Discussion Papers
Department of Economics
University of Copenhagen

No. 12-03

# Eating Behavior and Social Interactions from Adolescence to Adulthood

Luisa Corrado and Roberta Distante

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

# Eating Behavior and Social Interactions from Adolescence to Adulthood

Luisa Corrado*    Roberta Distante†

This Version: June 19, 2012

**Abstract**

This paper analyzes the importance of social ties for eating behavior of US youth. We propose a novel approach that addresses identification of social endogenous effects. We overcome the problem of measuring the separate impact of endogenous and contextual effects on individual Body Mass Index (BMI) in a dynamic linear-in-means model, where individual- and group-specific unobservable effects are controlled for. We show that the main drivers of eating behavior are habituation and imitation effects. Imitation effects explain most of the variation in BMI of individuals who were normal-weight and overweight during adolescence. Obese adolescents, instead, become future obese adults through wrong habits enforced by imitative behavior.

*Key-Words*: Overweight, Obesity, Peer Effects, Social Networks, Personal History, Dynamic Linear-in-means Model.

*JEL*: D10; D71; I19; J11; Z13

*Faculty of Economics, University of Rome Tor Vergata, Italy and University of Cambridge, CreMic and CIMF. E-mail: lc242@econ.cam.ac.uk. Luisa Corrado gratefully acknowledges the Marie-Curie Intra European fellowship 039326.

†Department of Economics, University of Copenhagen, Denmark.
E-mail:roberta.distante@econ.ku.dk.

# 1 Introduction

Overweight and obesity are social plagues of modern societies. According to the World Health Organization (WHO), there are more than one billion overweight adults in our globe, at least three hundred million of them obese, and figures are even worse for children and adolescents. This is a multifaceted condition with social and psychological dimensions in all ages and socioeconomic groups. The rising epidemic surely reflects significant changes in eating behavioral patterns of communities: over-consumption of carbohydrates and saturated fats as well as scarce physical activity are important causes. However, social and peer effects likely act as drivers of such a large-scale phenomenon, that seems to occur independent of cultural, economic, and environmental circumstances (Cohen-Cole, 2006; Christakis and Fowler, 2007; Trogdon et al., 2008; Fortin and Yazbeck, 2011). Indeed, many epidemic phenomena occur because they spread within social groups through the homogenization of behaviors among individuals of the same group (Manski, 2000): hence, the propensity of a person to behave in a certain way varies positively with the dominant behavior in her group, similar to informal enforcement mechanisms or social norms (Kandori, 1992; Bernheim, 1994).[1]

The aim of the present paper is to estimate the impact of social and peer effects on eating behavior of US adolescents who transition into adulthood.[2] By means of a novel, yet simple empirical approach we overcome the problem of identifying the impact of social endogenous effects on individual Body Mass Index (BMI). Furthermore, we examine to what extent adulthood BMI status depends on habituation and imitation during adolescence.

Adopting the notion first introduced by Manski (1993), we can identify three types of group effects impacting on individual behavior: endogenous effects, which occur when individual behavior varies with the behavior of the group; contextual effects, that arise when peer group characteristics directly affect individual behavior; correlated or group (unobservable) effects, that arise because group members share a common environment or common latent traits that affect their individual behavior. Analyzing the statistical effect of social interactions is generally challenging due to a special kind of identification problem, the so called reflection problem (Manski, 1993): in a linear-in-means model of social interactions, the distinct role of endogenous and contextual effects may be difficult

[1] Peer effects have been extensively examined both in education (Bénabou, 1993) and in psychology (Brown, 1990; Brown et al. 1996). For a review of the literature on social interaction effects see Brock and Durlauf (2001).

[2] The expression 'eating behavior' comprises all the actions having influence on body weight, e.g. quantity and quality of food, physical exercise, and lifestyle-related issues.

to disentangle because such effects co-move. Since the work of Manski (1993), many are the studies that tackle the estimation of peer effects (see Brock and Durlauf, 2001; Moffitt, 2001 for a review of the literature). Recent empirical work seems characterized by writhed frameworks that require information on the network structure (Bramoullé et al., 2009; Petacchini et al., 2010; Fortin and Yazbeck, 2011; Corrado and Fingleton, 2012) or on out of group effects (Cohen-Cole, 2006).[3] Different from this literature, we resort on a simple and reasonable framework to identify social endogenous effects in a linear-in-means model. Specifically, we estimate a dynamic linear-in-means model that allows individual behavior to linearly depend on individual past behavior as well as on group-specific effects, which include some group observable characteristics and the expected aggregate behavior of the others in the group. Such an assumption makes sense when not only choice is thought of as being the result of social and peer effects, but also of past behavior. Habituation as well as social behavior are possible determinants of individual choice, especially in the case of eating decisions. Furthermore, our econometric strategy allows us to control for endogeneity, individual and group heterogeneity by exploiting stationarity restrictions of a system GMM estimator augmented to control for individual- and group-specific unobservable effects.

We make use of the National Longitudinal Study of Adolescent Health (Add Health) dataset, a (US) representative sample of adolescents who transition into early adulthood for which information on demographic, health and socioeconomic status is registered along four waves, from 1994 to 2008.[4] We are able to study the behavioral causes of overweight and obesity among teenagers, and the effects of such behavior during their transition to adulthood. In contrast to Cohen-Cole and Fletcher (2008), who also focus on obesity and social interactions using the Add Health dataset, the results of our dynamic linear-in-means model show that the tendency of individuals to become overweight is the outcome of both social effects and past individual behavior. Cohen-Cole and Fletcher (2008) general finding is that social interactions with closest peers are not significant once fixed effects for social groups and individuals are accounted for.

[3] Specifically, Cohen-Cole (2006) uses out of group effects to identify endogenous and contextual effects. However the method does not take into account individual- and group-specific unobservable effects. In another recent paper De Giorgi et al. (2010) show that, in a context where peer groups do not overlap fully, it is possible to identify all the relevant parameters of the standard linear-in-means model of social interactions. The Instrumental Variable approach proposed in their paper, though, while properly accounting for correlated unobservable effect at the group level, does not provide a separate estimation of unobservable effects at the individual and group level.

[4] The time distance between each wave is not homogeneous. The first and the second waves are two consecutive years, the third is 6 years later than the second, and the fourth 6 years later than the third. We have weighted the sample in order to account for this difference in gaps between waves (Cf. Appendix B).

In our estimation, instead, social endogenous effects - i.e., the tendency to be affected by the behavior of others in the same school - are still present even after accounting for school and individual effects. We estimate that a 1% variation in average group BMI produces a 0.44% variation in current BMI status. Such finding is in line with the evidence reported in Christakis and Fowler (2007) and Trogdon et al. (2008). The former study analyzes data on a social network of people pertaining to the Framingham Offspring Study, finding that a person's chance of becoming obese increases by 57% if a friend became obese in a given interval. The latter makes use of Add Health data to estimate a coefficient of friends' average BMI of about 0.50. However, on the one hand Christakis and Fowler (2007) make use of a statistical method which ignores school and individual effects, and are criticized by Cohen-Cole and Fletcher (2008) as their results may not merely represent endogenous social interactions. On the other hand, Trogdon et al. (2008) base their identification of peer effects on an instrumental variable estimation in a cross-sectional analysis, where the instruments are obesity of (two self-nominated) friends' parents, and friends' birth weight; in their study a friend-selection effect might be driving results, rendering the appropriateness of the instrumental variables questionable.

Our study differs from the ones just mentioned in many aspects, which are considered crucial for results to be reliable. First, we believe that schoolmates represent the best approximation of a potential reference group, as these are individuals who adolescents compare and interact with in their everyday life, especially during meals. Second, looking at the OLS estimations proposed by Christakis and Fowler (2007) and Cohen-Cole and Fletcher (2008), the presence of a lagged dependent (or independent) variable in a social interactions model can lead to substantial biases in the estimation, unless properly addressed.[5] In this respect, misspecification of the model or of the error structure can lead to very large biases and thus incorrect inference. Our proposed GMM approach conciliates the different positions. Indeed, the econometric framework and estimation strategy used aim at overcoming the limits of the previous works, while relying on plausible hypotheses about habituation and reference groups. Third, we are able to trace out eating behavioral patterns from adolescence to adulthood. We show that obese teenagers become obese adults picking up wrong habits which are enforced by imitative behavior; the coefficient of autocorrelation for this category is 0.97 and the one related to group BMI is greater than one. The story seems different for adults who were normal-weight and overweight during adolescence: their adult outcome is not highly correlated with the past; rather, the role of peers at school has a crucial importance for

[5] Liu et al. (2006) find evidence of significant bias in estimation relating to the dynamic role of social interactions by making use of simulation techniques.

their current BMI.

Finally, we deal with missingness in the dependent variable (BMI) and in the most important economic variable, income, by replacing missing values using a multiple-imputation method. This has a dramatic effect on the coefficient of average BMI when we consider estimation results for different weight categories. Specifically, our results show that for individuals who were overweight and obese adolescents there is a marked increase in the group effect coefficient. Such result is supportive of the hypothesis that certain categories of individuals are less likely to report their weight so that average group effects will be downsized if missingness in the dependent variable is not properly accounted in the estimation.

The paper is structured as follows. Section two illustrates the main identification issue arising in linear-in-means models of social effects and shows how to resolve the identification problem in panel data using the lagged endogenous variable as an internal instrument. Section three describes the system GMM estimation strategy employed in the paper. Section four describes the Add Health dataset and present the main results. Section five concludes.

# 2 Linear-in-Means Models of Social Interactions

## 2.1 The Linear-in-Means Model and the Reflection Problem

The baseline LMM is conceptually very simple. Usually not derived from any predefined individual decision problem, this model allows individual behavior to linearly depend on some individual-specific characteristics as well as on group-specific factors, which include some group observable characteristics and the expected aggregate behavior of the others in the group. This makes it easily interpretable as a regression model, and therefore interesting to the econometrician. However, as pointed out by Manski (1993), the LMM suffers from a special kind of identification problem - the so called reflection problem - due to difficulties in disentangling two different group-effects, namely contextual and endogenous effects. Therefore, in such a framework measuring the impact of social interactions is typically challenging.

Consider the simple version of the model, where estimation concerns are not yet addressed. Assume to have $G$ non-overlapping, *a priori* determined groups, each of them made of $N^g$ individuals. Individual choice is assumed to be the result of the following process:

$$y_{ig} = a + y^e_{ig}\beta + \mathbf{x}'_g\boldsymbol{\gamma} + \mathbf{r}'_{ig}\boldsymbol{\delta} + \varepsilon_{ig} \ , \text{ where } \begin{array}{c} g = 1, \ldots, G \\ i = 1, \ldots, N^g \end{array} . \tag{1}$$

The individual-specific terms are defined by a $r \times 1$ vector of observable characteristics, $\mathbf{r}_{ig}$, and $\varepsilon_{ig}$, a random and unobservable scalar assumed to be independent and identically distributed across individuals. As to group-specific factors, these are divided into a $k \times 1$ vector of predetermined characteristics, $\mathbf{x}_g$, and the expected average choice in the group, $y^e_{ig}$. These two terms are conceptually different, the former being interpreted as contextual effects and the latter as an endogenous effect, and those exist under the condition that $\beta$ is non-zero and $\boldsymbol{\gamma}$ has at least a non-zero element. The key effect is exerted by $y^e_{ig}$, since it creates reciprocal reactions between individual decisions.

Using expected average behavior rather than the realized one is merely due to analytical convenience. This is a reasonable assumption when the behaviors of the rest of group are not directly observable - i.e., in large groups. When it comes to empirical analysis, such an assumption presupposes a restriction on the way individuals form expectations about the average choice in their group. Specifically, expectations are supposed to be consistent with the structure of the choices in the model, or self-consistent. This means that the perceived average choice is equivalent to the mathematical conditional expectation of the average choice, $y^e_g$, given the information set of each individual. The information set includes values of $r_{ig}$ for other individuals within $i$'s group, as well as the equilibrium expected choice level that occurs for her group. Individuals are assumed to be unable to observe the choices of others, $y_{-ig}$, or their random payoff terms $\varepsilon_{ig}$. Alternative information assumptions will not affect the qualitative properties of the model. For the LMM, self-consistency amounts to:

$$y^e_{ig} = y^e_g = \frac{a + \mathbf{x}'_g\boldsymbol{\gamma} + \mathbf{r}'_g\boldsymbol{\delta}}{1-\beta} = \frac{a + \mathbf{x}'_g\boldsymbol{\gamma}}{1-\beta} + \frac{\mathbf{r}'_g\boldsymbol{\delta}}{1-\beta}, \tag{2}$$

where $\mathbf{r}_g$ is the average of $\mathbf{r}_{ig}$ within group $g$.

Notice that such an assumption on the aggregate outcome implies a unique equilibrium: there exists only one expected average choice level that is consistent with the model, given individual and group characteristics. Therefore, equation (2) maps these characteristics into a single $y^e_g$.

An identification problem in this framework could arise because endogenous and contextual effects may co-move. Indeed, under the self-consistency assumption, the contextual variables determine the endogenous variable, as indicated by condition (2).

Given that the identification failure is a consequence of the correlation, by construction, between the endogenous and the contextual effects, Manski (1993) renamed it 'reflection problem', which is not too dissimilar from the basic identification problem in linear regressions with linearly dependent covariates. Manski's original argument is that every contextual effect might be defined as the average of a corresponding individual characteristic. For example, if one controls for student's maternal education one also introduces average (school) maternal education so that $\mathbf{x}_g = \mathbf{r}_g$. Condition (2) becomes

$$y_g^e = \frac{a + \mathbf{x}_g' \left(\boldsymbol{\gamma} + \boldsymbol{\delta}\right)}{1-\beta}, \tag{3}$$

meaning that the regressor $y_{ig}^e = y_g^e$ in (1) is linearly dependent on the regressors $a$ and $\mathbf{x}_g$ in (1), so the parameters are not identified. Substituting (3) into (1):

$$y_{ig} = \frac{a}{1-\beta} + \frac{\beta}{1-\beta}\mathbf{x}_g' \left(\boldsymbol{\gamma} + \boldsymbol{\delta}\right) + \mathbf{r}_{ig}'\boldsymbol{\delta} + \varepsilon_{ig}. \tag{4}$$

We can therefore state the following two remarks on the identification of social interaction effects in a LMM:

**Remark 1** *In the structural model (1) the set of regressors* $(1, y_g^e, \mathbf{x}_g, \mathbf{r}_{ig})$ *requires the estimation of* $2 + k + r$ *parameters.*

**Remark 2** *Assuming reflection* $\mathbf{r}_g = \mathbf{x}_g$ *in the reduced form (4) the set regressors* $(1, \mathbf{x}_g, \mathbf{r}_{ig})$ *allows us to identify* $1 + k + r$ *parameters. Hence, the endogenous effect parameter,* $\beta$*, remains unidentified.*

It is then clear why in the LMM framework identification of parameters is a major challenge. In the remainder of this section we show how to achieve identification of the endogenous effect parameter, $\beta$.

## 2.2 An AR(1) Linear-in-Means Model: Breaking the Reflection Problem

We discuss a dynamic LMM of social interactions, and show how the reflection problem can be broken. Consider a case in which the econometrician has access to a grouped panel, with $G$ non-overlapping groups $(g = 1, \ldots, G)$ of individuals and $N^g$ individuals

$(i = 1, \ldots, N^g)$ sampled in the $g^{th}$ group. The following autoregressive model generates the observed data:

$$y_{t,ig} = a + y_{t-1,ig}\varphi + y^e_{t,ig}\beta + \mathbf{x}'_{t,g}\boldsymbol{\gamma} + \mathbf{r}'_{t,ig}\boldsymbol{\delta} + \varepsilon_{t,ig} \tag{5}$$

In practice, the set of individual-specific attributes supposed to be determining individual behavior at time $t$ is assumed to depend on past period choice, $y_{t-1,ig}$. Such an assumption makes sense when not only is choice thought of as being the result of contemporaneous exogenous characteristics, but also of a certain past behavior that could play a role in actual choice. Extending the example on peer effects and students' obesity, we use student's body mass index in the previous period, $y_{t-1,ig}$, as an internal instrument to resolve the reflection problem since it will be orthogonal to the error term. The use of internal instruments to solve endogeneity problems is advocated for example by Lewbel (1997). Lewbel's idea is that when the endogenous regressor has a skewed distribution certain transformations of the data, including using lagged endogenous effects, provide a set of valid instruments.

The self-consistency condition in this case is:

$$y^e_{t,ig} = y^e_{t,g} = \frac{a + y_{t-1,g}\varphi + \mathbf{x}'_{t,g}\boldsymbol{\gamma} + \mathbf{r}'_{t,g}\boldsymbol{\delta}}{1-\beta} = \frac{a + \mathbf{x}'_{t,g}\boldsymbol{\gamma}}{1-\beta} + \frac{y_{t-1,g}\varphi + \mathbf{r}'_{t,g}\boldsymbol{\delta}}{1-\beta}. \tag{6}$$

The term $y_{t-1,g}$ is the average choice in the group in $t-1$, which enlarges the individual information set among the observable effects. Therefore, even under the assumption $\mathbf{x}_{t,g} = \mathbf{r}_{t,g}$, there is an additional element, $y_{t-1,g}$, which allows identification. Indeed, the social equilibrium equation is:

$$y^e_{t,g} = \frac{a + \mathbf{x}'_{t,g}(\boldsymbol{\gamma} + \boldsymbol{\delta})}{1-\beta} + \frac{y_{t-1,g}\varphi}{1-\beta}. \tag{7}$$

Substituting the social equilibrium into (5) yields:

$$y_{t,ig} = \frac{a}{1-\beta} + y_{t-1,ig}\varphi + \frac{\beta}{1-\beta}\mathbf{x}'_{t,g}(\boldsymbol{\gamma} + \boldsymbol{\delta}) + \frac{\beta\varphi}{1-\beta}y_{t-1,g} + \mathbf{r}'_{t,ig}\boldsymbol{\delta} + \varepsilon_{t,ig}. \tag{8}$$

Clearly, the model is now identified.

**Proposition 1** *In the structural model (5) the set of regressors* $(1, y_{t-1,ig}, y_{t,g}, \mathbf{r}_{t,ig}, \mathbf{x}_{t,g})$ *requires the estimation of* $(3 + r + k)$ *parameters.*

**Proposition 2** *Assuming reflection* $\mathbf{r}_g = \mathbf{x}_g$ *in the reduced form (8) the set regressors* $(1, y_{t-1,ig}, y_{t-1,g}, \mathbf{r}_{t,ig}, \mathbf{x}_{t,g})$ *allows to identify* $(3 + r + k)$ *parameters. Hence, all the parameters in the structural equation (5) are identified and the ratio of the two coefficients* $\frac{\beta\varphi}{1-\beta}$ *and* $\varphi$ *gives the endogenous effect* $\beta$*.*

The model avoids the linear dependence between $y_{t,g}$, $\mathbf{x}_{t,g}$ and $\mathbf{r}_{t,g}$ since we have the average action of the group in the previous period, $y_{t-1,g}$, as an additional regressor. This implies that $y_{t,g}$ depends on the entire history of $\mathbf{x}_{t,g}$ and $\mathbf{r}_{t,g}$ resolving the contemporaneous correlation with the same variables. Once the correlation is resolved, we can get an efficient and consistent estimation of all the parameters. Specifically, in the following section we illustrate how to estimate the social interaction parameters in the structural equation (5).

## 3 Estimation

We consider the following econometric framework:

$$\begin{aligned} y_{t,ig} &= y_{t-1,ig}\varphi + y^e_{t,ig}\beta + \mathbf{x}'_{t,g}\boldsymbol{\gamma} + \mathbf{r}'_{t,ig}\boldsymbol{\delta} + e_{t,ig}, \ |\varphi| < 1 \qquad (9) \\ e_{t,ig} &= \alpha_g + u_{t,ig}, \\ u_{t,ig} &= f_i + \varepsilon_{t,ig} \end{aligned}$$

where we allow for individual-specific effects, captured by $f_i$ as well as for group-specific effects, $\alpha_g$; $\varepsilon_{t,ig}$ is an individual-specific random disturbance. Notice that the system (9) allows us to decuple $a = \alpha_g + f_i$ in equation (5). Appendix A demonstrates that system (9) accounts for correlated effects both at the individual and group level so that $\alpha_g$ and $f_i$ can be treated as random.[6]

In order to account for the presence of endogeneity, we assume that:

$$E\left[\mathbf{h}_{t,ig}\ \varepsilon_{s,ig}\right] \neq 0, \quad \mathbf{h}_{t,ig} = \left[\mathbf{x}_{t,g}, \mathbf{r}_{t,ig}\right] \qquad (10)$$

$i = 1, ..., N^g$, $g = 1, ..., G$, and $s \leq t$. This assumption allows both for contemporaneous correlation between current disturbances and covariates and feedbacks from past shocks into the current value of the covariates. Moreover, the following assumptions hold:

[6] Most of the current research on social interaction effects (see De Giorgi et al. 2010 among others) is also accounting for potential correlated (unobservable) effects at the group level, without estimating these effects.

$$\begin{aligned} E\left[\varepsilon_{t,ig}\right|\mathbf{X}_t] &= 0 \\ Var\left[\varepsilon_{t,ig}\right|\mathbf{X}_t] &= \sigma_\varepsilon^2 \end{aligned}$$

where $\mathbf{X}_{t,ig} = [y_{t-1,ig}, y_{t-1,g}, \mathbf{x}_{t,g}, \mathbf{r}_{t,ig}]$.

We assume that $f_i$ and $\varepsilon_{t,ig}$ are independently distributed across individuals and have a familiar error structure in which:

$$E\left[f_i\right] = 0,\ E\left[\varepsilon_{t,ig}\right] = 0,\ E\left[f_i\ \varepsilon_{t,ig}\right] = 0 \quad \text{for } t = 2, ..., T, \quad i = 1, ..., N^g, \quad g = 1, ..., G$$

and

$$E\left[\varepsilon_{t,ig}\ \varepsilon_{s,ig}\right] = 0, \quad \forall t \neq s. \tag{11}$$

In addition, we impose the initial condition

$$E\left[y_{1,ig}\ \varepsilon_{t,ig}\right] = 0 \text{ for } t = 2, ..., T,\ i = 1, ..., N^g,\ g = 1, ..., G \tag{12}$$

Conditions (11) and (12) imply the following moment $m = 0.5\,(T-1)\,(T-2)$ conditions:

$$E\left[y_{t-s,ig}\ \varepsilon_{t,ig}\right] = 0 \text{ for } t = 3, ..., T,\ s \geq 3$$

First difference GMM can poorly behave when time series are highly persistent, as lagged levels of the series provide only weak instruments for subsequent first differences. In addition, first differencing would lead to loose substantial information from contextual effects, which are somewhat time-invariant. Therefore, we resort to a more efficient GMM estimator that exploits stationarity restrictions. Bond et al. (2001a) show that this system GMM estimator provides more reasonable estimates than first-differenced GMM.[7] Blundell and Bond (1998) consider the additional assumption that

$$E\left[f_i\ \Delta y_{2,ig}\right] = 0, \text{ for } i = 1, ..., N^g \text{ and } g = 1, ..., G \tag{13}$$

[7] We have four waves and 4443 respondents therefore we use the Arellano-Bond estimator which was designed for small T large N panels. The second lag is required, because it is not correlated with the current error term, while the first lag is. This is also shown by the Arellano-Bond test for autocorrelation which has a null hypothesis of no autocorrelation and it is applied to the differenced residuals. The test for AR (1) process in first differences usually rejects the null hypothesis (as in our results reported in Table 6).

This further assumption implies additional $T-2$ linear moment conditions:

$$E\left[u_{t,ig}\ \Delta y_{t-1,ig}\right]=0,\ \text{for}\ t=3,...,T,\ i=1,...,N^{g},\ \ g=1,...,G \tag{14}$$

These allow us to use lagged first-differences of the series as instruments for the equation in levels, as suggested by Arellano and Bover (1995).

Finally, given the assumption of endogenous regressors (10) the following moment conditions are also available:

$$E\left[\varepsilon_{t,ig}\ \Delta\mathbf{h}_{t,ig}\right]=0,\ \text{for}\ t=3,...,T,\ \ i=1,...,N^{g},\ \ g=1,...,G. \tag{15}$$

# 4 Data and Results

## 4.1 The National Longitudinal Study of Adolescent Health

The National Longitudinal Study of Adolescent Health (Add Health) is a (US) nationally representative, school-based survey of youth. The study was designed to determine how peers (within family, schools, neighborhoods, and communities) as well as individual characteristics influence health behaviors and therefore health outcomes.

While initially focused on adolescents only, in later phases the study analyzes health and health behaviors during the transition from adolescence into early adulthood. Indeed, in the first years of adulthood the young develop habits, and choose their lifestyle so that future health and well-being are strongly affected by such behaviors. It is therefore possible to study what happens during the transition to adulthood, as well as to explore early behavioral causes of adult chronic diseases.

The survey is made of four waves. In $1994-1995$ a random sample of $7th$ to $12th$ grade students from schools across the country was selected. About $90,000$ young individuals participated by filling out a brief questionnaire at school. Afterwards, at-home interviews with students and their parents were conducted. Students were interviewed again in their homes one year later (1996). School administrators provided information about the schools participants attended and existing data were compiled to describe neighborhoods and communities (in both waves $1994-1995$ and 1996). In the last two waves ($2001-2002$; $2007-2008$) participants in the first in-home interview were re-interviewed at ages 18 to 26, and again at ages 24 to 32.

The survey contains information on demographics, family life and background, school and academic outcomes, and health behaviors (drug use, smoking, pregnancy, etc.). For this research, the desired sample is the one relative to the in-home survey of the public-use data sets.

The reader is cross-referred to Appendices B and C for further details on design, weighting and missing information of data at hand.

### 4.2 The Dependent Variable

Our dependent variable is BMI, constructed using self-reported height and weight.[8] BMI is an index of weight-for-height which is age-independent and the same for both sexes. It is computed as weight in kilograms divided by the square of height in metres (kg/m$^2$) and it is standardly used to classify underweight, overweight and obesity.[9]

Table 1 shows the international classification of underweight, overweight and obesity according to BMI, as reported on the WHO studies.[10]

Table 1 about here

Based on this classification, we constructed a transition matrix for BMI in order to analyze the dynamic behavior of the variable in our sample.

Table 2 about here

First of all, we notice that more than 50 percent of individuals has a normal body weight, while the probability of facing an overweight individual is about 25% and an obese one about 17%.[11] Probabilities located on the main diagonal are quite high, meaning that BMI is highly autocorrelated, especially for normal-weight and heavily obese people. Such a finding strongly corroborates the validity of our empirical specification which includes lagged BMI among the set of regressors, given our hypothesis that habituation effects as well as imitation effects explain current BMI.

### 4.3 Descriptive Statistics

Tables 3 to 5 show the summary statistics of the sample under analysis.

[8] We make use of self-reported height and weight because Add Health wave 1 lacks information on measured height and weight. However, it has been shown that BMI computed using self-reported variables is highly correlated with BMI generated using measured height and weight ($r = 0.92$), and correctly classifies 96% as to obesity status (Goodman et al., 2000).

[9] BMI is not a direct measure of body fatness. However, it parallels changes obtained by direct measures of body fat such as underwater weighing and dual energy x-ray absorptiometry (DXA), therefore it can be considered as a proxy for measures of body fat.

[10] http://apps.who.int/bmi/index.jsp?introPage=intro.html

[11] Obtained by summing the percentages of all the obese categories, i.e. 11.03%, 4.28% and 3.31%.

Tables 3 – 5 about here

Table 3 displays the descriptive statistics relative to our dependent variable of interest, BMI, and control variables for the whole sample covering all waves. Figure 1 shows that average BMI is close to the threshold between normal weight and overweight - as predicted by the transition matrix. Having a look at the distribution of BMI (Figure 1) we realize that the modal bins are BMI=20−22 and BMI=22−24, meaning that normal-weight individuals are those for whom frequency is highest. Furthermore, the distribution appears to be right-skewed, signaling a majority of overweight and obese individuals in the sample observed. Household income levels (Figure 2) are in line with those reported by the US Census Bureau. Discrepancies between our data and the US Census Bureau data probably lie in the very high percentage of missing values, a problem addressed in the estimation.[12] Furthermore, in wave four self-reported income information comes in range format. Thus, in order to achieve a longer panel specification for household income, we take the average of each income brackets as the point information related to each individual. This might explain why we observe observations clustered around some values, as clearly visible in Figure 2.

Concerning the sample composition of some characteristics of interest, we observe that the proportion of females is slightly larger than males, registering 57% of counts. Also, the two wider ethnic groups are white and African American, while American Indian and Asian groups have a very small impact on the ethnical composition. The vast majority of individuals is in a good to excellent health status, while only 28% of the sample lives in a completely urban city, and 29% lives in a geographical area with low unemployment rate (though this variable shows a high percentage of missing values). Finally, the figure on parental education shows a low percentage of college graduate, both on the mother and on the father side.

All variables but income have a percentage of missing values of about 20%, which we consider acceptable and equally distributed across characteristics. We decide to deal with income missingness, instead, as it seems to be quite significant, and with BMI missingness, as it is our dependent variable. Specifically, given that peer-effects are derived as averages of individual BMI by dealing with missingness in this variable we explicitly take into account that certain categories may be less likely than others to report their weight, which can bias the definition of average (group) BMI. Details about

[12] Cf. Appendix C.

the procedure are described in Appendix C.[13]

Tables 4 and 5 are informative about variations in average BMI depending on certain characteristics, for both the entire sample and the adolescent subsample (first two waves only) respectively. The only difference for adolescents is that average BMI is in general lower. In both tables the most important figures are the correlation of poorer health statuses with higher average BMI, and the correspondence of lower BMI statuses to higher parental education. Also, those with a household income greater than the median show a higher BMI on average. Therefore, what comes out is that parental education, income (possibly correlated with parental education), and health are important factors for individual eating behavior.

### 4.4 Reference Groups

A crucial issue in the analysis of social endogenous effects in eating behavior is the definition of reference groups. Many papers attempting to address the complexities of social interactions in obesity rely on the nomination of adolescents' closest peers or on family history, and this is always subject to selection problems that the authors do not seem to address (e.g., the most important ones, Christakis and Fowler, 2007; Cohen-Cole and Fletcher, 2008; Fowler and Christiakis, 2008; Trogdon et al., 2008).[14] Besides, it could be restrictive to consider self-nominated friends or family as the only plausible reference group, especially for phenomena like overweight or obesity which may depend on social norms and acceptance in a broader context. Rather, we believe that schoolmates better fit the potential reference group adolescents compare and interact with. Indeed, interconnections between members of the same school may determine mutual influence through a variety of factors, e.g., food quality and quantity, time spent to exercise, appearance, etc. It is likely that contextual effects (those exerted by environmental factors) on eating behavior are common to schoolmates, and may drive similarities in individual behavior - therefore in their body weight.

Hence, our peer groups correspond to all the individuals belonging to the same school, meaning that endogenous effects measure the propensity to become overweight due to a direct interaction within the school. Such a choice is consistent with our dynamic analysis of eating behavioral patterns from adolescence to adulthood, mostly because it is believed that what affects the transition of body weight into early adulthood is

[13] Missingness here is not due to design effects, as data have been previously weighted and therefore adjusted to account for those.

[14] Cohen-Cole and Fletcher (2008) and Christakis and Fowler (2007) consider the data on obesity status for an individual (in their terminology, an "Ego") at a given point in time and estimate its relationship to the obesity status of a friend ("Alter").

behavior during adolescence (e.g., *inter alia*, Kemper et al., 1999; Sun Guo et al., 2002; Kvaavik et al., 2003; Gordon-Larsen et al., 2004) which in turn depends on schoolmates behavior (cf. Section 4.6).

### 4.5 Estimation Results

In this section we report results produced by estimating the system defined in (9).[15]

A premise is due at this stage. We make clear to the reader that the lagged dependent variable on the right hand side does not refer to the value of the dependent variable the year before, as the gaps between waves are not homogeneous; rather, that embeds all past history up to the previous wave. In particular, wave one and two are consecutive years registering information on adolescents, wave three is 6 years later than wave two and includes data on early adults, wave four is again 6 years later than wave three and contains information on adults. Data have been purposely weighted to account for uneven time gaps, and a dummy variable for being adolescent (observations registered in waves one and two) has been included to capture variation due to being part of the adolescent cohort versus belonging to the adult cohort.

Our specification allows us to investigate the hypothesis that obesity can spread through peers versus the claim that obesity is essentially an individual outcome linked to personal and family history. We also establish whether peer effects may be stronger for obese pupils compared to the non-obese counterparts.

As pointed by Cohen-Cole and Fletcher (2008), in order to avoid spurious conclusions on the role exerted by group behavior the estimation should include contextual effects.[16] In other terms both individual and group behavior can be affected by exposure to common influences: for example, the opening of a fast food, gym or recreational area near a school could simultaneously affect the weight of all pupils in the same school. Since access to such facilities may be linked to the socio-demographic characteristics of the adolescents in the same school, we include the average school values for household

[15] The Arellano–Bond estimators is available for Stata 9.0 as proprietor program written by Roodman (2006) (called `xtabond2`). See http://ideas.repec.org/c/boc/bocode/s435901.html.

[16] The paper by Cohen-Cole and Fletcher (2008) argues that previous studies on the spread of obesity (Christakis and Fowler, 2007) do not include a sufficiently broad set of contextual effects to account for a range of hypothesized causes of obesity, therefore overstating the endogenous effect. Corrado and Fingleton (2012) also suggest that the significance of a spatially lagged dependent variable involving network dependence and spatial externalities may be misleading, since it may be simply picking up the effects of omitted spatially dependent variables, incorrectly suggesting the existence of a spillover mechanism.

income, age, gender, ethnicity and parental education in the estimation.[17,18] Therefore, as in Cohen-Cole and Fletcher (2008), we consider a time-dependent set of school specific covariates, $\mathbf{x}_{t,g}$. These represent a much richer set of controls to absorb the average change in social context experienced by all individuals in the sample. They can also be interpreted as school-specific trends which account for environmental factors shared by adolescents in the same school. Clearly, more environmental confounders may exist which are positively correlated with an individual's BMI. We therefore enrich our instrumental variables set by adding two location-specific variables indicating whether the neighborhood where the individual resides is characterized by a low unemployment rate, and whether the adolescent lives in a completely urban area.[19,20] These environmental confounders reflect the social context of the geographical area where the respondents reside and represent a valid set of instruments since they are likely to be correlated with both individual and group BMI but neither with the unobserved individual propensity or tolerance to become overweight, nor with unobservable effects at the school level.

We employ a system GMM estimation which uses the levels equation (5) to obtain a system of two equations: one differenced and one in levels. Additional instruments can be obtained by adding the second equation so that variables in levels can be instrumented with their own lags. This usually increases efficiency. The set of endogenous variables $[y_{t-1,ig}, y_{t,g}]$ includes lag individual BMI, $y_{t-1,ig}$, and contemporaneous group BMI, $y_{t,g}$; these are instrumented with GMM style instruments, i.e., third and fourth lags of the endogenous variables $[y_{t-3,ig}, y_{t-4,ig}, y_{t-3,g}, y_{t-4,g}]$. The exogenous variables chosen as set of standard instruments $[\mathbf{r}_{t,ig}, \mathbf{x}_{t,g}, \mathbf{z}_t]$ include the exogenous controls, $\mathbf{r}_{t,ig}$, their school average, $\mathbf{x}_{t,g} = \mathbf{r}_{t,g}$, and two additional instruments, $\mathbf{z}_t$, characterizing the macro-area where each adolescent lives (urban and employment rate).

In order to compare the results at hand with previous findings by Cohen-Cole and Fletcher (2008) we also address the issue of missing data. In the dataset we register $3,372$ missing observations for income and $1,654$ missing observation for BMI. We use a

[17]In system GMM, one can include time-invariant regressors, which would disappear in difference GMM. Asymptotically, this does not affect the coefficients estimates for other regressors. This is because all instruments for the levels equation are assumed to be orthogonal to the fixed effects, thus to all time-invariant variables; in expectation, removing them from the error term does not affect the moments that are the basis for identification.

[18]We also performed estimation with average effects at the school level and centered effects at the individual level in order to account for potential collinearity among regressor and the results were similar. These additional results are available on request.

[19]The definition of neighborhood follows a geographical criterion as such community variables are based on state, county, tract, and block group levels derived from addresses.

[20]Our estimation also accounts of wave effects, through the inclusion of time and adolescent dummies.

Multiple Imputation method to estimate these missing values as described in Appendix C, because we expect missingness at random to be explained by covariates included in our model (e.g., ethnicity or gender).

The third column in Table 6 reports the estimates for the system defined in (9) where missingness in income and BMI are accounted for.[21] Results show that current BMI is affected both by past individual decisions and social behavior. In fact, an increase by 1% in past BMI leads to an increase in current BMI by 0.83%. This result is very much in line with the evidence from the transition probabilities in Table 2 where BMI is highly autocorrelated, especially for heavily obese people. Looking at peer effects, we can see that an increase by 1% in the average BMI leads to an increase in current BMI by 0.44%. Both the Sargan test and the Hansen test indicate that the instruments chosen as a group are exogenous. Looking at the significance of other controls, we find that adolescent of Asian ethnicity tend to experience a lower BMI than their White and Black counterparts. Other studies also show that the prevalence of overweight and obesity among Asian Americans is much lower than the national average and all other main racial/ethnic groups (Gordon-Larsen et al., 2003; Popkin and Udry, 1998). In addition, adolescents belonging to Black ethnic groups have a higher BMI. This result is also in line with other evidence using Add Health data showing that lower socioeconomic status and minority population groups have less access to physical activity facilities, which in turn is associated with decreased physical activity and increased overweight (Gordon-Larsen et al., 2006). We also find that obesity is less widespread among adolescents whose father gained a college education. There are different channels through which parental education can affect their children's health. Education might have a direct impact on child health because it helps parents to make better health investments for themselves and their children. Alternatively, education can affect child health indirectly. An increased level of education can give access to more skilled work with higher earnings and these resources could be used to invest in health (Case et al., 2002; Lindeboom et al., 2009). In the presence of assortative mating, individuals with a higher level of education also marry partners with higher levels of education, which positively affects family income. In this respect, public health strategies aimed at preventing obesity may need to target families of low socioeconomic status early in children's lives, in order to counteract the adverse effect of poor socioeconomic status on parental health and eating decisions.[22]

[21] We use a robust standard errors estimation where the standard covariance matrix is robust to panel-specific autocorrelation and heteroskedasticity. We also bootstrap the standard error and find no difference with the robust standard errors. The results are available on request.

[22] We also estimate a model omitting the health dummies and their group averages among the set

If we consider imputed data for income only the qualitative results do not change insofar past individual behavior still dominates. In this case an increase by 1% in past individual BMI leads to an increase by 0.67% in current individual BMI whereas an increase in average BMI leads to an increase in current individual BMI by 0.46%. It is worth stressing that when we impute data for BMI alongside income the coefficient for Black ethnicity is now significant. This seems to support the evidence that data are not missing completely at random, and that weight self-reported information might be dependent on individual ethnicity with Black being less likely to report their weight than individuals of other ethnicities. The results also show that dealing with missingness of both income and BMI increases by 76% the coefficient of average BMI (from 0.25 to 0.44). In Table 6 (Model 1) we might erroneously understate the effect exerted by peers if BMI missingness in not properly addressed.

### 4.6 The Role of Habituation and Imitation in Obesity Behavior

In this section we want to assess how habituation and imitative behavior influence the behavior of adults who were normal-weight, overweight and obese adolescents.[23]

We note from Table 2 that BMI is highly autocorrelated, especially for normal-weight and heavily obese people. In this instance, personal history and personality traits may dominate upon the influence of the reference group. We therefore estimate model (9) for each BMI category, paying attention to endogenous sample selection arising from selecting categories based on the dependent variable. Hence, we split the sample according to the BMI status in wave 1 and keep individuals in the same strata. This allows us to clearly understand how behaviors during adolescence contribute to adult outcomes. The results for normal, overweigth and obese adolescents are reported in Tables 7, 8 and 9 respectively.

We find that personal history (lagged BMI) does not matter for individuals who were normal and overweight adolescents; rather, they seem to be affected mainly by their reference group behavior. For those who were obese in adolescence, instead, habituation is certainly a fundamental driver of current BMI, though the effect exerted by social ties

of regressors (available on request). This exclusion has the effect of amplifying the impact of lagged individual BMI on individual BMI and to downsize the effect of average BMI. An increase by 1% in past BMI leads now to an increase in current BMI by 0.94%. Whereas an increase by 1% in average BMI leads to an increase by 0.19% in current BMI. We therefore opt to include the health dummy variable among the set of regressors since the significance of lagged individual BMI may be simply pick up an omitted variable problem. Note that the potential endogeneity of the health variables is controlled by the use of lagged endogenous instruments in the system GMM estimation.

[23] Specifically, we focus on the International Classification of Weight according to the WHO, as reported in Table 1.

is explosive. In practice, BMI status during adulthood is due to both past behavior and group behavior for individuals who experienced obesity when adolescent - therefore they are obese adults (Whitlock et al., 2005)- but peer effects outnumber habituation effects.

Table 7, Model 3 (benchmark model) shows that individual behavior is dominated by the influence from peers for the sample of normal-weight individuals when teenagers. In this instance, for any 1% increase in average BMI we expect about 0.37% increase in individual BMI, whereas the coefficient for past BMI is not statistically significant. Results show that normal-weight adolescents tend to develop a social behavioral pattern in eating, perhaps related to social inclusion (e.g., Falkner at al, 2001; Chen and Brown, 2005). Table 8 shows the same behavior for individuals who were overweight during their adolescence: a 1% increase in average BMI leads to an increase in individual BMI by 0.67% (Model 3). Interestingly, gender plays a role in explaining eating behavioral patterns of normal-weight adolescents, registering a negative relationship with individual BMI, while the coefficient for the gender dummy is insignificant for overweight adolescents. Ethnicity, instead, does not seem a decisive driver of differences in BMI status.

For obese adolescents the story is different. A pattern of self-weight-manteinance behavior is observed, possibly supported by patterns of wrong behavioral routines such as unhealthy eating habits and scarce exercise. However, the influence of peers at the school level is now stronger, even explosive. Table 9 (Model 3) reports that an increase by 1% in average BMI leads to an increase by 1.32% in current BMI. The habituation effect is lower - still very high in absolute terms - leading to a 0.97% increase in BMI for a 1% increase in aggregate BMI. This means that obese adolescents become future obese adults through wrong habits enforced by imitative behavior. As stressed by Christakis and Fowler (2007), having obese school contacts might change a person's tolerance for being obese or might influence his or her adoption of specific behaviors (e.g., smoking, eating, and exercising). The fact that adolescents' appearance and behaviors are influenced by the appearance and behaviors of those around them suggests that weight gain in one person might influence weight gain in others. In addition to such strictly social effects, it is plausible that physiological imitation might occur (Fogassi et al., 2005); areas of the brain that correspond to actions such as eating food may be stimulated if these actions are observed in others and this possibility is higher within restricted social environments, such as schools, where individuals spend most of their time. Moreover, the positive effect of the gender dummy on BMI of obese individuals could signal that efforts to prevent obesity should not ignore the central role of cognitive factors, as often obese young women lack motivation, and personality traits may dominate over external factors (Andajani-Sutjahjo et al., 2004).

Finally, it is also important to stress that dealing with missingness of both income and BMI has a dramatic effect on the coefficient of average BMI. In Table 7 (Model 1) we might erroneously think the habituation effect to be explosive and understate the effect exerted by peers if BMI missingness in not properly addressed. Table 8 shows that for overweight adolescents, when missing data for income and BMI (Model 3) are replaced by multiple imputation there is an increase by 33% in the coefficient of the group effect measured by average BMI (from 0.501 to 0.667). Table 9 (Model 1) shows that for obese individuals the peer effect would disappear when missingness is not taken into account, while it represents a key factor in delineating eating behavioral patterns for obese adolescents. In addition the results also show that the magnitude of lagged individual BMI in Models 1 and 2 of all tables may be misleading, since it may be simply picking up the effects of missing data. This evidence is generally not captured by a typical full sample estimation (Table 6) where we can observe that the coefficients of average BMI is rather stable across the three models.

## 5 Conclusions

Personal and family history, the impact of the social context where each individual lives as well as endogenous effects induced by interactions with peer groups are all possible determinants of eating behavior. One of the econometric challenges is to identify the separate impact of the endogenous and contextual effects, and to break the so called reflection problem (Manski, 1993). The dynamic linear-in-means model proposed in the paper allows us to estimate all social effects and to control for individual- and group-specific unobservable effects, by exploiting stationarity restrictions of a system GMM estimator. Our results show that individuals tend to become overweight mainly due to habituation and social effects, even properly accounting for contextual effects. In particular, imitative behavior seems to explain a relevant part of variation in body mass index of all individuals in the sample, though habituation plays the most important role. Individuals who were normal-weight and overweight during adolescence are not influenced by past behavior; rather, they are affected by average behavior in their reference group. Obese adolescents, instead, become future obese adults, showing a high persistence in their body mass index status; their wrong habits are enforced by imitative behavior, as peer effects impact dramatically on current weight status.

Such peer effects are of obvious policy significance, though rarely taken into account by policy makers or even by entities with a collective perspective. The implication is that group-level interventions may be more successful and more efficient than individual interventions, as a social multiplier effect takes place. This means that clinical and policy interventions may be more cost-effective than policy-makers have previously supposed. We are facing a health problem characterized by a imitative, therefore multiplicative, dimension. Public health policy makers should implement urgent and targeted actions for preventing this epidemic to spread further.

Despite the advantage of being able to identify peer groups at the level of schoolmates, data used in our study have some limitations. First, the longitudinal analysis is conducted in a panel dataset with non-homogeneous gaps relative to adolescents who become adults. Second, friends outside of school and romantic partners are not captured. Future research is needed to better understand the mechanisms behind the influence of peers on weight. Several candidates exist, such as peer influence on weight loss attempts, physical activity, and perceptions of own body weight. The causal mechanisms should also be consistent with effects of higher moment of the BMI distribution. Knowledge about the source of peer influence on weight and the size of any social multipliers will improve implementation and evaluation of policies aimed at reducing overweight or obesity in adolescence, hence in adulthood.

## References


[1] Anderson, T.W., and Hsiao, C. (1982): "Formulation and estimation of dynamic models using panel data," Journal of Econometrics, 18: 47–82.

[2] Andajani-Sutjahjo, S., Ball K., Warren, N., Inglis V. and Crawford D. (2004): "Perceived personal, social and environmental barriers to weight maintenance among young women: A community survey", International Journal of Behavioral Nutrition and Physical Activity: $1-15$.

[3] Arellano, M., and Bond, S. (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," Review of Economic Studies 58: 277–97.

[4] Arellano, M., and Bond, S. (1998): "Dynamic Panel data estimation using DPD98 for Gauss: A guide for users."

[5] Arellano, M., and Bover, O. (1995): "Another look at the instrumental variables estimation of errorcomponents models," Journal of Econometrics 68: 29–51.

[6] Baum, C., Schaffer, M., and Stillman, S. (2003): "Instrumental variables and GMM: Estimation and testing," Stata Journal 3(1): $1-31$.

[7] Beck, T., and Levine, R. (2004): "Stock markets, banks, and growth: Panel evidence," Journal of Banking and Finance 28(3): 423–42.

[8] Bénabou, R. (1993): "Workings of a city: Location, education, and production," Quarterly Journal of Economics, CVIII, 619–652.

[9] Blundell, R., and Bond, S. (1998): "Initial conditions and moment restrictions in dynamic panel data models," Journal of Econometrics 87: 11–143.

[10] Bond, S., Hoeffler, A., and Temple, J. (2001a): "GMM estimation of empirical growth models," Economics Papers $2001-$W21, Economics Group, Nuffield College, University of Oxford.

[11] Bond, S., Bowsher, C., and Windmeijer, F. (2001b): "Criterion-based inference for GMM in autoregressive panel data models," Economics Letters, Elsevier, 73(3): $379-388$, December.

[12] Bond, S. (2002): "Dynamic panel data models: A guide to micro data methods and practice," Working Paper 09/02. Institute for Fiscal Studies. London.

[13] Brock, W. A., and Durlauf, S. N. (2001): "Discrete choice with social interactions," Review of Economic Studies, 68(2), 235 – 260.

[14] Brown, B. (1990): "Peer groups and peer cultures," in At the Threshold, ed. by S. Feldman, and G. Elliott, Cambridge, MA. Harvard University Press.

[15] Brown, B., Clasen, D., and Eicher, S. (1986): "Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents," Developmental Psychology, 22(4): 521 – 530

[16] Bernheim, D. (1994): "A theory of conformity," Journal of Political Economy, 102: 841 – 877.

[17] Bramoullé, Y., Djebbari, H., and Fortin, B. (2009): "Identification of peer effects through social networks," Journal of Econometrics, 150(1): 41–55.

[18] Burke, M. A., and Heiland, F. (2007): "Social dynamics of obesity," Economic Inquiry, 45, 571–591.

[19] Case, A., M. Lubotsky and C. Paxson (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," American Economic Review, 92: 1308 – 1334.

[20] Chamberlain, G. (1982): Multivariate Regression Model for Panel Data, Journal of Econometrics, $18, 5 - 46$.

[21] Chamberlain, G. (1984): Panel Data in Handbook of Econometrics ed. by Z. Griliches and M.D. Intriligator, Amsterdam, North Holland Publishing Co, pp. 1247 – 1318.

[22] Chen, E.Y. and Brown, M. (2005): "Obesity stigma in sexual relationships," Obesity Research, 13(8): 1393 – 7.

[23] Christakis N. A., and Fowler, J. (2007): "The spread of obesity in a large social network over 32 years," The New England Journal of Medicine, 357: 370 – 379.

[24] Cohen-Cole, E. (2006): "Multiple groups identification in the linear-in-means model," Economic Letters, 92(2): 157–162.

[25] Cohen-Cole, E., and Fletcher, J. (2008): "Is obesity contagious? Social networks VS. environmental factors in the obesity epidemic," Journal of Health Economics, 27(5): 1382 – 1387.

[26] Corrado, L. and B. Fingleton (2012): "Where is the Economics in Spatial Econometrics?", Journal of Regional Science, Volume 51, Issue 2, pp. 1 – 30.

[27] Crawford, D. and Ball, K. (2002): Behavioural determinants of the obesity epidemic", Asia Pacific Journal of Clinical Nutrition, 11(Suppl 8): S718 – 721.

[28] De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010): "Identification of social interactions through partially overlapping peer groups," American Economic Journal: Applied Economics, 2(2): 241 – 75.

[29] Falkner N.H., Neumark-Sztainer, D., Story, M., Jeffery, R.W., Beuhring, T. and Resnick, M.D. (2001): "Social, educational, and psychological correlates of weight status in adolescents," Obesity Research, 9(1):32 – 42.

[30] Fogassi, L., Ferrari, P.F., Gesierich. B., Rozzi, S., Chersi, F., Rizzolatti, G. (2005) "Parietal lobe: from action organization to intention understanding," Science, 308: 662 – 7.

[31] Fortin, B. and Yazbeck M. (2011): "Peer Effects, Fast Food Consumption and Adolescent Weight Gain," Centre Interuniversitaire sur le Risque, les Politiques Èconomiques et l'Emploi, Cahier de recherche/Working Paper 11 – 03.

[32] Fowler, J., and Christakis, N. A. (2008): "Estimating peer effects on health in social networks; A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pail," Journal of Health Economics, 27, 1400 – 1405.

[33] Goodman, E., Hinden, B. R., and Khandelwal, S. (2000): "Accuracy of teen and parental reports of obesity and body mass index," Pediatrics, 106: 52 – 58.

[34] Gordon-Larsen, P., Adair, L.S., Nelson, M.C., and Popkin, B.M. (2004): "Five-year obesity incidence in the transition period between adolescence and adulthood: the National Longitudinal Study of Adolescent Health," American Journal of Clinical Nutrition, 80(3): 569 – 575.

[35] Gordon-Larsen, P., Nelson M.C., Page P., et al. (2006): "Inequality in the built environment underlies key health disparities in physical activity and obesity," Pediatrics, 117: 417 – 24.

[36] Harris, K. M. (2011): "Design Features of Add Health," Carolina Population Center. website: www.cpc.unc.edu/projects/addhealth/data/guides/.

[37] Kandori, M. (1992): "Social norms and community enforcement," Review of Economic Studies, 59: 63 – 80.

[38] Kemper, H.C.G., Post, G.B., Twisk, J.W.R., and van Mechelen, W. (1999): "Lifestyle and obesity in adolescence and young adulthood: Results from the Amsterdam Growth And Health Longitudinal Study (AGAHLS)," International Journal of Obesity, 23(3): s34–s40.

[39] Kvaavik, E., Tell, G.S., and Klepp, K. (2003): "Predictors and Tracking of Body Mass Index From Adolescence Into Adulthood. Follow-up of 18 to 20 Years in the Oslo Youth Study," Arch Pediatr Adolesc Med, 157:1212 – 1218.

[40] Lewbel, A. (1997): "Constructing Instruments for Regressions with Measurement Error when no Additional Data are Available with an Application to Patents and RD," Econometrica, 65(5): 1201 – 1214.

[41] Lindeboom M., Nozal A. L. and van der Klaauw, B. (2009): "Parental education and child health: Evidence from a schooling reform", Journal of Health Economics, 28(1): 109 – 131.

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

[43] Manski, C. F.(2000): "Economic analysis of social interactions," Journal of Economic Perspectives, 14(3): 115 – 136.

[44] Moffitt, R. A. (2001): "Policy interventions, low-level equilibria, and social interactions" in S. N. Durlauf & H. P. Young (Eds.), *Social dynamics*, pp. 45–82. Cambridge, MA: MIT Press.

[45] Gordon-Larsen P., Adair L.S., and Popkin, B.M. (2003): "The relationship of ethnicity, socioeconomic factors, and overweight in US adolescents." Obesity Research, 11: 121 – 9.

[46] Popkin, B.M. and Udry, J.R. (1998): "Adolescent obesity increases significantly in second and third generation U.S. immigrants: the National Longitudinal Study of Adolescent Health," Journal of Nutrition, 128: 701 – 6.

[47] Sun Guo, S., Wu, W., Chumlea, W.C., and Roche, A.F. (2002): "Predicting overweight and obesity in adulthood from body mass index values in childhood and adolescence," American Journal of Clinical Nutrition, 76(3):653 – 658.

[48] Trogdon, J., Nonnemaker, J., and Pail, J. (2008): "Peer effects in adolescent overweight," Journal of Health Economics, 27, 1388–1399.

[49] Tourangeau, R. and Shin, H. (1999): "National Longitudinal Study of Adolescent health: Grand Sample Weights," National Opinion Research Center and Carolina Population Center. website: http://www.cpc.unc.edu/projects/addhlth/.

[50] Whitlock E.P., Williams, S.B., Gold, R., Smith, P.R. and Shipman, S.A. (2005): "Screening and interventions for childhood overweight: a summary of evidence for the US Preventive Services Task Force," Pediatrics, 116(1).

# A GMM and Chamberlain's Correlated Effects Approach in Linear Panel Data Models

Consider system (9)

$$\begin{aligned} y_{t,ig} &= y_{t-1,ig}\varphi + y^{e}_{t,ig}\beta + \mathbf{x}'_{t,g}\boldsymbol{\gamma} + \mathbf{r}'_{t,ig}\boldsymbol{\delta} + e_{t,ig}, \ |\varphi| < 1 \\ e_{t,ig} &= \alpha_g + u_{t,ig}, \\ u_{t,ig} &= f_i + \varepsilon_{t,ig} \end{aligned} \tag{16}$$

By recursion we can write (for $t = 1, ..., T$) :

$$\begin{aligned} y_{t,ig} = {} & \left(1 + \varphi + ... + \varphi^{t-1}\right) f_i + \left(1 + \varphi + ... + \varphi^{t-1}\right) \alpha_g + \varphi^t y_{0,ig} + \\ & + [y^{e}_{t,ig}\beta + y^{e}_{t-1,ig}\beta\varphi + ... + y^{e}_{1,ig}\beta\varphi^{t-1}] \\ & + [\mathbf{x}'_{t,g}\boldsymbol{\gamma} + \mathbf{x}'_{t-1,g}\boldsymbol{\gamma}\varphi + ... + \mathbf{x}'_{1,g}\boldsymbol{\gamma}\varphi^{t-1}] + \\ & + [\mathbf{r}'_{t,ig}\boldsymbol{\delta} + \mathbf{r}'_{t-1,ig}\boldsymbol{\delta}\varphi + ... + \mathbf{r}'_{1,ig}\boldsymbol{\delta}\varphi^{t-1}] + \\ & + [\varepsilon_{t,ig} + \varphi\varepsilon_{t-1,ig} + ... + \varphi^{t-1}\varepsilon_{1,ig}] \end{aligned} \tag{17}$$

This transformation links system (9) to Chamberlain's method (1982, 1984) to deal with correlated effects in dynamic linear panel data models. In fact, we can write system (17) in compact form as:

$$E[y_{t,ig} \mid \mathbf{W}_i] = \mathbf{W}_i' \, \mathbf{\Pi} + \eta(f_i \ + \alpha_g)$$

where $\mathbf{W}_i = \left[y_{0,ig}, \ y^{e}_{t,ig}, ..., y^{e}_{1,ig}, \ \mathbf{x}_{t,g}, ..., \mathbf{x}_{1,g}, \ \mathbf{r}_{t,ig}, ..., \mathbf{r}_{1,ig}\right]$. The $\mathbf{\Pi}$ matrix is defined in terms of the coefficients of the linear predictors of the dependent variable at each period given all explanatory variables at all periods. For the individual effect, $f_i$, and the group effect, $\alpha_g$, we therefore have:

$$E[f_i, \alpha_g \mid \mathbf{W}_i] = 0$$

Given the equivalence with system (17) both $f_i$ and $\alpha_g$ can therefore be treated as random individual- and group-specific effects also in the original system (9).

## B Design and Weighting

The Add Health Study is a US representative, probability-based survey of adolescents in grades 7 through 12 conducted between 1994 and 1995, and extended to 2008 with three in-home interviews. The sample design used to collect the data embeds a certain degree of complexity which should be accounted for. Indeed, failing at considering such complexity may result in biased parameter estimates and incorrect variance estimates. Hence, we corrected for design effects and unequal probability of selection, according to what is suggested in the Add Health user guides.[24]

We exploit the longitudinal feature of the dataset, keeping the strength of its innovative design. With the longitudinal data from adolescence, the third and four in-home interviews allow "researchers to map early trajectories out of adolescence in health, achievement, social relationships, and economic status and to document how adolescent experiences and behaviors are related to decisions, behavior, and health outcomes in the transition to adulthood. The fundamental purpose of this [...] follow-up was to understand how what happens in adolescence is linked to what happens in the transition to adulthood when adolescents begin to negotiate the social world on their own and develop their expectations and goals for their future adult roles." (Harris, 2011). Data have been appropriately weighted to correct for time gaps in their longitudinal format. For details on the Add Health weighting scheme, the reader is cross-referred to Tourangeau and Shin (1999).

## C Missing data

There are several reasons why the data may be missing. We say that data are "missing completely at random" if the probability that an observation is missing is not related to the value of that observation or to the value of any other variable. In this case the design power is lower, but the estimated parameters are not biased. However, this data feature is not very common.

In other cases data may be classified as "missing at random". For data to be missing at random, missingness should not depend on the value of the missing observation after controlling for another variable. The type of missingness should be dealt with in order to produce relatively unbiased estimates. Both these types of missingness are said to be "ignorable", but the latter needs to be addressed in some way.

Finally, we could have "missing not at random" data, i.e. data for which missingness

[24] http://www.cpc.unc.edu/projects/addhealth/data/guides

depends on the value of the missing observation. Under such circumstances, the only way to obtain unbiased estimates is to write a model that takes missing data in due account. Clearly, this could be a rather difficult task as we rarely know what the missingness model is.

Concerning our case, we consider the variables of the dataset to display missing-at-random or missing-completely-at-random values. For example, on the one hand, income self-reported information might be dependent on individual ethnicity: black people could be less likely to report their income than white individuals. The black probably have lower incomes than the white, and it would at first appear that missingness on income is related to the value of income itself. But the data would still be missing at random if the conditional probability of missingness were unrelated to the value of income within each ethnic group. On the other hand, missing values on gender, for example, could be considered as being missing completely at random.

We decide to deal with missingness in two ways: by applying a Multiple Imputation method to two variables of interest, namely household income and household income/BMI (Models 2 and 3 of Estimation Tables).[25] In the first case, we are able to replace all the missing values, but such a replacement does not add new information, as the overall mean, with or without replacing missing data, will be the same. In addition, such a process leads to underestimating the error. Multiple Imputation, instead, involves estimating what the missing values would be, and then using those "imputed" values in the solution. Obviously, in this case we have selected the variables for which it could make sense to expect missingness at random to be explained by other variables included in our model. Income is the variable showing the highest number of missing values; moreover, for the reasoning just explained, we believe that such a missingness is not completely at random. Therefore, we perform a multiple imputation on income only first, and on income together with BMI later. Indeed, body weight is another variable showing a high level of missing cases. Therefore, we decide to impute BMI as it is our dependent variable, since peer effects are directly generated from it.

We make use of Multiple Imputation by Chained Equations (MICE)(Sterne et al., 2009). For a set of variables, $x_1; \ldots ; x_k$ some or all of which have missing values, the MICE algorithm initially fills all missing values at random. The first variable with at least one missing value, e.g., $x_1$, is then regressed on the other variables, $x_2; \ldots ; x_k$. The estimation is restricted to individuals with observed $x_1$. Missing values in $x_1$ are replaced

[25] Our criteria for choosing to deal with missingness are, first the (potentially reasonable) source of missingness -i.e., completely random or random-, second the percentage of missing cases – i.e., if greater than 20%-, and last the importance of such missing information for the model (e.g., our dependent variable). The percentage of missing values per variable are reported in Table 3.

by simulated draws from the posterior predictive distribution of $x_1$, an important step known as proper imputation. The next variable with missing values, say $x_2$, is regressed on all the other variables, $x_1; x_3; \ldots x_k$. Estimation is restricted to individuals with observed $x_2$ and uses the imputed values of $x_1$. Again, missing values in $x_2$ are replaced by draws from the posterior predictive distribution of $x_2$. The process is repeated for all other variables with missing values in turn (cycle), for about ten cycles.[26] The entire procedure is repeated independently $M$ times, yielding $M$ imputed datasets.[27]

[26]Because each variable is imputed using its own imputation model, MICE can handle different variable types (for example, continuous, binary, unordered categorical, ordered categorical).

[27]Standard texts on MI suggest that small numbers of imputed datasets ($M = 3$ to 5) are adequate.

Table 1: International Classification of Weight According to BMI

| Classification | Category | BMI (Principal Cut-off Points) |
|---|---|---|
| Underweight | 1 | <18.50 |
| Normal Range | 2 | 18.50-24.99 |
| Overweight (Pre-Obese) | 3 | 25.00-29.99 |
| Obese Class I | 4 | 30.00-34.99 |
| Obese Class II | 5 | 35.00-39.99 |
| Obese Class III | 6 | ≥40 |

Note. Source: Adapted from WHO, 1995, WHO, 2000 and WHO 2004.

Table 2: Transition matrix of BMI categories

| | | BMI in $t$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| | 1 | 45.09 | 52.32 | 2.18 | 0.24 | 0.12 | 0.06 | 100 |
| | 2 | 2.31 | 74.22 | 20.15 | 2.91 | 0.32 | 0.09 | 100 |
| BMI in $t-1$ | 3 | 0.31 | 13.55 | 55.74 | 24.96 | 4.64 | 1.07 | 100 |
| | 4 | 0.14 | 2.06 | 15.64 | 48.29 | 24.97 | 8.92 | 100 |
| | 5 | 0.00 | 0.81 | 5.81 | 19.92 | 37.80 | 35.57 | 100 |
| | 6 | 0.29 | 0.00 | 0.29 | 4.34 | 15.03 | 80.06 | 100 |
| | Total | 5.43 | 51.20 | 24.76 | 11.03 | 4.28 | 3.31 | 100 |

Note. Categories defined as in Table 1.

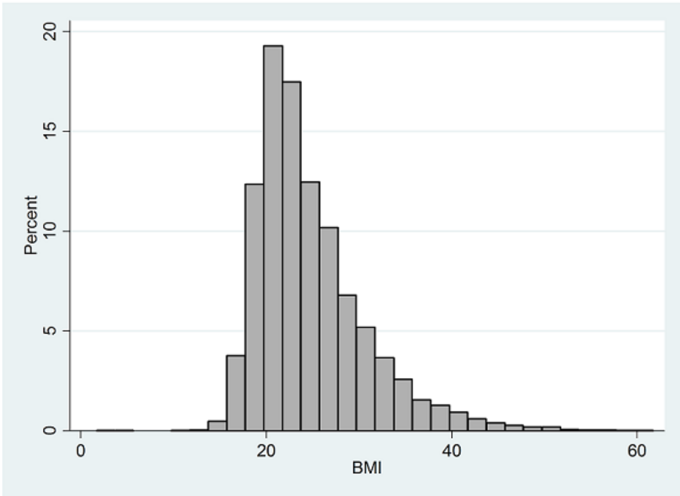


Figure 1: BMI distribution. Bin width: 2 units.

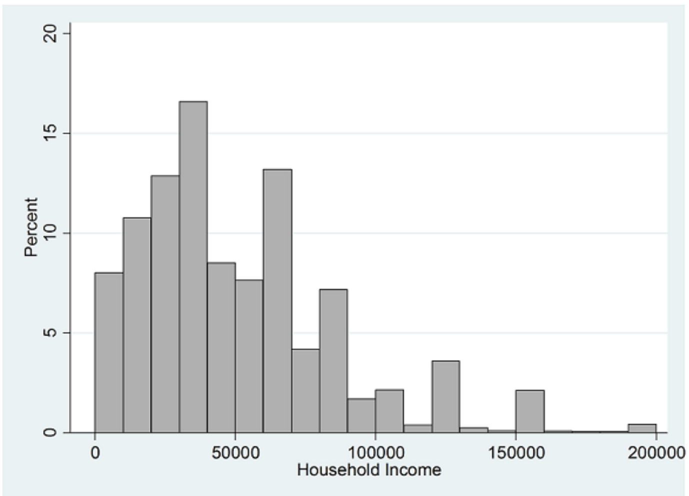


Figure 2: Household income distribution. Bin width: 10,000 US Dollars.

Table 3: Summary Statistics (Waves I-IV, total sample)

| Variable | Definition | Mean | SD | Missing Values |
|---|---|---|---|---|
| BMI | weight (kg)/height$^2$ (m) | 24.74 | 5.90 | 19% |
| Household Income | Total income before taxes | 52,320.90 | 50,626.62 | 39% |
| Median of Household Income | 40,000 | | | |
| 1st percentile | 70 | | | |
| 99th percentile | 200,000 | | | |
| Ethnicity (%) | Race as observed by interviewer | | | 11% |
| White | | 60.84 | 48.81 | |
| Black | African American | 21.94 | 41.39 | |
| American Indian | Native American. | 1.05 | 10.18 | |
| Asian | | 3.24 | 17.71 | |
| Other | | 2.31 | 15.02 | |
| Self-reported Health (%) | | | | 17% |
| Excellent | | 22.80 | 41.95 | |
| Very Good | | 33.04 | 47.03 | |
| Good | | 21.48 | 41.07 | |
| Fair | | 5.19 | 22.18 | |
| Poor | | 0.44 | 6.62 | |
| Age | Age from birth | 20.70 | 5.53 | 17% |
| Gender (%) | Proportion of females | 57.09 | 49.50 | 20% |
| Mother Education (% ) | Proportion of college graduate (or certified 4 years in college) | 9.63 | 29.50 | 18% |
| Father Education (% ) | Proportion of college graduate (or certified 4 years in college) | 9.45 | 29.95 | 18% |
| Urban | Completely urban city | 28.51 | 45.15 | 0% |
| Unemployment Rate | Low unemployment rate (vs Medium and High) | 29.24 | 45.49 | 46% |

Table 4: Average BMI by Subsample (Waves I-IV, total sample)

| Variable | Mean | SD |
|---|---|---|
| BMI if Male | 24.44 | 4.98 |
| BMI if Female | 23.95 | 4.99 |
| BMI if White | 24.70 | 5.75 |
| BMI if Black | 25.86 | 6.57 |
| BMI if American Indian | 29.37 | 8.91 |
| BMI if Asian | 23.28 | 4.85 |
| BMI if Other | 23.53 | 4.75 |
| BMI if Mother went to college | 22.58 | 4.70 |
| BMI if Father went to college | 22.45 | 4.60 |
| BMI if Excellent health | 23.31 | 4.61 |
| BMI if Very good health | 24.32 | 5.46 |
| BMI if Good health | 26.41 | 6.83 |
| BMI if Fair health | 28.31 | 8.06 |
| BMI if Poor health | 30.07 | 10.20 |
| BMI if Living in completely urban city | 22.95 | 4.76 |
| BMI if Living in not completely urban city | 22.67 | 4.73 |
| BMI if Low unemployment rate | 24.78 | 5.71 |
| BMI if Medium-High unemployment rate | 24.77 | 5.97 |
| BMI if Household income>Median | 24.39 | 5.50 |
| BMI if Household income<Median | 25.23 | 6.31 |

Table 5: Average BMI by Subsample (Waves I-II, adolescent sample)

| Variable | Mean | SD |
|---|---|---|
| BMI | 22.73 | 4.73 |
| BMI if Male | 22.57 | 4.60 |
| BMI if Female | 22.82 | 4.81 |
| BMI if White | 22.43 | 4.51 |
| BMI if Black | 23.43 | 5.13 |
| BMI if American Indian | 26.37 | 7.47 |
| BMI if Asian | 21.56 | 3.81 |
| BMI if Other | 23.28 | 4.65 |
| BMI if Mother went to college | 22.17 | 4.28 |
| BMI if Father went to college | 21.80 | 4.02 |
| BMI if Excellent health | 21.71 | 3.72 |
| BMI if Very good health | 22.35 | 4.24 |
| BMI if Good health | 23.78 | 5.35 |
| BMI if Fair health | 25.64 | 6.82 |
| BMI if Poor health | 26.92 | 8.55 |
| BMI if Living in completely urban city | 22.95 | 4.76 |
| BMI if Living in not completely urban city | 22.67 | 4.73 |
| BMI if Low unemployment rate | 22.53 | 4.46 |
| BMI if Medium-High unemployment rate | 22.78 | 4.81 |
| BMI if Household income>Median | 22.38 | 4.39 |
| BMI if Household income<Median | 23.14 | 5.05 |
| BMI if Smoked at least 1 cigarette daily for 30 days | 24.26 | 5.41 |
| BMI if Smokers in household | 22.98 | 4.93 |
| BMI if Own decision in diet | 22.90 | 4.73 |
| BMI if Have dinner with parents frequently | 23.03 | 4.63 |

**Table 6: Estimates using full sample**

| Dependent variable: ln(BMI) | | | | | | |
|---|---|---|---|---|---|---|
| Variables | Model 1 | | Model 2 | | Model 3 | |
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| $\ln(\text{BMI})_{t-1}$ | 0.981*** | (0.214) | 0.675*** | (0.165) | 0.839*** | (0.154) |
| Average ln(BMI) | 0.259* | (0.151) | 0.469*** | (0.091) | 0.440*** | (0.084) |
| Household Income | 0.003* | (0.002) | 0.001 | (0.002) | 0.000 | (0.002) |
| Age | −0.003 | (0.003) | 0.001 | (0.003) | 0.000 | (0.003) |
| Woman | −0.005 | (0.004) | −0.002 | (0.003) | 0.004 | (0.004) |
| Ethnic Group: White | −0.019 | (0.046) | −0.060** | (0.028) | −0.094** | (0.047) |
| Ethnic Group: Black | −0.037 | (0.049) | −0.047 | (0.029) | −0.082* | (0.048) |
| Ethnic Group: American Indian | −0.032 | (0.049) | −0.074* | (0.038) | −0.108** | (0.055) |
| Ethnic Group: Asian | −0.012 | (0.046) | −0.078*** | (0.029) | −0.102** | (0.047) |
| Ethnic Group: Other | −0.029 | (0.034) | −0.060** | (0.030) | −0.081 | (0.050) |
| Maternal Education (college) | 0.005 | (0.011) | −0.004 | (0.010) | 0.001 | (0.009) |
| Paternal Education (college) | −0.009 | (0.006) | −0.011* | (0.006) | −0.011** | (0.005) |
| Health Status (Excellent to Fair) | −0.019 | (0.012) | −0.031*** | (0.007) | −0.025*** | (0.007) |
| First-2-waves Dummy | −0.006 | (0.006) | −0.004 | (0.007) | 0.001 | (0.007) |
| Average Household Income | 0.003 | (0.005) | 0.000 | (0.003) | 0.002 | (0.003) |
| Average Age | −0.001 | (0.017) | −0.007*** | (0.002) | −0.009*** | (0.002) |
| Average presence of women | 0.001 | (0.017) | −0.003 | (0.016) | 0.006 | (0.018) |
| Ethnic Group: Average White | −0.220 | (0.320) | −0.180 | (0.273) | −0.211 | (0.301) |
| Ethnic Group: Average Black | −0.238 | (0.321) | −0.202 | (0.272) | −0.241 | (0.301) |
| Ethnic Group: Average American Indian | −0.183 | (0.320) | −0.188 | (0.274) | −0.245 | (0.301) |
| Ethnic Group: Average Asian | −0.200 | (0.325) | −0.158 | (0.274) | −0.188 | (0.302) |
| Ethnic Group: Average Other | −0.216 | (0.322) | −0.153 | (0.273) | −0.165 | (0.304) |
| Average Maternal Education (college) | −0.140** | (0.059) | −0.064 | (0.063) | −0.070 | (0.057) |
| Average Paternal Education (college) | 0.041 | (0.042) | 0.040 | (0.042) | 0.038 | (0.041) |
| Average Health Status (Excellent to Fair) | 0.038* | (0.021) | 0.035** | (0.017) | 0.054*** | (0.020) |
| Observations | 6,598 | | 10,700 | | 10,677 | |
| Number of individuals | 4,095 | | 4,655 | | 4,646 | |
| Number of Instruments | 31 | | 31 | | 31 | |
| Arellano-Bond Test for AR(1) in first Differences: Pr $> z$ | $9.16e^{-07}$ | | $2.06e^{-08}$ | | 0 | |
| Arellano-Bond Test for AR(1) in first Differences: $z$ | −4.909 | | −5.607 | | −6.665 | |
| Hansen test of overid. restrictions: Pr $> \chi^2$ | 0.298 | | 0.888 | | 0.341 | |
| Hansen test of overid. restrictions: Degrees of Freedom | 3 | | 3 | | 3 | |
| Hansen test of overid. restrictions: $\chi^2$ | 3.684 | | 0.635 | | 3.349 | |
| Sargan test of overid. restrictions: Pr $> \chi^2$ | 0.308 | | 0.683 | | 0.386 | |
| Sargan test of overid. restrictions: Degrees of Freedom | 3 | | 3 | | 3 | |
| Sargan test of overid. restrictions: $\chi^2$ | 3.600 | | 0.683 | | 3.034 | |

Standard errors in parentheses:.$^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$

NOTES. Models differ because of missing values treatment. Model 1: Missing values are not treated; Model 2: income missing values replaced by multiple-imputed values; Model 3: BMI and income missing values replaced by multiple-imputed values.
For further details on missing values treatment see Appendix A, section A.1.2.
Our estimation also accounts of wave effects, through the inclusion of time and adolescent dummies.

**Table 7: Estimates using sample of individuals who are normal-weighted during adolescence**

Dependent variable: ln(BMI)

| Variables | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| $\ln(\text{BMI})_{t-1}$ | 1.269*** | (0.446) | 1.183*** | (0.429) | 0.588 | (0.673) |
| Average ln(BMI) | 0.194* | (0.115) | 0.269*** | (0.079) | 0.369*** | (0.106) |
| Household Income | 0.004 | (0.003) | 0.002 | (0.002) | −0.002 | (0.002) |
| Age | −0.002 | (0.002) | −0.002 | (0.002) | −0.000 | (0.003) |
| Woman | −0.008 | (0.006) | −0.010** | (0.005) | −0.006* | (0.004) |
| Ethnic Group: White | −0.036 | (0.066) | −0.055 | (0.040) | −0.059** | (0.026) |
| Ethnic Group: Black | −0.023 | (0.066) | −0.051 | (0.040) | −0.049* | (0.027) |
| Ethnic Group: American Indian | −0.042 | (0.071) | −0.058 | (0.056) | −0.048 | (0.044) |
| Ethnic Group: Asian | −0.040 | (0.067) | −0.045 | (0.045) | −0.062** | (0.030) |
| Ethnic Group: Other | −0.033 | (0.067) | −0.093** | (0.042) | −0.058* | (0.030) |
| Maternal Education (college) | −0.004 | (0.012) | −0.001 | (0.010) | −0.014 | (0.011) |
| Paternal Education (college) | −0.008 | (0.008) | −0.004 | (0.006) | −0.008 | (0.010) |
| Health Status (Excellent to Fair) | −0.020** | (0.010) | −0.018** | (0.008) | −0.029*** | (0.008) |
| First-2-waves Dummy | 0.001 | (0.010) | −0.003 | (0.009) | −0.006 | (0.008) |
| Average Household Income | −0.002 | (0.007) | −0.003 | (0.004) | 0.001 | (0.004) |
| Average Age | −0.005* | (0.003) | −0.009*** | (0.002) | −0.008*** | (0.002) |
| Average presence of women | 0.035 | (0.026) | 0.026 | (0.023) | −0.001 | (0.020) |
| Ethnic Group: Average White | −0.356 | (0.438) | −0.353 | (0.468) | 0.001 | (0.357) |
| Ethnic Group: Average Black | −0.379 | (0.440) | −0.374 | (0.470) | −0.017 | (0.357) |
| Ethnic Group: Average American Indian | −0.287 | (0.436) | −0.336 | (0.469) | −0.010 | (0.351) |
| Ethnic Group: Average Asian | −0.376 | (0.447) | −0.366 | (0.472) | 0.012 | (0.364) |
| Ethnic Group: Average Other | −0.356 | (0.446) | −0.271 | (0.473) | 0.076 | (0.356) |
| Average Maternal Education (college) | −0.071 | (0.085) | −0.042 | (0.088) | −0.074 | (0.129) |
| Average Paternal Education (college) | 0.046 | (0.055) | −0.002 | (0.047) | 0.010 | (0.061) |
| Average Health Status (Excellent to Fair) | 0.043* | (0.026) | 0.019 | (0.024) | 0.047** | (0.021) |
| Observations | 4,183 | | 6,591 | | 6,620 | |
| Number of individuals | 2,581 | | 2,867 | | 2,881 | |
| Number of instruments | 31 | | 31 | | 31 | |
| Arellano-Bond Test for AR(1) in first Differences: $\Pr > z$ | 0.00369 | | 0.00186 | | 0.222 | |
| Arellano-Bond Test for AR(1) in first Differences: $z$ | −2.904 | | −3.111 | | −1.221 | |
| Hansen test of overid. restrictions: $\Pr > \chi^2$ | 0.754 | | 0.900 | | 0.737 | |
| Hansen test of overid. restrictions: Degrees of Freedom | 3 | | 3 | | 3 | |
| Hansen test of overid. restrictions: $\chi^2$ | 1.195 | | 0.583 | | 1.266 | |
| Sargan test of overid. restrictions: $\Pr > \chi^2$ | 0.777 | | 0.941 | | 0.702 | |
| Sargan test of overid. restrictions: Degrees of Freedom | 3 | | 3 | | 3 | |
| Sargan test of overid. restrictions: $\chi^2$ | 1.101 | | 0.396 | | 1.417 | |

Standard errors in parentheses:.$^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$

NOTES. Sample split according to the BMI status in the first wave (individuals are kept in the same strata).

Models differ because of missing values treatment. Model 1: Missing values are not treated; Model 2: income missing values replaced by multiple-imputed values; Model 3: BMI and income missing values replaced by multiple-imputed values.

For further details on missing values treatment see Appendix A, section A.1.2.

Our estimation also accounts of wave effects, through the inclusion of time and adolescent dummies.

**Table 8: Estimates using sample of individuals who are overweight during adolescence**

| Dependent variable: ln(BMI) | | | | | | |
|---|---|---|---|---|---|---|
| Variables | Model 1 | | Model 2 | | Model 3 | |
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| $\ln(\text{BMI})_{t-1}$ | 0.097 | (0.610) | −0.195 | (0.923) | −1.281 | (1.841) |
| Average ln(BMI) | 0.501** | (0.155) | 0.599*** | (0.122) | 0.667*** | (0.196) |
| Household Income | 0.007* | (0.004) | 0.005 | (0.003) | 0.002 | (0.010) |
| Age | 0.001 | (0.006) | −0.002 | (0.006) | 0.004 | (0.009) |
| Woman | −0.001 | (0.012) | −0.001 | (0.018) | −0.015 | (0.023) |
| Ethnic Group: White | 0.055 | (0.072) | 0.040 | (0.042) | 0.010 | (0.064) |
| Ethnic Group: Black | 0.080 | (0.063) | 0.075* | (0.039) | 0.064 | (0.058) |
| Ethnic Group: American Indian | 0.038 | (0.077) | 0.043 | (0.050) | −0.016 | (0.099) |
| Ethnic Group: Asian | 0.004 | (0.081) | 0.045 | (0.056) | 0.047 | (0.088) |
| Ethnic Group: Other | 0.082 | (0.071) | 0.073* | (0.044) | 0.023 | (0.069) |
| Maternal Education (college) | 0.051 | (0.034) | 0.039 | (0.053) | 0.069 | (0.072) |
| Paternal Education (college) | −0.036 | (0.036) | −0.042 | (0.053) | −0.112 | (0.113) |
| Health Status (Excellent to Fair) | −0.034** | (0.014) | −0.026** | (0.011) | −0.034*** | (0.012) |
| First-2-waves Dummy | −0.003 | (0.017) | −0.025 | (0.022) | 0.003 | (0.037) |
| Average Household Income | 0.012 | (0.012) | 0.005 | (0.009) | 0.011 | (0.014) |
| Average Age | −0.015** | (0.007) | −0.002 | (0.007) | 0.001 | (0.024) |
| Average presence of women | −0.054 | (0.069) | −0.012 | (0.045) | 0.023 | (0.089) |
| Ethnic Group: Average White | −0.418 | (0.938) | −0.061 | (0.654) | 0.396 | (0.988) |
| Ethnic Group: Average Black | −0.426 | (0.938) | −0.089 | (0.652) | 0.356 | (0.985) |
| Ethnic Group: Average American Indian | −0.346 | (0.944) | −0.084 | (0.672) | 0.463 | (1.058) |
| Ethnic Group: Average Asian | −0.252 | (0.946) | 0.021 | (0.662) | 0.528 | (1.007) |
| Ethnic Group: Average Other | −0.468 | (0.936) | −0.141 | (0.658) | 0.376 | (1.002) |
| Average Maternal Education (college) | −0.209 | (0.206) | −0.042 | (0.088) | 0.117 | (0.332) |
| Average Paternal Education (college) | 0.213 | (0.246) | −0.002 | (0.047) | 0.413 | (0.392) |
| Average Health Status (Excellent to Fair) | 0.086 | (0.057) | 0.021 | (0.045) | 0.047 | (0.065) |
| Observations | 942 | | 1,515 | | 1,542 | |
| Number of individuals | 577 | | 668 | | 669 | |
| Number of instruments | 31 | | 31 | | 31 | |
| Arellano-Bond Test for AR(1) in first Differences: $\Pr > z$ | 0.256 | | 0.706 | | 0.839 | |
| Arellano-Bond Test for AR(1) in first Differences: $z$ | −1.137 | | −0.377 | | 0.204 | |
| Hansen test of overid. restrictions: $\Pr > \chi^2$ | 0.933 | | 0.646 | | 0.997 | |
| Hansen test of overid. restrictions: Degrees of Freedom | 3 | | 3 | | 3 | |
| Hansen test of overid. restrictions: $\chi^2$ | 0.436 | | 1.659 | | 0.0548 | |
| Sargan test of overid. restrictions: $\Pr > \chi^2$ | 0.413 | | 0.334 | | 0.936 | |
| Sargan test of overid. restrictions: Degrees of Freedom | 3 | | 3 | | 3 | |
| Sargan test of overid. restrictions: $\chi^2$ | 2.864 | | 3.403 | | 0.420 | |

Standard errors in parentheses:.$^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$

NOTES. Sample split according to the BMI status in the first wave (individuals are kept in the same strata).
Models differ because of missing values treatment. Model 1: Missing values are not treated; Model 2: income missing values replaced by multiple-imputed values; Model 3: BMI and income missing values replaced by multiple-imputed values.
For further details on missing values treatment see Appendix A, section A.1.2.
Our estimation also accounts of wave effects, through the inclusion of time and adolescent dummies.

**Table 9: Estimates using sample of individuals who are obese during adolescence**

Dependent variable: ln(BMI)

| Variables | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| $\ln(\text{BMI})_{t-1}$ | 0.664* | (0.399) | 1.020*** | (0.333) | 0.968*** | (0.358) |
| Average ln(BMI) | 1.166 | (0.828) | 1.017 | (0.690) | 1.321** | (0.611) |
| Household Income | 0.006 | (0.011) | −0.001 | (0.010) | 0.017 | (0.013) |
| Age | −0.013* | (0.007) | −0.006 | (0.009) | −0.002 | (0.009) |
| Woman | 0.030 | (0.020) | 0.048** | (0.021) | 0.051** | (0.022) |
| Ethnic Group: White | −3.489 | (4.463) | 0.404 | (4.176) | −1.117 | (3.767) |
| Ethnic Group: Black | −3.486 | (4.445) | 0.383 | (4.174) | −1.116 | (3.767) |
| Ethnic Group: American Indian | −3.483 | (4.396) | 0.359 | (4.140) | −1.139 | (3.705) |
| Ethnic Group: Asian | −3.521 | (4.419) | 0.224 | (4.127) | −1.200 | (3.727) |
| Ethnic Group: Other | −3.493 | (4.478) | 0.403 | (4.186) | −1.132 | (3.791) |
| Maternal Education (college) | −0.059 | (0.060) | 0.023 | (0.031) | 0.025 | (0.032) |
| Paternal Education (college) | 0.066 | (0.044) | 0.034 | (0.032) | 0.013 | (0.032) |
| Health Status (Excellent to Fair) | −0.031 | (0.032) | 0.016 | (0.034) | −0.003 | (0.036) |
| First-2-waves Dummy | −0.043 | (0.031) | −0.035 | (0.035) | −0.034 | (0.038) |
| Average Household Income | −0.005 | (0.019) | −0.002 | (0.015) | −0.016 | (0.013) |
| Average Age | −0.002 | (0.009) | −0.006 | (0.012) | −0.013 | (0.012) |
| Average presence of women | −0.130 | (0.095) | −0.067 | (0.102) | −0.023 | (0.095) |
| Ethnic Group: Average White | 0.902 | (2.801) | −3.245 | (2.168) | −2.809 | (2.564) |
| Ethnic Group: Average Black | 0.883 | (2.806) | −3.305 | (2.161) | −2.874 | (2.568) |
| Ethnic Group: Average American Indian | 0.802 | (2.759) | −3.458 | (2.134) | −3.019 | (2.500) |
| Ethnic Group: Average Asian | 1.032 | (3.048) | −3.318 | (2.319) | −2.738 | (2.730) |
| Ethnic Group: Average Other | 0.935 | (2.891) | −3.390 | (2.182) | −2.922 | (2.626) |
| Average Maternal Education (college) | −0.166 | (0.300) | −0.105 | (0.324) | −0.203 | (0.295) |
| Average Paternal Education (college) | 0.168 | (0.200) | 0.133 | (0.224) | 0.336 | (0.215) |
| Average Health Status (Excellent to Fair) | 0.635 | (0.980) | −0.311 | (0.686) | 0.222 | (0.746) |
| Observations | 399 | | 650 | | 657 | |
| Number of individuals | 247 | | 272 | | 270 | |
| Number of instruments | 31 | | 31 | | 31 | |
| Arellano-Bond Test for AR(1) in first Differences: $\Pr > z$ | 0.0135 | | 0.00627 | | 0.00214 | |
| Arellano-Bond Test for AR(1) in first Differences: $z$ | −2.470 | | −2.733 | | −3.070 | |
| Hansen test of overid. restrictions: $\Pr > \chi^2$ | 0.0589 | | 0.165 | | 0.222 | |
| Hansen test of overid. restrictions: Degrees of Freedom | 2 | | 2 | | 2 | |
| Hansen test of overid. restrictions: $\chi^2$ | 5.664 | | 3.601 | | 3.012 | |
| Sargan test of overid. restrictions: $\Pr > \chi^2$ | 0.0541 | | 0.325 | | 0.320 | |
| Sargan test of overid. restrictions: Degrees of Freedom | 2 | | 2 | | 2 | |
| Sargan test of overid. restrictions: $\chi^2$ | 5.833 | | 2.250 | | 2.276 | |

Standard errors in parentheses:.$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

NOTES. Sample split according to the BMI status in the first wave (individuals are kept in the same strata).

Models differ because of missing values treatment. Model 1: Missing values are not treated; Model 2: income missing values replaced by multiple-imputed values; Model 3: BMI and income missing values replaced by multiple-imputed values.

For further details on missing values treatment see Appendix A, section A.1.2.

Our estimation also accounts of wave effects, through the inclusion of time and adolescent dummies.