living in the North East
$Central$	Household living in the Centre
$typfam$	Family type: single parent
$typfam1$	Family type: both parents
$typfam2$	Family type: single
$n\_members$	Household size
$n\_members0\_5$	# persons 0-5 years old
$n\_members6\_17$	# persons 6-17 years old
$n\_members18\_34$	# persons 18-34 years old
$n\_members35\_65$	# persons 35-65 years old
$Gender\_RP$	Gender of the reference person
$Mstatus\_RP$	Marital status of the reference person
$Prof\_pos\_RP$	Professional position of the reference person
$Home$	Home ownership
$Rooms$	# of rooms
$Ec\_resource$	Adequacy of economic resources
P_index	Gereral Price Index at regional level
PI_food	Gereral Price Index of Food at regional level

Table 2: Final matching variables.

The next step is imputation from the donor to the recipient. The chosen imputation method is the Sequential Regression Multiple Imputation (Raghunathan et al., 2001) implemented by the software IVE-ware that allows imputation of missing data both in the recipient variables and in the set of matching variables. IVE-ware has a number of desirable properties that make it particularly well suited for imputing missing data in large datasets. For example, each imputation model can be specified according to the nature of the variable to be imputed. The software easily handles arbitrary missing data patterns with categorical and continuous variables. Finally, a sequential method is used to impute missing values: the variable with the least amount of missing data is imputed first and then used in subsequent imputations; the next variable with the second least amount of missing data is then imputed and used in subsequent imputations. The resulting quality of the matching can be assessed by comparing the marginal distribution of the target variable (*ltotexp*) in the observed data (i.e. in the donor dataset) and in the dataset obtained after the matching, as shown in Table 1.

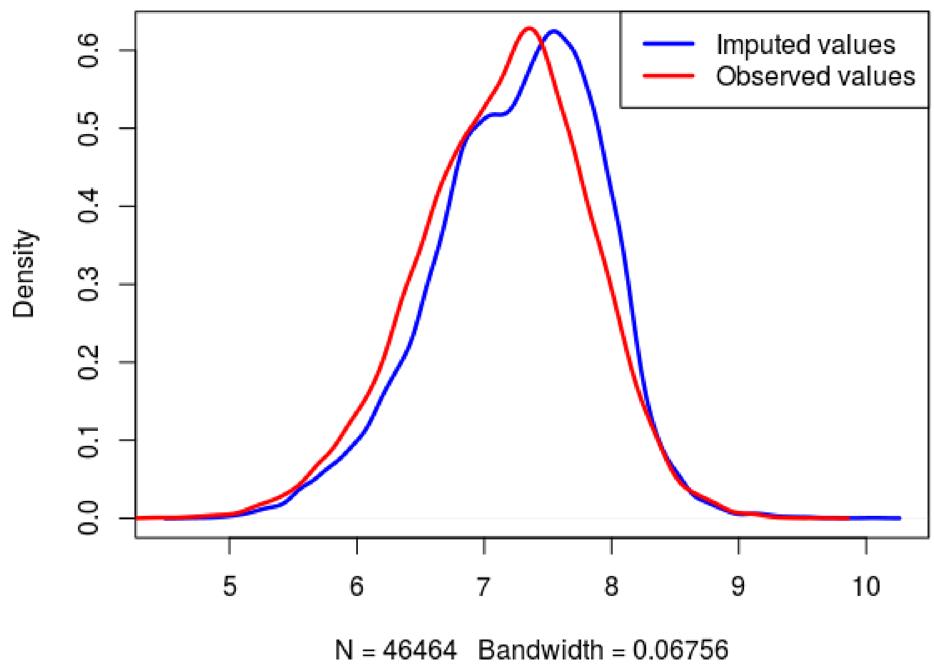

Figure 2: Pdf of the target imputed variable (*ltotexp*) in the original donor dataset (HBS) and in the final resulting dataset. Distribution of the observed target variable (*ltotexp*) collected in the HBS survey (2012) and imputed ltotexp in the datasets obtained as result of the SM procedure.

The distributions of the observed and imputed data are very close, an expected result given the high explanatory power of the predictors included in the model of the target variable. This closeness increases the reliability of the statistical matching. In addition, it reduces the problem of conditional independence, a required hypothesis for the validity of SM (see Appendix A). The empirical analysis that follows only involves the dataset resulting from SM, including all the MSH (the recipient) variables plus the imputed target variable.

4 Identification

We aim at assessing whether the presence of other overweight/obese family members, either children in the same age group or adults, has a positive and significant effect on the probability of a child being overweight/obese. While this is an interesting problem per se, the fact that the impact of children's weight outcomes on the weight outcome of other children in the same family may be considered as a type of peer effect makes this issue even more relevant. We use a narrow peer-group definition that includes all children aged 6 to 14 years belonging to the same family (whether siblings or not) (see e. g. Trogdon et al., 2008).

Christakis & Fowler (2007) were the first to offer support to the social nature of any induction of obesity. They showed that people embedded in social networks are influenced by the behaviors of those around them and as such weight gain in one person might influence weight gain in others. One of their main findings is that, among pairs of adult siblings, if one sibling became obese, the chance that the other would become obese increased by 40%. Obesity in socially close alters might influence a person's obesity through different psychosocial channels such as changing a person's social norm regarding obesity acceptability or, as it might be the case for children, affecting nutritional habits and other health related behaviors. Even though the literature has found strong evidence of social interactions in a variety of contexts, identifying such effects remains a formidable challenge. Manski (1993) showed that empirical identification of peer effects in linear models may fail for a number of reasons. First, peers' weight can directly influence individual weights (endogenous effect). Since we adopt a narrow measure of peer group the endogenous effect might be stronger in our context compared to broader peer-group measures. Second, the exogenous characteristics of the peer group can indirectly influence individual weights (contextual effects). For example, peers' characteristics such as their parents' income, education or occupation influence children's weight. Since children in the same family also share the same parents with high probability, we control for these contextual effects using a number of variables accounting for the characteristics of the family and its members. Third, unobservable variables common to both groups can affect individual weights (correlated effects). The endogenous and contextual effects reflect social interactions. The identification problem in the linear model of social interaction can be avoided by weakening the linearity assumption. Brock & Durlauf (2007) show that, under certain conditions, binary choice models of social interactions are identified (see also Brock & Durlauf, 2001). Borrowing the notation from Blume et al. (2011) we can describe our model of social interactions as

$$\omega_{ig} = k + c'X_i + d'Y_g + Jm_{ig} + \varepsilon_{ig} + \alpha_g \tag{1}$$

where ω_{ig} is a binary indicator that takes value one if individual i in group g is obese according to a body mass index score, X_i is a vector of exogenous effects, Y_g indicates the so called contextual effects, while m_{ig} is the share of obese individuals in family g after excluding individual i. The group fixed effect α_g contains unobserved family characteristics such as genetic features, unobserved habits of the group members, neighborhood characteristics, proximity to fast food restaurants. The identification of the spillover effects due to social interactions is clearly complicated by the presence of the group fixed effect variable. One possible way of dealing with α_g is to resort to instrumental variables. Similarly to other identification aspects, the use of instrumental variables is often subject to criticism. Blume et al. (2011) stress the fact that instrumental variables are difficult to justify given that the group fixed effect is itself undertheorized. A notable exception is the paper by Cooley (2010), where α_g represents teacher quality, while the considered instrumental variable is based on the effect of a school reform in North Carolina. Another interesting exception is the paper by Bramoullé et al. (2009). In this case, valid instruments are generated by exploiting the network structure of peer relationships. Other papers assume that contextual effects Y_g have no direct effect on ω_{ig}. The excluded variables are then used as instruments (see, e.g., Trogdon et al., 2008; Gaviria & Raphael, 2001). In our case, due to the special features of the reference group, the choice of an instrumental variable seems to be even more complicated. Our main problem is to deal with the potential correlation between the unobserved genetic factors that may trigger obesity in the group fixed effect α_g and the endogenous effect m_{ig}.

In order to instrument the endogenous effect m_{ig} we consider the percentage of obese adults per region denoted as *adult_obese_byreg*. We may reasonable assume that such a variable has no direct impact on the weight outcome of individual i and on the characteristics determining overweight that are hidden in the fixed effect α_g. We assume such a variable to be related to the endogenous effect m_{ig}. Given the genetic factors that determine overweight, the probability of being obese, at family level, is higher in an area where the percentage of obese/overweight people is higher (Lauria et al., 2016). This may well reasonably be caused by lifestyle choices and prevailing social norms, independently of the specific genetic characteristic of individual i. The first stage regression is defined as

$$m_{ig} = \pi_0 + \pi_1' X_i + \pi_2' Y_g + \pi_3' Z_{ig} + v_{ig} = \pi' W_{ig} + v_{ig} \tag{2}$$

where Z_{ig} includes *adult_obese_byreg* and other instruments as reported in Table 4.

5 Estimating the probability of child obesity

The dependent variable of our empirical model is a binary variable for a child being overweight/obese (*obechild*). Table 3 shows the summary statistics of the relevant variables in the final dataset. Such variables are related either to the child or to the group he or she belongs to. More specifically, the child related variables are *obechild*, *gender* (equal to one if male) and *age*. With respect to such variables we notice that, on average, overweight and obese children are 29% of Italian children aged 6-14 years, the proportion of male and female children is approximately the same and that their mean age is 10 years.The remaining variables refer to the children's reference group. The mean size of their household is 4 persons (*n_members*). The mean weight and height of adults (aged 17 or older) in the family is 70.6 kg and 170 cm, respectively (*mweight* and *mheight* variables). There is on average 1 adult overweight/obese in the household (*n_obeadult*), while the mean number of additional overweight or obese children in the same household (*n_othobechild*) is 0.159. The estimated models include, instead of the number of other (obese) children in the family (excluding the considered child) and obese adults, their shares, this is, *s_othobechild*, *s_othchild* and *s_obeadult* respectively. Other included individual and family controls are whether the parents consume soda drinks (*soda_drinks*), whether the household lives in a central or northern region (*cn*), the employment status of the household head (*D_student* or *housewife*; *D_unemployed*; *D_retired*), whether there is only one parent in the household (*typfam*), the general price index at regional level (*p_index*) and the logarithm of total monthly consumption expenditure of the household in Euros (*ltotexp*).[4]

Variable	Mean	St. Dev.	Min	Max
obechild	0.292	0.455	0	1
gender(=1 if male)	0.495	0.500	0	1
age	9.987	2.595	6	14
s_othobechild	0.072	0.174	0.000	0.667
s_othchild	0.244	0.266	0.000	0.750
s_obeadult	0.420	0.351	0.000	1.000
n_othobechild	0.159	0.392	0	2
n_othchild	0.537	0.636	0	3
n_obeadult	0.922	0.793	0	3
mweight (kg)	70.676	9.159	35.5	117.5
mheight (cm)	168.959	0.27	110	193
n_members	4.120	1.018	2	11
ltotexp	7.478	0.630	5.469	9.741
p_index	106.022	0.613	104.600	108.100
cn	0.578	0.494	0	1
typfam	0.119	0.324	0	1
adult_obese_byreg	24.954	5.476	17.700	36.100
soda_drinks	0.155	0.362	0	1
s_sibl_lunch	0.186	0.246	0.000	0.750
s_sibl_sport	0.176	0.244	0.000	0.750
prop_tv_oth	0.810	0.317	0.000	1.000
aver_fruit_oth	1.135	0.756	0.000	5.500

Table 3: Summary statistics for the variables included in the estimation problem.

We address the issue of potential endogeneity of the presence of other overweight/obese children in the same family using IV estimation.[5] The instruments we consider are the

[4]The three occupational dummies are defined starting from a categorical variables identifying four possible types of employment positions for each member of the household: 1 = *employed*, 2 = *unemployed or in search of first employment*, 3 = *houseworker or student*, 4 = *retired or not employable*.

[5]Binary choice models in the presence of endogeneity may be estimated in various ways and there is no clear consensus on what the best choice may be. A number of authors (Angrist & Pischke, 2009;

share of siblings (excluding the considered child) aged between 6 and 18 who have lunch at home (*s_sibl_lunch*), the share of siblings aged between 6 and 18 who practice physical activity (*s_sibl_sport*), the proportion of persons watching TV every day excluding children (*prop_tv_oth*), the average portion of fruit for the other members of the family (*aver_fruit_oth*), the percentage of obese adults per region (*adult_obese_byreg*) (see Table 3).[6] For the sake of comparison, we estimate a linear probability model (see columns 1, 3, 4 in Table 4) and a model that adequately captures the features of the dependent variables, i.e. a probit model (columns 2, 5 in Table 4). The presence of endogeneity introduces further complications and needs to be addressed in a suitable manner. The two stage least squares (2SLS) is probably the most popular estimation method for endogenous linear models (see columns 6 in Table 4 for the first stage estimation). It is worth mentioning that the errors of a linear probability model are naturally heteroskedastic. Therefore we will conduct our t tests using robust standard errors. In addition to that we estimate our model with a many instruments consistent estimator that is robust to the presence of heteroskedasticity, the symmetric jackknife estimator with Fuller correction (SJEF, Bekker & Crudu, 2015). As a benchmark estimator we also use the ordinary least square estimator (OLS). Furthermore, we estimate a two-step probit model (Rivers & Vuong, 1988) that takes care not only of our binary dependent variable but also of endogeneity. A standard probit model is estimated as a benchmark. In order to carry out the estimation of our models we use a sample of 3,906 children between 6 and 14 year old. Table 4 shows different sample sizes due to the presence of missing data. Moreover, it is important to notice that the instruments that we use are not weak as suggested by the F test in the first stage regression. Furthermore, the Sargan test associated to the 2SLS is not rejected. This implies that our moment conditions are globally satisfied. We are interested in the coefficients of

Wooldridge, 2010) advocate the use of simple linear probability models. Their claim is that, even though predicted values may fall outside the zero-one interval, the linear probability model is able to consistently estimate marginal effects, which is what most researchers are interested in. On the other hand, some authors (Lewbel et al., 2012) criticize this approach suggesting the linear probability model is not only unable to provide valid predicted values but also fails to recover the correct marginal effects.

[6]We introduce variables associated to siblings aged between 6 and 18 years as they may influence the original target units, i.e. children aged between 6 and 14.

s_othobechild and *s_obeadult*, i.e. the number of overweight children and adults in the household, respectively. In all the empirical models, both variables have a positive and strong significant effect on the probability of a child being overweight and the coefficient of *s_othobechild* is consistently larger than the *s_obeadult* coefficient. Thus, the presence of overweight family members has a strong impact on children weight outcomes, but the presence of peer overweight children is more relevant than that of overweight adults. A brute force method to compare our estimated coefficients in terms of magnitude is to multiply the coefficients associated to the linear probability model by a factor 2.5. Interestingly the signs of the estimates are consistent across all estimators. Moreover, the linear probability model estimated with 2SLS (with robust standard errors) and the SJEF estimator deliver broadly comparable results and are different from the OLS results. In general, we can say that having either an additional obese child or an obese adult in the family increases the probability of observing an obese child.

	Dependent variable:					
	obechild					s_othobechild
	OLS	Probit	2SLS	SJEF	IV Probit	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
constant	3.799***	9.282**	2.132	2.389*	5.648	0.050
	(1.234)	(3.743)	(1.648)	(1.767)	(3.993)	(0.465)
s_othobechild	0.315***	0.935***	2.144***	2.421***	6.725***	
	(0.046)	(0.138)	(0.460)	(0.508)	(1.190)	
s_othchild	−0.094***	−0.312***	−0.627***	−0.709***	−2.012***	0.301***
	(0.030)	(0.095)	(0.135)	(0.151)	(0.365)	(0.021)
s_obeadult	0.169***	0.522***	0.075**	0.053*	0.205**	0.047***
	(0.021)	(0.064)	(0.035)	(0.039)	(0.089)	(0.008)
gender	0.196	0.554	0.549**	0.560**	1.535***	−0.177***
	(0.168)	(0.520)	(0.221)	(0.243)	(0.589)	(0.062)
ltotexp	0.0003	−0.004	0.025	0.027	0.065	−0.010*
	(0.016)	(0.050)	(0.020)	(0.022)	(0.055)	(0.006)
gender × ltotexp	−0.018	−0.049	−0.065**	−0.065***	−0.174**	0.023***
	(0.022)	(0.069)	(0.029)	(0.032)	(0.078)	(0.008)
p_index	−0.034***	−0.095***	−0.019	−0.022*	−0.066*	−0.001
	(0.012)	(0.035)	(0.015)	(0.016)	(0.037)	(0.004)
age	0.028	0.114	0.028	0.032	0.127	−0.001
	(0.024)	(0.076)	(0.029)	(0.032)	(0.080)	(0.009)
age2	−0.003**	−0.009**	−0.003*	−0.003**	−0.010***	0.0001
	(0.001)	(0.004)	(0.001)	(0.002)	(0.004)	(0.0004)
cn	−0.111***	−0.338***	−0.054**	−0.049**	−0.167***	0.010
	(0.015)	(0.045)	(0.022)	(0.024)	(0.060)	(0.009)
typfam	0.061***	0.190***	0.078***	0.064**	0.196***	−0.0002
	(0.022)	(0.069)	(0.025)	(0.028)	(0.074)	(0.008)
$\widehat{v}_{ig}$					−0.901***	
					(0.182)	
adult_obese_byreg						0.004***
						(0.001)
soda_drinks						0.016**
						(0.007)
sibl_pranzo_p						0.043**
						(0.019)
sibl_sport_p						−0.055***
						(0.017)
prop_tv_oth						0.005
						(0.009)
aver_fruit_oth						−0.006*
						(0.003)
Observations	3,906	3,906	3,788	3508	3,508	3,508
R^2	0.070					0.242
Adjusted R^2	0.068					0.239
Log Likelihood		-2,217.176			-1,993.836	
Akaike Inf. Crit.		4,458.353			4,013.672	
Residual Std. Error	0.439 (df = 3894)		0.519 (df = 3776)			0.153 (df = 3491)
F Statistic	26.795*** (df = 11; 3894)					69.723*** (df = 16; 3491)
Wu-Hausman			24.096*** (df = 1; 3775)			
Sargan			7.161 (df = 5)			
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4: Columns (1) to (5) display the estimates for the model in Equation 1, while Column (6) is the estimate of the first stage regression described in Equation 2. Standard errors are found within parentheses.

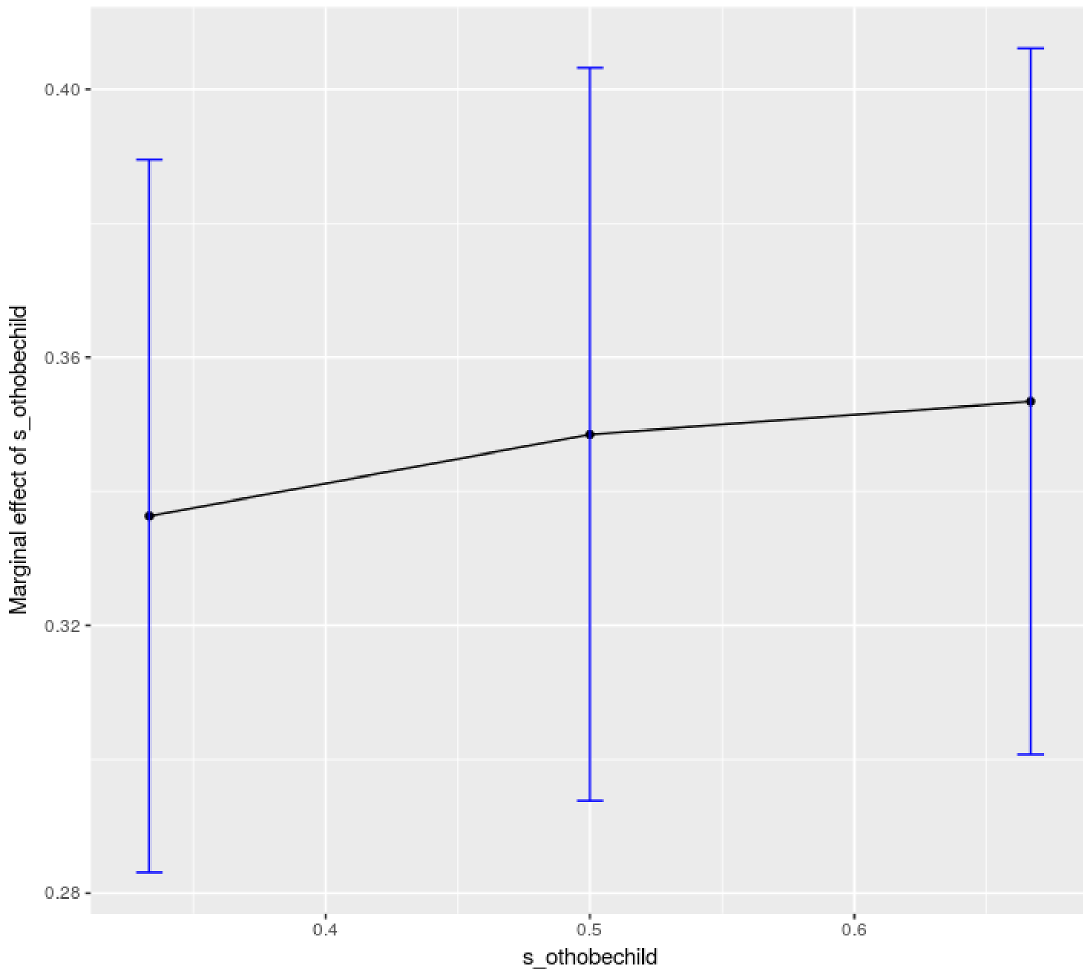

Figure 3: Marginal effect of the probit model with respect to the endogenous variable ($s\_othobechild$) with confidence intervals.

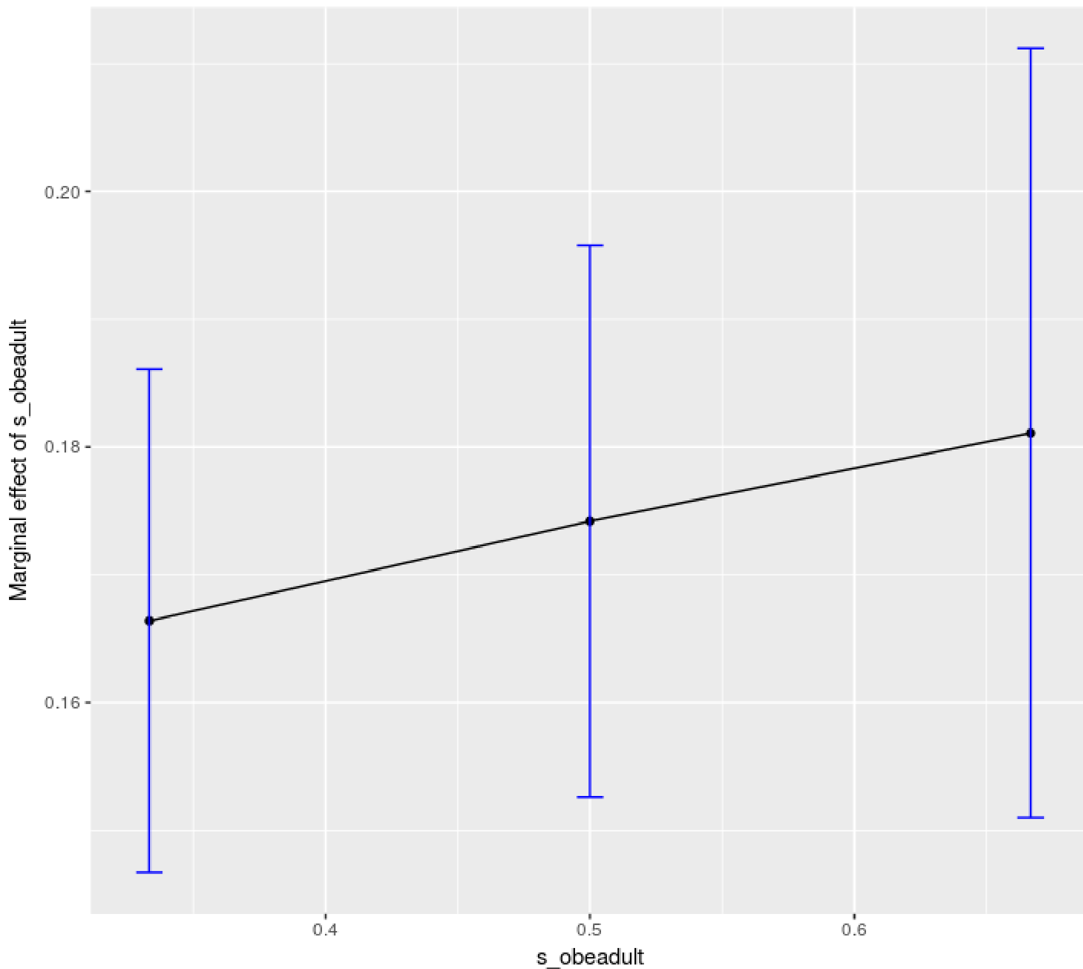

Figure 4: Marginal effect of the probit model with respect to the variable ($s\_obeadult$) with confidence intervals.

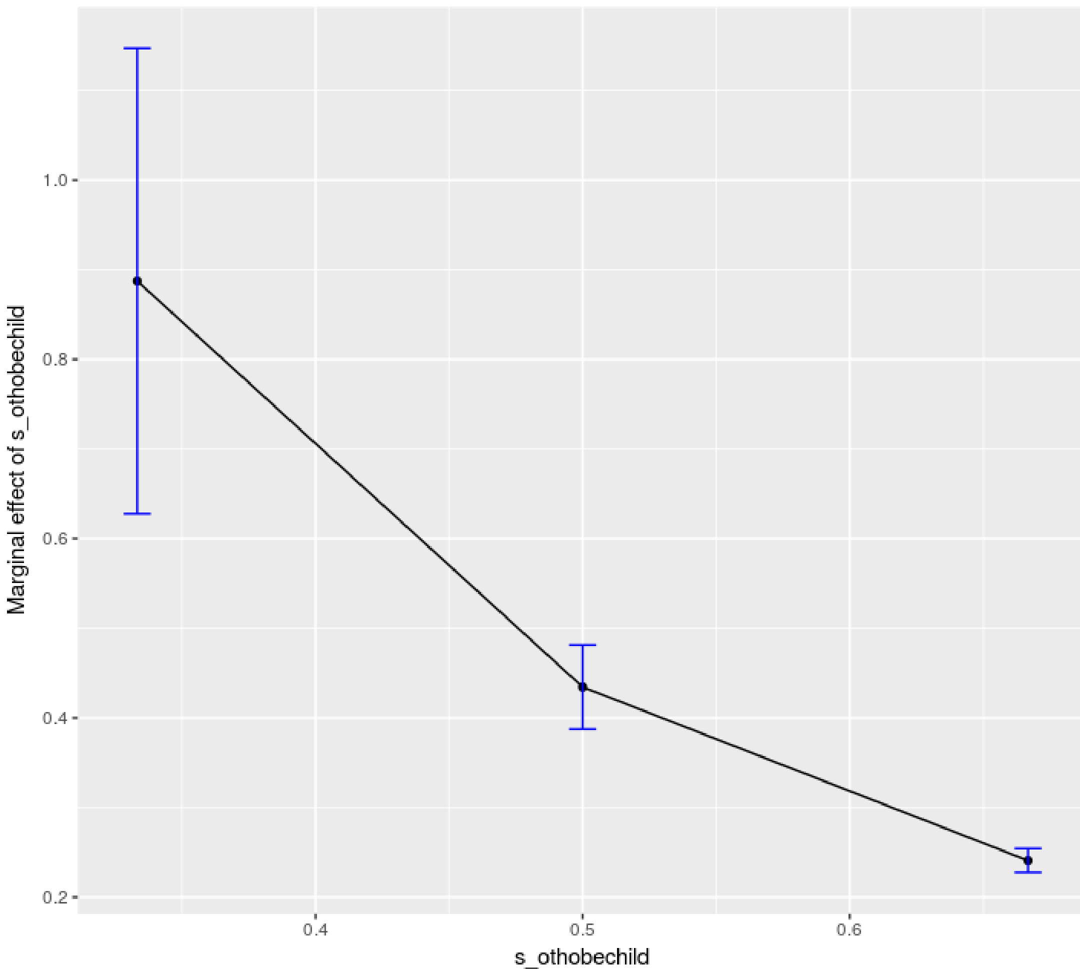

Figure 5: Marginal effect of the IV probit model with respect to the endogenous variable ($s\_othobechild$) with confidence intervals.

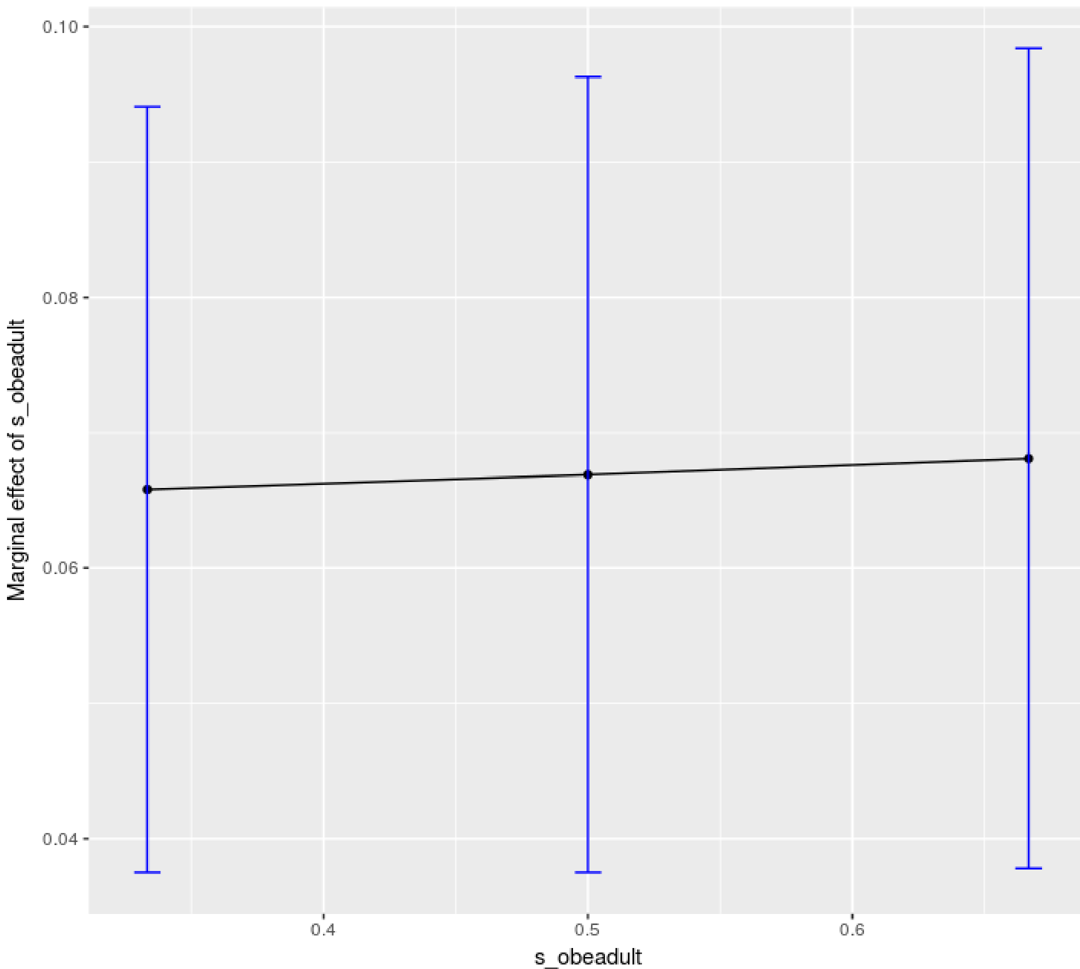

Figure 6: Marginal effect of the IV probit model with respect to the variable ($s\_obeadult$) with confidence intervals.

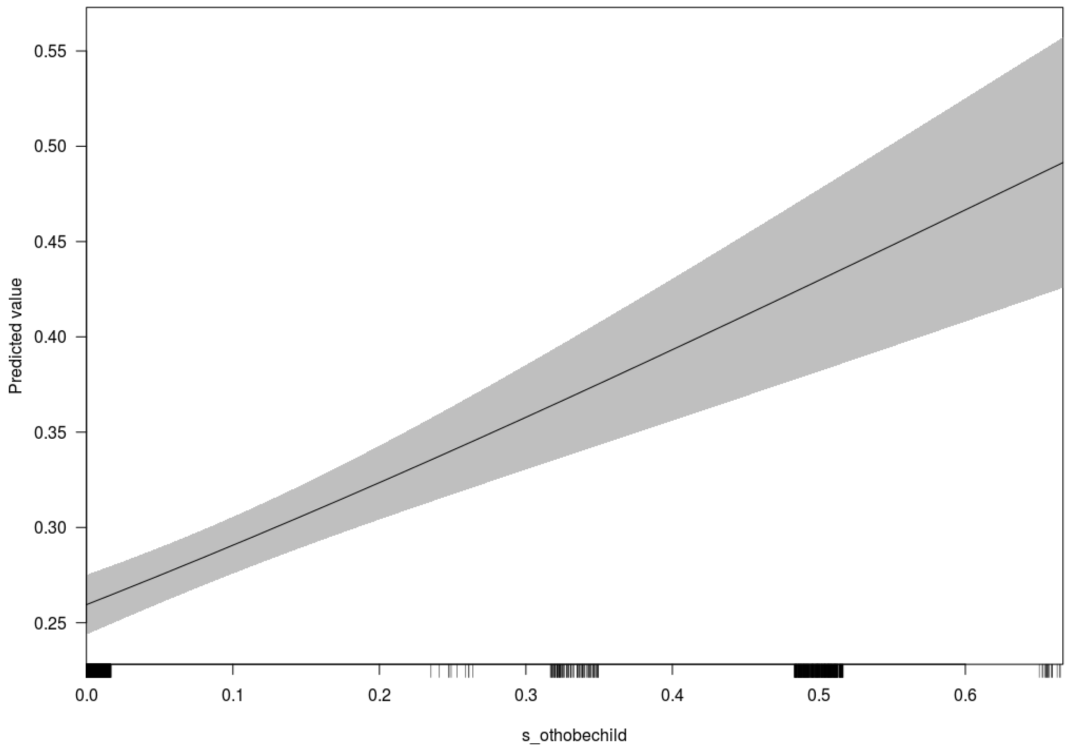

Figure 7: Predicted value of the probit model with respect to the endogenous variable (*s_othobechild*) with confidence intervals.

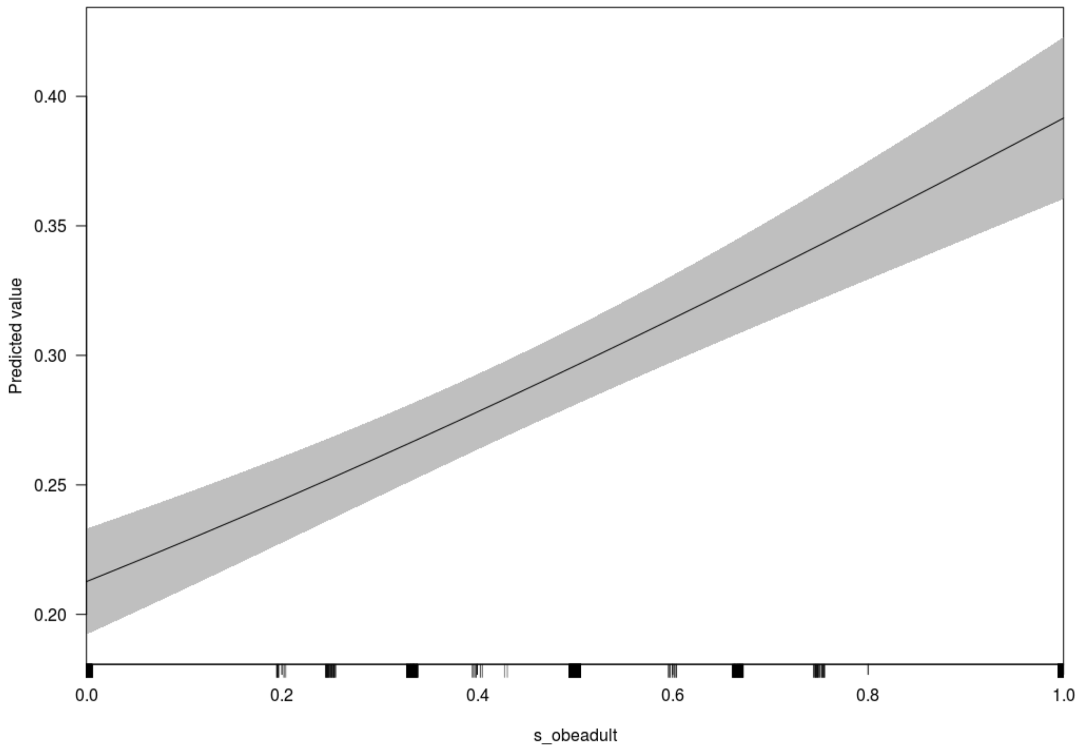

Figure 8: Predicted value of the probit model with respect to the variable (*s_obeadult*) with confidence intervals.

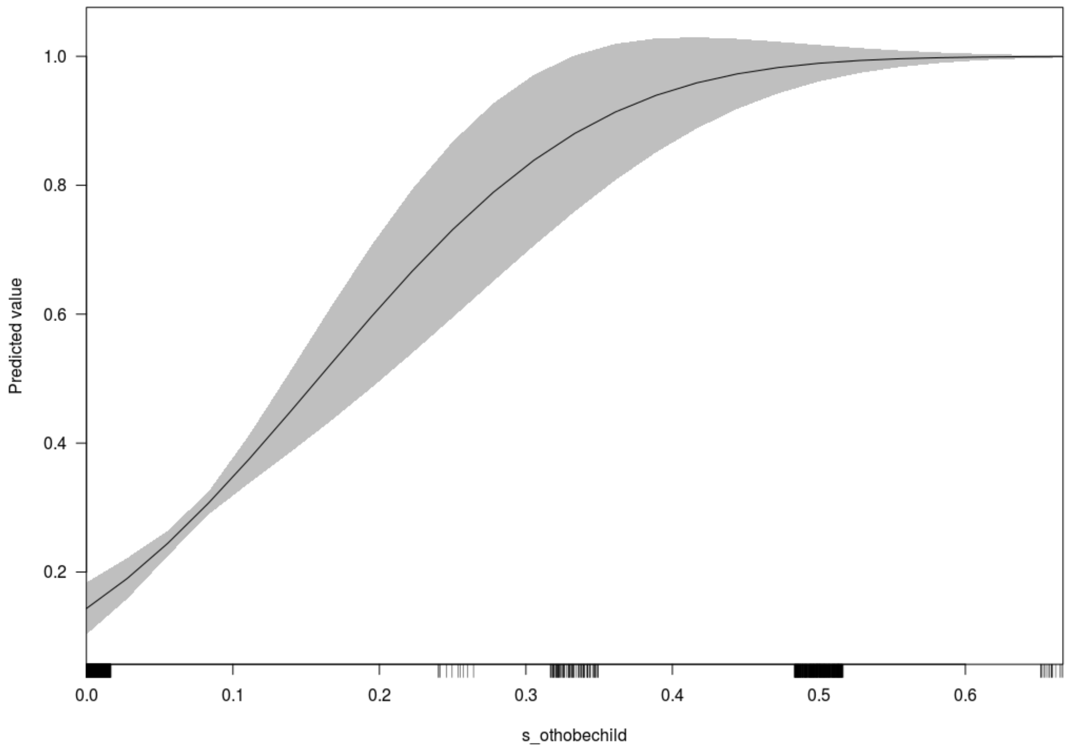

Figure 9: Predicted value of the IV probit model with respect to the endogenous variable ($s\_othobechild$) with confidence intervals.

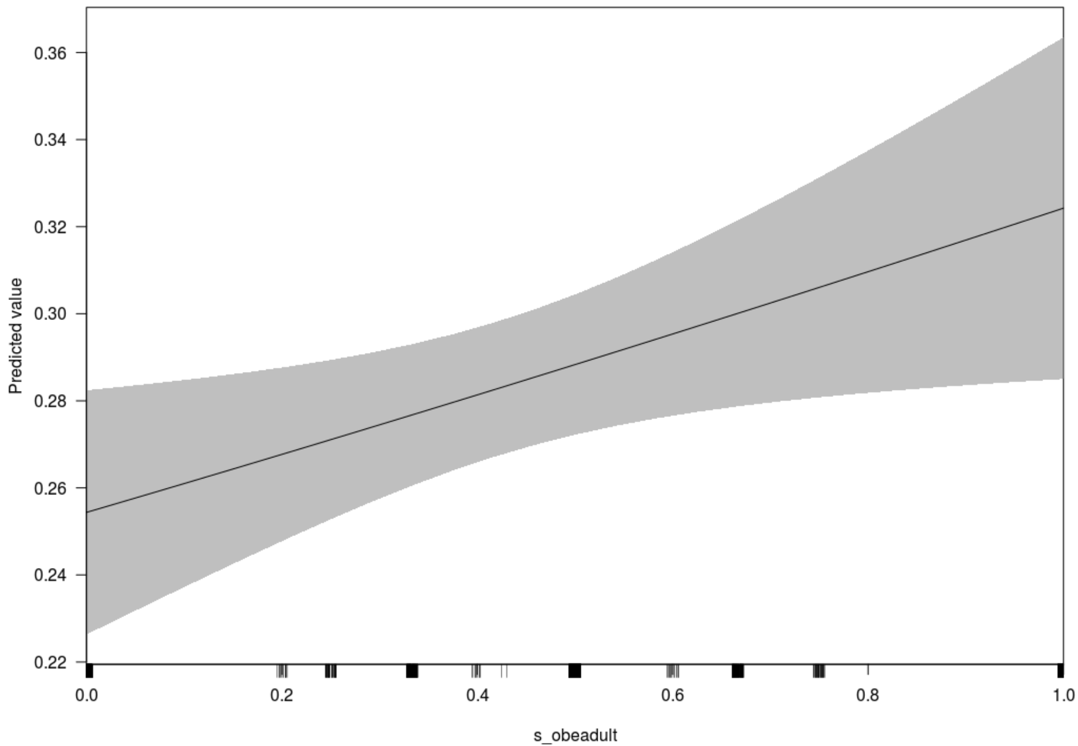

Figure 10: Predicted value of the IV probit model with respect to the variable ($s\_obeadult$) with confidence intervals.

When we compare probit and IV probit results, in particular with respect to $s\_othobechild$, we observe some peculiar differences. The marginal effect resulting from the simple probit is nearly flat, between 0.30 and 0.35, while the IV probit displays a more complex marginal effect structure, as it can be seen in Figure 3 and Figure 5. More specifically, with respect to Figure 5, for credible values of $s\_othobechild$, say, $\frac{1}{3}$, $\frac{1}{2}$ and $\frac{2}{3}$, the probability of individual i to be obese is always positive, with a clear decreasing pattern. When we consider $s\_obeadult$, we notice a significantly different behaviour. The prediction, both in the case of probit and IV probit, is nearly a straight line. Hence, the marginal effect tends to be nearly flat and it takes values between 0.16 and 0.18 in the case of the standard probit and between 0.06 and 0.07 in the case of the IV probit.

As to the other explanatory variables, living in Central or Northern Italy has a negative and significant impact on the probability of a child being overweight in all of our empirical models, as expected, thus confirming the existence of a North-South divide in Italy, at least as far as children obesity is concerned. Similarly, single parent families have a positive and significant effect on child obesity. The coefficients associated to age and age squared define an inverted-U function suggesting that age will have a marginally decreasing effect on child obesity. A further interesting aspect is the interaction between $ltotexp \times gender$ suggesting that girls may have a lower probability of being obese when compared to boys. However, the first stage regression suggests that the effect of *ltotexp* on the number of other obese children in the family is negative and non negligible, highlighting that children living in low income families are more likely to be obese.

6 Conclusions

This paper extends the recent literature on children obesity on two fronts. First, we assess whether the presence of overweight family members – other children and adults – affects children's weight outcomes. To our knowledge no study has yet analyzed the impact on children obesity of the obesity status of other members of their own family. Second, as the rate of overweight children in Italy is one of the highest in OECD and non-OECD countries (OECD, 2015) we offer additional insights into the determinants of Italian children overweight and obesity. We use a new dataset resulting from statistical matching of the 2012 cross sections of two surveys, the Multipurpose Household Survey and the Household Budget Survey, both supplied by ISTAT. Endogenous peer groups within the family are accounted for using a set of instrumental variables. We find evidence of a strong, positive impact of overweight adults and of peer obese children in the family on children weight outcomes. Interestingly, in all empirical models we find that the impact of overweight peer children in the household is larger than the impact of adults, a result that deserves further investigation. Despite the need to develop targeted approaches for obesity prevention in Italian children most at risk, empirical evidence on the factors affecting Italian children weight outcomes remains poor. We show that social interaction within the family is one important factor in explaining the prevalence of Italian children weight outcomes. Further exploration of causal pathways linking social interaction within the family and children obesity is thus desirable.

References

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics.* Princeton.

Asirvatham, J., Nayga, R. M., Jr. & Thomsen, M. R. (2014). Peer-effects in obesity among public elementary school children: A grade-level analysis. *Applied Economic Perspectives and Policy*, *36*(3), 438-459.

Baker, M. & Milligan, K. (2008). Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates. *Journal of Health Economics*, *27*(4), 871 - 887.

Bekker, P. A. (1994). Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, *62*(3), 657-681.

Bekker, P. A. & Crudu, F. (2015). Jackknife instrumental variable estimation with heteroskedasticity. *Journal of Econometrics*, *185*(2), 332-342.

Bekker, P. A. & van der Ploeg, J. (2005). Instrumental variable estimation based on grouped data. *Statistica Neerlandica*, *59*(3), 239-267.

Binkin, N., Fontana, G., Lamberti, A., Cattaneo, C., Baglio, G., Perra, A. & Spinelli, A. (2010). A national survey of the prevalence of childhood overweight and obesity in italy. *Obesity Reviews*, *11*(1), 2–10.

Blume, L. E., Brock, W. A., Durlauf, S. N. & Ioannides, Y. M. (2011). Identification of social interactions. In *Handbook of social economics* (Vol. 1, pp. 853–964).

Bracale, R., Milani, L., Ferrara, E., Balzaretti, C., Valerio, A., Russo, V., … Carruba, M. O. (2013). Childhood obesity, overweight and underweight: a study in primary schools in milan. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, *18*(2), 183–191.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*(1), 41–55.

Brilli, Y., Del Boca, D. & Pronzato, C. (2016). Does child care availability play a role in maternal employment and children's development? evidence from italy. *Review of Economics of the Household*, *14*(1), 27-51.

Brock, W. A. & Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, *68*(2), 235–260.

Brock, W. A. & Durlauf, S. N. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics*, *140*(1), 52–75.

Cawley, J. & Liu, F. (2012). Maternal employment and childhood obesity: A search for mechanisms in time use data. *Economics & Human Biology*, *10*(4), 352-364.

Champion, S. L., Rumbold, A. R., Steele, E. J., Giles, L. C., Davies, M. J. & Moore, V. M. (2012). Parental work schedules and child overweight and obesity. *International Journal of Obesity*, *36*, 573–580.

Christakis, N. A. & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine*, *357*, 370–379.

Cohen, M. L. (1991). Statistical matching and microsimulation models. In C. F. Citro & E. A. Hanushek (Eds.), *Improving Information for Social Policy Decisions – The Uses of Microsimulation Modeling: Volume II, Technical Papers* (pp. 62–88). National Academy Press.

Cole, T. J., Bellizzi, M. C., Flegal, K. M. & Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal*, *320*(7244), 1240.

Cooley, J. (2010). Desegregation and the achievement gap: Do diverse peers help? *Unpublished manuscript, University of Wisconsin-Madison*.

Dishion, T. J. & Tipsord, J. M. (2011). Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology*, *62*(1), 189-214.

D'Orazio, M., Zio, M. D. & Scanu, M. (2006). *Statistical matching: Theory and practice (Wiley series in survey methodology).* John Wiley & Sons.

Fertig, A., Glomm, G. & Tchernis, R. (2009). The connection between maternal employment and childhood obesity: Inspecting the mechanisms. *Review of Economics of the Household*, *7*(3), 227-255.

Fowler, J. & Christakis, N. (2008). Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher and Trogdon, Nonnemaker, and Pais. *Journal of Health Economics*, *27*(5), 1400-1405.

Gaina, A., Sekine, M., Chandola, T., Marmot, M. & Kagamimori, S. (2009). Mother employment status and nutritional patterns in japanese junior high schoolchildren. *International Journal of Obesity*, *33*(7), 753–757.

García, E., Labeaga, J. M. & Masagué, A. C. O. (2006, July). *Maternal Employment and Childhood Obesity in Spain* (Working Papers No. 2006-17). FEDEA.

Gaviria, A. & Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, *83*(2), 257–268.

Greve, J. (2011). New results on the effect of maternal work hours on children's overweight status: Does the quality of child care matter? *Labour Economics*, *18*(5), 579-590.

Gwozdz, W., Sousa-Poza, A., Reisch, L. A., Ahrens, W., Eiben, G., Fernandéz-Alvira, J. M., … others (2013). Maternal employment and childhood obesity – A European perspective. *Journal of Health Economics*, *32*(4), 728–742.

Gwozdz, W., Sousa-Poza, A., Reisch, L. A., Bammann, K., Eiben, G., Kourides, Y., … others (2015). Peer effects on obesity in a sample of European children. *Economics & Human Biology*, *18*, 139–152.

Herbst, C. M. & Tekin, E. (2011). Child care subsidies and childhood obesity. *Review of Economics of the Household*, *9*(3), 349–378.

Hubbard, M. (2008). The effect of mothers' employment and child care decisions on the body mass status of young children. *The University of North Carolina at Chapel Hill Working Paper*.

Lauria, L., Pizzi, E., Andreozzi, S. & Galeone, D. (2016). Il Sistema di sorveglianza OKkio alla SALUTE: risultati 2014. *Istituto Superiore della Sanità Technical Report*.

Lewbel, A., Dong, Y. & Yang, T. T. (2012). Comparing features of convenient estimators for binary choice models with endogenous regressors. *Canadian Journal of Economics/Revue Canadienne d'Économique*, *45*(3), 809–829.

Liu, E., Hsiao, C., Matsumoto, T. & Chou, S. (2009). Maternal full-time employment and overweight children: Parametric, semi-parametric, and non-parametric assessment. *Journal of Econometrics*, *152*(1), 61–69.

Lobstein, T., Jackson-Leach, R., Moodie, M. L., Hall, K. D., Gortmaker, S. L., Swinburn, B. A., ... McPherson, K. (2015). Child and adolescent obesity: part of a bigger picture. *The Lancet*, *385*(9986), 2510–2520.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Mora, T. & Gil, J. (2013). Peer effects in adolescent bmi: evidence from spain. *Health Economics*, *22*(5), 501–516.

Morrill, M. S. (2011). The effects of maternal employment on the health of school-age children. *Journal of Health Economics*, *30*(2), 240–257.

Nie, P., Sousa-Poza, A. & He, X. (2015). Peer effects on childhood and adolescent obesity in china. *China Economic Review*, *35*, 47–69.

OECD. (2015). Overweight and obesity among children.

Powell, K., Wilcox, J., Clonan, A., Bissell, P., Preston, L., Peacock, M. & Holdsworth, M. (2015). The role of social networks in the development of overweight and obesity among adults: a scoping review. *BMC Public Health*, *15*(1), 996.

Radner, D., Allen, R., Gonzalez, M., Jabine, T. & Muller, H. (1980). Report on exact and statistical matching techniques. *Statistical Policy Working Paper 5. U.S. Department of of Commerce, Office of Federal Statistical Policy and Standards*.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J. & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*(1), 85–96.

Rivers, D. & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, *39*(3), 347–366.

Rodgers, W. L. (1984). An evaluation of statistical matching. *Journal of Business & Economic Statistics*, *2*(1), 91–102.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, *91*(434), 473–489.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A. & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759.

Trogdon, J. G., Nonnemaker, J. & Pais, J. (2008). Peer effects in adolescent overweight. *Journal of Health Economics*, *27*(5), 1388–1399.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, *16*(3), 219–242.

Von Hinke Kessler Scholder, S. (2008). Maternal employment and overweight children: does timing matter? *Health Economics*, *17*(8), 889–906.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Family Ties and Children Obesity in Italy

Running title: Family Ties and Obesity

Federico Crudu* Laura Neri† Silvia Tiezzi‡

University of Siena and CRENoS University of Siena University of Siena

March 2018

Abstract

This paper estimates the influence of overweight family members on weight outcomes of Italian children aged 6 to 14 years. We use a new dataset matching the 2012 cross sections of the Italian Multipurpose Household Survey and the Household Budget Survey. Endogenous peer groups within the family are accounted for using a set of instrumental variables. We find evidence of a strong, positive effect of both overweight adults and peer children in the family on children weight outcomes. The impact of overweight peer children in the household is larger than the impact of adults. These findings can help identifying the main factors driving the rise in Italian children obesity in the past few decades.

Key words: children obesity; family ties; IV probit; heteroskedasticity.

JEL classification: I12

*Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, federico.crudu@unisi.it, Tel.: +39 0577 232603.

†Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, laura.neri@unisi.it.

‡Department of Economics and Statistics, University of Siena, Piazza San Francesco, 7/8 53100 Siena, silvia.tiezzi@unisi.it.