On the persistence of obesity

J. J. Daouli, A. Davillas, M. Demoussis*, N. Giannakopoulos
Department of Economics, University of Patras

Abstract

While the prevalence of obesity is increasing worldwide, little is known about its dynamic behavior. We employ longitudinal data (NLSY79) for adult white males and appropriate dynamic discrete choice models. Assuming that obesity follows a Markov process we first identify patterns of obesity experiences. Then, we decompose the identified profound persistence into unobserved heterogeneity and genuine state dependence, while controlling for initial conditions. The results indicate that obesity has a lasting effect and is primarily affected by early-life biological endowments and family background conditions. Therefore, policy interventions designed to reduce obesity prevalence should focus on early adulthood.

Keywords: Obesity, persistence, dynamic discrete choice models
JEL classification: I12, C25

* Corresponding author: Department of Economics, University of Patras, University Campus, Rio, 26504, Greece, Tel.: +302610 996134, e-mail address: micdem@upatras.gr

I. Introduction

The prevalence of obesity is increasing worldwide, more so in developed western societies. Over the last two decades the average Body Mass Index (BMI)[1] in the U.S. has increased considerably while obesity prevalence more than doubled (Chou et al., 2004). As a rule, obesity is considered to be an individual-specific characteristic with significant social ramifications. Obesity is additionally identified as an economic phenomenon. Philipson (2001) argues -in a rational choice context- that obesity is a preventable state, which can be adjusted through diet and behavioural modifications. Thus, the obesity state can be altered through behavioural adjustments, which will be assumed if benefits exceed their costs. In this context, an improved understanding of the underlying process of obesity outcomes is expected to provide policy makers with the necessary tools for the design and implementation of an effective public policy in the fight against obesity. The objective of this paper is to analyze obesity dynamics and investigate the sources of obesity persistence if indeed exists.

Typically, an individual is classified as obese if his/her BMI index exceeds a certain threshold. A large body of literature has investigated the determinants of obesity using cross-sectional and pooled panel data sets and techniques (Chou et al., 2004; Huffman and Rizov, 2006; Komlos and Baur, 2004; Lakdawalla and Philipson, 2009; Rashad, 2006; Rashad et al., 2006). However, the available studies assume that obesity is determined only contemporaneously. That is, they ignore the fact that obesity could be an endogenous outcome of an intertemporal process regarding the costs and benefits of being in the desired obesity state[2]. In fact, recent studies suggest that body weight could be modeled through a dynamic human capital investment approach according to which individuals compare the lifetime costs and benefits of increased body weight (Lakdawalla and Philipson, 2009). However, it is well known that the costs of increased body weight are different at different points of the body-weight distribution. These costs are assumed to be greater at the upper tail of the distribution since approaching and eventually falling into the obesity state is associated with increased costs (e.g., health and work related).

Following this human capital approach, current body weight is considered to be a stock variable depending on former periods' body weight, (Ng et al., 2010). In particular, an individual is considered to be (or running the risk of becoming) obese if his body weight is permanently at the right end of the population distribution of weights (BMI≥30) or his

[1] The Body Mass Index (BMI) is defined as body weight in kilograms divided by the square of height in meters.
[2] For an analysis of the dynamics of childhood fatness measures see Bao (2007) and for an analysis of the joint dynamics of spousal obesities see Kano (2008).

intertemporal body-weight growth, conditional on his initial weight, is substantial. With regard to the latter, Cawley (1999) argues that “calorie intake” and thus weight gain constitutes a stochastic process, which within the confines of a rational addiction model implies that, since “caloric intake” is addictive, weight gain follows a snowballing pattern. Becker and Murphy (1988) suggest that addiction requires a lasting effect of past and current behavioural outcomes. In this context, current obesity is assumed to encompass a lasting effect (persistence), which is the result of genuine state dependence, unobserved heterogeneity and serial correlation of errors (Heckman, 1981).

We employed longitudinal data from the National Longitudinal Survey of Youth (NLSY79) and focused on adult obesity[3] of white males. We modeled obesity as a Markov process and estimated transition probability matrices for the period 1986 – 1994. Significant immobility seems to characterize movements in and out of the obese group, indicating the presence of significant persistence. In order to formally model and test this dynamic stochastic process and further decompose the observed persistence of obesity, we utilized dynamic discrete choice models. The decomposition refers to "genuine" state dependence and unobserved heterogeneity. Using alternative estimators and controlling for a number of observed individual and family background characteristics, the obtained results indicate that the identified effect of state dependence on obesity persistence is upward biased. In fact, most of the state dependent effect is due to the existing conditions at the beginning of the process that created the obesity outcomes (i.e., early life conditions). Thus, current obesity status is primarily affected by individual specific time-invariant characteristics and to a lesser extent by genuine state dependence.

The paper is organized as follows. Section II presents a brief description of the utilized dataset, definition of variables, transition probability matrices and synthetic mobility indicators. Section III presents the econometric specification and the empirical strategy. In Section IV we present and discuss the results of the econometric analysis of alternative model specifications and dynamic discrete choice estimators. The paper concludes with Section V which contains a summary of major findings and some policy implications.

II. Data and preliminary analysis

Microdata are extracted from the NLSY79, which is a US representative sample of 12,686 young men and women who were 14 to 22 years of age when initially interviewed in

[3] Adult obesity is characterized by lower variation than childhood obesity (WHO, 2000).

1979. Interviews were administered yearly through 1994 and biennially from 1996 through 2006. The working sample begins in 1985 (when all respondents were adults) and ends in 1994 (after this year the interviews became biennial). We exclude 1987 and 1991 because there are no available data on body weight. Following Cawley (2004), each survey year body-weight and the 1985 body-height were used for the BMI calculation. An individual is categorized as obese if his BMI is greater than or equal to 30. That is, an individual specific dummy indicator is created which takes the value of 1 if BMI≥30 and 0 otherwise.

We focus on adult obesity in white males since a) females experience significant variations in body weight related to pregnancy periods and b) significant differences in prevalence of obesity are observed for different ethnic/racial groups (Baum, 2007; Baum and Ruhm, 2009). Furthermore, we exclude individuals not continuously observed in the examined period. The adopted sample selection criteria and the resulting sub-sample allow us to work with a more homogenous group of individuals. The resulting eight-year balanced panel consists of 1,416 adult males observed in every wave of the survey period 1985-1994 (11,328 observations)[4]. The set of explanatory correlates includes the educational level, age, marital status, smoking behavior and residential choices. The vector also includes some pre-sample conditions such as, living with parents at the age of 14, the number of siblings and the average body-weight growth in the period 1981-1985 (see Chou et al., 2004; Rashad, 2006; Huffman and Rizov, 2007 and Lakdawalla and Philipson, 2009).

Figure 1 displays the evolution of the obesity prevalence in the examined period. It appears that the proportion of obese individuals more than doubled between 1985 (7.2 per cent) and 1994 (16.7 per cent) verifying the well known fact of increased prevalence of obesity over time (Chou et al., 2004; Lakdawalla and Philipson, 2009; Huffman and Rizov, 2007). A detailed picture of obesity dynamics can be obtained by looking at the historical sequence of an individual's obesity outcome. The sequence refers to an eight-calendar year period where each respondent may have followed any of the 256 ($=2^8$) different pathways regarding their obese/non-obese status (see Table 1). We observe that the vast majority of observations (82.3 per cent) classify individuals as either continuously obese ("11111111") or continuously non-obese ("00000000"). Furthermore, single transitions (e.g., "00000001", "10000000", "01111111", "11111110" , etc.) account for 7.9 per cent. This directly implies that the likelihood of multiple transitions (more than one) is relatively small (9.8 per cent).

[4] Balanced and unbalanced panels produced similar summary statistics, indicating that attrition is not a factor. Moreover, attrition due to deaths from obesity-related diseases is rare in the NLSY's age range (Courtemanche, 2009).

Since the majority of individuals are non-obese (00000000) in all sample periods (78.2 per cent), the process underlying the sequences is not independent over time, i.e., serial persistence is present. That is, if serial persistence has originated from a time-invariant personality trait, then one would have observed different number of individuals across each class of sequences but approximately the same number of individuals in each sequence within a class, ceteris paribus. For example, in the class of sequences in which individuals are obese 5 times in total, the most prevalent sequences are those in which an individual is obese in the last five survey years (00011111). The same holds for the classes of 6 (00111111) and 7 (01111111) times. Given that multiple transitions are observed with a low order of probability (9.8 per cent), we should resort to a model that accounts for both, unobserved heterogeneity (time-invariant personality traits) and state dependence (snowballing effect). Such a model is expected to fit better the observed obesity patterns.

--Insert Figure 1 about here --

--Insert Table 1 about here --

To further corroborate the above findings, transition probability matrices are estimated for consecutive time periods (survey years) using data on the same individual. This exercise allows us to estimate -through a Markov process- how obesity status in past periods affects the probability distribution of current obesity. Based on these estimates, we calculate next two mobility measures (i.e., Prais-Shorocks Index-PS[5] and Immobility Ratio-IR[6]). The corresponding results are presented at Table 2. Panel A presents results for subsequent survey- year periods whereas Panel B presents results for the entire period (1985-1994). At Panel A we see that more than 80 per cent of individuals observed in the obese state in one survey year were also observed in the same state in the subsequent survey year. It is interesting to note (Panel B) that the same holds for the entire period i.e., the survey pair 1985-1994. Both panels reveal that obesity is characterized by a lasting effect between past and current obesity experiences[7]. This lasting effect implies lack of mobility which is also

[5] The Prais-Shorrocks (PS) index (Shorrocks, 1978) was calculated as PS=[n-*trace*(P)]/[n-1] where, P=P[$p_{i,j}$] stands for the *nxn* transition matrix and $p_{i,j}$ denotes the conditional probability of moving into state *j* next period, given than state *i* is observed in the current period and *n* is equal to the number of different states (i.e., obese, non-obese). The PS index takes values in the interval [0, 1] where, zero stands for perfect immobility and one for perfect mobility.

[6] The Immobility Ratio (IR) is calculated as the average value of the entries on the main diagonal of the transition matrix (see ft.5) and shows the immobility of individuals between the obese and the non-obese states over a specified time interval. It takes values in the interval [0, 1] indicating perfect mobility if 0 and perfect immobility if 1.

[7] We have also used alternative heights for the calculation of BMI, i.e., the height of 1981, 1982 and the average of 1981, 1982 and 1985 and the results remained practically unchanged.

confirmed by the values of the calculated PS and IR ratios. This empirical evidence is followed in the next section by a formal investigation of the observed persistence.

--Insert Table 2 about here --

Table 3 presents descriptive statistics of the dependent and independent variables for the full sample and for different sub-samples of individuals, according to their obesity status. It is observed that 13.4 per cent of the whole sample over the examined period is classified as obese. Regarding the explanatory correlates we observe that- in comparison to the continuously non-obese sub-sample- continuously obese individuals are less educated, less likely to be permanent smokers[8], have different residential/geographical profiles, more likely to live in rural areas and to have lived with both biological parents at the age of 14, have more siblings and their average body-weight growth (in the period 1981-1985) was considerably higher (twice as high). Furthermore, at the lower part of Table 3 we present the distribution of the number of years in which individuals are obese in the examined period. For example, 2.54 per cent of individuals were classified as continuously obese 5 times in the 8-year period. As shown above, the persistence of the obesity status (continuously 0 or 1) is profound (76.55+4.59=81.14 per cent). Single and multiple transitions account for the remaining 18.86 per cent.

--Insert Table 3 about here --

III. Econometric Methodology

The obesity indicator (O_{it}) is modeled as a dummy variable, which takes the value of 1 if an individual i at time t is classified as obese and 0 otherwise. Since obesity is assumed to follow a first-order Markov process, the issue of serial persistence needs to be incorporated in the model specification. Given that the unobserved characteristics e_{it}, which affect obesity, are normally independently distributed, the dynamic pooled probit model *(Pooled)* takes the form:

$$\Pr(O_{it} = 1 \mid x_{it}) = \Phi(\gamma O_{it-1} + x_{it}\beta + e_{it}) \tag{1}$$

where, O_{it-1} is the lagged dependent variable, x_{it} is a set of individual characteristics, γ and β are coefficients to be estimated and $\Phi(.)$ is the correspondent cumulative normal distribution function. However, in equation (1) the longitudinal nature of the data is ignored and thus a more flexible specification is required, which can handle the issue of unobserved

[8] Questions regarding smoking were asked in 1984, 1992 and 1994. Following Levine et al. (1997) we identify three groups: "permanent smokers" (smoking daily in all years), "non-smokers" (not smoking in all years) and "occasional smokers" (those reporting transitions in and out smoking).

time-invariant heterogeneity. For this purpose, a random effect term (a_i) is included in the error component, i.e., $e_{it} = a_i + u_{it}$, which captures differences among individuals due to time-invariant unobserved factors. The resulting random effects probit model *(RE)* takes the following form:

$$\Pr(O_{it} = 1 | x_{it}, a_i) = \Phi(\gamma O_{it-1} + x_{it}\beta + a_i + u_{it}) \qquad (2)$$

where, a_i is assumed to be random and unrelated to x_{it}. Equation (2) is estimated using convectional maximum likelihood estimation (MLE) (see Greene, 2003). The presence of unobserved heterogeneity requires rejection of the hypothesis that the variance due to unobserved heterogeneity, i.e., $\lambda = \sigma_a^2 / (\sigma_a^2 + \sigma_u^2)$, is equal to zero. Further, the presence of obesity state dependence requires rejection of the hypothesis that $\gamma = 0$, while controlling for a_i.

However, the proposed approach suffers from the "initial conditions" problem, i.e., the obesity status in the beginning of the process could be correlated with a_i. Therefore, the effect of state dependence could be overestimated (Heckman, 1981). This can be resolved either by directly controlling for O_{i1} in (2) (Wooldridge, 2005, i.e., *RE-IC-W*)[9] or by using a flexible reduced form model for O_{i1} conditional on the initial period regressors x_{i1} and other pre-sample variables (Heckman, 1981, i.e., *RE-IC-H*)[10]. In the latter case, the pre-sample variables- presented at Table 3- were used as instruments. The number of siblings and living with parents are expected to capture the influence of family background during childhood on O_{i1}. Lastly, in order to account for the effects of metabolism, past consumption habits and earlier physical activity (Cawley, 1999; Baum and Ruhm, 2009), the average weight growth in the pre-sample period was also included.

III. Results

Table 4 presents the estimated results of two model specifications: the pooled probit (*Pooled*) and the dynamic random effects probit model (*DRE*). We observe that the effect of lagged obesity is positive and statistically significant in both models. In the first model and according to the estimated marginal effect, an individual classified as obese in t-1 has an 80

[9] In this model we also include the averages of the time-variant explanatory variables (Mundlak, 1978). We do not control for the time average of the years of education completed since it is highly correlated with the reported years of education variable (correlation coefficient about 98 per cent).
[10] We do not account for autocorrelated errors due to insurmountable computational difficulties.

per cent probability to be classified as obese in year t. In the second model, which accounts for unobserved time-invariant heterogeneity, the corresponding probability is only 11.9 per cent. This substantial reduction (the estimated coefficient in the pooled model was 2.771 while in the DRE was only 1.233[11]) is due to the decomposition of the error term and the explicit modeling of the time-invariant unobserved component, which is statistically significant. In addition, the quasi-likelihood test shows that -between the two models- the DRE is preferred. Turning now to the estimated marginal effects of the covariate vector (DRE model) we observe that there is a significant and negative correlation between current obesity and education, employment status and smoking on a permanent basis. For example, permanent smokers have a lower probability (around 6.5 percentage points) of being obese compared to non-smokers. In contrast, current obesity is positively and significantly correlated with age and marital status. For instance, married individuals have on average a higher probability (2.4 percentage points) of being obese compared to unmarried ones.

--Insert Table 4 about here --

Table 5 presents estimation results for the Wooldridge (DRE-IC-W) and Heckman (DRE-IC-H) model specifications, which account for the endogeneity of initial conditions. In the DRE-IC-W model the estimated coefficient (and its marginal effect) of lagged obesity is positive, significant and lower (8.6 per cent) than the corresponding coefficient in the DRE model (Table 4). Furthermore, the marginal effect of being classified as obese in the initial period is significant, positive and considerable in size (67.0 per cent). These findings imply the existence of a positive correlation between the obesity status in the initial period and the process that generates the obesity outcomes. It is worth noticing that the reduction in the model's variance due to unobserved heterogeneity (λ), compared with the DRE model (Table 4), is about 7 per cent. With regard to the effects of individual covariates, it appears that the employment status, age and residential choices lost their statistical significance, indicating possible endogeneities with respect to the initial obesity status. In addition, the remaining explanatory variables still exert a significant impact albeit of a smaller magnitude (e.g., smoking and education).

The DRE-IC-H model produced similar results[12]. The hypothesis regarding the exogeneity of the initial conditions was rejected (see the value of θ at the lower part of Table

[11] For comparison purposes with the pooled probit, the DRE estimate of lagged obesity has been multiplied by $\sqrt{1-\lambda}$.

[12] The inclusion of averages of the time-variant covariates, as in case of the DRE-IC-W model, does not alter the coefficient of the lagged dependent variable.

5 and the value of the quasi-likelihood test). Nevertheless, the advantage of this model is that it makes possible the examination of the correlation between the pre-sample variables and the obesity status at the initial period (see lower part of Table 6). The results of this exercise reveal that living with parents during childhood and the number of siblings are associated with a lower probability of obesity in the initial period. In contrast, the weight gain in the pre-sample period exerts a positive and statistically significant effect on the obesity status of the initial period.

--Insert Table 5 about here --

--Insert Table 6 about here --

Overall, the obtained results indicate that the probability of being obese is influenced by a) genuine state dependence, i.e., correlation between current and past obesity statuses, while controlling for initial conditions (family background and metabolism), b) various observed covariates (e.g., education, employment status, age, marital status, smoking behavior) and c) by time-invariant unobserved factors. With regard to observed correlates, the obtained results are in agreement with the accumulated evidence. For example, various studies report a negative correlation between obesity and educational attainment (Chou et al., 2004; Conti et al., 2010; Webbink et al., 2010 and Nakamura and Siciliani, 2010). Additionally, the negative correlation between employment and obesity is also identified by Morris (2007) and it is often attributed to various causes such as, discrimination in hiring (Rooth, 2009) and/or consumption of low quality food by the poorer individuals (Cawley, 2004). Age and marital statuses were also identified as significant contributors (Sobal et al., 1992; Chou et al., 2004; Lakdawalla and Philipson, 2009). Finally, medical reasons may explain the negative association between smoking and obesity (Williamson et al., 1991). In this context, a positive correlation between cigarette prices and obesity has been identified (e.g., Chou et al, 2004; Rashad et al., 2006 and Baum 2009). Nevertheless, our analysis also shows that unobserved heterogeneity plays an important role in obesity, probably reflecting its inheritable nature (Stunkard et al, 1986 and Allison et al., 1996).

With regard to genuine state dependence, our results imply that state dependence is overestimated when initial conditions are ignored. Furthermore, initial conditions do not seem to be random but instead they may constitute the result of pre-sample/historical factors. For example, sociological and clinical studies suggest that obese children are more likely to live with a single parent, which in turn could be associated with low food quality (Gable et al., 2000 and Gibson et al., 2007). Furthermore, growing up in families with many siblings could also be associated with poor food habits (Racelli and Belmont, 1975). Lastly, the utilized in this study variable of “average weight growth” in the pre-sample period constitutes apparently a good proxy for metabolism and past consumption habits.

IV. Conclusion

The objective of this paper was to explore the dynamics of obesity and to provide some plausible explanations of the observed obesity persistence. We utilized longitudinal data for white adult males from the NLSY79 (1985-1994) and appropriate dynamic discrete choice models. Obesity persistence is assumed to follow a Markov process which has been verified by the estimation of transition probability matrices and the calculation of appropriate mobility indices. Following this, we attempted to disentangle the sources of obesity persistence, i.e., unobserved heterogeneity and state dependence. Using appropriate econometric estimators we identified possible sources of unobserved heterogeneity and the genuine state dependence effect.

The obtained results show that obesity is systematically affected by observed factors such as education, employment, age, marital status, smoking behavior and residential choices. However, obesity persistence is endogenously determined and, according to our results, it is affected by unobserved heterogeneity in the process of generating obesity outcomes. This heterogeneity potentially encompasses the role of “inheritably known” factors which however are unobserved to the researcher, i.e., genes and genetic mechanisms. Controlling for these factors it is shown that genuine state dependence is also a significant contributor in obesity persistence.

From a policy perspective, the obtained results suggest that the implied long-term effect of obesity experiences is the result of “initial conditions” pertaining –among other things- to biological and family background endowments. These conditions lead to an obesity “trap”. That is, individuals with an obesity history can not easily escape from this status, which in a rational choice context could be the preferred one. Thus, policy interventions designed to reduce obesity prevalence should focus on altering the structure of incentives at early childhood regarding the costs and benefits of obesity at adulthood.

References

Allison, D. B., Kaprio J., Korkeila, M., Koskenvuo, M., Neale M.C., Hayakawa, K., 1996. The heritability of body mass index among an international sample of monozygotic twins reared apart. International Journal of Obesity 20 (6), 501-506.

Bao, Y., 2007. Effects of family structure and parental resources on child body weight. University of Illinois at Chicago, Ph.D. thesis.

Baum, C.L., 2007. The effects of race, ethnicity and age on obesity. Journal of Population Economics 20 (3), 687–705.

Baum, C.L., 2009. The effects of cigarette costs on BMI and obesity. Health Economics 18 (1), 3–19.

Baum, C.L., Ruhm C.J., 2009. Age, socioeconomic status and obesity growth. Journal of Health Economics 28(3), 635–648.

Becker, G., Murphy, K.M., 1988. A theory of rational addiction. Journal of Political Economy 96 (4), 675-701.

Cawley, J., 1999. Rational addiction, the consumption of calories, and body weight. University of Chicago, PhD. Thesis.

Cawley, J., 2004. The impact of obesity on wages. Journal of Human Resources 39 (2), 451-474.

Chou, S.Y., Grossman, M., Saffer, H., 2004. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. Journal of Health Economics 23(3), 565-87.

Conti, G., Heckman, J.J., Urzua, S., 2010. The Education-Health Gradient. American Economic Review 100 (2), 234-38.

Courtemanche, C., 2009. Rising cigarette prices and rising obesity: Coincidence or unintended consequence? Journal of Health Economics 28 (4), 781–798.

Gable, S., Lutz, S., 2000. Household, parent, and child contributions to childhood obesity. Family Relations 49, 293–300.

Gibson, L.Y., Byrne, S.M., Davis, E.A., Blair, E., Jacoby, P., Zubrick, SR., 2007. The role of family and maternal factors in childhood obesity. The Medical Journal of Australia 186 (11), 591-595.

Greene W.H., 2003. Econometric Analysis, 5th ed., Pearson Education Prentice-Hall, Upper Saddle River, NJ.

Heckman J.J., 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data With Econometric Applications. MIT Press, Cambridge, MA.

Huffman, S., Rizov M., 2007. Determinants of obesity in transition economies: The case of Russia. Economics and Human Biology 5(3), 379-391.

Kano, S., 2008. Like husbands, like wife: A Bivariate Dynamic Probit Analysis of Spousal Obesities. Unpublished manuscript.

Komlos, J., Baur, M., 2004. From the tallest to (one of) the fattest: the enigmatic fate of the size of the American population in the twentieth century. Economics and Human Biology 2(1), 57–74.

Lakdawalla, D., Philipson, T., 2009. The growth of obesity and technological change. Economics and Human Biology 7(3), 283–293.

Levine, P., Gustafson, T., Velenchik, A., 1997. More Bad News for Smokers? The Effects of Cigarette Smoking for Labor Market Outcomes. Industrial and Labor Relations Review 50(3), 493-509.

Morris, S., 2007. The impact of obesity on employment. Labour Economics 14(3), 413–433.
Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46(1), 69-85.

Nakamura, R., Siciliani, L., 2010. Education and Body Mass Index: Evidence from ECHP. Discussion paper 10/04, Department of Economics, University of York.

Ng, S.W., Norton, E.C., Guilkey D.K., Popkin B.M., 2010. Estimation of a dynamic model of weight. National Bureau of Economic Research Working Paper 15864.

Philipson,T., 2001. The world-wide growth in obesity: an economic research agenda. Health Economics 10(1), 1–7.

Racelli, G.P., Belmont, L., 1975. Obesity in nineteen year old men: family size and birth order associations. American Journal of Epidemiology 109(1), 66–77.

Rashad, I., 2006. Structural Estimation of Caloric Intake, Exercise, Smoking, and Obesity. Quarterly Review of Economics and Finance 46(2), 268-83.

Rashad, I., Grossman, M., Chou S.Y., 2006. The Super Size of America: An Economic Estimation of Body Mass Index and Obesity in Adults. Eastern Economic Journal 32(1), 133-148.

Rooth, D.O., 2009. Obesity, attractiveness and differential treatment in hiring - a field experiment. Journal of Human Resources 44(3), 710-735.

Shorrocks, A.F., 1978. The measurement of mobility. Econometrica 46(5), 1013–1024.

Sobal. J., Rauschenbach, B.S., Frongillo, E.A., 1992. Marital status, fatness and obesity. Social Science and Medicine 35(7), 915–23.

Stunkard, A. J., Foch, T. T., Hrubec, Z., 1986. A twin study of human obesity. Journal of American Medical Association 256(1), 51-54.

Webbink, D., Martin, N.G., Visscher, P.M., 2010. Does education reduce the probability of being overweight? Journal of Health Economics 29 (1), 29-38.

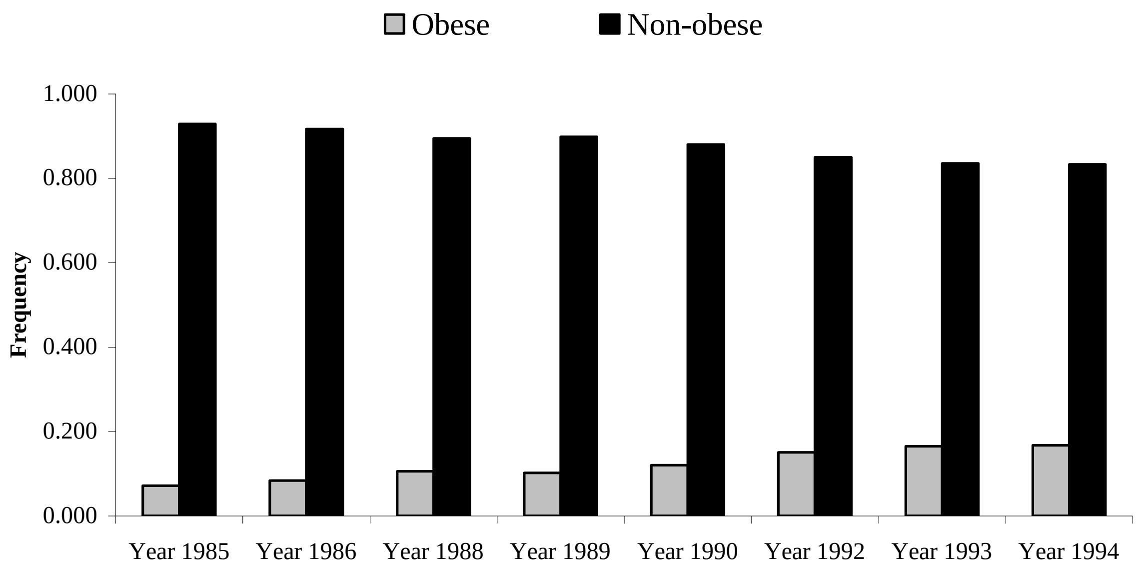
WHO, 2000. Obesity: Preventing and managing the global epidemic: report of a WHO consultation on obesity. World Health Organization, Geneva.

Williamson, D.F., Madans, J., Anda, R.F., Kleinman, J.C., Giovino, G.A., Byers T., 1991. Smoking cessation and severity of weight gain in a national cohort. New England Journal of Medicine 324, 739–745.

Wooldridge, M.J., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics 20(1), 39-54.

Figures

Figure 1. Evolution of obesity prevalence (1985-1994)



Source: NLSY79. Balanced panel (1985-1994) of white males with non-missing observations on BMI. Sample weights provided by the NLSY79 were applied.

Tables

Table 1. Observed sequences of obesity status

Sequence	Freq.	Sequence	Freq.	Sequence	Freq.	Sequence	Freq.	Sequence	Freq.
0000000	11688	00000111	224	00001111	128	11100001	0	00111111	152
subtotal	***11688***	00001011	48	00010111	0	11100010	0	01011111	24
00000001	216	00001101	16	00011011	8	11100100	0	01101111	8
00000010	96	00001110	16	00011101	8	11101000	0	01110111	32
00000100	88	00010011	8	00011110	8	11110000	8	01111011	0
00001000	64	00010101	0	00100111	48	***subtotal***	***296***	01111101	16
00010000	16	00010110	8	00101011	0	00011111	88	01111110	0
00100000	48	00011001	0	00101101	0	00101111	48	10011111	16
01000000	16	00011010	8	00101110	0	00110111	16	10101111	0
10000000	16	00011100	0	00110011	0	00111011	16	10110111	0
subtotal	***560***	00100011	0	00110101	0	00111101	0	10111011	0
00000011	160	00100101	0	00110110	0	00111110	40	10111101	0
00000101	48	00100110	24	00111001	0	01001111	0	10111110	0
00000110	56	00101001	0	00111010	16	01010111	8	11001111	16
00001001	8	00101010	0	00111100	0	01011011	8	11010111	0
00001010	8	00101100	0	01000111	24	01011101	0	11011011	0
00001100	32	00110001	24	01001011	8	01011110	0	11011101	0
00010001	16	00110010	0	01001101	0	01100111	0	11011110	0
00010010	0	00110100	0	01001110	0	01101011	0	11100111	8
00010100	0	00111000	0	01010011	8	01101101	0	11101011	0
00011000	8	01000011	0	01010101	0	01101110	16	11101101	0
00100001	8	01000101	8	01010110	0	01110011	8	11101110	0
00100010	0	01000110	0	01011001	0	01110101	8	11110011	0
00100100	0	01001001	0	01011010	0	01110110	0	11110101	8
00101000	8	01001010	0	01011100	0	01111001	0	11110110	0
00110000	0	01001100	0	01100011	0	01111010	0	11111001	8
01000001	0	01010001	0	01100101	0	01111100	0	11111010	0
01000010	0	01010010	0	01100110	0	10001111	16	11111100	8
01000100	0	01010100	0	01101001	0	10010111	0	***subtotal***	***296***
01001000	0	01011000	0	01101010	0	10011011	0	01111111	144
01010000	0	01100001	0	01101100	0	10011101	0	10111111	48
01100000	8	01100010	8	01110001	0	10011110	0	11011111	24
10000001	0	01100100	0	01110010	0	10100111	24	11101111	48
10000010	0	01101000	0	01110100	0	10101011	0	11110111	8
10000100	8	01110000	16	01111000	0	10101101	0	11111011	16
10001000	0	10000011	8	10000111	8	10101110	0	11111101	0
10010000	0	10000101	0	10001011	0	10110011	0	11111110	16
10100000	8	10000110	8	10001101	0	10110101	0	***subtotal***	***304***
11000000	16	10001001	0	10001110	0	10110110	0	11111111	600
subtotal	***392***	10001010	0	10010011	0	10111001	0	***subtotal***	***600***
		10001100	0	10010101	0	10111010	0		
		10010001	0	10010110	0	10111100	0		
		10010010	0	10011001	0	11000111	16		
		10010100	0	10011010	0	11001011	8		
		10011000	0	10011100	0	11001101	0		
		10100001	0	10100011	8	11001110	0		
		10100010	0	10100101	0	11010011	0		
		10100100	0	10100110	0	11010101	0		
		10101000	8	10101001	0	11010110	0		
		10110000	8	10101010	0	11011001	0		
		11000001	8	10101100	0	11011010	0		
		11000010	0	10110001	0	11011100	0		
		11000100	0	10110010	0	11100011	0		
		11001000	0	10110100	0	11100101	0		
		11010000	8	10111000	0	11100110	0		
		11100000	8	11000011	0	11101001	0		
		subtotal	***464***	11000101	0	11101010	0		
				11000110	8	11101100	0		
				11001001	0	11110001	0		
				11001010	0	11110010	16		
				11001100	0	11110100	0		
				11010001	0	11111000	0		
				11010010	8	***subtotal***	***336***		
				11010100	0				
				11011000	0				

Source: NLSY79. Balanced panel (1985-1994) of white males with non-missing observations on BMI.
Notes: In the "Sequence" columns the value of 1 corresponds to those individuals observed in the obese group in the t^{th} wave while the value of 0 to those observed in the non-obese group in the t^{th} wave.

Table 2. Transition probabilities and mobility measures of obesity dynamics

		Panel A			
		Non-obese	*Obese*	*PS*	*IR*
			1986		
1985	Non-obese	0.974	0.026		
	Obese	0.176	0.824		
	Obs.	1867		0.202	0.899
			1988		
1986	Non-obese	0.960	0.040		
	Obese	0.175	0.825		
	Obs.	1867		0.216	0.892
			1989		
1988	Non-obese	0.979	0.021		
	Obese	0.214	0.786		
	Obs.	1867		0.235	0.882
			1990		
1989	Non-obese	0.962	0.038		
	Obese	0.153	0.847		
	Obs.	1867		0.191	0.905
			1992		
1990	Non-obese	0.948	0.052		
	Obese	0.139	0.861		
	Obs.	1867		0.191	0.904
			1993		
1992	Non-obese	0.963	0.037		
	Obese	0.112	0.888		
	Obs.	1867		0.149	0.926
			1994		
1993	Non-obese	0.968	0.032		
	Obese	0.147	0.853		
	Obs.	1867		0.179	0.910
		Panel B			
			1994		
		Non-obese	*Obese*	*PS*	*IR*
1985	Non-obese	0.885	0.115		
	Obese	0.145	0.855		
	Obs.	1867		0.260	0.870

Source: NLSY79. Balanced panel (1985-1994) of white males with non-missing observations on BMI.
PS : Prais-Shorocks Index
IR : Immobility Ratio

Table 3. Descriptive statistics

	Full sample	Continuously Obese	Continuously Non-Obese	Single transition from obesity	Single transition to obesity	Multiple transitions
	[1]	[2]	[3]	[4]	[5]	[6]
Obesity (0/1)	0.134					
	(0.341)	-	-	-	-	-
Highest grade completed	13.254	12.517	13.432	12.507	12.783	12.684
	(2.390)	(1.691)	(2.470)	(1.450)	(2.178)	(2.019)
Employment status (0/1)	0.941	0.941	0.944	0.955	0.936	0.927
	(0.235)	(0.236)	(0.231)	(0.209)	(0.244)	(0.260)
Age	28.698	28.958	28.656	30.244	28.558	28.897
	(3.921)	(3.697)	(3.932)	(4.101)	(3.907)	(3.908)
Married (0/1)	0.542	0.529	0.535	0.509	0.533	0.614
	(0.498)	(0.500)	(0.499)	(0.504)	(0.499)	(0.487)
Other marital statuses (0/1)	0.458	0.471	0.465	0.491	0.467	0.386
	(0.498)	(0.500)	(0.499)	(0.504)	(0.499)	(0.487)
Permanent smokers (0/1)	0.213	0.106	0.235	0.000	0.119	0.183
	(0.409)	(0.308)	(0.424)	(0.000)	(0.324)	(0.387)
Non-smokers (0/1)	0.600	0.672	0.587	0.583	0.611	0.653
	(0.490)	(0.470)	(0.492)	(0.497)	(0.488)	(0.476)
Occasional smokers (0/1)	0.187	0.223	0.178	0.417	0.270	0.164
	(0.390)	(0.416)	(0.383)	(0.497)	(0.444)	(0.371)
North-East (0/1)	0.200	0.232	0.195	0.097	0.168	0.251
	(0.400)	(0.423)	(0.396)	(0.298)	(0.374)	(0.433)
North-Central (0/1)	0.370	0.469	0.365	0.278	0.360	0.370
	(0.483)	(0.500)	(0.481)	(0.451)	(0.480)	(0.483)
South (0/1)	0.271	0.200	0.275	0.397	0.263	0.268
	(0.444)	(0.400)	(0.447)	(0.493)	(0.441)	(0.443)
West (0/1)	0.160	0.098	0.165	0.229	0.209	0.111
	(0.366)	(0.298)	(0.371)	(0.423)	(0.407)	(0.314)
Urban (0/1)	0.761	0.626	0.779	0.744	0.681	0.750
	(0.426)	(0.484)	(0.415)	(0.440)	(0.466)	(0.433)
Rural (0/1)	0.239	0.374	0.221	0.256	0.319	0.250
	(0.426)	(0.484)	(0.415)	(0.440)	(0.466)	(0.433)
Pre-sample variables						
Living with parents at 14 years old (0/1)	0.815	0.699	0.827	1.000	0.843	0.754
	(0.388)	(0.459)	(0.379)	(0.000)	(0.364)	(0.431)
Number of siblings	3.037	2.650	3.058	2.517	3.169	2.987
	(2.015)	(1.779)	(2.033)	(1.619)	(1.968)	(2.019)
Average weight growth † (0/1)	3.013	5.856	2.696	5.189	2.807	4.084
	(4.547)	(5.682)	(4.141)	(4.888)	(5.832)	(5.049)
Number of waves in the obesity status						
Zero	76.55%	-	100%	-	-	-
One	3.88%	-	-	25%	18.92%	21.62%
Two	2.47%	-	-	12.5%	12.61%	13.51%
Three	3.18%	-	-	12.5%	18.92%	15.54%
Four	2.19%	-	-	12.5%	11.71%	11.49%
Five	2.54%	-	-	0%	8.11%	18.24%
Six	2.19%	-	-	12.5%	14.41%	9.46%
Seven	2.4%	-	-	25%	15.32%	10.14%
Eight	4.59%	100%	-	-	-	-
Obs. in every wave	1416	65	1084	8	111	148

Source: NLSY79. Balanced panel (1985-1994) of white males with non-missing observations on any of the variables. Sample weights provided by the NLSY79 were applied.
Notes: Standard deviations are in parentheses.
† Average of the percentage change of body weight in two-periods: 1981-1982 and 1982-1985.

Table 4. Estimated coefficients and marginal effects: pooled and dynamic random effects probit models

	Pooled		DRE	
	Coefficients	*Marginal effects*	*Coefficients*	*Marginal effects*
Obesity at t-1	2.771 (0.069) ***	0.800 (0.014) ***	1.233 (0.109) ***	0.119 (0.020) ***
Highest grade completed	-0.056 (0.010) ***	-0.008 (0.001) ***	-0.213 (0.036) ***	-0.016 (0.003) ***
Employment status	-0.162 (0.081) **	-0.025 (0.014) **	-0.298 (0.157)*	-0.025 (0.014)*
Age	0.010 (0.006) *	0.0014 (0.0008) *	0.106 (0.014) ***	0.008 (0.001) ***
Married	0.089 (0.046) *	0.012 (0.006) *	0.310 (0.099) ***	0.024 (0.007) ***
Permanent smokers	-0.279 (0.065) ***	-0.035 (0.007) ***	-0.982 (0.222) ***	-0.065 (0.012) ***
Occasional smokers	-0.038 (0.059)	-0.005 (0.008)	-0.129 (0.207)	-0.010 (0.015)
Northeast	0.025 (0.060)	0.004 (0.009)	0.107 (0.199)	0.008 (0.016)
South	-0.126 (0.056) **	-0.017 (0.007) **	-0.413 (0.181) **	-0.030 (0.013) **
West	-0.025 (0.069)	-0.003 (0.010)	-0.130 (0.211)	-0.010 (0.015)
Urban	-0.120 (0.051) ***	-0.018 (0.008) **	-0.256 (0.132) *	-0.020 (0.011)*
Constant	-1.050 (0.229) ***	-	-2.725 (0.612)***	-
λ	-		0.821 (0.025) ***	
θ	-		-	
Log-Likelihood	-1893.003		-1843.903	
Quasi-likelihood ratio test	-		98.20 ***	
Obs	9912		9912	

Source: NLSY79. Balanced panel (1985-1994) of white males with non-missing observations on any of the variables. Sample weights provided by the NLSY79 were applied.
Notes: Cluster robust standard errors in parentheses. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10%, respectively.
Pooled: Dynamic pooled probit model.
DRE: Dynamic random effects probit model.

Table 5. Estimated coefficients and marginal effects: modeling "initial conditions"

	DRE-IC-W		DRE-IC-H	
	Coefficients	*Marginal effects*	*Coefficients*	*Marginal effects*
Obesity at t-1	0.925 (0.095) ***	0.086 (0.016) ***	0.991 (0.095) ***	0.084 (0.014) ***
Obesity at t=1	4.473 (0.359) ***	0.670 (0.045) ***	-	-
Highest grade completed	-0.107 (0.033) ***	-0.012 (0.002) ***	-0.208 (0.039) ***	-0.015 (0.003) ***
Employment status	-0.256 (0.171)	-0.021 (0.015)	-0.317 (0.149) **	-0.024 (0.012) **
Age	0.133 (0.014) ***	0.010 (0.001) ***	0.127 (0.014) ***	0.009 (0.001) ***
Married	0.430 (0.114) ***	0.032 (0.008) ***	0.345 (0.103) ***	0.024 (0.006) ***
Permanent smokers	-0.931 (0.208) ***	-0.059 (0.011) ***	-1.055 *** (0.230)	-0.065 (0.012) ***
Occasional smokers	-0.089 (0.189)	-0.006 (0.014)	-0.194 (0.204)	-0.013 (0.014)
Northeast	0.014 (0.474)	0.001 (0.036)	0.105 (0.197)	0.008 (0.014)
South	-0.407 (0.372)	-0.029 (0.025)	-0.341 (0.206) *	-0.023 (0.013) *
West	0.068 (0.418)	0.005 (0.032)	0.111 (0.201)	0.008 (0.015)
Urban	-0.022 (0.176)	-0.002 (0.013)	-0.190 (0.128)	-0.014 (0.010)
Constant	-0.381 (1.116)		-3.719 (0.723) ***	
λ	0.767 (0.026) ***		0.850 (0.016) ***	
θ	-		1.031 (0.144) ***	
Log-Likelihood	-1651.378		-2000.236	
Quasi-likelihood ratio test	326.21 ***		345.674 ***	
Obs	9912		9912	

Source: NLSY79. Balanced panel (1985-1994) of white males with non-missing observations on any of the variables. Sample weights provided by the NLSY79 were applied.
Notes: Cluster robust standard errors in parentheses. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10%, respectively.
DRE-IC-W: Dynamic random effects probit model with initial conditions. The model also includes the mean values of the time-invariant independent variables. Their estimated coefficients and the correspondent marginal effects are not reported and are available from the authors upon request.
DRE-IC-H: Dynamic random effects probit model with a flexible reduced form specification for "initial conditions" (Heckman, 1981).

Table 6. Modeling initial period obesity status: DRE-IC-H model specification

Highest grade completed	-0.192
	(0.069) ***
Employment status	-0.902
	(0.324) ***
Age	0353
	(0.067) ***
Married	0.025
	(0.240)
Permanent smokers	-0724
	(0.337) **
Occasional smokers	-0.383
	(0.311)
Northeast	0.253
	(0.313)
South	-0.041
	(0.312)
West	-0.160
	(0.354)
Urban	-0.078
	(0.242)
Living with parents at 14 years old	-0.443
	(0.245) *
Number of siblings	-0.115
	(0.058) **
Average weight growth	0.102
	(0.022) ***
Constant	-8.570
	(1.911) ***
Obs.	1416

Source: NLSY79. The model has been estimated using the first survey year observations (1985) of the balanced panel (1985-1994) for white males with non-missing observations on any of the variables. Sample weights provided by the NLSY79 were applied.
Notes: Cluster robust standard errors in parentheses. Asterisks ***, ** and * indicate statistical significance at the 1%, 5% and 10%, respectively.
DRE-IC-H: Dynamic random effects probit model with a flexible reduced form specification for "initial conditions" (Heckman, 1981).