On the Nexus between Economic and Obesity Crisis in Spain: Parametric and Nonparametric Analysis of the Role of Economic Factors on Obesity Prevalence

Amr Radwan* and José M. Gil

Centre de Recerca en Economia i Desenvolupament Agroalimentaris (CREDA)-UPC-IRTA, Barcelona, Spain.

Contributed Paper prepared for presentation at the 88th Annual Conference of the Agricultural Economics Society, AgroParisTech, Paris, France

9 - 11 April 2014

*[Corresponding author] (CREDA-UPC-IRTA, Parc Mediterrani de la Tecnologia Edifici ESAB, C/Esteve Terrades, 8 08860 Castelldefels, Barcelona, Spain. e-mail: Amr.radwan@upc.edu)

Abstract

Poor diets and rising obesity rates dominate the current food, nutrition and health policy debate in many countries, including Spain. Despite the increasing obesity rate in Spain, there has been no known published research in Spain that has studied the economic factors affecting obesity prevalence. The main aim of our work is analysing the relevance of economic factors in obesity prevalence in Spain. This aim is especially relevant in shadow of the Economic crisis that hit Spain recently. Our methodological approach depending upon the estimation of a multinomial Logit Regression (MLR) Combined with a nonparametric model, the Multivariate Adaptive Regression Splines (MARS), to judge the role of different socioeconomic factors on the obesity prevalence. Despite the desirable advantages of using nonparametric models such as MARS, our paper is the first attempt to use this type of models to analyse the determinant factors of obesity prevalence. Our results suggest that Socio Economic factors seem to have a significant impact on obesity prevalence. MARS models outperform the traditional MLR and could be a helpful tool for understanding the nature of the relationship and moreover it could be helpful as pre estimate guides the estimation of its parametric counterpart.

Keywords: Obesity, Multinomial Logit, Multivariate Adaptive Regression Splines (MARS), Spain.

JEL code: I10; I18; Q11

1. Introduction

Obesity is partly a result of an energy imbalance caused by consumption of too many calories and/or low expenditure of calories (i.e., low physical activity) over a considerable period. Consequently, most published economic research has examined the increased growth of obesity rates by analyzing several factors that may contribute to this imbalance of caloric consumption and usage (Myton et al., 2012; Lin et al.,2011; Cutler et al., 2003; Chou et al., 2004; Lakdawalla and Philipson, 2002; Loureiro and Nayga, 2005; among others).

Due to rising concerns about obesity, the availability, accessibility and choice of foods to meet an adequate diet are becoming key challenges to our food system today. Good nutrition is essential to obtain optimum health and productivity and in reducing the risk of chronic and infectious diseases. Understanding factors influencing food consumption and obesity is needed to gain a clearer picture of the mechanisms that would cause individuals to eat unhealthful or become over weighted Especially, as it is observed that consumers tend to overeat despite quite obvious future health implications. Hence, knowledge about how people make food choices and how economic and non-economic factors influence food consumption and obesity is critically important to improve policy interventions and developing agricultural and food programs that can assure a safe, affordable, reliable and nutritious food supply and promote health.

Previous economic studies have analyzed the influence of income on health. In general, there seems to be a consensus about the positive effect of income on health (Smith, 1999). Consequently, we would expect, all things being equal, a negative effect on obesity due to: 1) the unavailability of healthy food in low income neighborhoods (Beaulac et al., 2009; Larson et al., 2009); 2) when healthy food is available it is usually more expensive, so poor families try to stretch their food budget by purchasing unhealthy cheap food (Drewnowski, 2010; Monsivais & Drewnowski, 2009); 3) Lower income neighborhoods have fewer physical activity resources making it difficult to lead a physically active lifestyle (Moore et al., 2008); 4) Low income families are more expected to face high levels of stress increasing the likelihood of being overweight or obese (Gundersen et al., 2011; Moore & Cunningham, 2012).

Apart from income, food prices also play an important role in obesity prevalence. Powell and Chaloupka (2009) in their literature review concluded that food prices had a significant but small effect on obesity and overweight prevalence concluding that fiscal pricing policies could help in reversing obesity trends and that small taxes or subsidies were not able to do so while nontrivial pricing interventions might have measurable effects on weight outcomes, especially for those belonging to low social and economic class.

Beside income and food prices, several recent economic studies explain the role played by different cultural and socio-demographic factors on obesity rates. Leaving genetics aside, obesity is caused by consumption of too much calories and/or low expenditures of calories (i.e. low physical activity). For example, Schlosser (2002) showed that the rapid growth of fast food and soda drinks consumption has increased the dietary intake of saturated fats, sugars, and calories and accordingly, the prevalence of obesity. Other researchers argue that female labor participation is a leading factor in increasing obesity rates (Garcia et al., 2006), mainly in childhood.

Most of the literature in Spain has concentrated on the adequacy of alternative instruments to measure obesity or on educational and environmental factors (i.e. food consumption) affecting obesity. However, limited attention has been paid to the role of economic factors (income and prices) on food choices, physical activity and, consequently, on the prevalence of obesity.

This study aims at analyzing the relevance of economic factors (mainly income and other socioeconomic characteristics of Spanish households and market prices) on the prevalence of obesity in Spain, which is one of the main contributions of this study.

Data come from the 2011-2012 National Health Survey (NHS) (INE, 2013). Information on most relevant variables is available at household level except for prices. Categorical body mass index (CBMI) has been calculated using the reported weight and height collected from participants. BMI has been used as a categorical variable to reduce the potential bias in BMI estimates as it is not measured but self-reported (Gil, 2011). The database has been extended by considering food at-home and Out-of-Home prices.

From a methodological point of view, this paper compares the results obtained from the use of parametric and nonparametric models to tackle this issue. While previous literature has focused on parametric methods, such as MLR, this study also has considered the estimation of a Multivariate Adaptive Regression Splines (MARS) model, which is flexible enough to provide more insight on how covariates interact with the prevalence of obesity.

Results from the MARS model will be used to increase the goodness-of-fit of the traditional parametric approach by allowing nonlinear covariates. The combined model clearly outperforms parametric approach while being easier to interpret than MARS. This is the second contribution of this paper.

To achieve the paper' objective, the rest of the paper is organized as follows. Section 2 provides a brief description on the obesity prevalence in Spain. The methodological approach applied in our analysis is explained in section 3. Our empirical application and the main results are discussed in sections 4 and 5, respectively. Finally, the paper ends with some concluding remarks.

2. The prevalence of overweight and obesity in Spain

Obesity is considered as a complex, multi factorial, chronic disease involving genetic, prenatal, socioeconomic, and environmental components. Worldwide prevalence of obesity nearly doubled between 1980 and 2008. In Europe, the regional office of the World Health Organization (WHO) and according to country estimates for 2008, over 50% of both men and women were overweight, and roughly 23% of women and 20% of men were obese. These estimates indicate that obesity prevalence in Europe in the last two decades has tripled affecting more than 150 million adults and 15 million children and adolescents in the region.

In Spain, the last National Health Survey (NHS) for 2011-12 (INE, 2013) indicated that the prevalence of overweight and obesity among Spanish adults aged 18 years old or more was 36.7% and 17.0% respectively. While the prevalence of obesity was quite similar between men and women (18% among men and 16% among women), the overweight prevalence was significantly higher among men (45%) than among women (28%). Obesity is associated with age as in the age segment between 18 and 24 years old its prevalence merely reaches 5.5% for both genders. On the contrary, in the segment between 65 and 74 years old the prevalence of obesity reaches 25.6% and 27.9% among men and women, respectively, although it is true that it decreases for the eldest segments. There is also a significant negative relationship between education level and obesity. In fact, the highest percentage (30.0%) is found among illiterate persons.

From a historical perspective, it is worth mentioning that, in spite of the up to now relative low percentages in relation to other EU countries, the prevalence of obesity in Spain has increased with a very alarming rate in 25 years increasing from 6.9% and 7.9% among men and women, respectively, in 1987, to the above mentioned 18% and 16%, in 2012.

The Spanish National Survey of Dietary Intake (ENIDE) (2011), concluded that obesity rates in Spain was not due to eat too much (daily energy intake was 2482 kcal, slightly lower than the recommended level between 2550 and 2600 calories, depending on the individual's physical activity), but to an unbalanced diet characterized by the overconsumption of red meat, sodas and pastries. According to the Spanish Food Safety Agency (AESA), food habits have changed with a significant reduction of both family meals and the time allocated to eat during weekdays.

Noticeably, the prevalence of obesity has increased during the financial crisis that started to affect Spanish households in 2009 and more intensively during 2010. Comparing the data from the last two National Health Surveys (2006 and 2012), the obesity rate significantly increased from 15.6% to 18%, among males and a lower increment observed among women (15.2% and 16%, in 2006 and 2012, respectively). This result has to do with a lower consumption of fresh foods, fruits and vegetables and a higher consumption of fast food, ready-to-eat meals and fatty foods, which have been relatively much cheaper (Rao et al., 2013). This situation seems to continue in the future as the OECD predicted that the number of overweight and obese people in Spain will rise by a further 10 per cent over the next decade.

An alarming 30 per cent of teenagers are overweight, putting Spain just behind USA and Scotland for obesity. Moreover, a stunning 40 per cent of youths aged between 13 and 18 never practice sport.

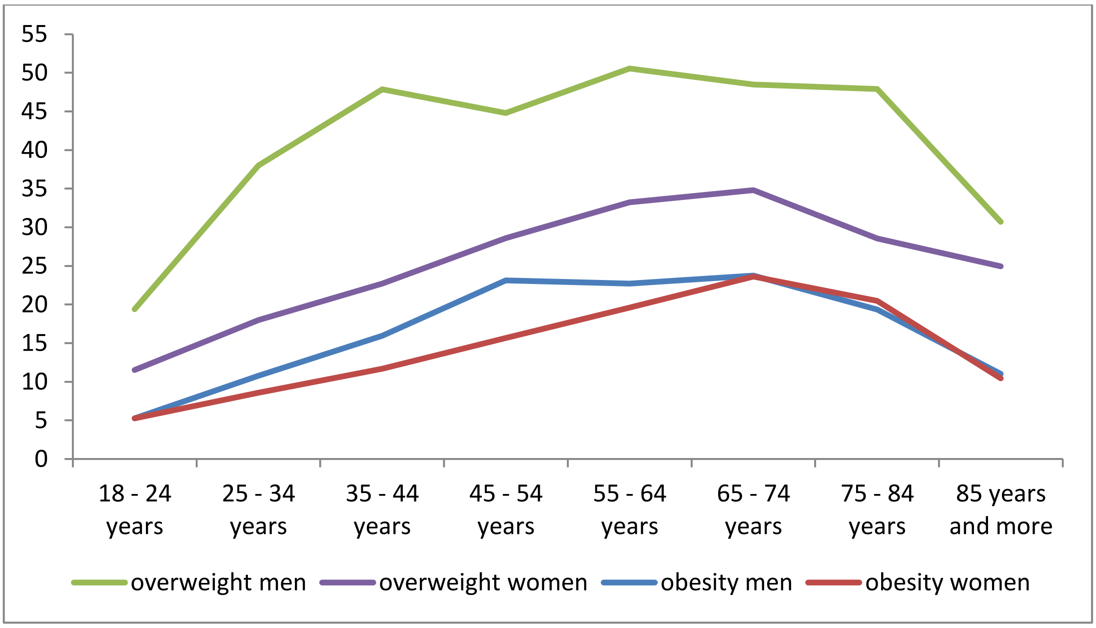
Although the WHO characterizes overweight and obesity as diseases, it is also well known that both (together with smoking) are key determinants in the incidence of the most important contemporary chronic diseases, such as cancer, cardiovascular problems, certain types of diabetes, etc. What is most worrying is that health disorders that were once almost exclusively associated with the elderly, such as type II diabetes, are now being diagnosed in children, mainly due to the increasing prevalence of childhood obesity. In fact, one in every five adolescents in Spain now runs the risk of suffering major cardiovascular problems in later life (Mora et al., 2012).

The economic costs associated with obesity are non-trivial as well. Obesity accounts for 7% of total health care costs (WHO, 2005) without considering other economic externalities, which, on the other hand, are difficult to estimate. In Spain, the Spanish Society for the Study of Obesity (SEEDO) estimates that direct and indirect obesity costs account for 7% of total health care costs (2.5 billion Euros/year).

3. The Multivariate Adaptive Regression Splines model (MARS)

Analyzing main determinants affecting the prevalence of overweight and obesity is not an easy task for two main reasons: 1) this is a complex phenomenon in which a large number of covariates could be considered; and 2) studies have shown that obesity response to socioeconomic covariates is frequently characterized by thresholds requiring flexible response functions (Cavaliere and Banterle, 2008). Moreover, the complexity of interactions between different socioeconomic covariates requires flexible multivariate models capable of dealing with the different ways covariates interact with the dependent variable. Let us take the age as an example. Figure 1 shows the relationship between age and the prevalence of overweight and obesity among men and women. As can be observed the relationship is not linear. Moreover, this nonlinear relationship differs between men and women underling an interaction between gender and age.

Figure 1 Overweight and obesity prevalence rates (%) by age group and gender



Source: National Health Survey (2011-2012)

The traditional approach to deal with the issue addressed in this paper has been the Multinomial Logistic Regression (MLR). Like any other classical parametric regression methods, MLR is ‘global’ in nature; that is, when a covariate enters the model all values of such covariate are considered to be relevant in explaining the variation of the dependent variable. This is not the case in reality in which the relationship is true only for certain values of the covariate. Also in the MLR, missing data are either dropped or replaced by mean values reducing the model performance.

The Multivariate Adaptive Regression Splines (MARS), first introduced as a data mining tool (Friedman, 1991), is able to address the above limitations of MLR and other classical regression methods. MARS is a nonparametric method hence, it is expected to perform as well as, or even better than, the classical regression techniques when distributional assumptions are not satisfied and it allows local models and thus a more accurate function approximation. MARS is not affected by any volume of missing data since it automatically introduces indicator functions for every variable that contains missing values. Furthermore, this method is designed to capture higher-order interactions, even in high-dimensional settings. But unlike other available nonparametric methods that can capture complex relationships among the variables such as the Classification and Regression Tree (CART) or Artificial Neural Networks (ANNs), MARS produces very simple and easy-to-interpret models.

MARS performance depends on data structure (Ture et al., 2005) but is generally known for predictive accuracy, computational speed and simplicity of interpretation. Leathwick et al. (2006) compared General Additive Models (GAM) and MARS models and highlighted the advantages of MARS in cases involving large data sets. MARS models are also parsimonious and provide more extensive predictions. Muñoz and Fellicisimo (2004) used two different ecological data sets to compare MARS over other modeling techniques such as MLR, Principal Component Regression and CART observing that MARS performed consistently well. Using a motor vehicle injury data consisting of 59 cases and 689 controls and with up to 3% missing values for some of the variables, Kuhnert et al. (2000) showed that MARS outperformed CART and MLR, in terms of accuracy and flexibility as a modeling tool. Haughton and Loan (2004) compared different statistical techniques to model vulnerability from a panel of 4,272 households. They showed that MARS, together with CART, were the most parsimonious model and were able to capture nonlinearities and interaction effects.

In MARS the data are left to reveal the variable knot locations while the user need not to input any specification into the model. The basis functions in MARS, which serve as independent variables, are truncated linear functions, which address the problem of discontinuity of recursive partitioning algorithms. Dissimilar with additive models, MARS allows interactions up to an order specified by the user, and trades off the interaction order and complexity of the additive functions and interactions (Frank, 1995; De Veaux et al., 1993).

MARS models nonlinear relationships through two ways: 1) each variable may have a cut-off point. In other words, the effect of each variable may vary across its range of values; and 2) there may be interaction effects between different variables which are significant even if the effect of the two individual variables is not.

Although MARS was initially suggested by Friedman (1991), Kooperberg et al. (1997) developed this approach especially for categorical covariates with some enhancement for continuous responses. MARS could be viewed as a generalization of the repeated discriminate method and the stepwise linear regression to improve the performance of a covariate set. The procedure first divides data into locales and then forms a regression equation for each one. Each obtained linear region is called “knot”.

Being y the dependent variable, which can be continuous or categorical, and $X = (X_1 \dots X_p)$ the set of potential predictive covariates and assuming that data are generated from an unknown model. In case of a continuous response this leads to:

$$
\begin{aligned}
y &= f(X1, ..., Xp) + \varepsilon \\
&= f(X) + \varepsilon
\end{aligned} \tag{1}
$$

The algorithm deals with this case as a classification problem so it determines a common set of basis functions in the predictors while estimates different coefficients for each dependent variable. This method seems quite similar to some neural networks where multiple outcome variables are predicted from common basis functions with different coefficients.

MARS allows covariates to enter in the model as a single variable or interacting with other covariates generating unbiased parameter estimations with strong algorithms. As the error is a member of exponential family and by introducing second order interactions leads to the following model.

$$f(x) = g_0 + \sum_{j1} g_{j1}(X_{j1}) + \sum_{j1 < j2} g_{j1,j2}(X_{j1}, X_{j2}) + \varepsilon_i \tag{2}$$

Linear splines and their tensor products are used to model the function g (·). A one-dimensional spline can be written as:

$$g(X) = b_{-1} + b_0 X + \sum_{k=1}^{K} b_k (X - t_k)_+ \tag{3}$$

An important characteristic of MARS model is the use of the so called Hinge functions or hockeystick-function which takes the form:

$$
(X - t_k)_+ = \begin{cases} X - t_k, & \text{if } X \geq t_k, \\ 0, & \text{else} \end{cases} \tag{4}
$$

where t_k is a constant called knot. Not only piecewise linear functions can be formed from hinge functions, but it can be multiplied together to form nonlinear functions.

The interaction $g_{j1,j2}$ can be modeled in the form:

$$\mathrm{g}_{12}(X_1, X_2) = \mathrm{g}_1(X_1) \times \mathrm{g}_2(X_2) \tag{5}$$

This resulted in the following model:

$$\begin{aligned}
&g_0 = \beta_0 \\
&g_{j1}(X_{j1}) = \sum_{j=1}^{M} \beta_i^{ji} B_i^{ji}(X_{ji}) \\
&g_{j1,j2}(X_{j1}, X_{j2}) = \sum_{j=1}^{M} \beta_i^{j1j2} B_i^{j1j2}(X_{j1}, X_{j2})
\end{aligned} \tag{6}$$

Considering M as the number of basis functions in the model, Bs the spline basis functions as described above and the βs are coefficients. The MARS model can be written as:

$$F(X) = \sum_{i=1}^{M} \beta_i B_i(X) \tag{7}$$

With the following possible types of basis functions

$$\begin{aligned}
&1 \\
&X_i \\
&(X_i - t_k)_+ \\
&X_i X_j \\
&(X_i - t_k)_+ X_j \\
&(X_i - t_k)_+ (X_j - t_l)_+
\end{aligned} \tag{8}$$

MARS is a stepwise process uses both forward and backward progresses for robust and unbiased parameter estimations. It starts by maximizing all possible effects of explanatory variables in the forward model and then removes the least effective functions in the backward model using Ordinary Least Squares method.

The main advantage of MARS comparing with other regressions such as logistic regression is that MARS is a data driven technique. Instead of fitting a single regression equation for the model, MARS get many piecewise regression equations which allow the researcher to obtain more consistent and unbiased estimates of the covariates.

The main principle of MARS is based on searching for every point where linearity breaks. Then this point is taken as a knot and predictive variables, which have an effect until that point, are modeled using a new regression equation. Then, the number of obtained regression equations is the same as the number of knots defined in the process.

MARS reaches the final model taking the obtained combination of basis functions into account (these functions are called Basis Functions) based on minimizing the Generalized Cross Validation (GCV) (Kayri, 2007). MARS uses GCV As measure for the degree of fit or lack of accuracy of the model to compare the performance of obtained models. Lower values of GCV are better.

$$\mathrm{GCV} = \frac{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left[y_i - \hat{f}_{M}(X_i)\right]^2}{\left[1 - \frac{d-M}{N}\right]^2} \tag{9}$$

with M being the number of basis functions in the model and $\hat{f}$ denotes the fitted values of the current MARS model. The numerator refers to the common residual sum of squares, which is penalized by the denominator, which accounts for the increasing variance as the model complexity increases. The penalizing parameter d can be chosen by the user, although the conventional value is d = 4. A smaller d generates a larger model with more basis functions; a larger d creates a smaller model with less basis functions. Thus, the GCV can be considered as a form of regularization by trading off between goodness-of-fit against model complexity. In MARS models we cannot use the

Residual Sum of Squares (RSS) for comparing models, because the RSS always increases as MARS terms are dropped, which means that if the RSS were used to compare models the backward step of model construction would always choose the largest model.

The main disadvantage of MARS is the low prediction power with insufficient sample size. This is not the case in our analysis as we have a quite big data set which consists of 19069 observations. Moreover, Briand et al. (2007) mentioned that the model may suffer from multicollinearities as MARS gets interaction between predictive variables involved in the model. Also, the MARS methodology has a risk of over fitting because of very exhaustive search that is conducted to identify nonlinearities and interactions. This drawback could be controlled through choosing the appropriate penalty term of the model.

4. Data and variable definition

Our data come from the 2011-2012 National Health Survey (NHS) (INE, 2013). The NHS is a cross-section survey that provides micro data on the health status of citizens and its determinants. It is carried out by the National Institute of Statistics (INE) in cooperation with the Spanish Ministry of Health, Social Services and Equality. The survey collects information on the individual socioeconomic characteristics, morbidity, food habits and the demand for health care of respondents. Food habits refer to two main issues: type of breakfast and frequency of consumption of selected food groups. However, the data set does not provide information on quantities consumed (or purchased) neither on prices. In order to take into account the effect of economic factors we augment the data set with regional average food prices (FAHP) (MAGRAMA, 2012), the regional price index for Food Out-of-Home (FOHP) relative to the Consumer

Price Index for Food (INE, 2012) and regional per capita expenditure as a proxy of household' disposable income (INE, 2012). Our sample consists of 19069 adults (18 years old or more). Table 1 show some descriptive statistics of the variables used in this study.

Table 1 Descriptive statistics of variables used in this study

Variable	Units	Mean	S. deviation
Categorical BMI (CBMI)	1= normal 2=overweighed 3=obese	1.72	0.74
Age	Years	50.18	18.41
Male	1=male 0=female	0.48	0.50
Food at Home Price (FAHP)	Regional average food prices (Euro)	2.25	0.18
Relative food out of home price(FOHP)	Consumer Price Index of Food Out-of-Home relative to the Global Consumer Price Index for Food by region	0.99	0.02
Income	Total per capita expenditure by region (Euro)	11101.51	1451.88
Physical exercise	1=doing intensive or moderate physical exercise 0=not doing	0.20	0.40
Perceived health status	1= perceive to have a good or very good health 0= perceive not to have a good health	0.70	0.46
Healthy breakfast	1= having breakfast at home 0= otherwise	0.86	0.35

To tackle with the objective of this paper, we have followed a three-step strategy. Firs, as a benchmark, we have estimated a MRL (Greene, 2003). As a second step, and trying to identify nonlinearities and covariate interactions, a multinomial MARS model has been estimated. MARS estimates will be used in the third step to improve the performance of the parametric MRL. In the following section, we will present the results from each of the three steps.

5. Results

Results from MLR model (Table 3) suggest that all variables except Food Out-of-Home Prices (FOHP) have a significant marginal effect on the probability of being obese. The effect of gender is positive; indicating that being male increases the likelihood to be obese, while the effects of perceived health status and physical exercise are negative. In the case of overweight, gender and age (doing sufficient physical exercise) have a positive (negative) significant marginal effect on its prevalence. These results are quite consistent with those reported in the literature. As mentioned above, however, one of the main shortcomings of the MLR model is that it fails to take into account the potential non-linearity that could exist between the dependent variable and its covariates. For instance, Table 2 suggest that age has a significant positive marginal effect on the prevalence of both overweight and obesity for all age groups while Figure 1 showed that this prevalence decreased from 74 years old onwards. Of course it would be possible to graph each covariate in order to detect nonlinearities but these can be conditioned by other covariates entering into the regression and by interactions among them. The MARS model simplifies this task.

Table 2 Parameter estimates and marginal effects from Multinomial Logistic Regression.

Overweight				
	Estimates		Marginal effects	
Constant	-2.181**	(0.838)		
Age	0.029**	(0.001)	0.005**	(0.000)
Male	1.028**	(0.035)	0.190**	(0.007)
Food At Home Price (FAHP)	0.023	(0.147)	0.037	(0.032)
Food Out-of-Home Prices (FOHP)	0.923	(0.832)	0.148	(0.179)
Income	0.000**	(0.000)	0.000	(0.000)
Physical exercise	-0.365**	(0.043)	-0.040**	(0.010)
Perceived health status	-0.139**	(0.041)	0.010	(0.009)
Healthy breakfast	-0.148**	(0.050)	-0.006	(0.011)
Obesity				
Constant	-0.936	(1.028)		
Age	0.032**	(0.001)	0.003**	(0.000)
Male	0.782**	(0.045)	0.041**	(0.006)
Food At Home Price (FAHP)	-0.491**	(0.191)	-0.070**	(0.024)

Food Out-of-Home Prices (FOHP)	1.089	(1.030)	0.092	(0.130)
Income	0.000**	(0.000)	0.000**	(0.000)
Physical exercise	-0.848**	(0.066)	-0.084**	(0.007)
Perceived health status	-0.610**	(0.048)	-0.082**	(0.007)
Healthy breakfast	0.405**	(0.062)	-0.050**	(0.009)
R^2	0.074			
AIC	36491			

Note: Standard Error in parentheses
** denotes statistical significance at 5 per cent significance level.

In order to keep the model as simple as possible, interactions of order 2 and a penalty term equaling 4 have been chosen.[1] As it was explained earlier, in the backward step the best model is reached by minimizing the GCV. The optimal model consisted of 17 basis Functions.

Table 3 shows the basis functions estimates for the final MARS model and equations 10-12 represent the final model. The MARS model outperformed the MLR (adjusted R2= 0.16).

CBMI (Normal Weight) = 0.407 + 0.011 * max (0, 46 - Age) - 0.415 * max (0, Male (1) - 0) + 0.119 * max (0, Perceived health (1) - 0) + 0.108* max (0, Physical exercise (1) - 0) + 0.013 * max (0, Age - 72) * max (0, Male (1) - 0) + 0.003 * max (0, 72 – Age) * max (0, Male (1) - 0) + 0.000 * max (0, Income - 9027) + 0.114 * max (0, Male(1) - 0) * max (0, Perceived health (0)-0) + 0.062 * max (0, Male(1) - 0) * max (0, Healthy breakfast(1) - 0) – 0.019 * max(0, Age - 46) * max (0, 2.14 - FAHP) – 0.004 * max(0, Age - 46) * max(0, Male(0) – 0) (10)

CBMI (Overweight) = 0.323 - 0.007 * max (0, 46 - Age) + 0.269 * max (0, Male (1) - 0) - 0.014 * max (0, Perceived health (1) - 0) - 0.038* max (0, Physical exercise (1) - 0) - 0.005 * max (0, Age - 72) * max (0, Male (1) - 0) - 0.002 * max (0, 72 – Age) * max (0, Male (1) - 0) + 0.000 * max (0, Income - 9027) - 0.074 * max (0, Male(1) - 0) * max (0, Perceived health (0)-0) - 0.003 * max (0, Male(1) - 0) * max (0, Healthy breakfast(1) - 0) – 0.003 * max(0, Age - 46) * max (0, 2.14 - FAHP) + 0.003 * max(0, Age - 46) * max(0, Male(0) – 0) (11)

CBMI (Obesity) = 0.270 - 0.004 * max (0, 46 - Age) + 0.146 * max (0, Male (1) - 0) - 0.015 * max (0, Perceived health (1) - 0) - 0.069* max (0, Physical exercise (1) - 0) - 0.009 * max (0, Age - 72) * max (0, Male (1) - 0) - 0.001 * max (0, 72 – Age) * max (0,

[1] Higher interaction orders (3 and 4) and different penalty terms (2, 3, 5 and 6) were considered. No significant differences were found in terms of basis functions, knots and variable importance.

$$\text{Male}(1) - 0) + 0.000 * \max(0, \text{Income} - 9027) - 0.040 * \max(0, \text{Male}(1) - 0) * \max(0, \text{Perceived health}(0) - 0) - 0.059 * \max(0, \text{Male}(1) - 0) * \max(0, \text{Healthy breakfast}(1) - 0) + 0.022 * \max(0, \text{Age} - 46) * \max(0, 2.14 - \text{FAHP}) + 0.001 * \max(0, \text{Age} - 46) * \max(0, \text{Male}(0) - 0) \quad (12)$$

Table 3 Parameter estimates from the MARS model

		Coefficients			Knots									
		CBMI (Normal Weight)	CBMI (Overweight)	CBMI (Obesity)	Age	FAHP	FOHP	Income	Male (0)	Male (1)	P. Health (0)	P. Health (1)	Healthy Breakfast (1)	Physical Exercise (1)
Intercept		0.407	0.323	0.270										
Basis functions	1	0.011	-0.007	-0.004	<u>46</u>									
	2	-0.415	0.269	0.146						0				
	3	0.119	-0.014	-0.105								0		
	4	0.108	-0.038	-0.069										0
	5	0.013	-0.005	-0.009	72					0				
	6	0.003	-0.002	-0.001	<u>72</u>					0				
	7	0.000	0.000	0.000				9027						
	8	0.114	-0.074	-0.040						0	0			
	9	0.062	-0.003	-0.059						0			0	
	10	-0.019	-0.003	0.022	46	<u>2.14</u>								
	11	-0.004	0.003	0.001	46				0					

Underlined cells indicate basis functions of type max (0, independent-knot), otherwise max (0, knot-independent)

The importance of each explanatory variable is calculated as the square root of the GCV value of a sub model from which all basis functions that involve this variable have been removed, minus the square root of the GCV value of the selected model. Resulting values are re-scaled by giving a value equals 100 for the variable with the largest importance. The variables Age, gender (Male) and perceived health status are the most important in predicting the prevalence of obesity and overweight (Table 4). Economic factors seem to be less important in explaining obesity and overweight with income being the most important among these economic factors.

Table 4 Variable importance for multinomial MARS model

Variables	Number of basis functions	Variable importance
Age	5	100
Male	6	60.27
Perceived health status	2	38.08
Income	1	27.85
Physical exercise	1	26.26
Healthy breakfast	1	12.36
FAHP	1	7.30
FOHP	0	0

As mentioned above, the MARS method is more flexible as, among other issues, it does not impose linearity between the dependent variable and its covariates. In fact, results obtained from the MARS model are quite similar to those obtained in the parametric approach, when the relationship is linear (i.e. Physical exercise), but significant differences have been found when this is not the case. Moreover, we have found more consistency with expected results in the case of variables that have a non-linear relationship (i.e. age). Moreover, while in the MLR the age has a significant but very small effect on the prevalence of obesity; in the MARS it is considered the most important variable.

As mentioned above, physical exercise is the only covariate in the MARS model entering without any transformation or interaction with other explanatory variable. As a consequence, both models (MLR and MARS) resulted in quite similar parameter estimates for this variable in terms of sign and magnitude. All other covariates enter the MARS model in a nonlinear way or interacting with other covariates, making this approach a very useful tool to provide better insights to policy makers about specific population segments towards they could address effective policies to tackle obesity.

Although both MARS and MLR indicate that being male increases the likelihood of being overweight and obese, the magnitude of parameter estimates are different as the MLR does not consider interactions between gender and other covariates. A self-perception of a good health status decreases the prevalence of overweight and obesity for the whole sample. However, a self-perception of a bad health status has a negative impact on the prevalence of obesity only among men. This is may be due to the fact that traditionally females are caring more about their appearance regardless their health status while having health problems could lead men to observe more their diets (Case and Menendez, 2009). Similarly, having a healthy breakfast reduces the risk of being obese or overweight only in the case of men.

Regarding economic variables, similar to the MLR model, Food Out-of-Home prices do not have any significant effect on the prevalence of overweight or obesity although this result should be interpreted with caution as we are using regional instead of household prices due to the unavailability of such prices at household level. Food at-Home prices and income seem to have a small significant effect but only on a specific groups of our sample. In fact, higher Food At-Home prices have a negative (positive) effect on the risk of being overweight (obese) but only when average food prices exceed 2.14 euros (approximately a value representing 90% of the average sample food prices) and for people younger than 46 years. Income has a very small positive effect on the prevalence of overweight and obesity but only in the case that total per capita expenditure exceeds 9027 euro (this value represents, approximately 80% of the average sample value).

Finally, while the MLR model fails to capture the nonlinear effect of age, this is not the case for MARS model. As can be observed, age is positively related with the prevalence of overweight and obesity only for women younger than 46. This positive relationship also holds for the whole sample between 46 and 72 years old. These results are more consistent with Figure 1.

As a third step in our study, the Parametric MLR model has been re-estimated allowing for nonlinearities and interactions obtained from the MARS model. Resulted in a quite similar results to those obtained from MARS model and with a better fitted model (R^2 = 0.084). Table 5 summarizes estimates and marginal effects of the different variables.

Table 5 Estimates and marginal effects from Multinomial logit model with transformation and interaction between covariates using MARS model as pre estimation.

Overweight				
	Estimates		Marginal effects	
Constant	-0.199**	(0.063)		
Male	1.368**	(0.115)	0.223**	(0.023)
Physical exercise	-0.381**	(0.045)	-0.045**	(0.010)
Perceived health status	-0.327**	(0.054)	-0.018	(0.012)
Spending<9027	0.000**	(0.000)	0.000	(0.000)
Age>46	-0.049**	(0.003)	-0.008**	(0.001)
Males with Age<72	-0.011**	(0.003)	-0.002**	(0.001)
Males with Age>72	-0.053**	(0.009)	-0.007**	(0.002)
Females with Age<46	0.019**	(0.002)	0.003**	(0.000)
Age>46 and FAHP<2.14	0.074**	(0.029)	0.009	(0.006)
Males having a healthy breakfast	-0.206**	(0.065)	-0.017	(0.014)
Males with a percieved bad health	-0.422**	(0.082)	-0.066**	(0.017)

Obesity				
Constant	-0.271**	(0.073)		
Male	1.439**	(0.136)	0.107**	(0.016)
Physical exercise	-0.846**	(0.067)	-0.081**	(0.006)
Perceived health status	-0.834**	(0.064)	-0.101**	(0.010)
Spending<9027	0.000**	(0.000)	0.000**	(0.000)
Age>46	-0.060**	(0.005)	-0.005**	(0.001)
Males with Age<72	-0.015**	(0.003)	-0.001**	(0.000)
Males with Age>72	-0.089**	(0.012)	-0.009**	(0.002)
Females with Age<46	0.015**	(0.003)	0.001**	(0.000)
Age>46 and FAHP<2.14	0.138**	(0.033)	0.014**	(0.004)
Males having a healthy breakfast	-0.533**	(0.081)	-0.058**	(0.009)
Males with a percieved bad health	-0.511**	(0.098)	-0.044**	(0.012)
R^2	0.084			
AIC	36091			

Note: Standard Error in parentheses
** denotes statistical significance at 5 per cent significance level.

6. Concluding remarks

Economic factors seem to have a significant but small impact on the prevalence of obesity. In general, increasing prices, being male and having a bad health increase the probability of being obese. On the other hand, doing sufficient physical exercise and having a more completed breakfast decrease the probability of being obese. The most limiting point of this type of analysis is the data availability.

MARS model outperforms the traditional MLR and could be a helpful tool for understanding the nature of the relationship and the importance of the different variables to be introduced into the model.

An interesting feature of MARS is that it offers specific results for determined groups which could be of a great importance for policy makers through allowing them to design specific policies to combat obesity targeting specific population groups assuring a higher effectiveness of such policies.

Although our study gives a first look on the effect of economic factors on obesity prevalence, at the same time it opens number of interesting future research lines such as: Using several editions of the NHS to estimate a panel data model, Combine the NHS with the continuous household budget survey so we can get a richer data base including individual values for the economic factors, Develop a food quality index to be included in the analysis and Compare results from the NHS and the Catalan Health Survey which include measured instead of reported weights and heights.

References

Agencia espanola de seguridad alimentaria y nutricion (AESAN), 2011, Encuesta nacional de ingesta dietetica espanola (ENIDE),http://www.west-info.eu/files/Report188.pdf. acced on October,1,2013.

Beaulac, J., Kristjansson E., Cummins S., 2009. A systematic review of food deserts, 1966-2007. Preventive Chronic Disease 6(3), 105.

Briand, L.C., Freimut, B., Vollei, F., 2007. Using multiple adaptive regression splines to understand trends in inspection data and identify optimal inspection rates. ISERN TR 00-07 2007. Avaliable from: http://www.salfordsystems.com.

Case, A., Menendez, A., 2009. Sex differences in obesity rates in poor countries: Evidence from South Africa. Economics & Human Biology, 7(3), 271-282.

Cavaliere, A., Banterle, A., 2008. Economic factors affecting obesity: an application in Italy. International Congress, August 26-29, 2008, Ghent, Belgium, European Association of Agricultural Economists.

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, 565-587.

Cutler, D.M., Glaeser, E.L., Shapiro, J.M., 2003. Why have Americans become more obese? Journal of Economic Perspectives 17, 93-118.

Deconinckb E., Zhang M.H., Petitet F., Dubus E., Ijjali I., Coomans D., Vander Heyden Y., 2008. Boosted regression trees, multivari-ate adaptive regression splines and their two-step combinations with multiple linear regression or partial least squares to pre-dict blood–brain barrier passage: a case study. Analytica Chimica Acta, 609, 13-23.

De Veaux, R. D., D. C. Psichogios, and L. H. Ungar., 1993. A Comparison of Two Nonparametric Estimation Schemes: MARS and Neural Networks. Computers and Chem. Engineering, 17(8), 819-837.

Drewnowski, A., 2010. Healthier foods cost more. Nutrition Reviews, 68, 184–185.

Frank, I. E., 1995. Tutorial: Modern Non-linear Regression Methods. Chemometrics and Intelligent Lab. Sys., 27, 1-19.

Frieadman, J.H., 1991. Multivariate additive regression splines. Ann. Of Stat., 19, 1-67.

Garcia, E., Labeaga, J.M.. and Ortega, C., 2006. Maternal Employment and Childhood Obesity in Spain. Documento de Trabajo, 2006-17. Fundación de Estudios de Economía Aplicada (FEDEA).

Gil,J., Mora,T., 2011.The determinants of misreporting weight and height: the role of social norms. Economics & Human Biology 9 (1), 78–91.

Greene, W.H., 2003. Econometric Analysis. Fifth Edition. Prentice Hall.

Gundersen, C., Mahatmya, D., Garasky, S., Lohman, B., 2011. Linking psychosocial stressors and childhood obesity. Obesity Reviews, 12(5), 54–63.

Haughton, D., Loan, L. T. T., 2004. Vulnerability of Vietnamese Households, 1992-1998. Working Paper (June 2004). Ford Foundation/General Statistics Office Project, Vietnam.

Inestituto nacional de Estadestica (INE), 2011, Encuesta nacional de salud(ENS) http://www.ine.es/jaxi/menu.do;jsessionid=E35B0D01B1416C7691485D1448FE740A.jaxi03?type=pcaxis&path=/t15/p419&file=inebase&L=0. Acceed on October, 1, 2013.

Jin, R., Chen, W., Simpson, T. W., 2000. Comparative Studies of Metamodeling Techniques Under Multiple Modeling Criteria. AIAA-2000-4801. American Institute of Aeronautics and Astronautics, Inc.

Kayri, M., 2007. Two-step clustering analysis in researches: A case study. Eurasian Journal of Educational Research 7(28), 89-99.

Kooperberg C., Bose S., Stone C.J., 1997. Polychotomous Regression. Journal of the American Statistical Association 92(437), 117 - 127.

Kuhnert, P. M., Do, K., McClure, R., 2000. Combining Non-Parametric Models with Logistic Regression: An Application to Motor Vehicle Injury Data. Comp. Stat. and Data Analysis 34, 371-386.

Lakdawalla, D., Philipson, T., 2002. The Economics of Obesity: A Theoretical and Empirical Examination. National Bureau of Economic Research Working Paper 8946.

Larson, N. I., Story, M. T., Nelson, M. C., 2009. Neighborhood environments: disparities in access to healthy foods in the U.S. American. Journal of Preventive Medicine, 36(1), 74-81

Leathwick, J. R., Elith, J., Hastie, T., 2006. Comparative Performance of Generalized Additive Models and Multivariate Adaptive Regression Splines for Statistical Modeling of Species Distributions. Ecol. Mod. 199, 188-196.

Lin, B.H., Smith, T.A., Lee, J.Y., Hall, K.D., 2011. Measuring weight outcomes for obesity intervention strategies: the case of a sugar-sweetened beverage tax. Economics & Human Biology 9, 329–341.

Loureiro, M., Nayga, R.M., 2005. International Dimensions of Obesity and Overweight Related Problems: an Economic Perspective. American Journal of Agricultural Economics, 87(5), 1147-1153.

Miljkovic, D., Nganje, W. , de Chastenet, H., 2008. Economic factors affecting the increase in obesity in the United States: Differential response to price. Food Policy, 33(1), 48-60.

Monsivais, P., Drewnowski, A. , 2009. Lower-energy-density diets are associated with higher monetary costs per kilocalorie and are consumed by women of higher socioeconomic status. J Am Diet Assoc, 9(5),814-22.

Mora,T., Gil, J., Sicras,A., 2012. The Influence of BMI, Obesity and Overweight on Medical Costs: A Panel Data Approach. FEDEA WP 2012-08, Oct. 2012.

Moore, C. J., Cunningham, S. A., 2012. Social position, psychological stress, and obesity: a systematic review. Journal of the Academy of Nutrition and Dietetics, 112(4), 518-526.

Moore, L. V., Diez Roux, A. V., Evenson, K. R., McGinn, A. P., Brines, S. J., 2008. Availability of recreational resources in minority and low socioeconomic status areas. American Journal of Preventive Medicine, 34(1), 16-22.

Muñoz, J., Fellicisimo, A. M. , 2004. Comparison of Statistical Methods Commonly Used in Predictive Modeling. J. of Vegetation Science , 15, 285-292.

Mytton, O., Clarke, D. , Rayner, M., 2012. Taxing unhealthy food and drinks to improve health, BMJ 344, 2931.

Philipson, T.J., Posner, R.A., 2003. The long-run growth in obesity as a function of technological change. Perspectives in Biology and Medicine, 46(3), 87-607.

Powell, L.M., Chaloupka, F.J., 2009. Food prices and obesity: evidence and policy implications for taxes and subsidies. Milbank Q. 87(1),229-57.

Rao, M., Afshin,A., Singh,G., Mozaffarian,D., 2013. Do healthier foods and diet patterns cost more than less healthy options? A systematic review and meta-analysis. BMJ Open. 3, 42-77.

Schlosser, E., 2002. Fast Food Nation: The Dark Side of the All-American Meal. New York: Perennial .

Schroeter, C., Lusk, J., 2005. Economic factors and body weight: an empirical analysis. Journal of agricultural and applied economics. 40(2), 523-38.

Smith, J. P., 1999. Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status. Journal of Economic Perspectives 13(2), 145-166.

Ture, M., Kurt, I., Kurum, A. T., Ozdamar, K., 2005. Comparing Classification Techniques for Predicting Essential Hypertension. Expert Systems with App 29, 583-588.

Warner, M. L., 2003. The incidence of overweight and obesity in southern adolescents. Presented at the Food Assistance and Nutrition Research Small Grants Program Conference, Washington, DC.