UNIVERSITÀ DEGLI STUDI DI VERONA

DIPARTIMENTO DI

Scienze Economiche

SCUOLA DI DOTTORATO DI

Economia

DOTTORATO DI RICERCA IN

Economia e Finanza

CICLO /ANNO (1° anno d'Iscrizione) XXIV/2009

TITOLO DELLA TESI DI DOTTORATO:

Health Concerns and Socio-Economic Determinants: Analysis and Evaluation in Italy and England

S.S.D. SECS-P/01

Coordinatore: Prof. Giam Pietro Cipriani

Tutor: Prof. Luca Pieroni

Dottorando: Dott. Luca Salmasi

15 Marzo 2012

Contents

Introduction **10**

1 Food Prices and Overweight Patterns in Italy **11**

1.1 Introduction 11

1.2 Basic facts 11

1.3 Theoretical background 15

1.4 Econometric framework, data issues and elasticities 18

1.4.1 Methods 19

1.4.2 Data and elasticities 22

1.5 Results and discussion 26

2 The Effect of Smoking Bans on Weight in Italy **41**

2.1 Introduction 41

2.2 Related literature 41

2.2.1 Clean indoor air laws: direct effects on smoking and business 41

2.2.2 Smoking reductions, BMI and obesity 43

2.3 Framework of the analysis 47

2.3.1 A model for the causal relationship between smoking and BMI 47

2.3.2 Data and related empirical issues 49

2.4 Results 53

2.4.1 IV and RD estimates 53

2.4.2 Robustness, specification checks and conditional estimations 62

2.4.3 Body weight estimates to changes in smoking habits of smokers 66

3 Socio-Economic Determinants of Body Weight in the UK **73**

3.1 Introduction 73

3.2 Background: body weight and socio-economic variables . . . 73
3.3 Data and methods . . . 77
3.4 Empirical strategy . . . 79
3.4.1 Theoretical remarks . . . 79
3.4.2 The empirical model . . . 80
3.5 Results . . . 85
3.5.1 Preliminary results . . . 85
3.5.2 Estimates and discussion . . . 88

4 Smoking and Body Weight: Causal Estimates from the UK 101
4.1 Introduction . . . 101
4.2 An econometric approach in evaluating smoking reductions on BMI . . . 102
4.2.1 Preliminaries: standard regression and average treatment effect (ATE) . . . 102
4.2.2 Quantile regression and quantile treatment effect (QTE) . . . 104
4.3 Identification strategy, data and empirical models . . . 107
4.3.1 Identification from the BHPS . . . 107
4.3.2 Empirical models . . . 110
4.4 Results . . . 115
4.4.1 Robustness by control group . . . 119
4.4.2 Investigating smoking effects on BMI of obese people . . . 123
4.4.3 IV estimates . . . 128
4.5 A primer conclusion: the effects on body weight . . . 132

Conclusions 139

Bibliography 153

List of Tables

1.1 Cointegration rank test statistics for AIDS system 27
1.2 Estimated cointegrating vectors and theoretical restrictions imposed 28
1.3 Residual serial correlation of the vector error correction model 29
1.4 Long-run estimated elasticities of demand system (equation 1.7) 30
1.5 VECM estimates, diagnostics and elasticity of substitution by sub-samples 35
1.A.1 Food classification 36
1.A.3 Sensitivity analyses of estimated elasticities from the demand system 39
1.A.4 Johansen's Cointegration Rank Test Statistics for AIDS system 40

2.1 Variables definition and descriptive statistics 54
2.2 Causal effect of smoking on BMI 58
2.3 Causal effect of smoking on BMI, employed adults 59
2.4 Causal effect of smoking on BMI, men 60
2.5 Causal effect of smoking on BMI, women 61
2.6 Causal effect of smoking on BMI - IDW estimates 63
2.7 OLS estimates of effect of smoking on BMI 64
2.8 Causal effect of smoking on BMI in overweight and obese individuals 66
2.9 BMI elasticities to number of cigarettes and percentage of smokers 67
2.A.1 Causal effect of smoking on BMI, employed adults - IDW estimates 69
2.A.2 Causal effect of smoking on BMI, men - IDW estimates 70
2.A.3 Causal effect of smoking on BMI, women - IDW estimates 71

3.1 Data definitions and sources 84
3.2 Means and medians of BMI and share of obese people 87
3.3 Test for equality of coefficients across quantiles 95
3.4 BMI OLS and quantile regressions: model (1) 96

3.5 BMI OLS and quantile regressions: model (2) 97
3.6 BMI effect of a 1% increase of selected variables: model(1) 98
3.A.1BMI OLS and quantile regressions: model(1), people over 50 99
3.A.2BMI OLS and quantile regressions: model (2), people over 50 100

4.1 BMI absolute variation and ATE, by smoking status 112
4.2 BMI absolute variation and ATE, by BMI classes and smoking status 113
4.3 Causal effect of quitting smoking on BMI, CG_S 117
4.4 Interaction effect with socio demographics covariates on BMI, using CG_S 118
4.5 Causal effect of quitting smoking on BMI, CG_{ALL} 120
4.6 Causal effect of quitting smoking on BMI, CG_{NS} 121
4.7 Causal effect of quitting smoking on BMI, CG_{08} 122
4.8 Causal effect of quitting smoking on BMI, obese individuals CG_S 124
4.9 Causal effect of quitting smoking on BMI, obese individuals CG_{ALL} 125
4.10 Causal effect of quitting smoking on BMI, obese individuals CG_{NS} 126
4.11 Causal effect of quitting smoking on BMI, obese individuals CG_{08} 127
4.12 Causal effect of quitting smoking on BMI, IV estimates 131
4.13 Process tests for the BMI equation. Subsample size = $5n^{2/5}$ 131
4.A.1Descriptive statistics of covariates 135

List of Figures

1.1 Share of obesity and overweight and price ratio of healthy/unhealthy foods . . . 13
1.2 Ratio of unhealthy/helathy food consumption and relative energy index 13
1.3 Healthy and unhealthy food prices, consumption and relative energy index 16
1.4 Budget shares (deseasonalised data, 1997 - 1 : 2005:12). 24
1.5 Deviations of observed expenditure shares from long-run equilibrium levels 29
1.6 Long-run patterns of substitution elasticities of healthy and unhealthy food. . . . 31

2.1 BMI discontinuity around the clean indoor air law. 50
2.2 Smoking discontinuities around the Italian clean indoor air law (2005). 51
2.3 BMI discontinuities around the clear indoor air law, by subgroups. 53
2.4 Cohort patterns of weight, BMI and smoking indicators 55
2.5 Causal effect of smoking on weight. Discontinuity sample 68

3.1 Percentage of obesity in UK by gender: 1993 - 2007 74
3.2 Consumption of calories in UK: 1995 - 2007. 76
3.3 Changes in UK relative food price: 1987 - 2007 76
3.4 Kernel density estimates of BMI by gender 86
3.5 Quantile regression estimates: model (1), male 89
3.6 Quantile regression estimates: model (1), female 90
3.7 Quantile regression estimates: model (2), male 91
3.8 Quantile regression estimates: model (2), female 92

4.1 BMI kernel density estimate, by smoking status 111
4.2 Quitting smoking effect on weight (kg), by control group 133
4.3 Quitting smoking effect (kg), IV estimates by control group 134

Introduction

This thesis analyses various aspects related to the economic determinants of obesity in two EU countries, Italy and the UK, which are of interest for different reasons. Although much later with respect to the United States and some continental European countries, the issue of the rise of weight has also become significant in Italy, where, 34.2% of adults were classified as overweight and 9.8% as obese in 2005. These percentages were found to differ greatly according to age, gender, years of education and income level. Moreover, the percentage of adults classified as obese has risen by 4 percentage points in the last decade as shown by ISTAT. This measure becomes more relevant when we consider that the obesity and overweight rate increased in the same period by 9 percentage points. In the UK, the increase in obesity is found to be similar to that of the United States although it started from a lower level. The percentage of obese individuals is about 24% in 2007 and is among the highest in Europe. The trends of obesity in aggregate have constantly risen over the last fifteen years (15% since 1993), similarly for both men and women. This persistent growth suggests that, at least, some causes may have become structural in determining obesity in the UK, but also in Italy.

The reasons for the significant increase in the obesity levels for the two countries analysed may be of various nature. We decided to focus our attention on the role of food prices in Italy, where the healthy properties of the Mediterranean diet have, up to now, mostly influenced the low obesity rates in the country. It becomes then of crucial importance to document any significant variation, in terms of food consumption, that may be responsible for the general increase in obesity and overweight rates. While for the UK, this research is related to a number of empirical papers testing overweight as the result of several socio-economic changes which have altered people's lifestyle choices. In particular, we examined the consequences of changes in relative prices and in the density of different types of restaurants on obesity, as well as the influence of cigarette consumption. We also refer to the effects of technological changes, which are responsible for shifts over time in employment from agricultural and manufacturing to services, implying a decrease in the strenuousness of jobs, and in the number of hours dedicated to physical exercise. Agricultural innovations are also responsible for reductions in the price of food

and consequently of calories.

Another very important determinant of the obesity epidemic is related, by the health economic literature, to the recent drastic reduction in smoking habits. Many previous works have documented a very clear inversion between the trends of smoking, which have constantly reduced, and BMI, which documented a sharp increase. However all these works did not provide a general agreement on the sign and the magnitude of these effects, and most importantly they did not provide a clear estimate of these effects on obese individuals. We take advantage of the introduction of the Clean Indoor Air Law, which prohibits smoking in public places, implemented in Italy as from 10 January 2005, to identify the relationship between smoking behavior and body weight within a regression discontinuity design.

In the case of the UK, we took advantage of the longitudinal framework of the BHPS and estimate a difference-in-differences (DID) model to account for individual fixed-effects related to individual health concerns and estimate the parameters of the DID model using a battery of control groups. We also performed IV and IVQR estimates for the average treatment effects (ATEs) and quantile treatment effects (QTEs) estimators to take into account properly issues related to endogeneity.

Chapter 1

Food Prices and Overweight Patterns in Italy

1.1 Introduction

In this chapter, the role of relative food prices in determining the recent increase in body weight in Italy is examined. Cross-price elasticities of unhealthy and healthy foods, estimated by a demand system, provide a consistent framework to evaluate substitution effects, when a close association is assumed between unhealthy (healthy) foods and more (less) energy-dense foods. A dataset constructed from a series of cross-sections of the Italian Household Budget Survey (1997–2005) was used to obtain the variables of the demand system, which accounts for regional price variability. The relative increase in healthy food prices was found to produce nontrivial elasticities of substitution towards higher relative consumption of unhealthy foods, with effects on weight outcomes. In addition, these changes were unevenly distributed among individuals and were particularly significant for those who were poorer and had less education.

1.2 Basic facts

We begin this section by illustrating the patterns of the prices and quantities of healthy and unhealthy foods purchased in Italy, based on Gelbach et al. (2009) classification, adapted for the features of Italian food consumption. Although this classification is built

to divide food with a higher and lower contents of fat and sugar, linked with calorie intakes, we exclude typical foods of the Mediterranean diet like pasta and olive oil from food macro-categories. We motivate this choice by observing that these categories are those that have shown small changes in consumption during the period analysed. Thus, if Italian consumers preserve their usual consumption behavior, the substitution effects with other foods may also be assumed to be negligible. The empirical section provides a sensitivity analysis of the estimates of substitution effects by including bread, pasta and olive oil in unhealthy and healthy food macro-categories. We report the details of this robustness check in Appendix 1.A.1.

Figure 1.1 shows the overlapping paths of the fractions of overweight and obese adults and the relative changes in healthy and unhealthy food prices. For this purpose, we match two sources of data. First, we use the annual multipurpose survey of Italian households (2002-2005) conducted by the Italian Institute of Statistics (ISTAT) to obtain an average of the body mass index (BMI) for the Italian population. We then integrate these measures of individual BMI with data from two more general ''multipurpose surveys" of 1994-1995 and 1999-2000, and interpolate the values for the missing years. Note that these breaks do not allow us to use the BMI indicator as a direct (long-run) body weight response to changes in food price categories [1].

The annual average of expenditure and prices for these two consumption categories of consumption are obtained by extracting micro data from the Italian household expenditure survey (HES) released annually by ISTAT. The latter surveys are the main sources for the empirical section, and we will discuss this issue further, anticipating now that the relative prices of unhealthy and healthy foods are obtained by the ratios between nominal and real expenditure, and that real expenditure, as a consumption index, aggregates the ratios between the current expenditure for each individual item and its price index (at national level). The graph, shows a constant increase in BMI over time. As an aggregation of the repeated household surveys, the picture which emerges shows that the prices

[1]The body mass index (BMI) is a measure of body fat based on height and weight, which applies to both adult men and women. Four categories are generally used to classify adults: i) Underweight BMI $\leq$ 18.5; ii) Normal weight = 18.5-24.9; iii) Overweight = 25-29.9; iv) Obesity = 30 or over. It is known that BMI is not the most accurate measure of body fat and that self-reported weight produces measurement errors for young and adult people. For a critical discussion of this indicator, see Burkhauser and Cawley (2008).

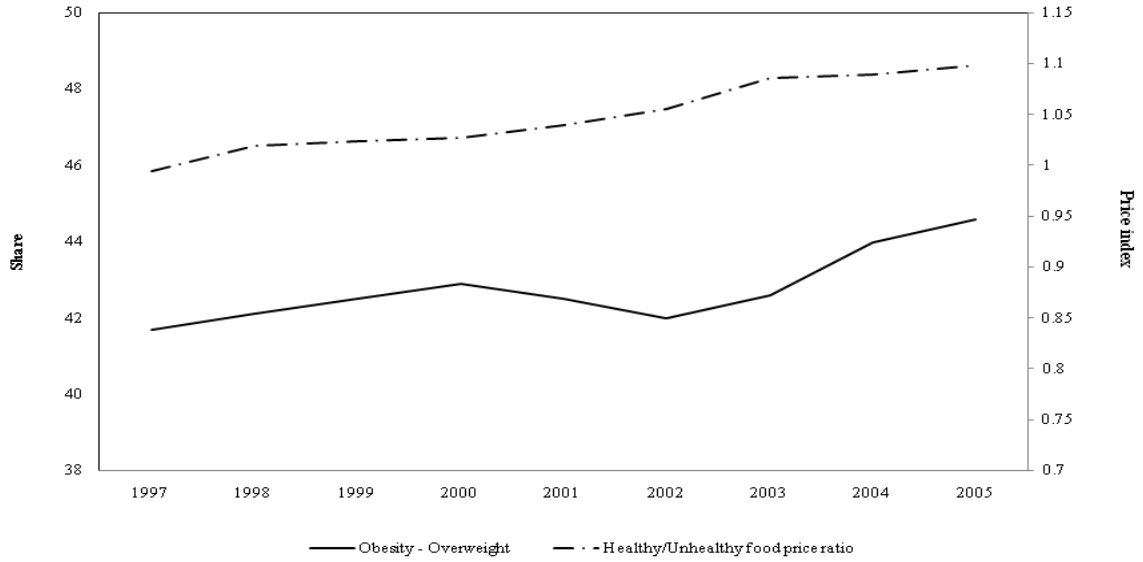


Figure 1.1: Share of obesity and overweight and price ratio of healthy/unhealthy foods

of healthy foods rise more quickly than those of unhealthy ones (1 percentage point per year).

As unhealthy food becomes relatively cheaper, people are expected to substitute healthy foods with unhealthy ones,. Relative consumption, expressed as a constant expenditure, is reported in the continuous line of Figure 1.2 (*i.e.*, vice versa with respect to relative prices). The predictions of the theory are supported by the slight upward positive trend in the growth of unhealthy food consumption during the period 1997-2005.

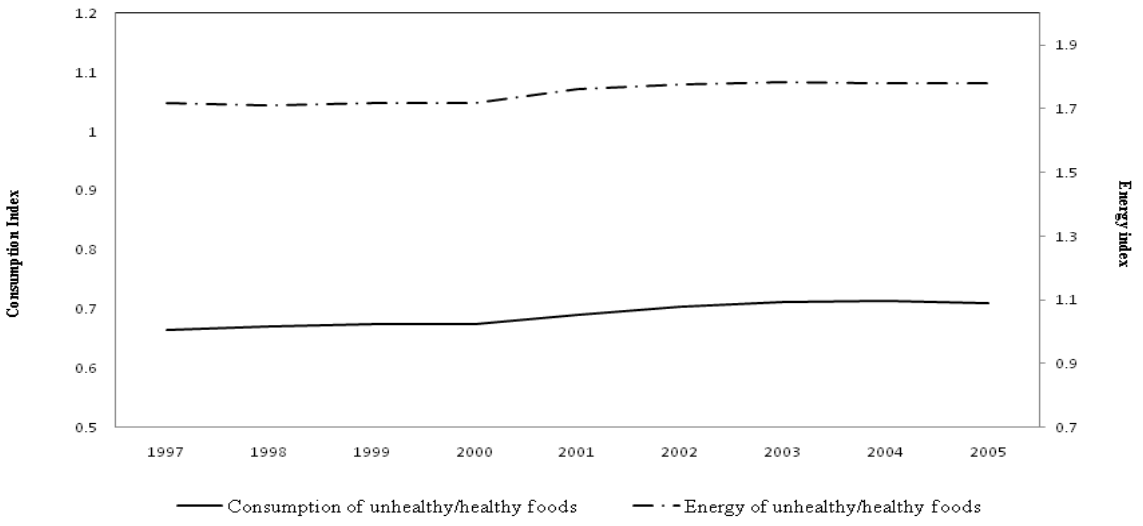


Figure 1.2: Ratio of unhealthy/helathy food consumption and relative energy index

Although we cannot directly retrieve the effects of food prices on weight outcomes, we follow the schemes used by Chesher (1998), transforming the categories of food consumption into energy (*i.e.*, calories). With information from the Italian Institute of Nutrition (INN, 1997), we can aggregate food categories, construct relative patterns, and associate the patterns of body weight. In particular, this source provides data for over 1000 food items, which were first aggregated into 62 categories - the same employed by ISTAT in its surveys - and further aggregated into the categories of healthy and unhealthy foods considered in this study. The dotted line of Figure 1.2, highlights the evolution of energy changes (expressed in calories) of unhealthy with respect to healthy foods. These patterns are close to those of the ratio between unhealthy and healthy food consumption, indicating that changes in food prices translate into changes in quantities of calories consumed by individuals, which may contribute to explaining the increase of the share of overweight individuals, as suggested by Drewnowski and Darmon (2005).

Although the patterns of the figures seem to be consistent with the working hypothesis, our approach differs conceptually from that of works which use representative time-series data to measure the elasticity of substitution *e.g.*, Zheng and Zhen (2008). For this reason, the rest of this study focuses on measuring the elasticity of substitution by grouping the Italian household expenditure surveys and estimating them according to socio-economic and demographic characteristics.

At descriptive level, the annual multipurpose surveys of Italian households, ISTAT (2007), illustrates significant changes in the recent prevalence of obesity and overweight for the selected socio-economic and demographic groups. The problem of overweight specifically affects males and increases with age, although it should be noted that it is important at all ages. The rise in BMI disparities among social groups is also reflected according to different educational levels. It is generally found that less educated adults have a greater risk of obesity.

Since the income is not collected in the HES, we cannot compare the patterns of relative prices of consumed quantities and the relative transformed energy intakes. However, to disentangle how, for example, individuals with lower incomes (i.e., lower economic status) cover calorie requirements, we use a relative poverty line following the ISTAT procedure according to which, a household is considered under the poverty line if the

total expenditure of a household composed by two individuals is lower than the average expenditure of the per-capita expenditure. An equivalence scale was used to determine the relative poverty line for households with a number of components other than two.

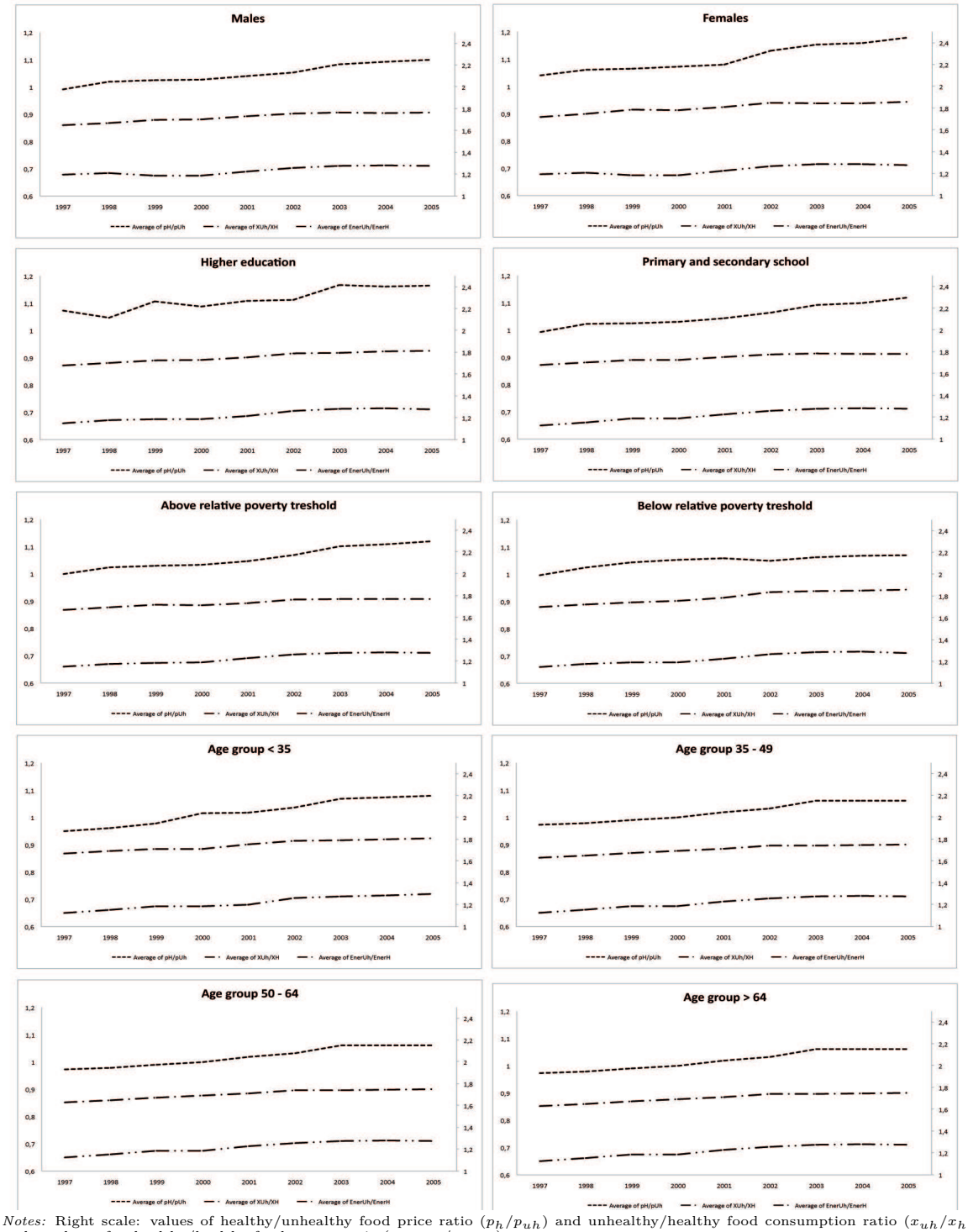
Figure 1.3 illustrates these points by dividing the sample into several socio-economic characteristics - gender, age, education and income - and collecting household data for each subsample. Naturally, gender is composed by males and females, education by household's head with university degree, master or PhD, secondary education or less; age by four classes (18-34; 35-49; 50-64; ≥ 65) and income by households below or above the relative poverty line.

Panel a) shows that shifts in relative prices (healthy/unhealthy) for men are related to an increase in unhealthy food consumption and, on average, of total calories; the influence on energy is less important for women. These results also find confirmation for educational and income groups. Greater sensitivity to (positive) healthy/unhealthy price changes regarding the consumption of unhealthy foods is shown for households with lower education (panel b) and lower income (panel c), with higher growth in the last few years. Panel d) shows the patterns of price changes by age. Although all the classes respond significantly to changes in healthy food prices, the age classes in which the potential substitution of healthy foods with unhealthy ones is more evident are those between 35-49 and 50-65.

These descriptive results are testable through the use of demand models. The existence of specific patterns of the elasticity of substitution may be assessed within demographic and socio-economic groups. In this way we are able to provide a general framework for testing the hypotheses empirically and support policy interventions. The rest of this work is devoted to applying this framework to one of the more used demand systems proposed in the literature, the $AIDS$ model of Deaton and Muellbauer (1980).

1.3 Theoretical background

In this section we briefly review the static $AIDS$ model. The specification of this demand system arises from a class of preferences in the logarithm of total expenditure. It satisfies the necessary and sufficient conditions for consistent aggregation across consumers,



Notes: Right scale: values of healthy/unhealthy food price ratio (p_h/p_{uh}) and unhealthy/healthy food consumption ratio (x_{uh}/x_h). Left scale: values of unhealthy/healthy food energy ratio ($ener_{uh}/ener_h$).

Figure 1.3: Healthy and unhealthy food prices, consumption and relative energy index

Deaton and Muellbauer (1980), Muellbauer (1976) and allows the estimation of demand elasticities with limited restrictions Deaton (1986).

It is assumed that there are n goods which can be purchased by consumers and which

can potentially be included in the demand system. We index these categories of goods (and services) by $i = 1.., n$. Because we are interested in the reaction of relative consumption between two macro categories of foods, *i.e.*, healthy and unhealthy, this classification determines what sub-categories of goods are included in the $AIDS$.

Analytically, the budget share of a certain good is equal to the expenditure generated by the good divided by the total expenditure. We use w_i to represent the budget share of a good i, $i = 1, ..., n$. Under the $AIDS$ specification, w_i takes the form of:

$$w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \log p_j + \beta_i \log (Y/P) + \nu_i \tag{1.1}$$

where p_j with $j = 1, ..., n$, are the prices of goods, Y is the total expenditure in the demand system, P is an overall price index, ν_i is a stochastic error, and α_i, β_i and γ_{ij} are parameters to be estimated. Note that the last term of (1) is based on the real expenditure ($Y/P = y^*$) devoted to good i. The budget share of product i increases as the total real expenditure of the category increases if β_i is positive, and decreases if β_i is negative. The second term represents the price effects of the various goods. We will return to this point later, after introducing all the ingredients of this flexible demand system, which will be used to estimate the (cross) price elasticities of demand and to identify the patterns of substitution between goods.

We then define w_i^* as the ''optimal" level of the expenditure budget share w_i for commodity i and $\log P = \alpha_0 + \sum_{k=1}^{n} \alpha_k \log p_k + \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \gamma_{kj} \log p_k \log p_j$. As commonly done in empirical papers, we also employ a linear approximation in this price index, *i.e.*, the Stone's price index, defined as $\log P = \sum_{i=1}^{n} w_i \log p_i$.

As discussed in the review proposed by Barnett and Serletis (2008), the $AIDS$ flexible model has a number of desirable properties. It derives an expenditure function from a second-order approximation to any expenditure function and provides the possibility of including the theoretical restrictions of adding up, homogeneity and symmetry in order to respect the predictions of the demand theory. Because the expenditure function must be linearly homogeneous and strictly increasing in p, adding up and homogeneity can be obtained by imposing $\sum_{i=1}^{n} \alpha_i = 1$ and $\sum_{i=1}^{n} \gamma_{ij} = \sum_{j=1}^{n} \gamma_{ij} = \sum_{i=1}^{n} \beta_i = 0$, respectively, while symmetry requires $\gamma_{ij} = \gamma_{ji}$ for all i, j.

Another property of a robust expenditure function (and demand system) is that it must be concave in prices. This means that the matrix of the second-cross partial derivatives must be negative semi-definite. In turn, this property gives rise to the matrix of the substitution effects of Slustky, $S_{ij} = \partial h_i(p,u)/\partial p_j$, with non-positive own-price effects , where $h_i(.)$ is Hicksian demand. Formally, it can be shown that the Slutsky substitution coefficients of model (1.1) are given as:

$$S_{ij} = \frac{y}{p_i p_j}[\gamma_{ij} + w_i w_j - \delta_{ij} w_i] \tag{1.2}$$

where δ_{ij} is the Kronecker parameter ($\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$). The matrix of substitution effects for the $AIDS$ model varying with data determines that negativity conditions must be evaluated (and eventually imposed) locally at a specific point in the sample[2]. That is, by scaling the data at a representative point (e.g. the mean of the sample) in which $P = y^* = 1$, we can obtain the local substitution term $\theta_{ij} = S_{ij}(P = y^* = 1)$[3].

Equation (1.1) is singular by construction, as the expenditure shares sum to 1. A frequently employed procedure to avoid econometric problems consists of dropping one equation from the system and, although the budget share demand system with a $n-1$ rank must be empirically confirmed, it provides complete characterisation of consumer preferences. Consequently, it can be used to estimate the income, own- and cross-price elasticities as well as the elasticities of substitution.

1.4 Econometric framework, data issues and elasticities

This subsection provides econometric support for modelling a long-run demand system which includes a gradual adjustment over time of consumption in response to shifts in relative prices. However, when time-series have a significant dimension, empirical demand

[2]See Cranfield and Pellow (2004) for a more thorough discussion of the role of global and local negativity in functional form selection.

[3]In order to remark the properties of the demand system, the $AIDS$ provide a reasonably accurate approximation at any set of prices not too far from the point of approximation.

system studies suffer from severe econometric flaws, because the time-series of budget shares, prices and real income are non-stationary.

One way of solving these issues is to use linear cointegration models Attfield (1997, 2004) although they may not be completely consistent, since the error terms in demand systems tend to be autocorrelated Lewbel and Ng (2005). Standard asymptotic theory may provide a poor guide to finite-sample inference when the errors are persistent in a cointegrated demand system. As one aim of the empirical strategy, we complement statistical analysis by investigating and testing the non-stationarity behaviours of the time-series residuals.

There is also a profound policy interest in obtaining parameter estimates from a cointegration framework. They are inextricably linked with the notion of long-run estimation, Pesaran (1997). Because we are specifically interested in analysing substitutability effects of healthy foods with respect to unhealthy ones, an important question for policy-makers is whether these trends will continue in the future. Indeed, a measure of the long-run elasticity of substitution is a powerful tool in assessing potential government intervention in preventing obesity, for example by taxes imposed on unhealthy foods or subsidies applied to healthy ones, Powell and Chaloupka (2009). In subsection 4.1, we report the conditions for the identification of a long-run demand system based on the cointegration rank of a vector autoregressive (VAR) model under the theoretical constraint of adding-up. We then show that, in this framework, the other theoretical restrictions of homogeneity and symmetry can be imposed and tested, and the estimated parameters recovered to calculate the price elasticities.

1.4.1 Methods

We formalise the specification of the equations of the demand system in (1.1) as a cointegrated demand system. Firstly, we consider the vector autoregressive (VAR) formulation of a demand system and describe the corresponding vector error correction (VECM) representation, following Johansen (1995). Formally, the data-generating process for $X_t = (X_{1t}, X_{2t})$ is assumed to belong to the class of VAR models:

$$X_t = \mu_0 + \mu_1 T + \phi_h D_{th} + \sum_{i=1}^{p} A_i X_{t-i} + \varepsilon_t \tag{1.3}$$

where $X_{1t} = (w_{1t}, , w_{2t}, ...w_{n,t})$ is the $n*1$ vector of budget shares and $X_{2t} = (p_{1t}, p_{2t}, ...p_{nt}, y_t)$ is the $n+1*1$ vector containing price indices and real expenditure. μ_0 is a $n*1$ constant term, μ_1 is a $n*1$ vector of coefficients related to the deterministic trend T, D_{th} is a vector containing deterministic variables (in our application centered seasonal dummies) and ϕ_h the corresponding $n*h$ matrix of parameters. A_i is a matrix of unknown parameters for the lags of X_t, ε_t is a Gaussian white noise process with covariance matrix Ω and p the lag order of the VAR. Equation (1.3) may be re-written in a VECM form as:

$$\Delta X_t = \mu_0 + \Pi X^*_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \tag{1.4}$$

where $\Pi = (\sum_{i=1}^{p} A_i - I_p)$, $\Gamma_i = -\sum_{j=i+1}^{p} A_j$,with $j = 1, ...p-1$. The matrix of parameters Π describes the long-run relationships of the VECM among the variables in vector $X^*_{t-1} = [X_{t-1}; D_t; T]$. Γ_i, with $i = 1, ...k-1$, is a vector of parameters which refers to the short-run dynamics of the system ΔX_{t-i}. In known general conditions, VECM equation (1.4) is formulated as:

$$\Delta X_t = \mu_0 + \alpha \beta^{*'} X^*_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \tag{1.5}$$

where α is a $n*r$ matrix, $\beta^* = (n + \Upsilon)*r$ matrix and $r(0 < r < Q = 2n)$ is the cointegration rank of the demand system. Υ is the matrix containing deterministic and seasonal components.

Pesaran and Shin (2002) show that, to recover exactly the long-run structural parameters of model (1.5), r restrictions on cointegrating relationships must be imposed on each non-singular demand equation, expressed in budget shares. In this context, the adding-up theoretical constraint executes a crucial role in identifying the structural model, implying a further implicit restriction of the rank of cointegration of the VAR model, i.e., $r = n-1$. Formally, disregarding deterministic terms, the matrix of cointegration vectors for the demand system is specified as:

$$\beta' = [-W_{n-1}, B] \tag{1.6}$$

where W is the $n-1 * n-1$ budget share matrix, and B the matrix of the parameters for log prices and real income. The long-run identifying restrictions can be imposed in the W matrix. A necessary and sufficient condition indicates that the number of identifying conditions k, is at least equal to r^2, and that the exact identification for a long-run AIDS demand system requires $r^2 = (n-1)^2$ restrictions. This implies that, for a demand system with three good categories, a diagonal framework for identification may be imposed as:

$$\beta' = [-I_{n-1}, B] \tag{1.7}$$

where I_{n-1} is the unit matrix, and $k - r^2$ over-identifying restrictions can be imposed and tested directly on cointegrating vectors. In the demand system context, the restrictions are derived from the theory and concern the hypotheses of homogeneity and symmetry, Deaton and Muellbauer (1980). Appendix 1.A.2 lists the matrices of the dynamic demand system described in (6) and (7) and those with imposed the theoretical over-identifying restrictions.

The maximum likelihood estimations of cointegration matrix β' are carried out by the ML estimator, which is super-consistent and mixed normal. This allows us to test over-identifying restrictions by a log-likelihood ratio statistic which is asymptotically distributed as an χ^2, with degrees of freedom equal to the number of over-identifying restrictions imposed.

As already observed, the cointegration rank for identifying the patterns of adjustment in each budget share implies $r = n-1$. Thus, the rank condition excludes all the cases in which $r < n-1$. Although in the empirical analysis of demand systems it may happen that the rank of cointegration is higher than the number of cointegration relations, i.e., $r > n-1$, in our study this hypothesis is economically inadmissible. This should imply additional cointegration relations involving price variables in the cointegration space. However, the consequence of this over-dimensionality in equilibrium relationships is in contrast with the causal impact of relative prices on the evolution of budget share consumption and in the measure of effects of substitution.

1.4.2 Data and elasticities

The HES provides information about the socio-demographic characteristics and expenditure levels of Italian households. Although this survey is monitored weekly and published on a monthly basis, it does not provide any information about the quantities purchased and the prices relative to the consumption of each good or service. To obtain estimates of price and (substitution) elasticities, empirical works use aggregate national price indexes. However, this approach requires a long span of cross sectional data to estimate a demand system with sufficient price variation - features which are almost never available in empirical applications. Aggregate price indexes are highly correlated, due to the restricted number of categories of consumption normally analysed in the demand system, leading to the rejection of theoretical restrictions and making estimated elasticities highly uncertain. This problem was examined and discussed by Coondoo et al. (2004).

For these reasons, surveys only gathering expenditure data have limited applicability in modern demand and welfare analysis, unless if researchers are able to include sufficient variability in prices. We believe that, by combining a regional price index (RPI) - instead of a national price index (NPI) - with the expenditure, we can respond to the issues discussed above satisfactorily, and consistently estimate the demand system of interest made up of three aggregate categories: a) healthy foods; b) unhealthy foods; and c) other foods and nondurables.

The use of RPI is also important for identifying price elasticities. Above we identify the long-run AIDS model from a technical standpoint. However, a demand system also requires that variations in prices should be a result of supply shifts. In this context, it is assumed that regional price index differences reflect supply shifts rather than movement along the demand curve. Following the line of argument of Gelbach et al. (2009) and Zheng and Zhen (2008), we also assume for Italy that marketing healthy foods like fruit and vegetables is more expensive than marketing unhealthy ones like fats and oils, because the former entail higher costs in transport, refrigeration, labour and packaging, and are much more prone to spoilage. But, in addition, the use of a regional price index allows us to account for heterogeneous differences in food production and distribution across the Italian regions, and enable us to identify the preference parameters for healthy food

separately from unhealthy food.

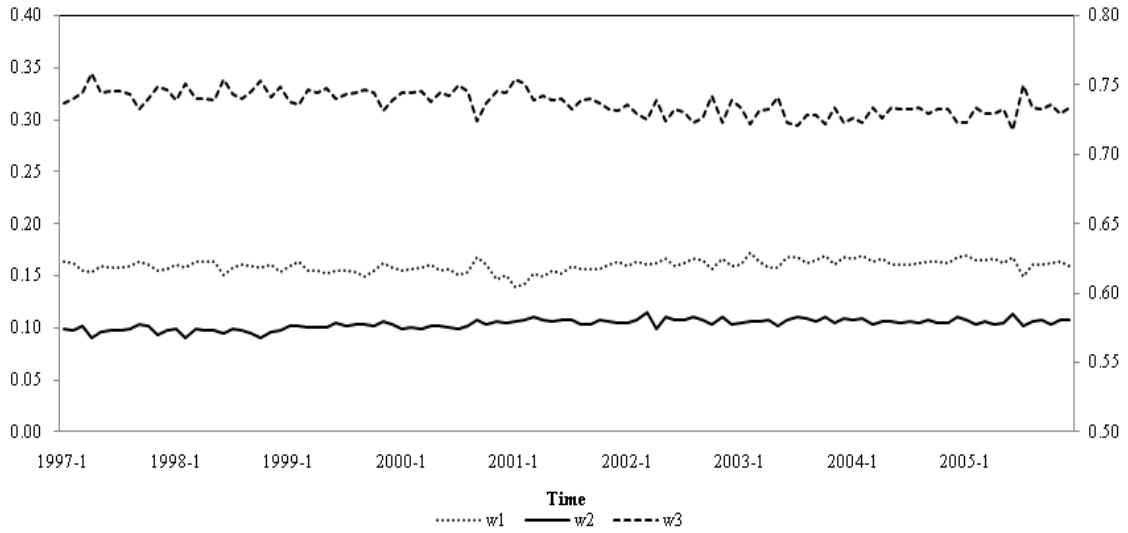
Data on the consumer price index for the whole collectivity (NIC in ISTAT methodology) are published by ISTAT on a monthly basis including more than 100 goods and services and allow us to aggregate sub-categories of goods consistently. This (chained) price index has been available at regional level since January 1999 and spans until December 2005, the last survey currently available.

In addition, since the survey was revised in 1996, we were obliged to start our analysis from 1997. However, to avoid any restrictions on the time-dimension of the sample, we approximate a monthly consumer price index at regional level for 1997 and 1998. We matched the monthly regional elementary price index with the households interviewed in a given month and region, grouping them according to the categories of goods selected in the demand system.

Figure 1.4 shows the graphs for aggregate budget shares w_{it}, with seasonal adjustments obtained by the X12 census procedure, highlighting the fact that at least some of these shares also show non-stationary behaviors. This study confirms the systematic evidence of non-stationary in the variables of the demand system Attfield (1997, 2004), Lewbel (1999), Ng (1995). But, as discussed in Lewbel and Ng (2005), because budget shares must, by construction, lie between 0 and 1, they cannot remain non-stationary forever. The small changes that take place from month to month imply that budget share changes can therefore approximate a non-stationary process for a long time, as is the case of budget shares in the Italian non-durable data.

Although the demographic and socio-economic samples seem to have some specificity, we extend the assumption that the variables of the demand system are non-stationary because the patterns of the variables related to the subsamples are close to those of the full sample and, as noted above, changes in the budget share sub-samples are very small[4]. We therefore proceed with the empirical analysis by testing the cointegrating rank of system (5), as a test for identifying the long-run demand system, irrespective of whether it is applied to the whole population or only to subsamples.

[4]Results of formal tests of non-stationarity with the FDFGLS test of Elliott et al. (1996) and KPSS of Kwiatkowski et al. (1992) for the full sample and subsamples are not presented here, but they are consistent with the hypothesis that budget shares (and also prices and real expenditure) contain unit roots. Both estimations and graphs of budget share subsample patterns are available upon request.



Notes: Right scale: budget share of other foods and non-durables (w_3). Left scale: values of budget shares of healthy (w_1) and unhealthy foods (w_2).

Figure 1.4: Budget shares (deseasonalised data, 1997 - 1 : 2005:12).

In order to include in the demand system analysis non-durable goods and to estimate elasticities properly, we assumed a multistage decision process, as proposed by Edgerton (1997). The estimated system is conditional on the choice of purchasing non-durable goods in a previous non-modelled stage that is determined as part of the consumers' overall decision process regarding how to allocate expenditure across the full range of goods. In fact as the expenditure is allocated within non-durables, only in the second (conditional) stage consumers decide how to allocate across healthy and unhealthy foods[5]. Under these assumptions income elasticity and Hicksian price elasticities are computed as follows:

$$\eta_i = \frac{\partial q_i}{\partial y}\frac{y}{q_i} = \frac{\partial w_i}{\partial \log y}\frac{1}{w_i} + 1 \tag{1.8}$$

with $i = 1, 2, 3$. The expression for expenditure elasticity indicates that a good is a luxury if $\eta_i > 1$ and a necessity if $\eta_i < 1$. The compensated price elasticities are given as:

$$\eta_{ij} = \frac{\partial h_i}{\partial p_j}\frac{p_j}{h_i} = \frac{\partial w_i}{\partial \log p_j}\frac{1}{w_i} + w_j - \delta_{ij} \tag{1.9}$$

[5]In a complete demand system, we should consider previously at least the choice of how to allocate total expenditure between goods and services for consumption Gorman (1995).

where the partial derivatives in (1.8) and (1.9) are obtained from equation (1.7) and $i, j = 1, 2, 3$. Then, the uncompensated price elasticities are obtained from the Slutsky equation, $\varepsilon_{ij} = \eta_{ij} - w_j\eta_i$.

One formal way of testing the substitution effects between healthy and unhealthy foods is to calculate their cross-price elasticities. According to Hicks (1986), $\eta_{ij} > 0$ indicates substitutability among goods, $\eta_{ij} < 0$ complementarity, and $\eta_{ij} = 0$ independence.

One important property of the Slutsky equation is that the estimated parameters of the matrix are symmetric. But unlike Allen's elasticity of substitution Allen (1938), the Hicksian framework does not impose symmetry restrictions in elasticity terms. In line with the aims of this chapter, we can verify how changes in the prices of unhealthy versus healthy foods simultaneously affect the relative cost of purchasing quantities of these categories. The framework is therefore close to that of Auld and Powell (2009). If, as shown in section 2, changes in healthy food prices are higher than in unhealthy food prices during the last decade in Italy, we can predict greater total calorie intake, generated by a substitution towards unhealthy foods, which determines an increase in the consumption of energy-dense foods. In addition, because the elasticity of substitution is the percentage change in the budget share allocated to good w_i, divided by the percentage change in price p_j, we can compute the net effects of healthy and unhealthy budget share responses to changes in prices. We use η_{12} to represent the elasticity of substitution of the healthy food to changes in unhealthy food prices; η_{21} represents the response of the unhealthy food budget share to changes in healthy food prices. Recalling that the food prices of both categories increased in the period 1997-2005, the net effect of this elasticity on the sample mean is given as:

$$\Delta\eta = \eta_{21} - \eta_{12} \tag{1.10}$$

If the cointegrated AIDS is exactly identified, when we extend estimations at each (monthly) point of the sample, we can use the estimations of the elasticity of substitution to evaluate the dynamics of the response of healthy and unhealthy food consumption to relative price changes by the scaling procedure of estimation discussed above. In the condition that the concavity condition is satisfied, we consistently reconstruct the

elasticities of substitution (η_{21} and η_{12}) for both the full sample and for the subsamples and their confidence intervals by bootstrapping the standard errors of elasticities.

1.5 Results and discussion

Following the discussion in Section 4, cointegration test procedures were used to identify the number of long-run empirically important relationships in our data. We only note that exact identification requires two cointegrating relations associated with n-1 (3-1) budget shares among the variables of the model, so that we specify a VAR(3) to evaluate it[6].

The test results of the trace statistic, Johansen (1995), and test (SL), Saikkonen and Lütkepohl (2000), are listed in Table 1.1.

We report the results of both tests from data obtained by aggregations of good categories using the national and regional price index. At the five percent significance level, neither the trace statistic nor the SL test reject the hypothesis that there are two cointegrating vectors among the variables of the demand system, irrespective of the price index used. Thus, by assuming that r=2, the long-run model (1.5) provides a consistent representation for assessing the significance of the effects of substitution between healthy and unhealthy food expenditure categories.

The parameter estimates of the demand model, which aggregates non-durable goods according to the regional price index (RPI), are listed in Table 1.2a; those obtained from the data aggregated by the national price index (NPI) are shown in Table 1.2b[7].

Although almost all the parameters of the two models estimated, with imposed homogeneity and symmetry restrictions, are significant and with similar size, we find that the theoretical restrictions are not jointly rejected at 1 percent only for the data obtained by RPI[8]. Thus we proceed below to the estimation of elasticities of the long-run demand system by using variables obtained from the regional price index.

[6]To select the order of the VAR, we used the sequential modified likelihood ratio (LR) test as in Lütkepohl (1991), while estimations are carried out by including centred seasonal dummies.

[7]The third cointegrating vector for other foods and non-durables is then recovered by the adding-up constraint.

[8]The differences in the results of the theoretical restriction tests are emphasized when small sample statistics are performed. In this case, although the model estimated with a national price index is still rejected at one percent, data which use a regional price index do not reject homogeneity and symmetry at five percent. These results are available upon request.

Table 1.1: Cointegration rank test statistics for AIDS system
Specification 1a - Regional price index.

H_0	Johansen Trace Statistic Test		S&L Cointegration Test	
r = 0	137.19	[0.000]	108.43	[0.000]
r = 1	78.41	[0.037]	58.89	[0.061]
r = 2	50.95	[0.092]	34.99	[0.152]
r = 3	27.51	[0.267]	17.62	[0.277]
r = 4	9.97	[0.647]	5.03	[0.566]
r = 5	3.56	[0.493]	0.46	[0.556]

Specification 1b - National price index.

H_0	Johansen Trace Statistic Test		S&L Cointegration Test	
r = 0	144.68	[0.000]	92.14	[0.010]
r = 1	79.75	[0.028]	61.55	[0.036]
r = 2	38.40	[0.557]	33.36	[0.207]
r = 3	18.94	[0.787]	13.96	[0.547]
r = 4	9.96	[0.649]	6.16	[0.421]
r = 5	3.38	[0.523]	0.05	[0.867]

Notes: r is number of cointegrating relations. Johansen trace tests and Saikkonen & Lütkepohl tests are reported for identification of AIDS model. These values are estimated with a $VAR(3)$ with restricted intercepts and no trends in six endogenous variables w_1, w_2, $logp_1$, $logp_2$, $logp_3$ and $log(Y/p)$. Seasonal centred dummies are included. P-values of the are shown in square brackets.

Given the stationary and error serial correlation criticisms of Lewbel and Ng (2005), generally found in static demand models estimated with aggregate data, we show now the patterns of estimated error vectors and related residual serial correlation tests of VECM. Figure 1.5 plots the resulting estimated errors (w_i^*- w_i), with $i = 1, 2$. Consistent with the assumptions of stationarity of errors, a stable dynamic is found. Furthermore, both Q-statistics, adjusted Q-statistics and multivariate LM statistics, Table 1.3, indicate the absence of any significant autocorrelation in the vector of the errors.

Table 1.4 lists the estimated compensated price and expenditure elasticities computed

Table 1.2: Estimated cointegrating vectors and theoretical restrictions imposed

Specification 2a - Regional price index.

	w_1	w_2	$logp_1$	$logp_2$	$logp_3$	Income	Intercept
Vector of cointegration (1)	-1	0	0.011	0.045	-0.056	0.095	0.146
			(0.002)	(0.017)	(0.015)	(0.045)	(0.094)
			[4.938]	[2.644]	[-3.594]	[2.081]	[1.526]
Vector of cointegration (2)	0	-1	0.045	0.037	-0.082	-0.024	0.084
			(0.017)	(0.006)	(0.017)	(0.056)	(0.063)
			[2.644]	[5.954]	[-4.829]	[-0.422]	[1.331]

Theoretical restrictions:	LR test	d.f.	p-value
Symmetry	1.397	(1)	[0.237]
Symmetry and homogeneity	10.390	(3)	[0.015]

Specification 2b - National price index.

	w_1	w_2	$logp_1$	$logp_2$	$logp_3$	Income	Intercept
Vector of cointegration (1)	-1	0	0.012	0.050	-0.062	0.049	0.124
			(0.002)	(0.011)	(0.014)	(0.029)	(0.104)
			[4.909]	[4.657]	[-4.467]	[1.671]	[1.192]
Vector of cointegration (2)	0	-1	0.050	0.032	-0.082	-0.012	0.024
			(0.011)	(0.006)	(0.063)	(0.014)	(0.056)
			[4.657]	[5.693]	[-1.473]	[-0.903]	[0.422]

Theoretical restrictions:	LR test	d.f.	p-value
Symmetry	6.460	(1)	[0.011]
Symmetry and homogeneity	33.754	(3)	[0.000]

Notes: Standard errors in round brackets; student's t-test in square brackets. Degrees of freedom and p-values of LR tests in round and square brackets, respectively.

at the sample means[9].

A few aspects of these estimations should be noted. We estimate negative and large

[9]Typically, one chooses this point to hold concavity because it is the point with the highest sample ''information" and the data are scaled consequently. Asymptotic standard errors of elasticities and confidence intervals are derived from bootstrap replications of the estimated parameters and their standard errors.

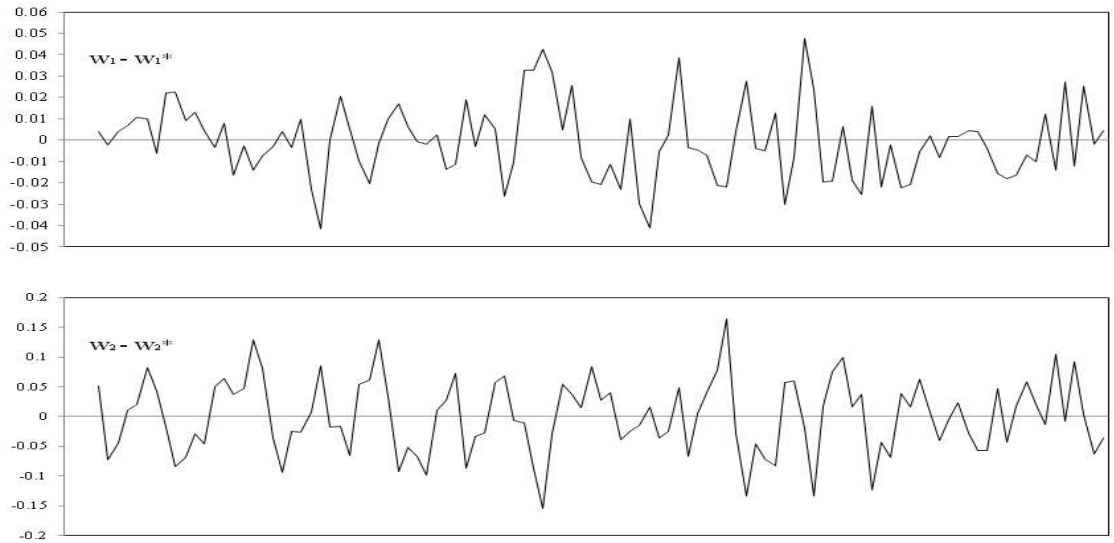


Figure 1.5: Deviations of observed expenditure shares from long-run equilibrium levels

Table 1.3: Residual serial correlation of the vector error correction model

	VEC Residual Portmanteau Tests					VEC Residual Serial Correlation LM Tests	
Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df	LM-Stat	Prob
1	7.646	-	7.719	-	-	46.424	0.114
2	20.253	-	20.571	-	-	42.455	0.213
3	42.478	0.212	43.451	0.184	36.000	37.933	0.381

Notes: Q-statistics and adjusted Q-statistics are shown with small sample correction for residual serial correlation up to specified order h (see Lütkepohl, 1991, for details). Also shown: LM test statistics for residual serial correlation up to specified order (see Johansen, 1995, for details). Under null hypothesis of no serial correlation, statistic tests are χ^2 distributed. Degrees of freedom are $k^2(p-h)$ and k^2, respectively, where k is number of endogenous variables of VAR and p is VAR lag order.

own price-elasticities. These results show that there are no violations of concavity and that consumers' demand for food responds to price changes. It is worth noting that the smaller compensated price elasticity in the residual component is strongly biased downwards by the inclusion of the expenditure categories for bread, pasta and olive oil (although the size is reduced when we compare it with the uncompensated price elasticities). We will return to the robustness of results below, to assess the sensitivity of these estimates. Demand for

Table 1.4: Long-run estimated elasticities of demand system (equation 1.7)

	Hicksian Price Elasticities			Income Elasticities
	(1)	(2)	(3)	
(1) Healthy foods	-0.774	0.402	0.371	1.59
	(0.134)	(0.106)	(0.092)	(0.801)
	[-5.776]	[3.793]	[4.032]	[1.988]
(2) Unhealthy foods	0.536	-0.573	0.037	0.802
	(0.154)	(0.201)	(0.015)	(1.428)
	[3.481]	[-2.855]	[2.466]	[0.560]
(3) Other foods and non-durable goods	0.082	0.006	-0.088	0.902
	(0.021)	(0.004)	(0.029)	(0.824)
	[3.904]	[1.668]	[-3.034]	[1.094]

Notes: Standard errors obtained by bootstrap procedure in round brackets; student's t-test in square brackets.

healthy food is a luxury, whereas unhealthy food is a necessity. The estimated expenditure elasticities are in line with the findings of Zheng and Zhen (2008) in the United States, although a different classification (and habits) were responsible for differences in impact measures. Lastly, the cross-price elasticities show that shifts in healthy or unhealthy food prices influence consumers' choices to substitute the relatively expensive food category for the cheaper one. Estimated cross-elasticities are statistically significant and have well-defined sizes.

However, we cannot make direct inferences regarding substitution effects on patterns of prevalent obesity in Italy because, as reported in Section 2, the prices of both unhealthy and healthy foods rose in the sample period. The asymmetric responses of the elasticities evaluated at the sample means show that a 1 percent increase in the price of healthy food increases the budget share in the unhealthy food category by 0.536 percent, whereas those of unhealthy food increases the budget share of healthy food by 0.402 percent. The implications are threefold. The net effects of changes in food consumption, given shifts in relative prices, indicate a slight but significant impact on the growth of unhealthy

food consumption. According to the net estimated elasticity of substitution, we can also reproduce the data reported in Figures 1.2 and 1.4, in which it was shown that quicker changes in healthy versus unhealthy food prices increased relative unhealthy food consumption and its expenditure share. Lastly, the increases in total calorie intakes and overweight patterns are, therefore, partly explained by the channel of convenient food purchases, less careful to the quality/price policies.

If we focus on the dynamics of the elasticities of substitution, η_{21} and η_{12}, an interesting implication of the non-stationarity of prices is that elasticities may change over time. This point is illustrated in Figure 1.6. The cross-price elasticity for healthy foods, η_{12}, appears to change little over time, whereas a slight recent increase in η_{21} is recorded since the end of 2001, when unhealthy foods became a much larger share of total spending (see Figure 1.4) or a rising category of food expressed in terms of relative quantities (see Figure 1.2).

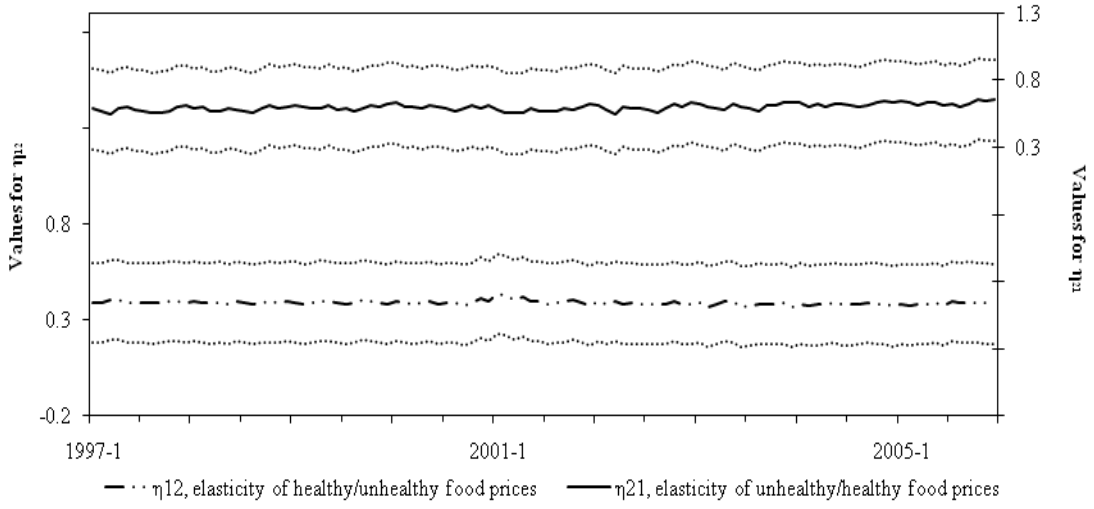


Figure 1.6: Long-run patterns of substitution elasticities of healthy and unhealthy food.

The robustness of our estimates are shown by moving bread, pasta and olive oil into the sectors of healthy (specification I) and unhealthy foods (specification II), respectively. The implicit price index was then used to estimate the parameters of the long-run demand system. Appendix 1.A.3 shows the results. Although these food categories are very demanding in terms of quantity, in views of their cultural importance in the Mediterranean area, estimations are close to the benchmark model. Note that, in both estimations, the dimension of the cross-price elasticities are reduced, revealing how these "new aggregate"

food categories show small changes in own or cross-prices. This result is in line with the low consumption responses to changes in prices of bread, pasta and olive oil founded by Conforti et al. (2001).

Clearly, the impact of changes in relative prices is not equally spread throughout the individuals of a society. As described above, inequality may arise from a heterogeneous consumption response among individuals as relative healthy and unhealthy food prices evolve. This implies a growing disparity in food access.

Table 1.5 lists the results of the long-run demand systems specified according to gender and socio-economic group, obtained by aggregating household expenditures for food and residual categories by RPI.

Focusing attention on the dimension of elasticities of substitution, η_{21k} and η_{12k}, only for individuals who belong to the high education group, we find that the cointegration rank is not exactly identified. Appendix 1.A.4, reports the results of the cointegration tests for each sub-group. It should also be noted that the computed elasticities of substitution for people above the relative poverty line and for the younger age class (age less 35) appear to be not statistically significant.

Besides these exceptions, our estimations contain several points of interest. First, cross-price elasticities for female household heads (0.08) indicate that the (net) effect of substitution of unhealthy foods for healthy foods, given changes in relative prices, is less than half in basis points of the effect for male household heads (0.18). This result partly emerges in Figure 1.3 (first panel), in which the sharp increase in relative healthy versus unhealthy food prices for female household heads does not affect changes in consumption or total calorie intake. Women appear to be able to keep their previous eating habits. These findings are confirmed by the ISTAT (2007) annual report, in which an increase in body weight is mainly found in males[10].

Second, the non-exact identification of the long-run demand system for highly educated individuals does not allow us to discuss the net substitution effects of changes in healthy over unhealthy food prices for this category. However, those who achieved, at maximum,

[10]The results of a greater propensity towards healthy food purchases find indirect confirmation by the greater control of women's weight with respect to those recorded for men, and by the low perception of being underweight of Italian women. These findings are in line with those obtained in France de Saint Pol (2009), Etilé (2007).

the level of secondary school are significant and of the expected (higher) dimensional effect with respect to the average. Individuals in this subgroup respond to changes in the rise of relative prices of healthy foods by consuming more unhealthy foods, 10 basis points more sensitive with respect to the whole sample.

Thus, in line with the results of Ball and Crawford (2005), less educated Italians tend to prefer eating unhealthy foods, with higher calorie intake. The estimated elasticity of substitution is also in libe with the remarkable growth rate of the share of overweight and obesity prevalence for less educated individuals (2.8) with respect to those with higher levels of education (0.8)[11].

Third, among individuals grouped according to age, the insignificant cross-price responses of young adults are in line with data shown in Figure 1.3, in which individual responses of quantities are not very sensitive to changes in relative food prices.

Bearing in mind that healthy food prices rise more than those of unhealthy ones, the other age groups were found to make a key contribution to the (net) elasticity of substitution by favouring the consumption of unhealthy foods. Higher elasticities of substitution for middle and advanced ages with respect to those estimated for the general population, leads to the fact that this group contributes more to body weight increases.

Fourth, the substitution effects towards unhealthy foods which interest people living below the relative poverty line are, remarkably, 6 basis points higher than that of the mean of the population. As found in other countries by Komlos and Baur (2004), the relative price mechanism for adults with lower income passes to specific allocation of their disposable income on food that covers calorie requirements more easily by substituting low-calorie products with cheaper high-calorie foods.

These results may have a consequence on policy-makers' actions. As advocated by some, taxing unhealthy foods to arrest the dynamics of the prevalence of obesity may not only be of limited impact due to low price responses but may also increase the inequality of groups with lower economic or social status. Chouinard et al. (2007) show that a ''fat tax" instrument may be extremely regressive at lower income levels. Thus, the cost of a percentage increase in price by buying an extra unit of unhealthy food is borne to a

[11]It is worth noting that these results obtained for the sample of less educated individuals, were computed by averaging the growth of each group from 1999/2000 to 2005.

greater extent by lower income groups (and lower educated groups).

Conversely, subsidies for healthy foods are likely to be more successful, especially for the disadvantaged subpopulation Lin and Guthrie (2007)[12]. That is, because more indigent groups react more to changes in relative food prices, by targeting subsidies for each Euro spent on healthy foods, participants may not only have monetary benefits but also, as an externality, make an improvement by reducing the prevalence of overweight.

In addition, estimations suggest that consumption income responses for individuals below the relative poverty line is higher for healthy foods by 10 basis points than otherwise assessed so far for representative individuals. Instead, we can project a demand response for unhealthy foods, other foods and non-durable goods which is 4 and 5 basis points lower, respectively, than previously estimated in the full sample model.

These results are crucial when we attempt to address complementary health and fiscal policies. Let us consider the introduction, with the Budget Act for 2009, approved at the end of 2008 by the Italian Parliament, of the ''social card" for indigent people aged over 65 and poor families with children under the age of three. Although the benefits for health are constrained by the limited numbers of potential beneficiaries and the dimension of the subsidy of the program which, in intent, follows that of the US food stamps, this welfare program could be used as a strategy to trigger policies to prevent the rise in the prevalence of overweight and obesity. This social program, which provides 40 euro a month and potentially involves less than the 20 percent of indigent people, may be extended, with very small administrative costs, to individuals below the poverty line, by targeting subsidies for purchasing healthy foods within the social card program - as proposed, for example, by the pilot project of the state of California Guthrie et al. (2007) [13].

[12]An implicit reason exists for the ineffectiveness of unhealthy taxation of energy-dense foods, determined by market competition in developed countries. As discussed by Powell and Chaloupka (2009), the presence of large quasi-competitive non-taxed high-calorie foods sold by groceries can potentially substitute taxed foods, making the impact on individual or aggregate body weight limited or irrelevant.

[13]The Italian Minister for Economy and Finance stated that in 2007, there were about 1,300,000 indigent people which were eligible for the ''social card", and over 7,400,000 poor individuals were estimated to be below the poverty line in the same year. A first ex-post evaluation reports that only 42 percent of indigent people complied with the ''social card" program.

Table 1.5: VECM estimates, diagnostics and elasticity of substitution by sub-samples

VARIABLES		VAR SPECIFICATION		THEORETICAL RESTRICTIONS		VECM DIAGNOSTIC	ELASTICITY OF SUBSTITUTION		
		Lag order	Rank of cointegration	Symmetry	Homogeneity and symmetry	Residuals serial correlation	η_{21}	η_{12}	Net effect $\Delta(\eta_{21} - \eta_{12})$
Gender	Male	3	2	3.484 (1) [0.062]	10.840 (3) [0.012]	39.660 [0.309]	0.533 (0.063)	0.354 (0.044)	0.178
	Female	3	2	10.723 (1) [0.001]	3.517 (3) [0.318]	39.280[0.325]	0.222 (0.021)	0.140 (0.013)	0.081
Education	High	2	3				not significant		
	Low	3	2	0.237(1) [0.625]	9.693 (3) [0.021]	38.828[0.343]	0.689 (0.201)	0.456 (0.127)	0.233
Relative poverty line	Above	3	2				not significant		
	Below	4	2	0.170 (1) [0.679]	11.91 (3) [0.007]	47.711[0.092]	0.591(0.189)	0.394 (0.124)	0.196
Age	< 35	2	2				not significant		
	$35 - 49$	4	2	0.509 (1) [0.475]	12.438 (3) [0.006]	76.077[0.348]	0.393 (0.192)	0.267 (0.130)	0.126
	$50 - 64$	3	2	1.878 (1) [0.170]	2.947 (3) [0.407]	38.043[0.376]	0.543 (0.064)	0.359 (0.043)	0.184
	> 65	3	2	0.133 (1) [0.714]	6.583 (3) [0.086]	52.453[0.037]	0.321 (0.039)	0.201 (0.024)	0.119

Notes: High education stands for people that achieved a degree, master or PhD. Low education for the others. The criterium used to select the optimal lag order is the sequential modified LR test statistic. The test of the rank of cointegration of the unrestricted VAR use the Johansen's procedure. Cointegration test inferences of the AIDS model applied to each subsample are reported in Appendix. Degree of freedom and p-values of the LR tests for the theoretical restrictions are reported in round and square brackets, respectively. Diagnostic autocorrelation test and estimations of substitution elasticities are obtained by imposing symmetry and homogeneity in the VECM. Q-statistics distributed as a χ^2 is reported with the p-values in square brackets. The degree of freedom are 36 for VAR lag order of 3 and 72 for lag order of 4. In round brackets are reported bootstrap standard errors of the estiamated elasticities.

APPENDIX 1.A.1

Table 1.A.1: Food classification

Healthy foods	**Unhealthy foods**
Meat	**Ice creams and Sweets**
Beef (lean)	Sugar
Poultry	Jam, marmelade, chocolate
Other low-fat meats	Ice creams
Fish	**Meat and cold cuts**
Fresh or forzen fish	Cold cuts
Preserved dry smoked fish	Pork
Fresh or frozen shellfish and other seafoods	Other fatty meats
Other fish	
Vegetables and legumes	**Oils and fat**
Dry or tinned legumes	Vegetable oil (except olive oil)
Dry or tinned vegetables	Butter
	Lard
Fruit	
Fresh fruit	**Milk**
Tinned fruit	Whole milk
Preserved fruit	Other whole-milk derivatives

Notes: Monthly price data at national and regional levels are available from ISTAT.

APPENDIX 1.A.2 Identification of long-run AIDS

The cointegration relationships in the VECM equation (1.6), subject to reduced rank restrictions on the $\Pi = \alpha\beta^{*'}$ matrix, are not identified. Following Pesaran and Shin (2002), the identification of the long-run parameters in $\beta^{*'}$ requires the imposition of r restrictions on each cointegrating vector, although a necessary and sufficient condition (order condition) for the identification is that the number of the identifying restrictions, k, should be at least equal to r^2.

In order to explain these fundamental identifying conditions in our demand system with three categories of goods, we note first that adding up reduces the rank to two, i.e. $r = (n-1)$. As a formal extension of the ECM vectors in equation (1.6), let us consider two non-identified cointegrating vectors made up of the variables w_{1t}, w_{2t}, lnP_{1t}, lnP_{2t}, lnP_{3t}, $ln(y_t/p_t)$ and the intercept. The associated parameters are:

$$\widetilde{\beta}'_U = \left\{ \begin{array}{ccccccc} \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} & \beta_{51} & \beta_{61} & \beta_{71} \\ \beta_{12} & \beta_{22} & \beta_{32} & \beta_{42} & \beta_{52} & \beta_{62} & \beta_{72} \end{array} \right\} \tag{1.A.1}$$

The exact identifying restrictions $r^2 = (n-1)^2 = 4$ assume a diagonal structure because theory suggests that budget shares responds mainly respond to own and cross-price changes and income impulses, but not to (endogenous) changes in other budget shares. Formally,

$$\left\{ \begin{array}{cc} \beta_{11} = -1, & \beta_{12} = 0 \\ \beta_{21} = 0, & \beta_{22} = -1 \end{array} \right\} \tag{1.A.2}$$

so that the cointegrating vectors may be written as:

$$\widetilde{\beta}' = \left\{ \begin{array}{ccccccc} -1 & 0 & \beta_{31} & \beta_{41} & \beta_{51} & \beta_{61} & \beta_{71} \\ 0 & -1 & \beta_{32} & \beta_{42} & \beta_{52} & \beta_{62} & \beta_{72} \end{array} \right\} \tag{1.A.3}$$

In order to test theoretical restrictions, long-run parameter restrictions should be included. As discussed in the text, the property of symmetry may be imposed as a cross-equation restriction, $\beta_{32} = \beta_{41}$. The cointegrating vectors thus assume the following structure:

$$\widetilde{\beta}'_S = \left\{ \begin{array}{ccccccc} -1 & 0 & \beta_{31} & * & \beta_{51} & \beta_{61} & \beta_{71} \\ 0 & -1 & \beta_{32} & \beta_{42} & \beta_{52} & \beta_{62} & \beta_{72} \end{array} \right\} \tag{1.A.4}$$

Estimations of cointegrating vectors subject to symmetry is tested by the LR statistic distributed as a χ^2 with one degree of freedom. This restriction is not rejected when the loglikelihoods of this restricted model is compared with the exact-identified model in equation (B.3), the differences are not significant.

Lastly, as suggested by the demand theory, we impose and test in the cointegration vectors the properties of both symmetry and homogeneity. In addition to the symmetry restriction, $\beta_{32} = \beta_{41}$, the restriction of homogeneity for each equation is added, that is, $(\beta_{31} + \beta_{32} = -\beta_{51})$ and $(\beta_{32} + \beta_{42} = -\beta_{52})$. Thus, the cointegration vector is given as:

$$\widetilde{\beta}'_{SH} = \left\{ \begin{array}{ccccccc} -1 & 0 & \beta_{31} & * & * & \beta_{61} & \beta_{71} \\ 0 & -1 & \beta_{32} & \beta_{42} & * & \beta_{62} & \beta_{72} \end{array} \right\} \tag{1.A.5}$$

As shown in the text, the LR statistic, distributed as a χ^2 with three degrees of freedom, is then used to test these joint theoretical restrictions.

APPENDIX 1.A.3

Table 1.A.3: Sensitivity analyses of estimated elasticities from the demand system

	Hicksian Price Elasticities			Income
	(1)	(2)	(3)	Elasticities
Specification A				
(1) Healthy foods + bread, pasta and olive oil	-0.547	0.262	0.285	1.091
	(.091)	(.072)	(.012)	(.601)
(2) Unhealthy foods	0.331	-0.523	0.192	0.852
	(.082)	(.181)	(.085)	(1.502)
(3) Other foods and non-durables goods	0.242	0.086	-0.328	1.022
	(.131)	(.046)	(.179)	(.075)
Specification B				
(1) Healthy foods	-0.657	0.281	0.376	1.381
	(.099)	(.084)	(.018)	(.642)
(2) Unhealthy foods + bread, pasta and olive oil	0.327	-0.253	0.026	0.632
	(.144)	(.076)	(.051)	(.899)
(3) Other foods and non-durables goods	0.187	0.056	-0.243	1.101
	(.102)	(.031)	(.125)	(.080)

Notes: in the specification A, bread, pasta and olive oil is moved from other food and non-durables goods to healthy food while in the specification B, bread, pasta and olive oil is moved to unhealthy food. Standard errors obtained by bootstrap procedure are shown in round brackets.

APPENDIX 1.A.4

Table 1.A.4: Johansen's Cointegration Rank Test Statistics for AIDS system

	Gender				Education				Income				
	Male		Female		High		Low		Above pov. thresh.		Below pov. thresh.		95% critical values
H_0	LR		LR		LR		LR		LR		LR		λ trace
r = 0	157.80	[0.000]	136.02	[0.000]	162.69	[0.000]	132.75	[0.000]	136.50	[0.000]	126.97	[0.000]	103.68
r = 1	90.89	[0.002]	75.98	[0.058]	106.86	[0.000]	77.64	[0.043]	76.37	[0.050]	76.20	[0.058]	76.81
r = 2	39.74	[0.488]	41.84	[0.386]	56.88	[0.026]	50.76	[0.095]	49.22	[0.126]	46.74	[0.193]	53.94
r = 3	19.40	[0.761]	21.48	[0.633]	27.62	[0.262]	25.72	[0.362]	27.42	[0.279]	21.91	[0.604]	35.07
r = 4	9.95	[0.649]	9.77	[0.667]	11.56	[0.497]	9.59	[0.683]	9.08	[0.730]	11.61	[0.492]	20.16
r = 5	3.44	[0.512]	3.68	[0.473]	2.27	[0.724]	3.86	[0.445]	3.81	[0.453]	4.21	[0.394]	9.14

	Age								
	<35		35-49		50-65		>65		95% critical values
H_0	LR		LR		LR		LR		λ trace
r = 0	127.01	[0.000]	134.93	[0.000]	167.21	[0.000]	131.27	[0.000]	103.68
r = 1	80.33	[0.025]	81.26	[0.021]	79.98	[0.027]	79.82	[0.028]	76.81
r = 2	45.46	[0.236]	49.86	[0.112]	51.58	[0.081]	45.43	[0.237]	53.94
r = 3	19.85	[0.735]	27.57	[0.264]	30.17	[0.158]	23.84	[0.477]	35.07
r = 4	8.21	[0.802]	12.42	[0.420]	10.16	[0.629]	9.94	[0.650]	20.16
r = 5	2.62	[0.659]	5.05	[0.288]	3.51	[0.500]	3.86	[0.444]	9.14

Notes: r is number of cointegrating relations. We report Johansen trace tests for evaluating the identification condition of AIDS model. These values are estimated by optimal length lags with restricted intercepts and no trends in six endogenous variables w_1, w_2, $logp_1$, $logp_2$, $logp_3$ and $log(Y/p)$. Seasonal centred dummies are included. P-values of tests are shown in square brackets.

Chapter 2

The Effect of Smoking Bans on Weight in Italy

2.1 Introduction

This chapter examines the causal effects of smoking behavior on body weight in Italy. In 2005, the Italian government introduced a smoking ban in all indoor public places. We use a regression discontinuity design, which exploits this exogenous variation across cohorts to achieve identification in our model. Our estimates indicate that the smoking ban reduced cigarette consumption and the smoking participation rate. Most interestingly, we estimate a significant, although not very large, effect of nicotine reduction on weight increases. Heterogeneous effects are also estimated, with smaller impact on men and employees and, conditionally on BMI distribution, overweight and obese people.

2.2 Related literature

2.2.1 Clean indoor air laws: direct effects on smoking and business

Although isolated examples of policies restricting smoking have been recorded earlier, the 1973 law in Arizona was the first state intervention achieving smoke-free aims in a number of public places. This law certified that ''nonsmokers had as much a right to clean air

and wholesome air as smokers had to their so-called right to smoke". The motivations for state intervention, thereafter flexibly applied during the 1970s to other US states, were followed by the Surgeon General's Report[1], which emphasized the adverse health effects of passive smoking on public health.

With the exception of Finland[2], European Clean Indoor Air Laws are relatively more recent. This explains why most studies of the economic impact of smoking bans are based on US data, as more detailed European smoke-free interventions adopting comprehensive smoking bans in virtually all public places and private worksites followed from the US experience.

This legislation not only protected non-smokers from the dangers of passive smoking, but also encouraged smokers to quit or reduce their consumption. However, until now, there has been limited evidence about the consequences of non-smokers' exposure to cigarette smoke. Several studies have shown that these laws help prevent young people from starting smoking, reduce the number of cigarettes smoked, and encourage some smokers to quit. One classic health economics paper analyzing the effects of cigarette consumption and the prevalence rate of smoking bans is that of Chaloupka and Saffer (1992). The authors, according with Evans et al. (1999), emphasize that, while prohibiting smoking at the work place is effective in reducing prevalence and consumption, these effects are much smaller than those obtained when restrictions are introduced only in public places, because in the first case the amount of time spent without being able to smoke is greater.

However, more detailed contributions about the effects of smoking regulation appear in the epidemiological literature. In a survey conducted in Ireland after the 2004 introduction of smoke-free legislation, Anonymous (2005) and Fong et al. (2006) found that, among Irish smokers who quitted after the ban, 80% reported that the law had helped them to quit and 88% that it had helped them to remain non-smokers. Gallus et al. (2006), evaluating the 2005 Italian law for smoke-free public places, estimated that betweeen 2004 and 2006 smoking prevalence decreased by 1.9% and that the daily number of cigarettes

[1]US Department of Health and Human Services (1972).

[2]The Finnish Tobacco Control Act (TCA) was first implemented in 1976 and progressively extended in more recent years.

decreased by 9.5%[3]. These results are in line with those found in other countries, such as the USA, Australia, Canada and Germany (Fichtenberg and Glantz, 2004). Note that this literature also emphasized differences by gender. For example, Chaloupka (1992) finds that clean indoor air laws are more effective for male smokers.

Another strand of the literature indicates that smoke-free policies and regulations have been used to evaluate the negative economic impact on the catering and hotel industries, a position largely supported by the tobacco industry (Scollo et al., 2003). Following the reviews of Scollo et al. (2003) and Eriksen and Chaloupka (2007), the position of the tobacco industry was clearly contradicted by almost all the scientific evidence, indicating that there was no negative economic impact of clean indoor air policies on revenues for bars, pubs and restaurants. This evidences is in line with that presented in the 2006 Surgeon General's Report by the US Department of Health and Human Services (2006). Eriksen and Chaloupka (2007) also review studies, finding a positive effect of smoke-free restaurant and *café* laws on economic activity, employment and revenues[4]. In the European context, the issue is still under debate and conflicting evidence on the subject is recorded. Italian data showed higher numbers of people frequenting restaurants and *cafés* after the implementation of the smoking ban, a prediction in line with the US context (Gallus et al., 2006). While, a drop in the number of customers, at least in the short-run, was observed after the introduction of the clean indoor air law in Scotland (Adda et al., 2006).

2.2.2 Smoking reductions, BMI and obesity

The medical perspective

Why should quitting (or reducing) smoking increase body weight? The medical literature gives two main reasons: (1) a direct change in metabolic rates; (2) a life-style change in food consumption. In this section, we briefly discuss these reasons. We anticipate that these effects have been measured and discussed so far mainly for quitters, in which the

[3]Gallus et al. (2006) also found that the drop in smoking prevalence and consumption is particurarly significant for younger generations.

[4]See, for example Luk et al. (2011) for the positive findings of the smoke-free law on the restaurant and bar sales in Ottawa, Canada.

addictive effects of nicotine consumption have been found to affect weight generally with extended consequences to other related health outcomes.

Quitting smoking may increase body weight because changes in nicotine assumption produce effects on human metabolism. Thus, the observed weight increase generally recorded after quitting, as already discussed in Keys et al. (1966), Karvonen et al. (1959) and Higgins (1967), does not turn out to be very large when empirical analysis also accounts for dietary habits. In particular, Grunberg (1985), Klesges et al. (1989), French and Jeffery (1995) explained that the weight gap between smokers and non-smokers was entirely due to differences in metabolic rates and to the more efficient ability of smokers to burn calories during the day. This result was also confirmed by the regularity with which the heart may beat 10-20 more times per minute after a cigarette has been smoked, whereas, after quitting, the metabolic rate slows down and returns to its average level (for pioneering studies, see Dill et al. (1934), Jacobs et al. (1965) and Glauser et al. (1970)). Thus, reductions in cigarette consumption may lead to gains in terms of smoke-related illnesses, with additional indirect positive effects on body weight[5].

Secondly, quitting smoking is often associated with changes in eating habits and preferences. A common symptom after quitting is an increase in food intake, which affects weight for longer than other symptoms. Increased appetite has traditionally been attributed to the fact that eating is a substitute for smoking; eating or snacking is similar to the action of smoking and can be used as a means of oral gratification Jacobs and Gottenberg (1981). In addition, preferences may also change. The sense of taste and smell return to be close to those of non-smokers, implying a larger propensity to move toward unhealthy food, which usually has higher calorie contents (Drewnowski and Darmon, 2005). Conversely, psychiatric studies have shown that quitters tend to be less depressed and exhibit fewer negative effects when they successfully quit than subjects who continue smoking (Cinciripini et al., 2003). Emotional states have therefore been associated with both weight loss and weight gain (Barefoot et al., 1998, Wurtman, 1993), and this may partly explain the variability of results and possible unexpected findings.

[5]This prediction matches the findings of Sargent et al. (2004) and Juster et al. (2007).

Smoking habits and BMI: estimates

One strand of economics literature has tested the hypothesis that stopping (or reducing) smoking causes weight gain, generally measured in terms of BMI. As a special focus, these works evaluated this relationship in the obese sub-group. For example, Chou et al. (2004) produced two new important perspectives using: (i) a large dataset for the US, the Behavioral Risk Factor Surveillance System (BRFSS), which includes health and socio-economic variables; (ii) and state-wide policies as a proxy instrument for smoking habits. Their estimates showed a positive relationship between cigarette prices and body weight, indicating that a decrease in smoking is responsible for increased obesity rates. Conversely, Gruber and Frakes (2006) found an adverse relationship between cigarette taxes and BMI, with a non-significant effect on obesity. Lastly, Baum (2009), after carefully controlling for state-specific time trends, suggested that both cigarette taxes and prices had positive effects on BMI and obesity prevalence.

Although these studies are comparable as regards the dataset used and state-wide policy instruments, there is no agreement in terms of the magnitude of the effects, especially in obese individuals. Clearly, the limitation emerging from the analyses is that state cigarette prices (or taxes) contain few within-state variations and are highly collinear with state dummy variables. This implies that their results may not only have high standard errors for the estimated coefficients, but may also be affected by unobserved confounders. Since smoking is an (individual) endogenous decision, the estimated parameters may turn out to be biased, since unobserved factors may be correlated with other risk behavior also affecting body weight (Aristei and Pieroni, 2010, Viscusi and Hersch, 2001)[6]. To clarify the importance of this point, we refer to the meta-analysis of Klesges et al. (1989), according to which weight gains range from 0.2 to 8.2 kg that are mainly explicited in the short-term. Klesges et al. (1989) estimate that, six months after quitting, body weight increases on average between 2 and 5 kg. However, Courtemanche (2009) shows that a rise in cigarette prices may also lead to a long-term reduction in body weight, questioning the positive effect of smoking reduction in weight gains.

In order to control for these issues, Liu et al. (2010) employed workplace smoking bans

[6]Reverse causality is also plausible, since overweight people choose smoking as a method of weight control (Cawley et al., 2004).

at US state level as an instrument for showing reduction in smoking or participation rate. The idea that we share with this study is using the discontinuity introduced by the policy intervention as a natural instrument. The anti-smoking ban is exogenous since: (i) it is enforced by the state, and (ii) is a universal program. With data from the BRFSS for the years 1998—2006, the above authors found that, when compared with IV estimates, OLS underestimated the impact on BMI of all and obese individuals. The magnitude of OLS estimates indicated that current smokers had between 1.2 (fewer controls) and 1.8 (more controls) lower BMI, than never and former smokers, but these more than doubled when the estimates were carried out with the IV estimator [-3.6 (fewer controls) and -4.2 (more controls)]. Although these results provided evidence that reducing smoking leads to a rise in body weight, they could not estimate the effect for a fixed post–ban period, because in the BRFSS data (both pooled or pseudo-panel) smoking bans were implemented at different periods for each state and thus did not have an homogeneous comparable period. Thus, the interpretation of estimated parameters should be related to a weighted-time post-reform outcome, an issue which limits comparisons of weight effects obtained from randomized evaluation programs in the medical literature (see, for example, Eisenberg and Quinn (2006)).

This work contributes to the literature reviewed in three ways. First, we explain weight gains by reductions in nicotine assumption determined either by quitters or smokers who reduce their cigarette consumption, on the basis of the trade-off between the expected benefits and the fixed costs associated with quitting or reducing smoking. The former behavior is usually particularly associated with high costs due to nicotine addiction and withdrawal[7]. Second, we estimate the effects of smoking habit changes on weight in the short term (i.e., after one year) and medium term (after three years), so that our quasi-experiment analysis is consistent with time different responses generally investigated in the medical framework. Third, we add empirical evidence of the causal effect of smoking on body weight. Despite a growing literature, in European countries little is known to what extent individuals' smoking choices affect body weight and even less about causal interpretations of model parameters.

[7]This theory refers to the rational addiction model of Becker and Murphy (1988), in which the development of nicotine dependence can be characterized in terms of tolerance, reinforcement and withdrawal effects.

2.3 Framework of the analysis

We focus attention here on the need to draw causal estimates of the effect of smoking on body weight. As in Liu et al. (2010), we isolate the effect of smoking from other confounding factors which may also affect BMI by using a structural model. In subsection 3.2, we motivate the use of the regression-discontinuity design from the evidence of significant differences in BMI and smoking indicators before and after the implementation of the ban.

2.3.1 A model for the causal relationship between smoking and BMI

Estimating the magnitude of the causal effect of reducing smoking on weight is a non-trivial challenge. In particular, many empirical studies have documented a negative association between cigarette consumption and BMI, although it is not clear whether BMI increases are also unintended consequences of reducing smoking and if so, what portion of this increase may be attributed to them (Nonnemaker et al., 2009). At this stage, we write a simple reduced form equation for a direct estimation of the relation between smoking variables (S_{it}) and body mass index (BMI_{it}) for each individual i at time t. That is:

$$BMI_{it} = \gamma_0 + \gamma_1 S_{it} + \sum_{j=0}^{J} \psi_j X_{jit} + \epsilon_{it}; \tag{2.1}$$

where X_{jit} is a set of j control variables and ϵ_{it} is an error term.

However, identification of the causal effect requires being able to control for heterogeneity in individuals' smoking and weight choices, so that we are sure that the estimated effect on BMI is not correlated with personal or social factors. If individuals' unobserved characteristics influence BMI changes, as well as smoking behavior, then least square estimates (OLS) of γ_1 will be biased. That is, unobserved variables affecting smoking behaviors may also be correlated with those influencing decisions to change body weight. For example, γ_1 may be estimated to be negative, although smoking variable S has no causal effect on BMI. In this case, OLS estimates of γ_1 would be confounded by the existence

of omitted variable bias. One would expect, for instance, that quitters (or individuals who reduce smoking) are more likely to adopt some other behavior such as eating more, which may increase BMI. Thus, we may observe a negative correlation between smoking and BMI, even when we do not have any causal effect between variables.

To estimate causal parameters of the investigated relationship, we need of an IV setting (Hahn et al., 2001). The implementation of clean indoor air laws allows us to identify the relationship between smoking behavior and body weight within a regression discontinuity design[8]. According to the 2005 Italian law, smokers are not permitted to smoke in public and sometimes not even in private places. Therefore, individuals of the same age living in pre- and post-ban periods experience different smoking restrictions, or treatments, and assignment to treatment was only determined by individuals' birth cohort.

However, unlike other models in the health literature already described in Section 2, we use a structural model derived from combination of the reduced forms. Identification is achieved by including a dummy variable (SB) in the first-stage equations of smoking behavior and BMI, to record the exogenous change in smoking introduced by the law. We define the smoking ban dummy variable as one for cohorts of individuals interviewed after the introduction of the smoking ban and zero for individuals of the same age but belonging to cohorts interviewed before the introduction of the smoking ban discontinuity. The discontinuity is generated for individuals belonging to each birth cohort after the year in which the ban took effect. The relevant reduced forms for smoking variables and BMI are the following:

$$BMI_{it} = \beta_0 + \beta_1 SB + \sum_{j=0}^{J} \phi_j X_{jit} + u_{it}; \tag{2.2}$$

and

$$S_{it} = \delta_0 + \delta_1 SB + \sum_{j=0}^{J} \psi_j X_{jit} + v_{it}. \tag{2.3}$$

We then follow the structural model proposed by Machin et al. (2011)[9] to derive BMI causal estimates. Formally, this is given as:

[8]Classic references are Trochim (1984, 2001) and Trochim and Campbell (1960).
[9]The authors develop a structural model of crime reducing effect of education.

$$BMI_{it} = \theta_0 + \theta_1 S_{it} + \sum_{j=0}^{J} \sigma_j X_{jit} + \eta_{it}; \tag{2.4}$$

where IV estimates of the coefficient on the BMI variable in equation (2.4) is the ratio of the reduced form coefficients in (2.2) and (2.3), $\theta_1 = \beta_1/\delta_1$.

This strategy identifies the average causal effect for those individuals (smokers or quitters after the ban), subjected to restrictions in smoking by the virtue of the ban, and allows us to estimate the local average treatment effect (LATE)[10]. Note that, the variation induced by the instrument is local in nature, as it has an impact only for smokers who quit or reduced smoking in the post-ban period.

Thus, the estimated effect according to our empirical approach is obtained by variations in smoking habits and BMI of those subjects who alter their status (treatment), because they react to the ban (instrument). This implies two important consequences in the estimation process. First, IV estimates exceed the OLS ones, at least because the instrument used is based on a policy intervention which only affects the choices of the smoking group. The ban does not affect smokers who continue to smoke or non-smokers. Second, we can calculate the full contribution of nicotine reduction in affecting body weight because, in addition to the quitters, there are individuals who may reduce cigarette consumption. For this reason, we consider two different smoking indicators: number of cigarettes smoked, and participation rates, to examine respectively the effect of changes in nicotine consumption and quitting smoking.

2.3.2 Data and related empirical issues

The dataset used in this chapter is the ELA survey, conducted in Italy by the Italian Statistical Institute (ISTAT). The ELA survey is a representative cross-section sample of the Italian population and provides detailed information on the demographics, social characteristics and health of 20,000 households each year, corresponding to approximately 50,000 individual records yearly.

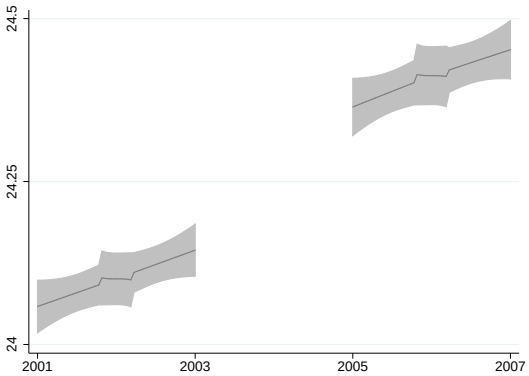
For the aims of the present study, the importance of this survey lies in the detailed section devoted to analysis of current and past smoking habits of individuals aged 18 and

[10]See Imbens and Angrist (1994). For an application see Angrist (1995).

over. In particular, we focus on individual smoking behavior, in terms of both participation rate and cigarette consumption. Contrary to the ISTAT ''Italian Household Budget Survey", ELA provides information on individuals rather than on households, allowing through analysis of socio-demographic and gender effects, without approximating them with the characteristics of the household head.

We used six rounds of this survey, corresponding to the years 2001 to 2007 (excluding 2004, for which data were not available). We selected a sample of individuals aged from 20 to 60 years in the pre-ban period 2001 - 2003, and compare them with individuals of the same age in the post-ban period, 2005 - 2007. Thus, we can compare smoking and BMI patterns for individuals of the same age around the discontinuity generated through the ban, where pre-ban periods observations constitute a proper counterfactual, with the most similar observable and unobservable characteristics, for treated individuals.

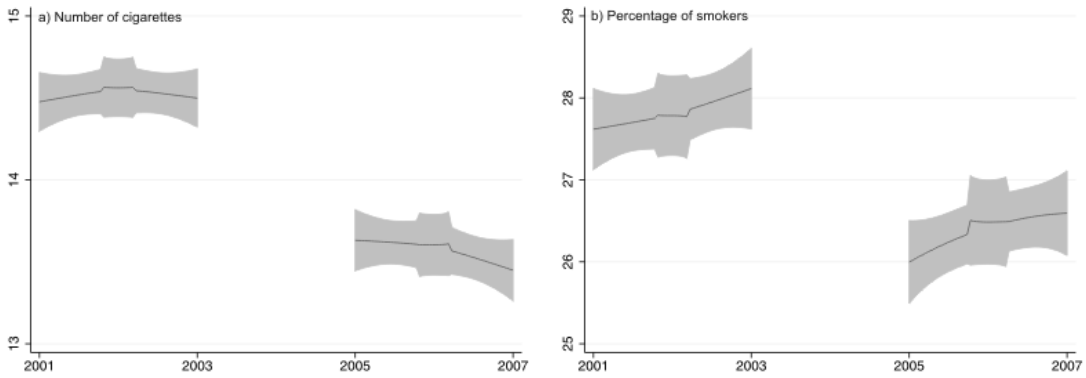
Following Hahn et al. (2001) and Oreopoulos (2006), to guarantee that we correctly estimate the causal impact of nicotine reduction on BMI, one condition is that the average effect of key variables is not null around the discontinuity. We illustrate descriptively the average effect of the smoking ban using the mean of total nicotine consumption (e.g., number of cigarettes) and the percentage of smokers for the years available in our sample. We complement the descriptive analysis of BMI variables around the discontinuity (see plot in Figure 2.1). We see that, although BMI seems to make a significant jump in the ban year (2005), it then continues along its (positive) long-run growth path.



Notes: BMI values for each year were obtained aggregating individual level data in pre- and post-reform periods. The solid line is estimated by the standard Kernel function.

Figure 2.1: BMI discontinuity around the clean indoor air law.

Figure 2.2 (panel a) shows the average number of cigarettes smoked and (panel b) the average percentage of smokers per year. Also in this case, we find a clear deviation from the long-term pattern of these variables in the year of the ban. The mean number of cigarettes smoked felt from 14.5 to almost 13.5, whereas the percentage of smokers decreased by almost 2 percentage points near the discontinuity. These findings indicate that the reform likely played a significant role in reducing smoking, considering both cigarette consumption and percentage of smokers.



Note: Estimations for number of cigarettes and percentage of smokers are obtained as in Figure 2.1.

Figure 2.2: Smoking discontinuities around the Italian clean indoor air law (2005).

Clearly, there are a number of other issues involved in the analysis between smoking behavior and BMI. First, without the Italian ban, the average BMI for smokers would have undergone the same variation as that for non-smokers. This assumption may be implausible, if treated and control subjects are unbalanced in the covariates which are believed to be associated with heterogeneity in unobservable characteristics. We therefore also include a set of control variables for gender, education, employment status, physical activity and job strenuousness (see Table 2.1).

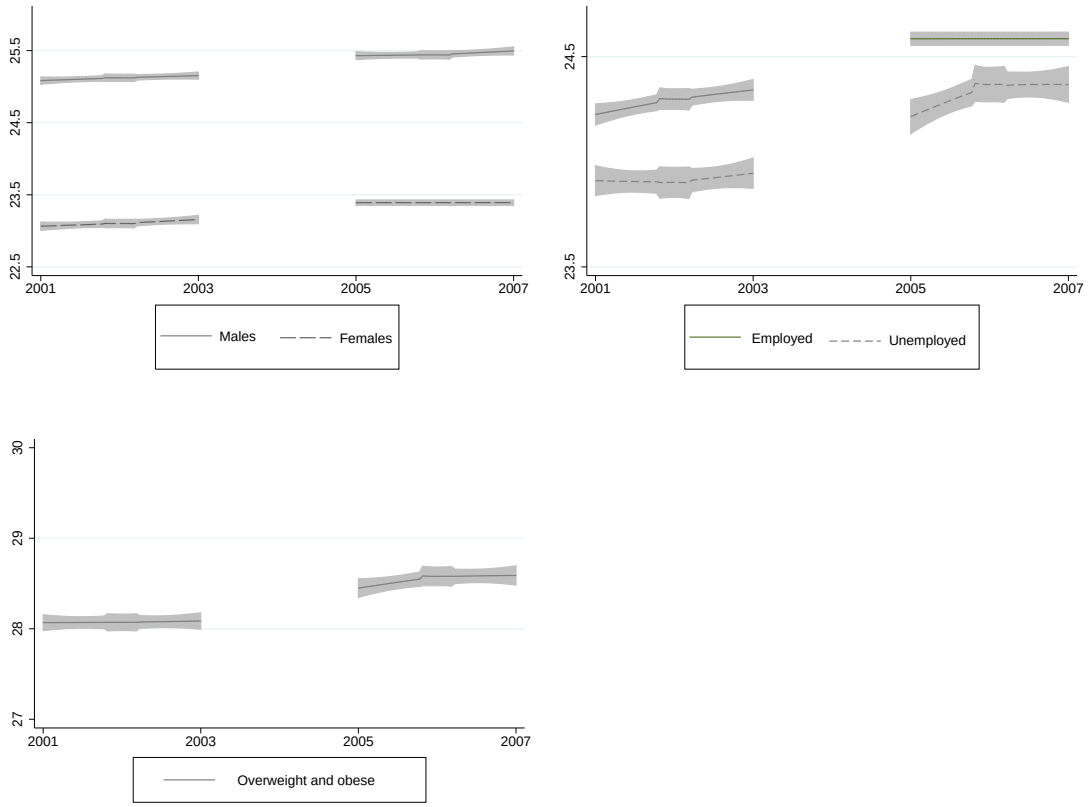
In particular, the literature findings discussed in Section 2 indicate that the smoking ban affected differently cigarette consumption through gender and occupation, and that these changes heterogeneously influenced the BMI distribution (Figure 2.3). Although women (dashed line) have lower BMI than men (solid line), both recorded significant changes in 2005. According to the descriptive statistics of Table 2.1, the greater response in terms of smoking habits after the ban was associated with men rather than women. In addition, although a larger impact in cigarette consumption affected the employee group

rather than the unemployed one, BMI changes in these groups do not appear to be very different. This implies that the effect of smoking on BMI for these sub-groups is not predictable a priori and should be further investigated empirically.

In section 4, we also extend evaluation of the impact of the smoking ban on BMI to these specific sub–groups. We also analyze individuals located at the top of the BMI distribution (e.g., overweight and obese people), who represent a particular risk group in terms of public health policies. We devote our attention to the quantile treatment effect at the mean of overweight and overweight and obese groups, corresponding respectively to the 77^{th} and 81^{st} percentiles of the BMI distribution. Note that, following Imbens and Rubin (1997), consistent estimates of the quantile treatment effect can be obtained under the LATE identifying assumption (and regression discontinuity design) from the marginal distribution of potential outcomes of smokers. In Table 2.1, the results indicate significant changes in nicotine consumption (and quitters) whereas non-parametric estimates show a very small BMI variation for such sub-groups after the introduction of the smoking ban (Figure 2.3).

Second, the identification based on the discontinuity introduced by the ban exploits the assumption that individual BMI differences at the same age are only attributable to changes in smoking behavior, whereas all the other weight determinants are stable. This assumption does not exclude the possibility that smokers and non-smoker may have heterogeneous BMI across different birth cohorts, before and after the discontinuity. Figure 2.4 shows a 5-year cohort reconstruction for the key variables of our model. Each line represents the evolution of a different indicator for individuals belonging to the same cohort, showing that individuals belonging to younger cohorts tend to have higher BMI (and consequently weight) than those belonging to older ones, age being constant. Thus, to control for these inter-generational differences, we include $n-1$ dummy cohorts in our model.

Third, the estimates of the impact of changes in smoking behavior on BMI may be sensitive to how distant the data are from the discontinuity. In line with the results of the review by Eisenberg and Quinn (2006) this relationship assumes a concave form whether short term responses of smokers are more sensitive in terms of weight gains. Under this hypothesis, our baseline model is estimated with the RD design (2003-2005), in which the



Note: Estimations for BMI patterns of the subsamples are obtained as in Figure 2.1

Figure 2.3: BMI discontinuities around the clear indoor air law, by subgroups.

structural parameters from this sample (one year post-reform variation) may be greater than those obtained from the whole sample, which considers a three years post-reform (2001-2003 compared to 2005-2007).

In the robustness section, we also provide estimates for the whole period by means of an estimator that weights observations inversely to distance (i.e., IDW, inverse distance weighted) of each year from the smoking ban to eventually investigate the magnitude of unobserved heterogeneity.

2.4 Results

2.4.1 IV and RD estimates

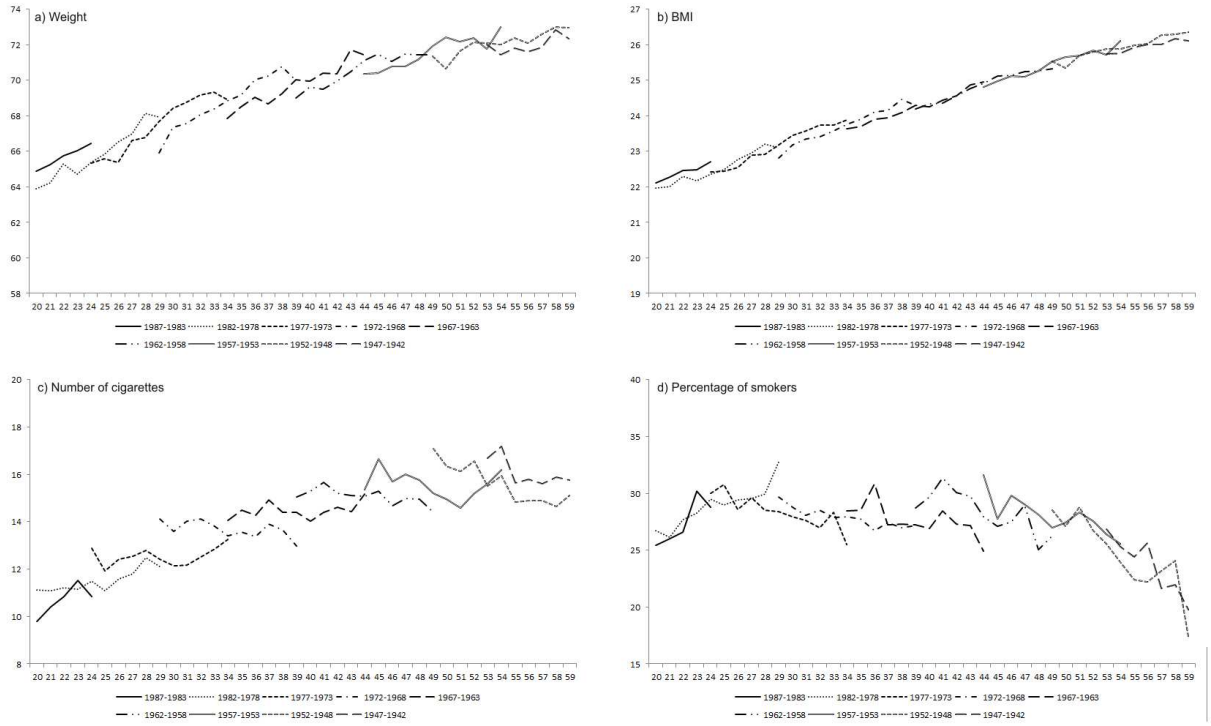
In this section, we present the results of our empirical analysis. Table 2.2 lists the estimated coefficients of the reduced form for equation (2.1) and the structural framework

Table 2.1: Variables definition and descriptive statistics

			Percentage of smokers				Cigarettes consumption (nr.)			
	Category		whole sample		discontinuity		whole sample		discontinuity	
		%	2001-03	2005-07	2003	2005	2001-03	2005-07	2003	2005
Covariates										
Gender	Male	0.49	0.35	0.33	0.35	0.32	16.31	15.16	15.88	15.03
	Female	0.51	0.22	0.21	0.22	0.2	11.95	11.17	11.61	11.07
Occupation	Employed	0.65	0.31	0.29	0.31	0.28	15.09	13.91	14.65	13.83
	Unemployed	0.35	0.23	0.23	0.23	0.21	13.51	12.78	13.12	12.57
Education	Degree or more	0.54	0.25	0.24	0.25	0.24	13.39	12.23	13	12.1
	Secondary or less	0.46	0.31	0.3	0.32	0.29	15.74	14.92	15.3	14.81
Physical activity	No	0.8	0.29	0.28	0.29	0.27	15.11	14.05	14.69	13.91
	Yes	0.2	0.25	0.23	0.25	0.22	12.32	11.35	11.96	11.41
Work at home	Low	0.31	0.31	0.29	0.32	0.3	15.3	14.02	14.85	14.06
strenuousness	Moderate or high	0.69	0.25	0.23	0.25	0.22	13.47	12.57	13.13	12.53
Work strenuousness	Low	0.24	0.25	0.22	0.25	0.22	13.74	12.59	13.28	12.87
	Moderate or high	0.76	0.32	0.3	0.32	0.3	15.26	14.12	14.9	13.94
Marital status	Married	0.6	0.26	0.24	0.26	0.23	15.19	14.13	14.68	14.03
	Single	0.4	0.32	0.31	0.32	0.3	13.91	12.97	13.65	12.82
Subsamples										
Overweight	No	0.7	0.27	0.26	0.28	0.25	13.97	12.99	13.56	12.85
	Yes	0.3	0.3	0.28	0.29	0.27	16.03	14.8	15.65	14.74
Overweight and	No	0.62	0.28	0.26	0.28	0.26	13.58	12.57	13.15	12.43
obese	Yes	0.38	0.29	0.27	0.29	0.26	16.32	15.08	15.97	15.03
All		1	0.28	0.27	0.28	0.26	14.62	13.58	14.21	13.47

Notes: We report percentage of cases and averages variables in pre- and post-reform periods, for the whole sample (2001-2003 and 2005-2007) and for the discontinuity sample (2003-2005) in each variable used to estimate the causal relationship between smoking and body weight.

from equation (2.2-2.4), separately for the whole (years 2001-2003 and 2005-2007, columns 1-5) and RD (years 2003 and 2005, columns 6-10) samples. Column (1) shows a positive correlation between the implementation of the clean indoor air law in Italy and body



Note: A cohort is defined as a group with fixed membership made up of individuals, who can be identified as they show up in the surveys (See, for example, Deaton (1985)). In the figures, each connected line represents the key variable behaviors (body weight, BMI, number of cigarettes and percentage of smokers) of a cohort over the years of observation. This representation permits some preliminary consideration about the presence of age and cohort effects (Kapteyn et al., 2005). The vertical difference between lines measures the cohort-time effect: differences between the key variables observed at the same age, but with different year of birth, highlight the presence of generational (or cohort) effects. On the other hand, differences along the same line measure the age-time effects.

Figure 2.4: Cohort patterns of weight, BMI and smoking indicators

weight increases, with an estimated change of 0.59 unit points of BMI [s.e.=0.024]. As expected, the ban is significantly correlated with reduced nicotine consumption in terms of number of cigarettes smoked [(-0.41; s.e.= 0.062)] and percentage of smokers [(-1.65; s.e.= 0.346)]. Given this exogenous policy shock, our causal estimates indicate the adverse impact of cigarette consumption on BMI [ΔBMI=-1.45; s.e.=0.253]. It is worth noting that the BMI variation induced by the percentage of smokers substantially accounts for all smoking-related variations in BMI. To obtain this result we must compare BMI changes in unitary points. In other words, starting from the estimated coefficient of -0.36 [s.e.=0.08], associated with the share of smokers in the whole sample (i.e. 0.27; see Table 2.1), in order to obtain the effect of this variable on BMI for the entire population we must multiply this coefficient by the inverse of the share of smokers in the sample (Baker et al., 2008, Havnes and Mogstad, 2010). In our example, the impact of participation rate (quitters) on BMI is -1.33 (i.e., $-0.36/.27$) and, according to this estimate,

the residual contribution to BMI by smokers that reduce cigarette consumption is quite small and limited.

Another important fact to be noted is that, as Figures 2.1 and 2.2 also show, the effect of changes in smoking habits estimated by RD on BMI, is substantially expected to be greater when measured per-period (i.e., one year): the estimated correlation between the smoking ban dummy and BMI, reported in column (6), is significant [(0.36; s.e= 0.04)], and is propedeutic in explaining that much of the causal variation in BMI by quitting smoking happens in the short term, as determined by the structural coefficients reported in columns (8) and (10), respectively. Our estimates indicate that 56% of the total BMI variation in the period (2005–2007) occurs in the first year (i.e., 0.82/1.45). This result derives from the fact that the coefficients estimated from the reduced forms for smoking rise in absolute magnitude. Unsurprisingly, these findings are in line with those of many empirical works in the US which are in accord with the adverse effect of public policies promoting a smoke-free environment and reduced cigarette consumption (e.g., Tauras (2005), Wasserman et al. (1991), Yurekli and Zhang (2000)). In this chapter, we also conclude that the immediate reaction to the Italian ban on smoking led to substantial short-term decreases in the smoking participation rate, a phenomenon which weakens over the post-ban period of three years. This result also helps us to clarify the interpretation of the causal effect estimated for our response variable.

The models presented in Table 2.2 account for differences in observable and unobservable heterogeneity, including a set of covariates listed in Table 2.1. All covariates are generally significant in our estimates. This allows us to test, according to the economic health literature, the possibility of heterogeneous behavior between subgroups as, for example, employed subjects. We investigate whether the implementation of the smoking ban caused heterogeneous drops in smoking behavior for employees with respect to the entire population. One potential concern with these estimates is that extending the already existing prohibition to smoke in offices also to common areas, our results may underestimate the effect of employed smokers on BMI.

Reduced form estimates, listed in columns (2) and (4) (whole sample) and (7) and (9) (RD sample) of Table 2.3 indicate that the ban significantly affects smoking habits for employed subjects. Interestingly, the effect for this group is larger than that obtained

from baseline estimates, for each smoking variable and irrespective of the sample used. In contrast, the coefficient associated to the BMI variation induced by the ban for employees (equation 2.2), is close to that obtained from the baseline estimates (columns 1 and 6). Consequently, a limited BMI increase due to reduced smoking is found in the structural equation, in which the unitary reduction in cigarette consumption increases BMI by 0.96 (s.e.= 0.113) in three years and 0.63 (s.e.= 0.129) one year after the ban generated discontinuity.

The fact that estimates in terms of BMI for the employed subsample are smaller than our baseline results makes an important link with the epidemiological literature. Unemployment status is known to be associated with persistence in smoking consumption and resistance to quit. For these reasons, unemployed individuals, represent a particularly interesting target group for government health promotion through anti-smoking policies (Kriegbaum et al., 2010, Schunck and Rogge, 2010).

Since empirical findings have shown that gender characteristics may also lead to significant differences in smoking behavior, we report in tables 2.4 and 2.5 alternative specifications of our baseline model, in which we measure the impact of smoking habit changes for male and female sub-groups and compare them with the baseline estimates of Table 2.2. The smoking ban turns out to have a higher correlation with smoking habits in men: this very probably also depends on the larger numbers of Italian male smokers and the number of cigarettes they consume (Aristei and Pieroni, 2010). Although column (1) of Table 2.5 clearly shows that women's BMI is more sensitive, leading to an increase of 0.68 unit points in the three years after the ban (0.59 from baseline model and 0.54 from male group), the causal estimates of the different specification reported in the tables remain equally distant, above and below the baseline results of Table 2.2, confirming the heterogeneous adverse effect on BMI. It does seem reasonable that the heterogeneous mechanisms discussed in the literature justify the greater response from men, as high participation rate or free time and income also apply in explaining gender responses to the ban argued here (see, for example Chaloupka (1992)). We discuss below the results of our model specification by gender, to explain heterogeneous changes in smoking habits in terms of body weight. We only note now that, in all previous subsample estimates, the general finding that the weight gain depends almost completely on quitting is confirmed.

Table 2.2: Causal effect of smoking on BMI

	Adults born between 1941-1989, years 2001 - 2007					**Adults born between 1943-1987, discontinuity sample**				
		Number of cigarettes		**Percentage of smokers**			**Number of cigarettes**		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Smoking Ban	0.59***	-0.41***		-1.65***		0.36***	-0.44***		-2.24***	
	(0.024)	(0.062)		(0.346)		(0.040)	(0.065)		(0.367)	
Number of cigarettes			-1.45***					-0.82***		
			(0.253)					(0.169)		
Percentage of smokers					-0.36***					-0.16***
					(0.080)					(0.034)
Constant	23.85***	3.96***	29.58***	29.30***	34.31***	23.94***	4.32***	27.48***	31.87***	29.09***
	(0.050)	(0.088)	(0.943)	(0.412)	(2.268)	(0.066)	(0.136)	(0.660)	(0.639)	(1.017)
Observations	170,702	170,702	170,702	170,702	170,702	57,409	57,409	57,409	57,409	57,409
R-squared	0.21	0.05		0.03		0.21	0.05		0.03	
Adj. R-squared										

Notes: Column (1) lists estimates of reduced form in equation (2.2)), i.e. effect of the smoking ban in January 2005 on BMI. Columns (2) and (4) also list estimates of reduced form of effect of ban on number of cigarettes and smoking participation rate, respectively. Columns (3) and (5) list estimates of structural model from equation (2.4) of causal effect of smoking habits on BMI. Estimates are for 2001 - 2007 (pre-ban, 2001 - 2003; post-ban, 2005 - 2007). Columns (6)-(10) list same estimates around discontinuity introduced by ban (2003 - 2005). All estimates include covariates described in Table 2.1 as controls .
*Standard errors in round brackets. Significant levels reported as: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 2.3: Causal effect of smoking on BMI, employed adults

	Adults born between 1941-1989, years 2001 - 2007					**Adults born between 1943-1987, discontinuity sample**				
		Number of cigarettes		**Percentage of smokers**			**Number of cigarettes**		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Smoking ban	0.53***	-0.55***		-2.40***		0.32***	-0.51***		-2.54***	
	(0.022)	(0.054)		(0.317)		(0.041)	(0.077)		(0.449)	
Number of cigarettes			-0.96***					-0.63***		
			(0.113)					(0.129)		
Percentage of smokers					-0.22***					-0.13***
					(0.032)					(0.026)
Constant	23.68***	4.19***	27.70***	32.59***	30.92***	23.77***	4.94***	26.88***	40.75***	28.93***
	(0.041)	(0.092)	(0.461)	(0.487)	(1.033)	(0.067)	(0.153)	(0.574)	(0.758)	(1.011)
Observations	110,559	110,559	110,559	110,559	110,559	37,280	37,280	37,280	37,280	37,280
R-squared	0.22	0.05		0.03		0.22	0.04		0.03	
Adj. R-squared										

Notes: Column (1) lists estimates of reduced form in equation (2.2)), i.e. effect of the smoking ban in January 2005 on BMI. Columns (2) and (4) also list estimates of reduced form of effect of ban on number of cigarettes and smoking participation rate, respectively. Columns (3) and (5) list estimates of structural model from equation (2.4) of causal effect of smoking habits on BMI. Estimates are for 2001 - 2007 (pre-ban, 2001 - 2003; post-ban, 2005 - 2007). Columns (6)-(10) list same estimates around discontinuity introduced by ban (2003 - 2005). All estimates include covariates described in Table 2.1 as controls .
*Standard errors in round brackets. Significant levels reported as: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 2.4: Causal effect of smoking on BMI, men

	Adults born between 1941-1989, years 2001 - 2007					**Adults born between 1943-1987, discontinuity sample**				
		Number of cigarettes		**Percentage of smokers**			**Number of cigarettes**		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Smoking ban	0.54***	-0.67***		-2.83***		0.36***	-0.65***		-2.98***	
	(0.030)	(0.093)		(0.435)		(0.039)	(0.106)		(0.540)	
Number of cigarettes			-0.80***					-0.54***		
			(0.142)					(0.120)		
Percentage of smokers					-0.19***					-0.12***
					(0.036)					(0.028)
Constant	23.60***	4.54***	27.22***	36.03***	30.44***	23.83***	5.37***	26.75***	41.43***	28.77***
	(0.048)	(0.153)	(0.547)	(0.800)	(1.151)	(0.060)	(0.196)	(0.577)	(0.925)	(1.096)
Observations	84,164	84,164	84,164	84,164	84,164	28,327	28,327	28,327	28,327	28,327
R-squared	0.12	0.03		0.02		0.13	0.03		0.03	
Adj. R-squared										

Notes: Column (1) lists estimates of reduced form in equation (2.2)), i.e. effect of the smoking ban in January 2005 on BMI. Columns (2) and (4) also list estimates of reduced form of effect of ban on number of cigarettes and smoking participation rate, respectively. Columns (3) and (5) list estimates of structural model from equation (2.4) of causal effect of smoking habits on BMI. Estimates are for 2001 - 2007 (pre-ban, 2001 - 2003; post-ban, 2005 - 2007). Columns (6)-(10) list same estimates around discontinuity introduced by ban (2003 - 2005). All estimates include covariates described in Table 2.1 as controls .
*Standard errors in round brackets. Significant levels reported as: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 2.5: Causal effect of smoking on BMI, women

	Adults born between 1941-1989, years 2001 - 2007					**Adults born between 1943-1987, discontinuity sample**				
		Number of cigarettes		**Percentage of smokers**			**Number of cigarettes**		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Smoking ban	0.68***	-0.23***		-1.06***		0.36***	-0.26***		-1.65***	
	(0.032)	(0.044)		(0.313)		(0.056)	(0.066)		(0.451)	
Number of cigarettes			-2.97***					-1.41***		
			(0.581)					(0.398)		
Percentage of smokers					-0.64***					-0.22***
					(0.313)					(0.065)
Constant	21.16***	1.72***	26.29***	19.23***	33.48***	21.65***	1.12***	23.23***	14.22***	24.80***
	(0.039)	(0.048)	(0.875)	(0.353)	(3.408)	(0.095)	(0.127)	(0.336)	(0.791)	(0.842)
Observations	86,538	86,538	86,538	86,538	86,538	29,082	29,082	29,082	29,082	29,082
R-squared	0.17	0.01		0.01		0.16	0.01		0.01	
Adj. R-squared										

Notes: Column (1) lists estimates of reduced form in equation (2.2)), i.e. effect of the smoking ban in January 2005 on BMI. Columns (2) and (4) also list estimates of reduced form of effect of ban on number of cigarettes and smoking participation rate, respectively. Columns (3) and (5) list estimates of structural model from equation (2.4) of causal effect of smoking habits on BMI. Estimates are for 2001 - 2007 (pre-ban, 2001 - 2003; post-ban, 2005 - 2007). Columns (6)-(10) list same estimates around discontinuity introduced by ban (2003 - 2005). All estimates include covariates described in Table 2.1 as controls .
*Standard errors in round brackets. Significant levels reported as: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

2.4.2 Robustness, specification checks and conditional estimations

In this section, we test for evidence that identification of the effects of smoking habits on weight gain comes from variations close to the discontinuity determined by the Italian smoking ban (Section 4.2.1). We also discuss the potential bias related to omitted variables and compare OLS estimates of average weight gain from changes in smoking habits with those obtained with IV regression (Section 4.2.2), and examine the causal effects of nicotine reduction, at different points of the BMI distribution, corresponding to specific quantiles of overweight and overweight and obese people (Section 4.2.3).

IDW estimations

As discussed above, the causal effects of BMI changes are predicted by changes in smoking habits (via the smoking ban) with the accuracy of implementing an estimator which measures LATEs. Note that these differences in measuring causal effects with the entire and RD samples, corresponding to periods of three years and one year before and after the discontinuity, respectively explain the relations tested in the medium and short-term for BMI and smoking habits. Clearly, the model presented in equations (2.2) and (2.3) is ensured by a predominantly identification strategy which only considers variations close to the discontinuity, so that estimates on the entire sample may be affected by trends in unobservable characteristics in different cohorts. This justifies using the IDW estimator, which assigns higher weights to those observations closer to the year of the ban. Table 2.6 shows the IDW estimates results. These estimates are qualitatively the same as those of Table 2.2, although BMI coefficients tend to decrease slightly in magnitude in absolute terms. To consider further these differences in specific sub-groups, the results for men, women and employees are reported in Appendixes 2.A.1-2.A.3. For the above mentioned sub-groups, the comparison between IDW estimates and those unweighted from the whole sample are found to be even closer than those of the baseline model in Table 2.2. Thus, we can conclude that our identification strategy in the medium term is not significantly affected by the distance of observations from the discontinuity.

Table 2.6: Causal effect of smoking on BMI - IDW estimates

	Adults born between 1941-1989, years 2001 - 2007				
		Number of cigarettes		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)
Smoking ban	0.54***	-0.40***		-1.66***	
	(0.025)	(0.059)		(0.322)	
Number of cigarettes			-1.37***		
			(0.236)		
Percentage of smokers					-0.33***
					(0.068)
Constant	23.90***	3.87***	29.21***	28.86***	33.34***
	(0.052)	(0.084)	(0.874)	(0.410)	(1.907)
Observations	170,702	170,702	170,702	170,702	170,702
R-squared	0.21	0.05		0.03	
Adj. R-squared	0.21	0.05	.	0.03	.

Notes: Column (1) lists estimates of reduced form in equation (2.2)), i.e. effect of the smoking ban in January 2005 on BMI. Columns (2) and (4) also list estimates of reduced form of effect of ban on number of cigarettes and smoking participation rate, respectively. Columns (3) and (5) list estimates of structural model from equation (2.4) of causal effect of smoking habits on BMI. Estimates are for 2001 - 2007 (pre-ban, 2001 - 2003; post-ban, 2005 - 2007). Columns (6)-(10) list same estimates around discontinuity introduced by ban (2003 - 2005). All estimates include covariates described in Table 2.1 as controls .
*Standard errors in round brackets. Significant levels reported as: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Endogeneity caveats and average effects on population

In contexts where causal parameters are not identified by a regression discontinuity design, the estimated relationship may be biased upwards or downwards by omitted variables. On one hand, quitters have generally been found to be less concerned about weight gain than subjects continuing to smoke. For example, Cawley et al. (2004) argue that the latter category have more self-control with respect to weight changes, so that quitters' behavior may include unobservable characteristics which overestimate the true causal effect. On the other hand, an underestimated impact on BMI increases may arise if smokers' choice to quit induces them to adopt other healthy behavior, which takes into account of general concerns about health or are more oriented towards the future. In this case, a downward bias may be produced estimating weight gains by significant positive correlations between smoking reductions and error terms, since being an ex-smoker for long period of time is associated with making large investments in health and well-being .

The estimates that use regression discontinuity design not only produce an impact which

is expected to be unbiased, but they are expected to exceed those from OLS. As argued above, the IV estimator yields the marginal causal effect of smoking on BMI of smokers affected by the policy. Table 2.7 shows the estimates for smoking habits on BMI, obtained from OLS. Although smoking coefficients are significant at the conventional 1% level, the magnitude of these estimates on weight gains are really reduced with respect to the models in which the IV estimator is used. For example, in the RD sample, the impact of nicotine consumption on the BMI of the entire population falls from -0.82 to -0.015 (s.e.=0.000) with OLS estimator. As expected, this result is proportionally confirmed for the impact of the percentage of smokers. These marked differences also explain why measuring precisely the effect of smoking on weight is still considered open, and make the use of linear regression questionable (see, for example, the critical discussion of Baum (2009)).

Table 2.7: OLS estimates of effect of smoking on BMI

Variables	**Whole sample**	**RD sample**
Nicotine consumption	-0.015***	-0.012***
	(0.003)	(0.003)
Percentage of smokers	-0.001***	-0.001***
	(0.000)	(0.001)

*Notes: Standard errors shown in round brackets. Significant levels reported as follows: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Conditional BMI estimation: the effects for overweight and obese people

One limitation of our analysis is that nicotine effects are estimated only at the average of BMI distribution. We also want to examine other points in BMI distribution mainly for two reasons. First, estimates at different quantiles may provide an opportunity to trace smoking habit changes for portions of the overall BMI distribution. Second, in terms of welfare, we are interested in particular points of BMI distribution, because we can highlight significant differences in the magnitude of the effects for particularly interesting sub-groups of the population (e.g. obese and overweight people). Since the overweight

dimension is more important than the dimension of obesity in Italy[11], we estimate the effects of smoking habits on points of the BMI distribution corresponding to the mean of the sample of overweight (OV) (77^{st} percentile) and overweight and obese (OVOB) (81^{st} percentile) individuals. We use the IV quantile treatment effect estimator, which can be obtained from the potential outcome framework (Imbens and Rubin, 1997) under LATE identification assumptions. Correpsonding to these points of BMI distribution, we propose a local discontinuty quantile regression estimate based on the RD sample.

Instrumental variable estimates are shown in Table 2.8 for the sample means of OV and OVOB. Like the IV estimates of the average BMI, the quantile estimates for these groups have the same (negative) sign and are statistically different from zero. To help interpretation of the magnitude of these estimated effects, let us consider that the difference between OV and the sample mean effect is about -0.26 (i.e., -0.82 at the sample mean, against -0.56 for the ov group). Our estimates also suggest that including obese individuals mitigate the adverse effects of smoking habits on BMI. Estimates at the 81^{th} percentile (e.g., average of OBOV group) indicate a further reduction of the causal effect on BMI growth, attributed to nicotine decreases, measured at about 40% less than the estimated coefficient at the sample mean (from -0.82 to -0.49, respectively). Although fewer observations in the tail of our sample do not allow us to trace estimates across the obese distribution, convincing differences in the results support the validity of our main findings, in the Italian case, of a significant gain in weight through nicotine reduction in people with high BMI, although smaller when compared to the effect at the average of the sample.

It is not surprising that, more than in the full sample, almost the whole contribution to BMI increases in these subgroups depends on variations attributable to quitters. One plausible interpretation regards the generally poor individual health condition of obese people, who are incentivated to stop smoking rather than to reduce cigarette consumption. We find that cuts in nicotine consumption are mainly due to quitters, although some interesting heterogeneous effects on BMI distribution emerge. First, the impact of quitting smoking on BMI is higher in the group which includes obese individuals. Second, unlike the economic literature discussed in Section 2, which focuses on empirical US analyses,

[11]Note that observations above 90% percentiles are sparse. See Pieroni et al. (2011).

Table 2.8: Causal effect of smoking on BMI in overweight and obese individuals

	Number of cigarettes	**Percentage of smokers**
Variables	**77^{th} (Overweight)**	**81^{st} (Overweight and Obese)**
Nicotine consumption	-0.56***	-0.49***
	(0.067)	(0.064)
Percentage of smokers	-0.11*	-0.12*
	(0.0612)	(0.067)

Notes: Estimates obtained by instrumental variable quantile regression (IVQR). Standard errors shown in round brackets. Significant levels reported as follows:
*p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

we find the unexpected result of a decreasing impact of the effect of smoking on BMI for people with critical weight levels. Especially in Europe, this phenomenon has been explained by the fact that obese people may reduce their weight and smoking habits as a consequence of changes in their life-style towards better health (Brunello et al., 2011).

2.4.3 Body weight estimates to changes in smoking habits of smokers

We now turn to a more policy-oriented analysis of the effects of smoking reduction on weight gain. Our first results, based on the RD sample, are shown in Table 2.9, which lists estimates of BMI changes expressed as elasticities, making the magnitude of the smoking indicators used in our analysis more easily interpretable. The results emphasize the finding that body weight changes due to smoking reductions are attributable mostly to quitters, irrespective of whether the estimates refer to the complete sample or to subgroups of individuals. Qualitatively, these results are fully consistent with all the estimates presented above. For example, higher variations in body weight are confirmed to occur in response to smoking changes in women, and the estimated elasticity is larger than that obtained in the full sample.

Most of the estimated BMI effects seem fairly small, although there is no difficulty in achieving statistical significance at the conventional levels. However, whether or not these smoking effects are considered large enough depends on the context. To simplify this

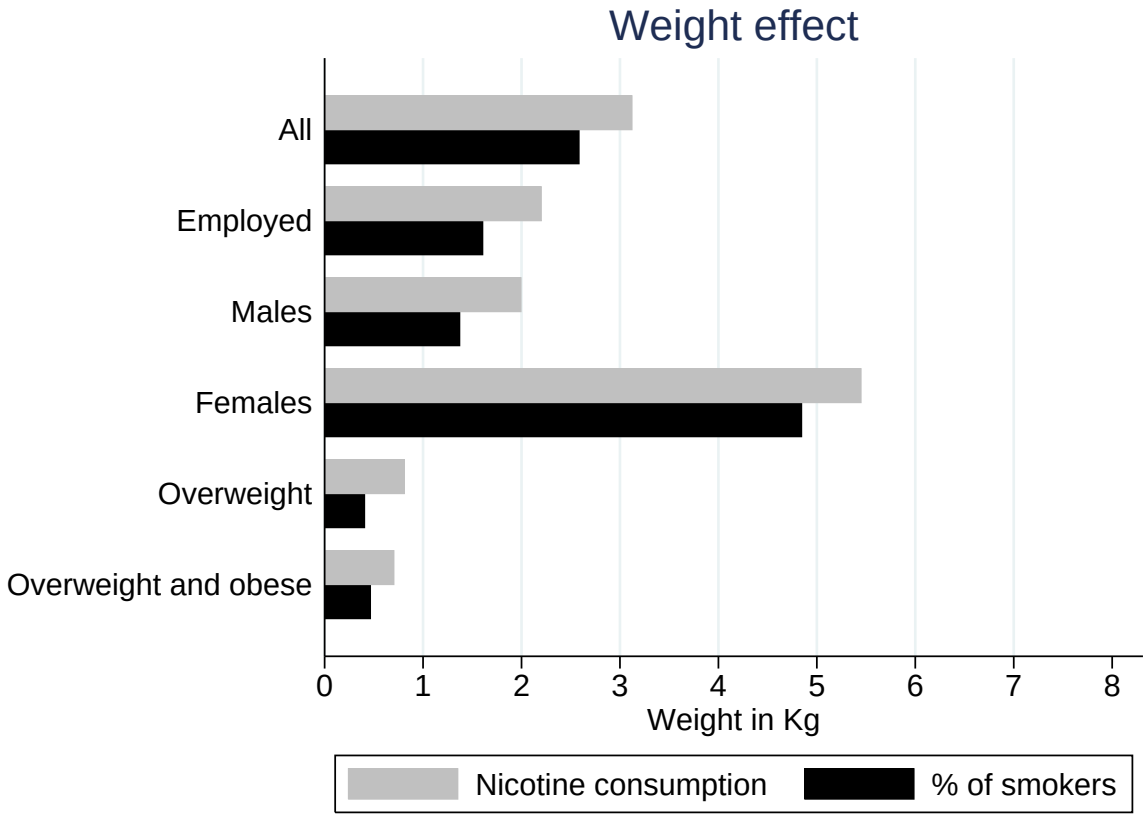
Table 2.9: BMI elasticities to number of cigarettes and percentage of smokers

Sample	**Number of cigarettes**	**Percentage of smokers**
All subjects	-0.36***	-0.51***
	(0.062)	(0.166)
Employed	-0.23***	-0.28***
	(0.042)	(0.069)
Men	-0.31***	-0.31***
	(0.062)	(0.081)
Women	-0.43***	-0.64*
	(0.165)	(0.362)
Overweight	-0.08***	-0.11*
	(0.009)	(0.064)
Obese and Overweight	-0.06***	-0.12*
	(0.009)	(0.067)

*Notes: Standard errors shown in round brackets. Significant levels reported as follows: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

analysis, we translate the estimated effects on BMI in terms of weight, multiplying each coefficient by the squared height measured as a mean of population (or subsamples). Let us consider the implications of estimated weight changes for the Italian smoking population which was about 14,000,000 in 2005. We would expect from Figure 2.5 a unitary reduction in cigarette consumption to determine a rise in body weight of 3 kg, of those affected by the smoking ban. Clearly, we estimate that more than 90% of this variation depends on the effect of quitters, implying that, in the average smoking population, the increase in weight accounts for 2.6 kg.

The estimates in Figure 2.5 also indicate that weight gains in employed people are lower with respect to the entire population, whereas gender differences appear to be more relevant; in addition, women's weight gains (5.6 kg) are more than double those of men (2 kg). This result can be explained by the fact that, because of reverse causality, some subjects and particularly women (see Cawley et al. (2004)) tend to use smoking to control their weight. The result listed above was also found by other studies on Italy and France (Gallus et al., 2006). Our results are inconsistent with the hypothesis, generally



Notes: Lighter bars: weight changes of nicotine consumption; darker bars: impact of quitters only. Effect on weight of nicotine consumption and percentage of smokers was retrieved multiplying coefficients obtained from RD estimates reported in columns 8 and 10 of each sample by squared height. To compare change in BMI unit points, estimated coefficient of percentage of smokers was multiplied by inverse of fraction of smokers (see details in Section 4.1).

Figure 2.5: Causal effect of smoking on weight. Discontinuity sample

predicted in the economic literature, that the effects of quitting smoking are greater in obese subjects. Overweight (or overweight and obese people) appear to respond poorly to changes in nicotine consumption, although this finding matches the results of Fang et al. (2009) for China and Flegal (2007) for the United States.

Although reductions in smoking may theoretically not be desiderable, particularly if weight gains are large and have high social costs that offset the benefits of quitting, the estimates are reassuring for Italy. The limited effect on weight changes in the groups at ''weight risk" leads to predict that future policies implementing cuts in smoking should maintain limited the cost in terms of health losses paid in order to achieve smoking reforms, even if the Italian patterns of overweight and obesity are increasing.

APPENDIX 2.A

Table 2.A.1: Causal effect of smoking on BMI, employed adults - IDW estimates

	Adults born between 1941-1989, years 2001 - 2007				
		Number of cigarettes		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)
Smoking ban	0.48***	-0.54***		-2.39***	
	(0.023)	(0.051)		(0.288)	
Number of cigarettes			-0.89***		
			(0.104)		
Percentage of smokers					-0.20***
					(0.027)
Constant	23.73***	4.08***	27.37***	32.14***	30.24***
	(0.042)	(0.095)	(0.417)	(0.492)	(0.853)
Observations	110,559	110,559	110,559	110,559	110,559
R-squared	0.22	0.05		0.03	
Adj. R-squared	0.22	0.04	.	0.03	.

Notes: Notes: see, Table 2.6.

Table 2.A.2: Causal effect of smoking on BMI, men - IDW estimates

	Adults born between 1941-1989, years 2001 - 2007				
		Number of cigarettes		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)
Smoking ban	0.49***	-0.64***		-2.73***	
	(0.029)	(0.090)		(0.416)	
Number of cigarettes			-0.77***		
			(0.137)		
Percentage of smokers					-0.18***
					(0.034)
Constant	23.66***	4.52***	27.13***	35.74***	30.14***
	(0.046)	(0.149)	(0.546)	(0.808)	(1.098)
Observations	84,164	84,164	84,164	84,164	84,164
R-squared	0.12	0.03		0.02	
Adj. R-squared	0.12	0.03	.	0.02	.

Notes: see, Table 2.6.

Table 2.A.3: Causal effect of smoking on BMI, women - IDW estimates

	Adults born between 1941-1989, years 2001 - 2007				
		Number of cigarettes		**Percentage of smokers**	
	BMI	Smoking	BMI	Smoking	BMI
	(reduced form)	(reduced form)	(structural form)	(reduced form)	(structural form)
	(1)	(2)	(3)	(4)	(5)
Smoking ban	0.62***	-0.23***		-1.16**	
	(0.034)	(0.042)		(0.297)	
Number of cigarettes			-2.68***		
			(0.484)		
Percentage of smokers					-0.54***
					(0.135)
Constant	21.21***	1.75***	19.47***	25.91***	31.68***
	(0.043)	(0.087)	(0.046)	(0.736)	(2.473)
Observations	86,932	86,932	86,932	86,932	86,932
R-squared	0.16	0.01		0.01	
Adj. R-squared	0.16	0.01	.	0.01	.

Notes: see, Table 2.6.

Chapter 3

Socio-Economic Determinants of Body Weight in the UK

3.1 Introduction

This chapter examines the socio-economic determinants of body weight in the United Kingdom by means of two recent waves from the British Household Panel Survey. While the patterns of overweight and obesity have drawn economists' interest in recent years, the main contribution of this work is to examine the weight determinants on the conditional distribution of body weight across individuals. Are there differing socio-economic causes for gaining weight in highly overweight people compared with underweight ones? Our results support some findings in the literature, but also point to new conclusions.

3.2 Background: body weight and socio-economic variables

In the last few decades, obesity has become an important risk factor for a number of severe and chronic diseases which constitute the main causes of death, including heart disease, stroke, and some types of cancer. It also contributes to other serious life-shortening conditions such as Type 2 diabetes. Data from the United States show that the prevalence of overweight and obesity began to increase around the mid-1980s and has continued to

increase dramatically. The increase in obesity in the UK is similar to that of the United States although it starts from a lower level (Brunello et al., 2009).

Figure 3.1 shows the trends of obesity in aggregate for all adult men and women, and indicates that obesity has constantly risen over the last fifteen years (15% since 1993), with similar trends for both men and women. This persistent growth suggests that, at least, some causes may have become structural in determining obesity in the UK.

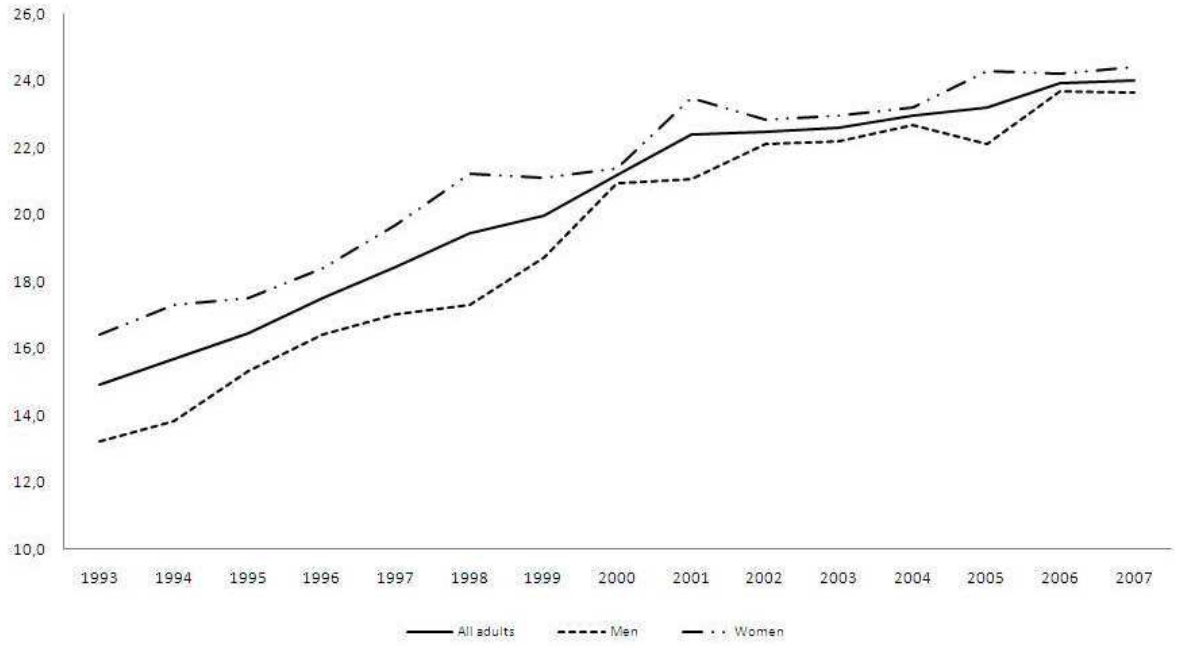


Figure 3.1: Percentage of obesity in UK by gender: 1993 - 2007

The consequences of adult overweight are also growing in the UK. The National Audit Office (NAO, 1998) stresses that 6% of total deaths in the UK can be associated with obesity, and increased to 6.8% in a few years according to a research conducted by a House of Commons Health Committee (2004). In addition, the number of Finished Consultant Episodes (FCEs), providing a primary diagnoses of obesity, has increased consistently from 1996 to 2006. As a consequence, the burden associated with obesity on the National Health Service (NHS) was estimated to have increased between 1998-2006 from 1.5% to 2.6% of total health expenditure. More recently, estimates by the Department of Health (2006) forecast that the NHS cost attributable to the obesity epidemic may rise to 5.3 billion sterling by 2025[1].

[1]These estimates are discussed in the report Department of Health (2006)

In contrast with the data reported for the United States and Italy, the increase in body weight in the UK does not seem to be associated with a significantly increasing pattern in calorie consumption[2]. Figure 3.2 shows the *per capita* calorie consumption, subdivided for home and eating out on an annual basis from 1995 to 2007. These patterns are stable over that period, showing a slight decrease in the last part of the sample. In Figure 3.3 is shown the path of food prices with respect to the aggregate price index: changes in relative food prices have decreased constantly each year by 1%, making food (calorie) consumption potentially more convenient. This is in line with the findings of Lakdawalla and Philipson (2002) in the United States who, while reporting a reduction in the price of food, also noted that the market demand for food did not seem to increase[3]. However, the determinants of obesity may not affect individuals equally. This argumentation is partly supported by addictive behaviour in gaining calories, which, include the effects of different (hidden) individual characteristics (Cawley et al., 2004). That is, overweight individuals may ''feed on themselves", so as obesity issues tend to become more entrenched in already obese individuals[4]. This implies that an increase in the relative demand for food by overweight people, but not necessarily an increase in aggregate food demand, may explain why obesity is increasing.

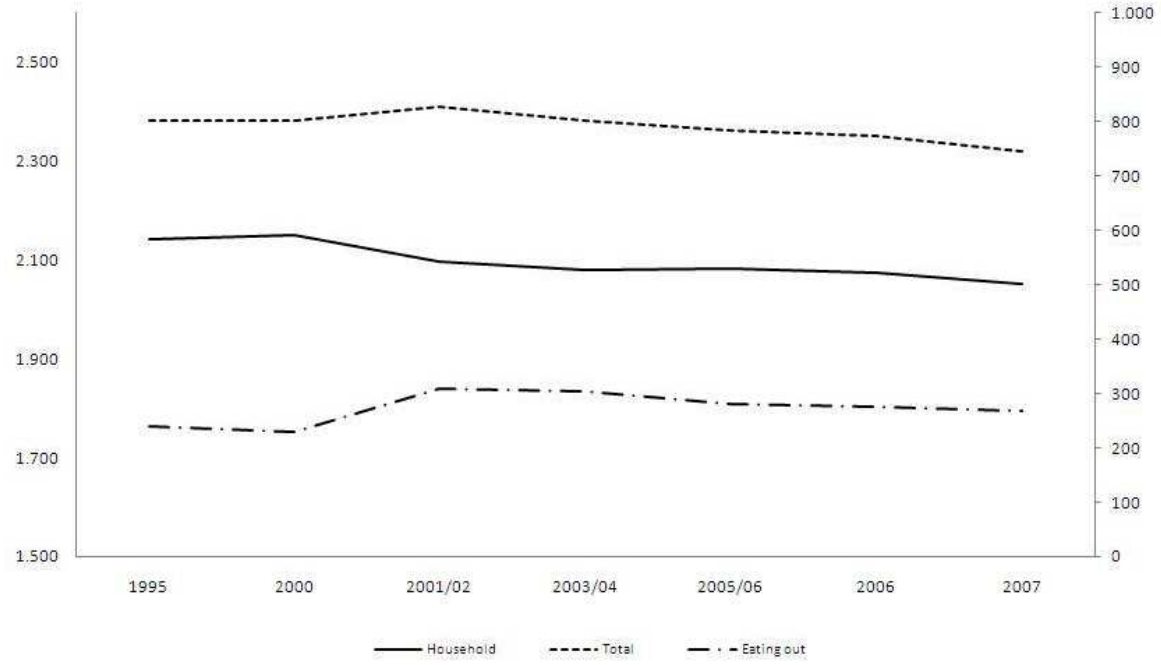
As will be argued below, if we rely on the energy accounting framework, in which body weight increases when more calories are taken in than are consumed, lack of physical activity seems to be a supplementary candidate in explaining the dimension of weight in the UK. The percentages of both men and women undertaking physical exercise has increased constantly and considerably over the last ten years[5], but even in 2007, one-third of the population had not kept up with the Government guidelines for physical exercise. Also in this context, the different roles played by men and women in the family and society seems to be a constraint for physical exercise and to affect gender body weight non-equally. The main reasons for not taking exercise, as they emerged from the survey,

[2]Bleich et al. (2007), Pieroni et al. (2011) respectively.

[3]Using historical data Costa and Steckel (1995) show frequently coinciding declines in calories and prices, and growth in weight. For example, the increasingly larger portions at fastfood outlets and restaurants should also be interpreted as responses to the growing food supply and consistent with the prediction of falling relative food prices.

[4]Blanchflower et al. (2009) provide cross-sectional evidence for Germany that overweight perceptions and dieting are influenced by a person's relative BMI.

[5]From 32% in 1997 to 40% in 2007 for men and from 17% to 21% for women.



notes: On the right scale: household and total calories; on the left scale: eating-out calories

Figure 3.2: Consumption of calories in UK: 1995 - 2007.

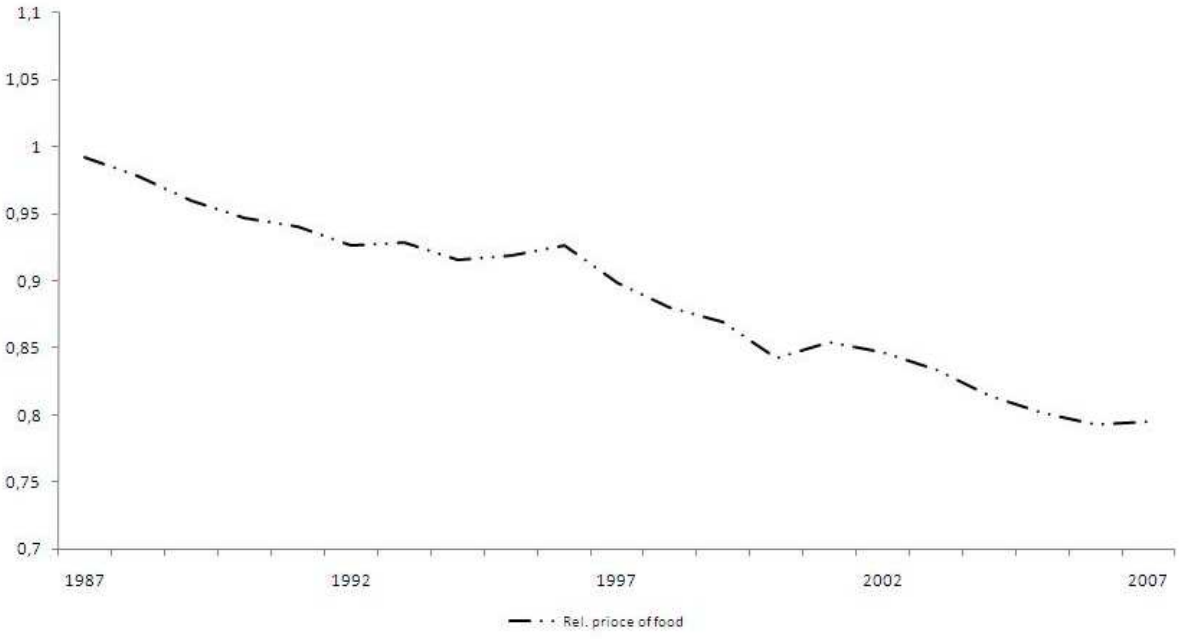


Figure 3.3: Changes in UK relative food price: 1987 - 2007

were work commitments and lack of leisure time for men, and exactly the opposite for women (Service, 2009). Since our aim is to explain why a given individual is overweight in a time-constant framework, which implies that the strenuousness of work both at home and in the market are constant, the cross-individual variability of the cost of physical activity can be used as a direct measure of the propensity (or frequency) of participating in leisure-based exercises, such as jogging or gymnastics or, more in general, of substituting extra hours of work with physical exercise.

Another explanation for increasing body weight was given by Chou et al. (2004), who argued that it was the result of several economic changes which have modified people's lifestyle choices. In particular, the main changes proposed to affect weight are: i) changes in relative prices, favouring meals in fast-food and full-service restaurants; ii) the increasing female work participation rate, which has reduced the amount of time spent on housework and cooking meals with basic ingredients, has determined growing weight, even when the relative prices of eating at home have declined. Within this framework traditional meals are assumed less dense in calories, and the demand for convenience unhealthy food increases in response to the increasing value of women's time spent in the household; iii) increases in the relative price of cigarettes - as well as the effects of legislation (Clean Indoor Air Laws) - may have contributed to increasing average weight, because smokers may have higher metabolic rates than non-smokers[6].

3.3 Data and methods

The dataset used in this chapter was extracted from the British Household Panel Survey (BHPS), a multi-purpose survey which reports information at both household and individual levels for a representative panel of the UK population. The original sample was composed of 5,500 households and 10,300 individuals, drawn from 250 areas of England and was subsequently enlarged to include Scotland and Wales in 1999 and Northern Ireland in 2002. The dataset has 18 waves: the first survey was conducted in 1980 but, for our purposes, we use a sample of two waves, the 14^{th} and 16^{th} waves, conducted re-

[6]In the medical literature ((French and Jeffery, 1995, Grunberg, 1985, Klesges et al., 1989), changes in dietary intake, physical activity and metabolic rate are some of the proposed mechanism through which body weight is affected by smoking.

spectively in 2004 and 2006, because two anthropomorphic characteristics of individuals (height and weight) were also collected for those particular waves. We then selected a balanced panel database on individuals from the two waves, and derived the BMI for a representative sample of the UK population (13,230 individuals for each year)[7].

As BMI is assumed to be non-normally distributed, we propose a quantile regression approach to estimate the relationship with its socio-economic determinants. The main empirical advantage is that the flexibility in estimating parameters at different quantiles does not require any assumptions regarding the error term (Koenker and Hallock, 2001, Koenker and Bassett, 1978). In addition, quantile regression allow to model unobserved heterogeneity and evaluate the effect of independent variables keeping the error term, which is assumed to be a measure of the unobserved heterogeneity, constant.

With this technique, we can examine the determinants of BMI throughout the conditional distribution, with particular focus on people with the highest and lowest BMI levels, which are arguably of the greatest interest. We follow the quantile regression formulation developed by Koenker and Bassett (1978), which yields parameter estimates at multiple points in the conditional distribution of the dependent variable[8]. One particular regression quantile is the solution to

$$\min_{\beta \in R^K} \left[\sum_{i \in \left\{ BMI_i \geq \bar{x}'\beta \right\}} \theta \left| BMI_i - \bar{x}'\beta \right| + \sum_{i \in \left\{ BMI_i \leq \bar{x}'\beta \right\}} (1-\theta) \left| BMI_i - \bar{x}'\beta \right| \right] \tag{3.1}$$

where $\theta \in (0,1)$. The estimates are obtained by minimising the weighted sum of absolute deviations, obtaining the n^{th} quantile by appropriately weighting the residuals. The conditional quantile of BMI_i, given the vector of explanatory $\bar{x}$, is

$$Q_{BMI}(\theta|\bar{x}) = \bar{x}'\beta_\theta \tag{3.2}$$

[7]BMI is a measure of body fat largely used in the social field because it is often recorded within socio-economic surveys. Although it has been shown that the self-reported weight generally used to estimate BMI produces measurement errors in young and adult people, the data obtained with adjusted BMI are very close to those obtained with self-reported indexes ((Burkhauser and Cawley, 2008, Zagorsky, 2005).

[8]A helpful introduction to quantile regression appears in Koenker and Hallock (2001). Applications of this method are increasingly common see for example Hartog et al. (2001) and Grg and Strobl (2002).

This formulation is analogous to OLS, $E(BMI|\bar{x}) = \bar{x}'\beta$, although OLS parameters are estimated only at the mean of the conditional distribution of the dependent variable. We can then calculate the elasticities associated with specific quantiles of the BMI to analyse the policy implications of socio-economic determinants on body weight.

3.4 Empirical strategy

3.4.1 Theoretical remarks

Despite the widespread use of BMI among social scientists, within the medical literature this index is considered to be an inaccurate measure of obesity, because it does not distinguish muscles, bones, and other lean body mass from body fat, (Gallagher et al., 1996, Garn et al., 1986, Gil and Mora, 2011, Jeffery, 1996, McCarthy et al., 2006, Smalley et al., 1990, Yusuf et al., 2005). As a result, BMI overestimates body fat among those who are muscular (Prentice and Jebb, 2001). Burkhauser and Cawley (2008) find in the US that obesity defined using BMI is only weakly correlated with obesity defined using more accurate measures of fatness, and that African Americans are particularly likely to be misclassified by this measure. Although, these findings suggest that social scientists should avoid using BMI uncritically and, preferably, use more accurate measures, our dataset does not report any other index to consistently correct it. Moreover the ethnic composition in the UK is not so prevalent as in the US and consequently this source of bias should be reduded[9].

To test the influence of socio-economic determinants of obesity growth in the UK, we applied the energy accounting approach, an appropriate multivariate framework to model body weight as a function of individual characteristics (Cutler et al., 2003, Michaud et al., 2007). This theory is useful because it is based on the excess of calories between energy intake and expenditure responsible for increases in individual weight at a given point in time over the life-cycle.

[9]Less problematic is the use of self-reported BMI that produces endogeneity by measurement errors in young and adult people. The estimates obtained with adjusted BMI are very close to those obtained with self-reported indexes (Burkhauser and Cawley, 2008, Zagorsky, 2005). Empirically, biases in self-reported measures of BMI can be corrected by the age variable which, particularly for height, tends to increase with age.

Note that, this framework should find its natural specification in a dynamic model in which calorie imbalance is adjusted over time, varying with the age effect. Unfortunately, as in the BHPS BMI is available only for two waves, we are not able to follow adequately individuals over time. An alternative approach involves the use of this panel as pooled data, interpreting differences in BMI across individuals as the effect of different health behaviours (healthy or unhealthy) on the entire BMI distribution. The underlying relationships are theoretically consistent if we are able to believe, or test, that the static reduced-form model, specified below, reflects the steady-state equilibrium conditions[10]. The steady-state assumption may not be valid for young people, that are still determining their optimal level of BMI, but it is surely satisfied if we consider a sample formed by older people in which BMI is assumed to be stable[11]. So, a relevant sensitivity analysis to verify the equilibrium assumption for our data is to compare the estimated parameters of the benchmark model with those obtained from a sub-sample which is assumed to be less age-sensitive (individuals aged 50 or more). If negligible differences are found between BMI estimates obtained from the full sample and those from older people, then this should mean that the results of the complete sample have a high degree of external validity in explaining the determinants of obesity in the UK.

In order fully to exploit information from the two BHPS waves, we account for short cyclical effects on variables by including a time dummy variable. Its inclusions is useful in identifying the unobserved time heterogeneity of individuals born in different periods.

3.4.2 The empirical model

Chou et al. (2004) list a number of hypotheses which link socio-economic determinants to body weight. Referring to their discussion and the literature they cite, we postulate that the following equation holds:

$$BMI_{i,t} = f\left(S_{i,t}, R_{j,t}, D_{i,t}\right) \tag{3.3}$$

[10]Within the context of Becker and Murphy (1988) household production model, we can also of think BMI as a health outcome, which is the result of choices made in a health production model (Lakdawalla and Philipson, 2002)

[11]As a by-product, the age variable can correct biases in self-reported measures of BMI, which tends to increase with age, particularly for height (Burkhauser and Cawley, 2008).

i is an index for individuals and t for time. $S_{i,t}$ denotes individual influences on body weight and $R_{j,t}$ the influence of specific macro-regional variables, which we are going to use to test our assumptions, while $D_{i,t}$ is a vector of socio-economic and demographic control variables.

The vector of individual variables $S_{i,t}$ contains the number of cigarettes smoked per day and whether or not in the household there is a woman who is invloved in a full-time job. We include in our model the number of cigarettes smoked because, although an inverse relationship between smoking and body weight has been documented in the clinical and economic literature, the effect of cigarette smoking on obesity remains inconclusive. Focusing on the economic literature, Chou et al. (2004), Rashad et al. (2006) and Baum (2009) have found that the decline in smoking rate by higher taxes or prices are associated with higher rates of obesity. Consistent with this finding, Flegal (2007) suggests that a decline in smoking increases obesity but these effects are estimated to be small. In contrast, Gruber and Frakes (2006) have found an opposite effect of smoking taxes on obesity using the same data. The evidence of this unexpected relationship was further supported by Cawley et al. (2004) when females groups were investigated. In addition, Nonnemaker et al. (2009) found no evidence between higher smoking taxes and obesity rates.

It has been widely argued that increased body weight is a response to expanded labour market opportunities for women which, by increasing the value of household time, have also increased the demand for prepared food. Although several studies have rejected this hypothesis (Cutler et al., 2003, Loureiro and Nayga, 2005), changes in the relative prices of prepared meals under increasing demand may indirectly be responsible for increased body weight. Under the hypothesis that, in a post-modern society, the marginal cost of an hour spent cooking at home is greater than the opportunity cost of an hour at work, the demand for prepared food increases as women, particularly mothers, tend to participate in work. Thus, average body weight is expected to increase as the female work participation rate rises. As previously stated we are going to discuss this hypothesis by including a dummy for the presence in the household of a woman involved in a full-time job.

While $R_{j,t}$ contains indicators for the price of fruit and vegetables, the price of take-

away restaurants, the density of restaurant and fast foods outlets and its squared value in the region of residence of each respondent. In his economic analysis of obesity, Philipson (2001) also emphasises the role of innovations. One such innovation, largely tested as a determinant in the obesity literature, concerns the growing availability of fast food and full-service restaurants. The spread of fast food is linked with an increase in less expensive food because, the greater food supply reduces the price of fast food with respect to other foods. In addition, the content of this food, more energy-dense, may corroborate the hypothesis of increases in body weight (Drewnowski et al., 2004, Schlosser, 2001). With respect to Auld and Powell (2009) and Chou et al. (2004), our data do not use separately the prices of fast food at regional level to test the hypothesis that reductions of these variables induce a substitution towards food consumption with higher calories. But, in the same way, we maintain the argument by including an index measuring the price of fruit and vegetables, at regional level, as a proxy behaviour of less energy-dense food, i.e. healthier food, so that we can examine whether price increases have significant effects on BMI growth. In these and all subsequent models, we also include the regional price of take-away meals and snacks as a control variable in $R_{j,t}$. Meeting household needs and work constraints, the great increase in take-away meals (and snacks) in the UK may have increased the proportion of energy-dense food in the diet and, on average, overweight. As argued in this literature, we are interested in testing this hypothesis in women[12].

In addition, the level of overweight has been found to be linked with the great increase in the *per capita* number of restaurants and fast-food outlets (see also Rashad et al. (2006)). It is known from studies in the United States that such outlets are located in areas where consumers put a relatively high price on their time. Currie et al. (2010) have found that, among pregnant women, the residence distance from fast food restaurants reduce the probability of gaining weight over 20 Kg. In our empirical analysis, we include the density of restaurant and fast food shops, which are assumed to be a proxy for unealthy food supply in each region and to be positively correlated with BMI, because of the reduction in time spent for searching this type of restaurants, while their likely non-linear

[12]The literature on food energy density did not confirm the concept that a decrease in the price of energy-dense food tends to increase total calorie consumption at aggregate level: if energy-dense foods become relatively cheaper, we may observe offsetting decreases in the consumption of less dense foods, so that total calories would change or even decrease (Auld and Powell, 2009).

influence is captured by the square of the same variable.

Table 3.1 lists all demographic variables $D_{i,t}$ as well as the variables included in the estimates. BMI is assumed to depend on (non-linear) age, race, marital status, education, and income. Schroeter et al. (2008) found in cross-country analyses that income changes could lead to weight gains, except in cases when all foods were inferior goods. However, the relationship between income and weight may differ given the narrow and small cross-country variability of work strenuousness. As argued by Lakdawalla and Philipson (2002), increases in income may raise or reduce weight. An increase in income could improve life conditions, granting access to larger quantities of food and thus increase BMI or could improve the eating habits of people allowing poor households to buy healthier food (with less calorie content) and thus have a negative effect on BMI. However, the magnitude of the income effect may be overestimated, due to reverse causality from obesity to income, i.e., endogeneity. Higher body weight may, indeed, lead to lower wages, due to effects on productivity or employment discrimination (Atella et al., 2008, Cawley et al., 2004). Weight and income may also be negatively correlated because of unobservable personal characteristics, such as self-discipline or impulsivity (Cutler et al., 2003). Also other variables may suffer from the same problem, but in this chapter we will not investigate this issue further. Thus our results must be interpreted only in terms of correlations between socio-economic determinants and BMI.

In our specification, we include three regional dummy variables, controlling for the effect of living in London, Yorkshire and the Humber and Scotland. These three regions are peculiar because, according to ''Statistics on Obesity, Physical Activity and Diet: England, February 2009", published by the Service (2009), and ''Obesity in Scotland: an epidemiology briefing", by the Scottish Public Health Observatory, inner and outer London are the areas with the lowest levels of obesity in the UK, while those of Yorkshire and the Humber and Scotland are the highest. In Scotland this result is true, especially for older women. However, excluding the possibility that the specific regional variables which we consider are correlated with genetic determinants, we examine therefore the socio-economic determinants of obesity in the UK net of the fact that regression disturbance terms may affect estimates.

Table 3.1: Data definitions and sources

Variable	**Definition**	**Source**
Job_hours	Number of hours normally worked per week, including overtime	BHPS
Phys_Activity	Dummy variable equal to one if respondents make physical activity at least once a week	BHPS
Strenuousness	Dummy variable that measures the strenousness of work in which respondents' are involved	BHPS
$Price_{F\&V}$	Price of fruits and vegetables	ONS
Price_{TA}	Price of take away and snacks	ONS
Rest/FF	Density of restaurants and fast food	ONS
Rest/FF^2	Squared density of and restaurants and fast food	ONS
N_Cigarettes	Number of cigarettes usually smoked per day	BHPS
Work_Mother	Dummy equal to one if the respondents' household mother is involved in a full time job	BHPS
Black	Dummy equal to one if respondents' ethnicity is black	BHPS
Age	Respondents' age	BHPS
Age^2	Respondents' squared age	BHPS
Net_Income	Net household income	BHPS
Net_Income^2	Squared net household income	BHPS
Couple	Dummy equal to one if respondents' marital status is couple	BHPS
Married	Dummy equal to one if respondents' marital status is married	BHPS
Divorced	Dummy equal to one if respondents' marital status is divorced	BHPS
Separated	Dummy equal to one if respondents' marital status is separated	BHPS
Widowed	Dummy equal to one if respondents' marital status is widowed	BHPS
Degree	Dummy equal to one if respondents' education is degree	BHPS
Diploma	Dummy equal to one if respondents' education is diploma	BHPS
Alevel	Dummy equal to one if respondents' education is Alevel	BHPS
Olevel	Dummy equal to one if respondents' education is Olevel	BHPS

Note: Data retrieved from British Household Panel Survey (BHPS) and Office for National Statistics (ONS)

In addition, we include in our analysis, accounting for the suggestions proposed by Lakdawalla and Philipson (2002) the number of hours worked (including overtime) to explain why a given individual may be overweight. We assume that workers who spend more extra hours at their jobs are more likely to be overweight than those who do normal job hours, because they have less time to devote to leisure and physical activity. This hypothesis is largely sustained by the increases in sedentary job in post-modern society.

Moreover, in Section 2, we showed how extra work commitments, but also lack of leisure time were the main reasons stated for not exercising, and that the latter explanation was mainly suggested by women. Alternatively to job hours we use a proxy for physical activity (dummy which takes value 1 when an individual exercises at least once a week and 0 otherwise) to measure the effect of the lack of leisure time on BMI. Since physical activity and job hours are both considered as proxies for the effect of reductions in leisure time on obesity we decided to specify two separate equations and avoid issues related to high correlation among our regressors. Formally, these specifications are given as:

$$BMI_{i,t,k} = f\left(W_{i,t,k}, S_{i,t,k}, R_{j,t,k}, D_{i,t,k}\right) \tag{3.4}$$

$k = 1$ identify the BMI reduced form that includes in the matrix $W_{i,t,k}$ the number of hours worked in a normal week (including overtime), while $k = 2$ the reduced form that includes the frequency of physical activity; as before $t = 2004, 2006$. If $k = 1$ we include also a control variable that measures the strenuosness of job (a dummy equal to one if the work is physically demanding), and $S_{i,t,k}$, $D_{i,t,k}$ and $R_{j,t,k}$ are matrices already described. Since gender is expected to influence BMI differently we will estimate separate gender models in order to highlight such differences.

3.5 Results

3.5.1 Preliminary results

Figure 3.4 shows the estimates of Epanechnikov kernel density functions for BMI distribution conditional on some covariates, below and above the median of our sample, and by gender.

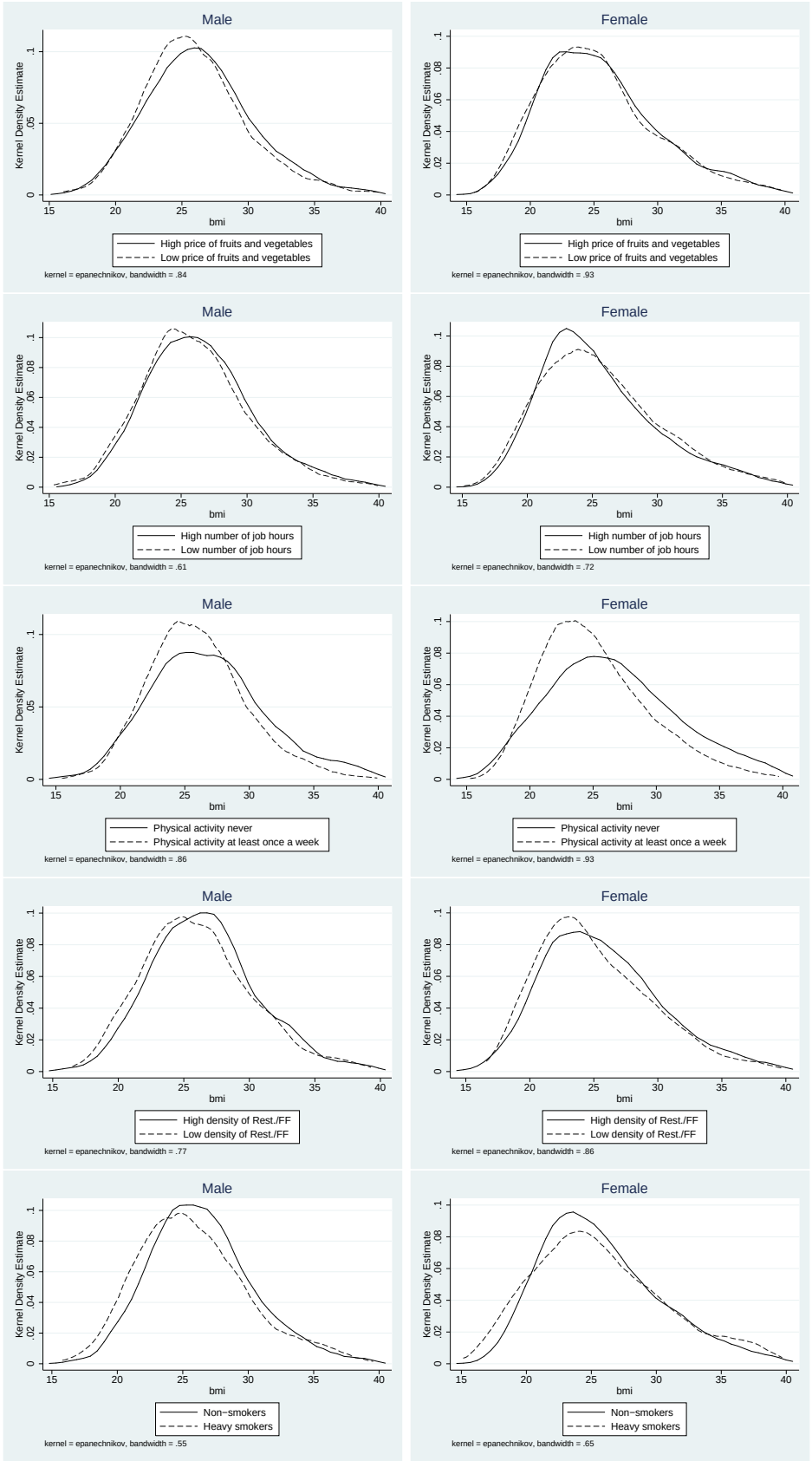


Figure 3.4: Kernel density estimates of BMI by gender

As a first result, we examine whether the error terms of OLS regression is normally distributed. Although the empirical conditional distributions for any panel in the figures are not very far from Gaussian distributions, they do not appear to meet the theoretical features required by BMI distributions and are skewed.

Table 3.2 lists BMI means and medians and measures the share of obese people at the threshold (i.e., BMI $\geq$ 30) for the covariates previously graphically analysed by kernel densities. For the price of fruit and vegetables, note that the average BMI for men living in an area with high prices is 1.59% higher than for those living in an area with lower prices. The situation is similar for women or when the median is taken into account. In line with our expectations, the proportion of obese people is estimated to be 17% of the distribution with respect to people living in areas with higher-priced fruit and vegetables, and 14% for lower-priced ones, respectively.

Table 3.2: Means and medians of BMI and share of obese people

	Male		Female		Total		Male	Female	Total
Variable	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**	**BMI $\geq$ 30**	**BMI $\geq$ 30**	**BMI $\geq$ 30**
High price of fruits & vegetables	26.41	26.15	25.54	24.85	25.93	25.54	17.51	16.94	17.27
Low price of fruits & vegetables	25.99	25.63	25.32	24.79	25.64	25.17	14.34	16.52	15.49
High number of job hours	26.42	26.11	25.46	24.62	26.02	25.54	17.38	16.19	16.88
Low number of job hours	26.02	25.63	25.61	24.94	25.76	25.23	15.72	13.69	14.73
Physical activity at least once a week	25.99	25.63	25.16	24.47	25.56	25.11	14.49	14.2	14.34
Physical activity never	26.85	26.52	26.63	26.17	26.72	26.31	22.19	24.98	23.84
High density of Restaurants and fast food	26.61	26.35	25.71	25.16	26.11	25.63	18.86	17.42	18.06
Low density of Restaurants and fast food	26.07	25.68	25.14	24.29	25.57	25.03	17.39	15.14	16.19
High number of cigarettes	25.79	26.11	25.65	25.04	25.73	25.61	17.65	17.83	17.75
Low number of cigarettes	26.46	25.38	25.73	24.85	26.07	25.12	15.08	19.89	17.41

Notes: The share of obese people has been obtained as $1 - F(BMI < 30)$*, where the probability of BMI lower than the obesity threshold has been calculated from the cumulative kernel density function of BMI conditioned to testing variables.*

Men working more than 30 hours a week (part-time work threshold) are more likely to have an average BMI higher than those working 30 hours or less (1.51% and 1.84% for the median). Instead, women do not reveal strong differences in the means and medians of empirical distributions. If we look at the share of obese adults, it is easy to note the fall (about 2%) for both men and women working less than 30 hours.

When we look at the variable which records physical activity habits, we observe huge differences between the BMI means and medians of people exercising at least once a week and those who never take any physical exercise: 3.31% for the mean and 3.47% for the median of men and 5.84% for the mean and 6.95% for the median of women. The quota

of estimated obese people for both men and women, is the 8% higher in the case of nophysical exercise, and this result is largely consistent with our expectations.

Lower densities of restaurants and fast-food outlets are associated with decreased BMI means and medians in people resident in such areas. However, the magnitude of the effects on BMI of the density of restaurants is not as large as expected. Consistently, the shares of obese people living in areas with lower densities of restaurants and fast-food shops decrease by 1% and 2% for men and women, respectively.

In order to understand the different impact of cigarette consumption on BMI, we functionally split our sample between "non-smokers", and "heavy smokers"[13] adults. Kernel densities, plotted by gender, show that the mean and median BMI of "heavy smokers" are smaller than those of "non-smokers". Moreover, the percentage of obese "heavy smokers" is smaller than that of obese "non-smokers", for men, although this relation is not supported by the graph for women. Although based on a descriptive approach, the impact of cigarette consumption seems to be significant on underweight and normal weight women, progressively falling in influence when we consider overweight and obese ones.

3.5.2 Estimates and discussion

Table 3.3 lists the values of the test of equality across quantiles for the covariates included in equation (3.4), separately for the equation which includes job hours (hereafter, model (1)) and physical activities (model (2)). This test is valid if, at least, one estimated percentile coefficient has a different effect with respect to the others. For the equations for women, we find larger differences in quantile estimates (e.g., physical activity habits, strenuousness of job, price of fruit and vegetables, density of restaurants and fast-food shops and its square, number of cigarettes smoked, black ethnicity, net income and net income squared, age and age squared, marital status, and education). For men, these differences in covariates are less marked (effects are significant for: physical activity habits, age and age squared, marital status and education). Thus, we proceed to estimate models by quantile regressions, and use OLS estimates to compare results.

Tables 3.4 and 3.5 list the BMI estimates of models (1) and (2) for selected quantiles

[13] "Heavy smokers" are adults smoking more than 20 cigarettes per day.

between the 10^{th} and 90^{th} percentile of the distribution. The parameter estimates of quantile regressions by gender are also shown in Figures 3.5-3.8[14].

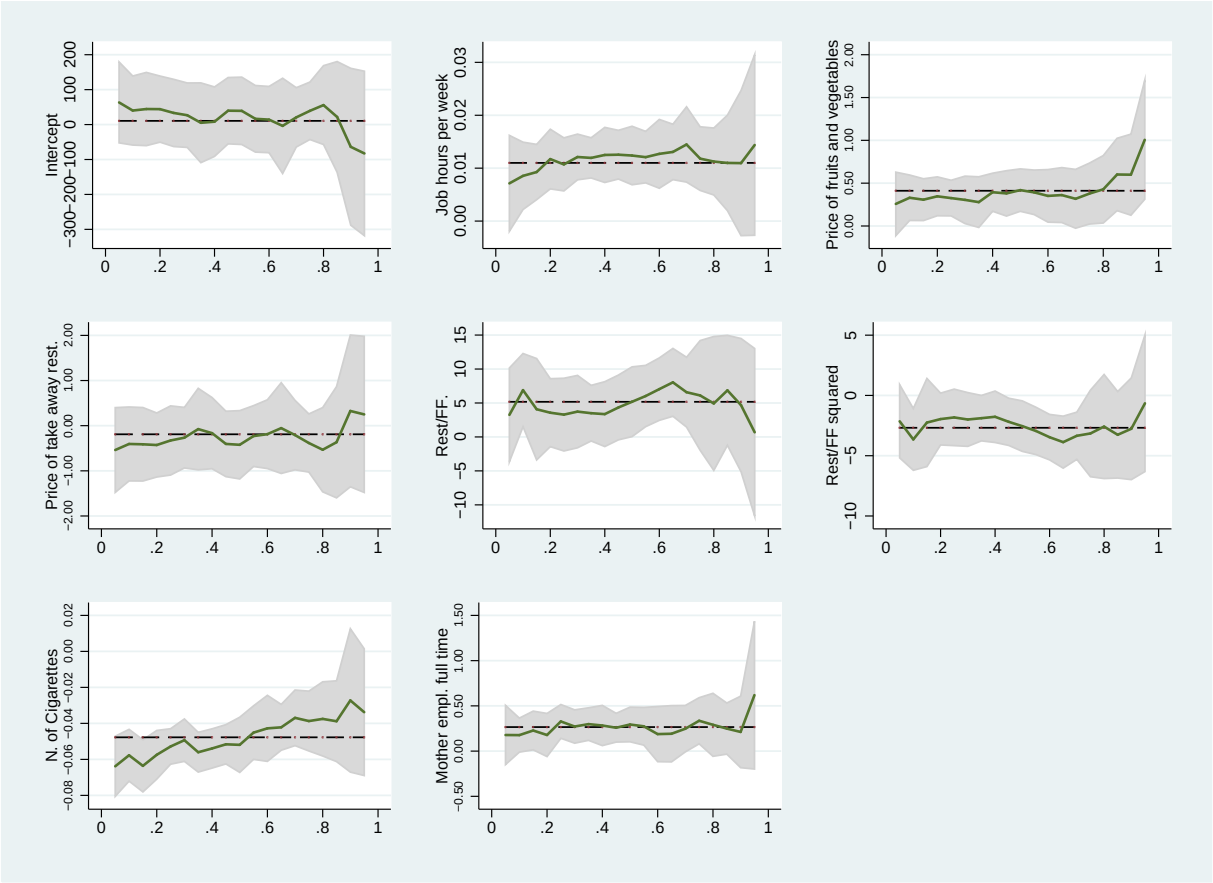


Figure 3.5: Quantile regression estimates: model (1), male

Irrespective of the model used, the estimated parameters of the (socio-demographic) control covariates are generally of the expected signs. Black respondents have a higher BMI than white respondents and, mostly for women, the coefficients vary across quantiles. Higher education is associated with a lower BMI. In addition, income effects are not significant for UK male respondents but are negative for female ones, for both OLS and quantile regressions after the median of the BMI distribution, but with very different effects. Married respondents have a BMI similar to that of couples, but greater than divorced, separated or widowed people.

[14]We report the empirical BMI distribution which corresponds to some points of quantile estimates. The 10^{th} percentile of BMI distribution corresponds to a BMI of 20.65 Kg/m^2 for men and 20.72 Kg/m^2 for women, the 25^{th} to 22.62 Kg/m^2 for men and 23 Kg/m^2 for women, the 50^{th} to 25.23 Kg/m^2 for men and 25.62 Kg/m^2 for women, the 75^{th} to 28.48 Kg/m^2 for men and 28.81 for women, and the 90^{th} to 31.95 Kg/m^2 for men and 32.50 Kg/m^2 for women.

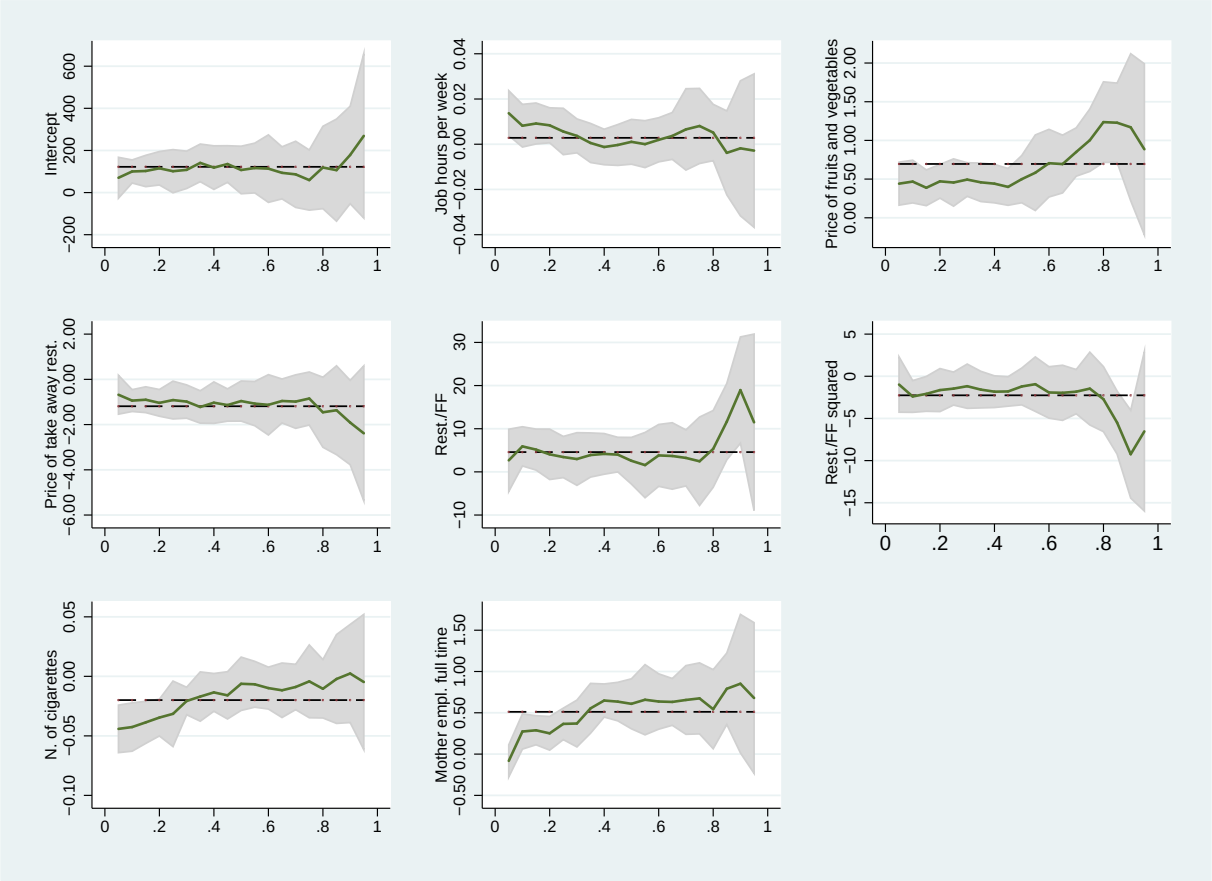


Figure 3.6: Quantile regression estimates: model (1), female

One first result of our estimates is that we find that body weight in men, is significantly and positively affected by job hours and that the effect is evenly distributed across quantiles (Table 3.4). Instead, women's BMI is not affected by job hours except for those at 10^{th} percentile. We conclude that employed men gain weight if they spend more hours at work, irrespective of whether OLS or quantile regressions are used. Comparing these results with the specification estimated directly by including the frequency of physical activity (Table 3.5), we find significant effects for higher quantiles with respect to the median value of BMI.

Below, if not specified, the estimated coefficients should be considered to produce similar effects through models (1) and (2). The price of fruit and vegetables is responsible for changes in BMI, with a larger effect on women. Although quantile estimates are very close to OLS up to the median, they become larger when estimated for overweight people. Gender differences are found in explaining food price effects of take-aways and restaurants

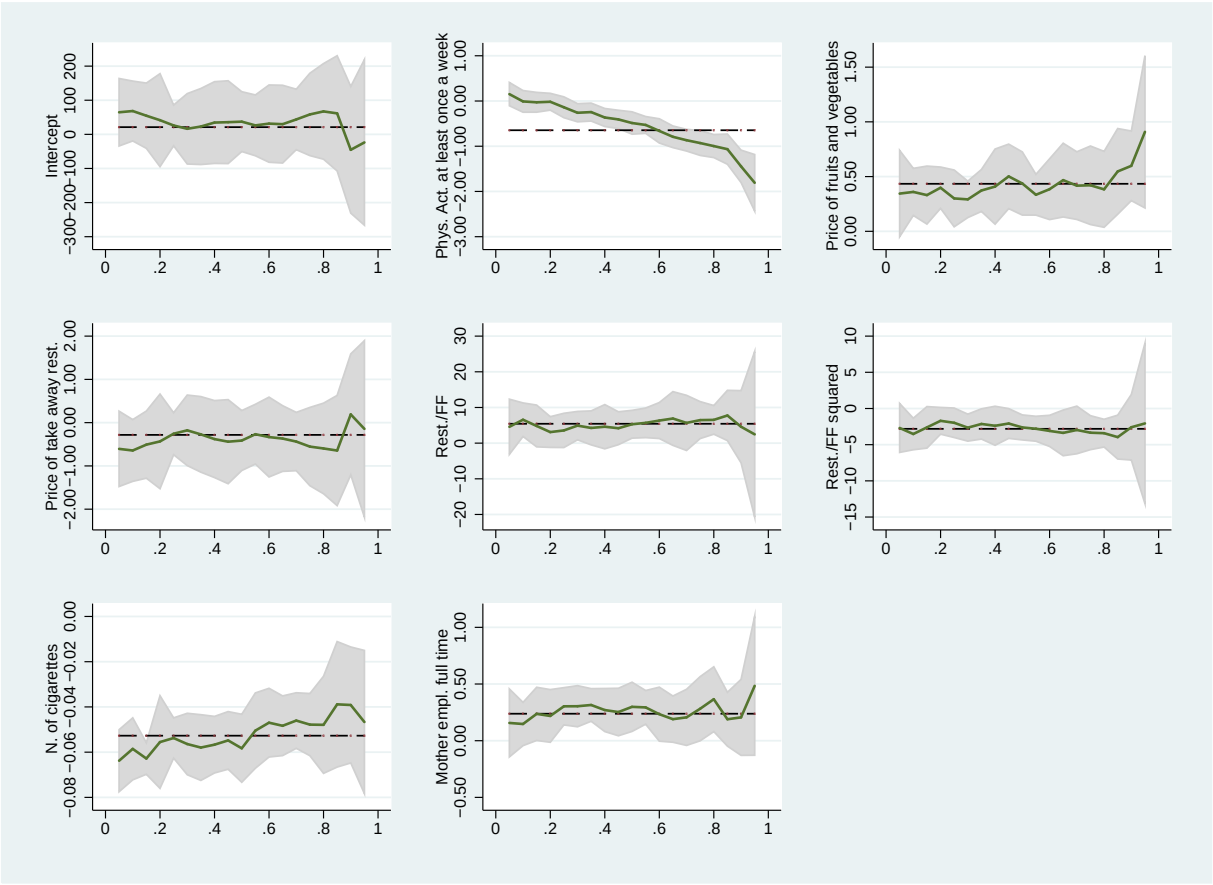


Figure 3.7: Quantile regression estimates: model (2), male

on BMI. Under the hypothesis of a greater propensity to supply more energy-dense food in take-aways and restaurants, for the women group we note the significant and largely negative impact on body weight at the 90^{th} percentile. The dimension of these effects is also confirmed by including the variable related to the presence in the household of a working mother.

The density of restaurants and fast-food outlets is significant for some quantiles of the samples analysed. Their growing availability positively affects men's BMI, with positive and significant coefficients in the 10^{th}, 50^{th} and 90^{th} quantiles, and is barely significant for the OLS model. The coefficient is almost the same across quantiles, except for the 90^{th} percentile, where its measure is three times larger than that of OLS. Apart from the 90^{th} quantile parameter, none of the others is significant for women. These estimates are consistent with the results obtained by Chou et al. (2004). The density of restaurants and fast-food shops induces an increase in the BMI in men who spend more time at

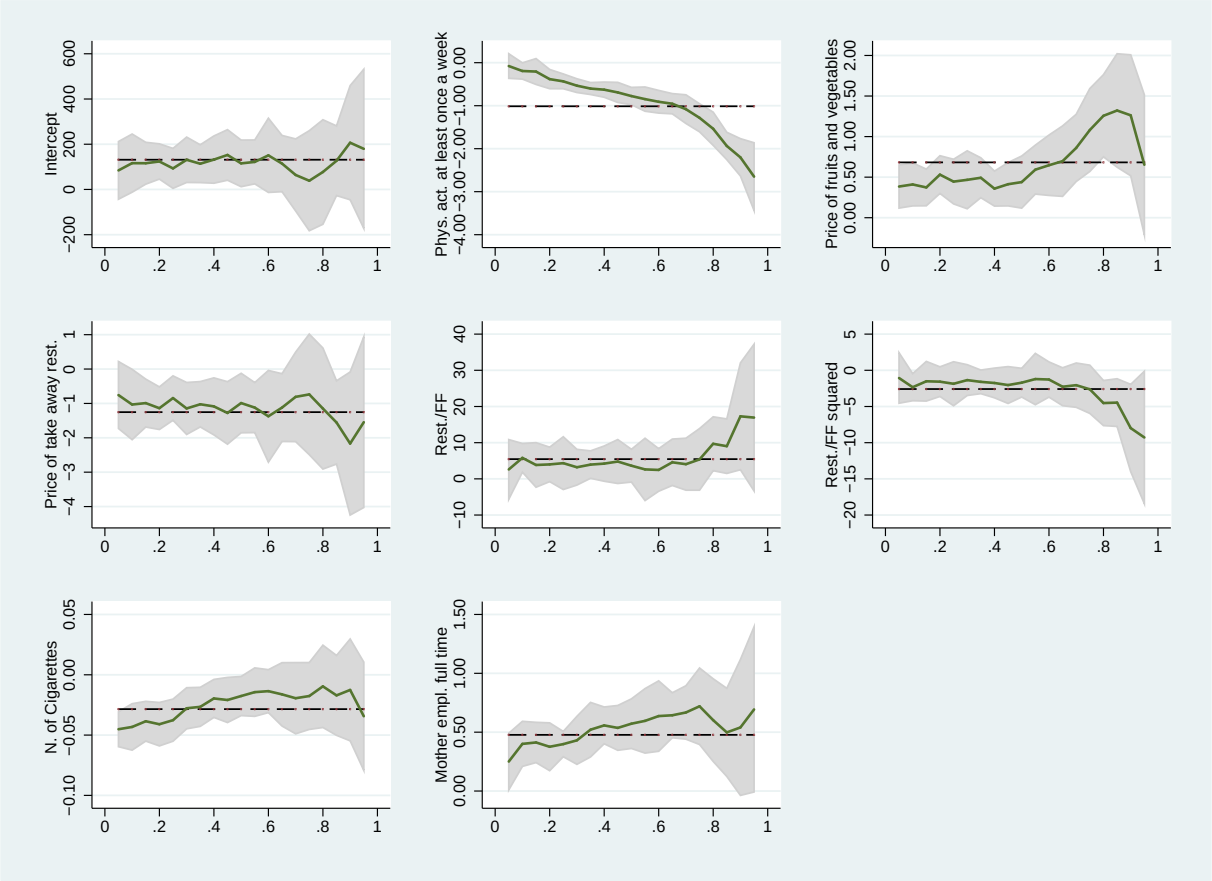


Figure 3.8: Quantile regression estimates: model (2), female

work while, on average, it is less responsible for increased BMI in women. This result is contradicted by the estimates at the 90^{th} quantile, where the values for overweight women become statistically significant.

In line with the explanation for men, the effect of the spread of restaurants and fast-food outlets on high BMI seems to depend positively on extra time worked, stimulating a demand for outside food, mainly fast-food, which increases calorie intake[15]. Lastly, also in the UK a negative association between cigarette consumption and BMI is empirically confirmed, and this is true for each estimated quantile except the most extreme ones.

As discussed in section 3, we estimate quantile regressions for the subsample of people aged over 50, assumed to be stable to long-term imbalances in energy intake and expenditure. The results are listed in Appendixes 3.A.1 and 3.A.2 With respect to the estimates for the complete sample showed in Table 3.4 and 3.5, we do not find remarkable

[15] Although data are not reported, the dataset does show a positive relationship between extra job hours and larger share of women's BMI. This additional analysis is available from the authors upon request.

differences in the coefficients of covariates affecting weight distribution. Only for some quantiles around the median, we denote slightly larger differences of the estimated parameters between the two samples. This implies that we cannot reject a BMI's steady-state condition for the entire sample[16].

Table 3.6 lists the estimated BMI effects of a 1% increase in the covariates described above. As previously stated, health policies based on OLS results would not efficiently measure the effect of some variables involved in the analysis on overweight and obese individuals. For example, if we focus on the price of fruit and vegetables, we note low estimated elasticity from OLS for men and women, but the coefficient associated to this variable becomes larger in quantile regressions when we consider women located beyond the 75^{th} percentile. The Table shows that most of these effects are quite minor, and often fail to be large or enough precisely measured to achieve statistical significance. The results thus indicate that changes in the price of fruit and vegetables affect each quantile of the BMI distribution with moderate effects for overweight people. Restaurant and fast-food densities have a significant effect on weight for men and women over the 50^{th} percentile. As expected, the number of cigarettes has a significant negative effect on body weight for much of the empirical distribution.

However, these effects may have more intuitive implications when they are expressed as changes in body weight due to policy interventions. Let us consider a representative adult at the average of the sample and at the 90^{th} percentile of the conditional BMI distributions for men and women. Admit a subsidy which decreases the price of fruit and vegetables and encourages the consumption of these healthier food. The value of these ''thin subsidies" is assumed to be 10% of the market price. Following our OLS estimates in Table 3.6 (model 1) carried out by gender, BMI would decrease by about 0.16 for men and 0.235 for women but would increase to 0.19 and 0.29, respectively, when we measure the effects for people at the 90^{th} percentile. This means that a man 1.75 m tall, weighting 80.23 kg at the mean of the sample, corresponding to the average BMI (26.2) could expect to be lighter by 0.5 Kg per year, whereas a representative woman (height 1.61 m and BMI 25.43) could expect a decrease of 0.75 kg if price subsidies for

[16]We also performed estimations that included higher polynomial orders of age covariate. The estimates were close to those reported in Table 3.4 and 3.5 and Appendix 3.A.1 and 3.B.2, that included the covariate age and age squared.

healthy food were available. This reduction is emphasised when we evaluate people at the 90^{th} percentile. In this case, the effects of reduced body weight are 0.66 kg for men and 0.79 kg for women. We do not have a specular proxy for evaluating the effects of taxation on unhealthy foods. We note that, as an alternative impact on body weight, several countries plan to impose or broaden sales taxes on soft drinks and other food items (for a discussion, see Uhlman, 2003). This is in line with several recent laws passed to discourage the consumption of unhealthy foods by increasing their effective prices to consumers. The UK has considered the introduction of various value-added taxes for food of poor nutritional value (Kuchler et al., 2005, Schroeter et al., 2008) although this has been recognised as a progressive burden for low-income families which spend a large portion of their income on food (e.g., Cash et al. (2004)).

We can repeat the exercise for changes in income. In addition to ''fat" taxes and ''thin" subsidies, several studies have determined that income has a major influence on obesity (e.g. Deaton and Paxson (1999), Drewnowski and Darmon (2005)). In developed economies, households with higher incomes tend to consume higher-quality diets consisting mainly of low-calorie foods, whereas low-income households, which generally use more energy-dense foods, have problems of overweight. Note that from our estimates this evidence is only partly sustained. Only non-working women show significant reductions in overweight and obesity as a response to increases in income. Consequently, any policy that reduces inequalities in the income distribution across women can reduce overweight. As a quantitative example, a hypothetical increase in income of 5%, generated through public intervention, is reflected in a decrease in women's weight by 0.70 Kg, which more than doubles (1.16 kg) when obese women at the 90^{th} percentile are taken into account.

Table 3.3: Test for equality of coefficients across quantiles

	Models			
	(1)		**(2)**	
Variable	M	F	M	F
Job_hours	0.28	1.54	-	-
	(0.889)	(0.188)	-	-
Phys Act	-	-	28.59	17.28
	-	-	(0.000)	(0.000)
Strenuousness	1.74	1.48	1.91	2.04
	(0.138)	(0.204)	(0.107)	(0.086)
$\text{Price}_{F\&V}$	0.45	1.95	0.51	2.46
	(0.774)	(0.099)	(0.729)	(0.043)
Price_{TA}	0.22	0.69	0.61	0.83
	(0.924)	(0.597)	(0.663)	(0.507)
Rest/FF	0.94	3.26	0.81	1.71
	(0.442)	(0.011)	(0.517)	(0.145)
Rest/FF^2	1.07	3.87	1.01	2.79
	(0.369)	(0.004)	(0.411)	(0.024)
N_Cigarettes	1.28	5.61	0.61	4.75
	(0.275)	(0.002)	(0.662)	(0.000)
Work_Mother	0.29	2.14	0.82	1.07
	(0.886)	(0.073)	(0.513)	(0.369)
Black	1.04	3.47	0.18	2.71
	(0.384)	(0.007)	(0.951)	(0.029)
Age	12.17	21.03	12.34	11.93
	(0.000)	(0.000)	(0.000)	(0.000)
Age^2	14.37	18.72	17.38	12.41
	(0.000)	(0.000)	(0.000)	(0.000)
Net_Income	0.65	4.06	0.36	3.88
	(0.627)	(0.002)	(0.839)	(0.003)
Net_Income^2	0.03	2.41	0.28	2.49
	(0.999)	(0.047)	(0.888)	(0.041)
Couple	0.65	0.591	1.01	0.23
	(0.627)	(0.672)	(0.408)	(0.922)
Married	2.42	0.91	1.82	0.28
	(0.046)	(0.463)	(0.125)	(0.891)
Divorced	1.84	3.14	3.18	3.53
	(0.118)	(0.014)	(0.012)	(0.007)
Separated	1‘.14	1.11	1.22	1.23
	(0.335)	(0.355)	(0.301)	(0.296)
Widowed	2.27	2.8	3.27	1.06
	(0.059)	(0.024)	(0.010)	(0.376)
Degree	3.81	4.55	7.66	3.81
	(0.004)	(0.001)	(0.000)	(0.007)
Diploma	3.95	7.51	4.19	1.91
	(0.003)	(0.000)	(0.002)	(0.105)
Alevel	4.09	2.05	7.68	2.03
	(0.002)	(0.085)	(0.000)	(0.087)
Olevel	8.64	1.82	8.59	3.53
	(0.000)	(0.122)	(0.000)	(0.007)

Note: p-values are shown in brackets and significant levels are reported with the following notation:
Model (1) includes in the vector of the explanatory variables job hours while model (2) uses physical activity.

Table 3.4: BMI OLS and quantile regressions: model (1)

	OLS		**Q 0.1**		**Q 0.25**		**Q 0.5**		**Q 0.75**		**Q 0.9**	
Variable	M	F	M	F	M	F	M	F	M	F	M	F
Job_hours	0.011***	-0.004	0.009***	0.006**	0.011***	-0.001	0.012***	-0.005	0.012***	-0.006	0.010**	-0.007
	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)	(0.007)	(0.006)	(0.011)
Strenuousness	-0.169	-0.294.	-0.049	0.074	0.013	-0.116	-0.008	-0.122	-0.228	-0.251	-0.560	-0.756**
	(0.135)	(0.172)	(0.121)	(0.078)	(0.111)	(0.135)	(0.113)	(0.151)	(0.242)	(0.185)	(0.382)	(0.323)
$\text{Price}_{F\&V}$	0.412***	0.631***	0.330***	0.329***	0.326**	0.389***	0.419***	0.514***	0.378***	0.851***	0.599***	0.937***
	(0.148)	(0.166)	(0.107)	(0.097)	(0.128)	(0.120)	(0.131)	(0.128)	(0.121)	(0.207)	(0.241)	(0.285)
Price_{TA}	-0.190	-1.077***	-0.404	-0.697**	-0.327	-0.990**	-0.423	-0.923**	-0.384	-1.07	-0.3268	-1.767**
	(0.283)	(0.309)	(0.403)	(0.273)	(0.363)	(0.419)	(0.388)	(0.371)	(0.702)	(0.344)	(0.734)	(0.870)
Rest/FF	5.175*	4.684	6.893**	4.568**	3.288	3.502	5.182**	4.099	6.111**	5.083	4.684	17.694**
	(3.129)	(3.355)	(2.909)	(1.875)	(2.408)	(2.301)	(2.488)	(3.001)	(3.115)	(4.254)	(7.208)	(8.055)
Rest/FF^2	-2.680**	-2.349*	-3.647***	-1.883**	-1.821.	-1.467	-2.547**	-1.884	-3.153**	-2.519	-2.763	-8.701**
	(1.363)	(1.418)	(1.328)	(0.821)	(1.040)	(0.978)	(1.110)	(1.306)	(1.305)	(1.700)	(2.885)	(3.433)
N_Cigarettes	-0.048***	-0.036***	-0.058***	-0.057***	-0.053***	-0.051***	-0.052***	-0.020**	-0.038***	-0.023	0.027	-0.007
	(0.009)	(0.010)	(0.007)	(0.007)	(0.006)	(0.008)	(0.007)	(0.009)	(0.011)	(0.02)	(0.018)	(0.022)
Work_Mother	0.266**	0.548***	0.177	0.287***	0.328***	0.418***	0.294**	0.644***	0.335***	0.744***	0.211	0.898***
	(0.119)	(0.147)	(0.121)	(0.078)	(0.102)	(0.137)	(0.120)	(0.162)	(0.148)	(0.164)	(0.184)	(0.326)
Black	0.690	0.584	0.809	-0.399	0.706	-0.574	0.701	0.804***	-0.054	0.322	0.501	3.983***
	(0.597)	(0.733)	(0.702)	(0.340)	(0.476)	(0.848)	(0.572)	(0.306)	(0.891)	(0.679)	(1.581)	(1.115)
Age	0.321***	0.291***	0.207***	0.178***	0.240***	0.216***	0.281***	0.250***	0.334***	0.353***	0.429***	0.446***
	(0.022)	(0.023)	(0.018)	(0.015)	(0.018)	(0.017)	(0.016)	(0.023)	(0.021)	(0.025)	(0.033)	(0.050)
Age^2	-0.003***	-0.003***	-0.002***	-0.002***	-0.002***	-0.002***	-0.003***	-0.002***	0	-0.003***	0	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Net_Income	-0.107	-1.31***	0.376	0.170	0.428	-0.724	0.158	-1.422***	-0.684	-1.239	-0.446	-2.631**
	(0.557)	(0.507)	(0.642)	(0.281)	(0.657)	(0.446)	(0.518)	(0.526)	(0.634)	(0.805)	(0.839)	(1.136)
Net_Income^2	-0.163	0.126	-0.213	-0.179	-0.162	0.189	-0.163	0.209	-0.035	0.042	-0.029	0.239
	(0.320)	(0.170)	(0.407)	(0.120)	(0.544)	(0.237)	(0.360)	(0.223)	(0.337)	(0.346)	(0.401)	(0.440)
Couple	0.483**	0.570**	0.845***	0.462.	0.751***	0.662***	0.623***	0.682***	0.411	0.673**	-0.002	0.477
	(0.233)	(0.238)	(0.184)	(0.251)	(0.190)	(0.231)	(0.189)	(0.228)	(0.278)	(0.316)	(0.552)	(0.522)
Married	0.554**	0.536**	1.219***	0.411**	1.017***	0.463**	0.806***	0.404**	0.528***	0.730**	-0.343***	0.987***
	(0.228)	(0.243)	(0.172)	(0.199)	(0.210)	(0.187)	(0.167)	(0.165)	(0.201)	(0.307)	(0.445)	(0.344)
Divorced	0.222	0.566	1.005***	-0.265	0.585.	0.253	0.739**	0.448	0.119	0.893**	-0.517	1.022*
	(0.368)	(0.350)	(0.308)	(0.268)	(0.315)	(0.239)	(0.324)	(0.346)	(0.319)	(0.431)	(0.881)	(0.605)
Separated	-0.398	0.282	0.243	0.096	-0.023	0.116	0.084	-0.169	-0.477	0.669	-1.979	1.688
	(0.492)	(0.497)	(0.561)	(0.229)	(0.387)	(0.407)	(0.436)	(0.410)	(0.459)	(0.770)	(1.544)	(1.131)
Widowed	0.684**	0.736**	1.529***	0.167	1.405***	0.361	0.898***	0.601**	0.621*	0.903***	-0.416	1.977***
	(0.333)	(0.338)	(0.389)	(0.205)	(0.297)	(0.426)	(0.276)	(0.264)	(0.341)	(0.324)	(0.575)	(0.449)
Degree	-1.177***	-1.881***	-0.786***	-0.972***	-0.667***	-1.414***	-0.855***	-1.746***	-2.104***	-2.408***	-2.229***	-2.647***
	(0.230)	(0.237)	(0.204)	(0.164)	(0.183)	(0.209)	(0.209)	(0.197)	(0.284)	(0.332)	(0.547)	(0.498)
Diploma	-0.589***	-1.138***	-0.137	-0.598***	-0.278.	-1.026***	-0.416***	-1.082***	-1.105***	-1.156***	-0.968***	-1.170***
	(0.184)	(0.191)	(0.159)	(0.107)	(0.144)	(0.203)	(0.147)	(0.178)	(0.255)	(0.250)	(0.309)	(0.341)
Alevel	-0.597**	-0.769***	-0.100	-0.571***	-0.396**	-1.060***	-0.532***	-1.025***	-1.050***	-0.820.	-1.708***	-0.375
	(0.235)	(0.262)	(0.178)	(0.108)	(0.155)	(0.246)	(0.189)	(0.202)	(0.233)	(0.429)	(0.514)	(0.538)
Olevel	-0.433**	-0.78***	0.283	-0.552***	0.109	-0.831***	-0.244	-0.692***	-0.710**	-0.859**	-1.580***	-0.634
	(0.205)	(0.218)	(0.176)	(0.127)	(0.167)	(0.163)	(0.176)	(0.185)	(0.217)	(0.406)	(0.431)	(0.391)
D_{2004}	-0.805***	-1.332***	-0.634***	-0.370***	-0.644***	-0.785***	-0.862***	-1.153***	-0.791***	-1.793***	-1.210***	-2.182***
	(0.187)	(0.206)	(0.181)	(0.143)	(0.192)	(0.176)	(0.196)	(0.245)	(0.216)	(0.329)	(0.347)	(0.592)
Cons.	10.692	108.293***	40.178	78.087**	33.388	112.956*	39.375	97.073**	39.346	91.655	-63.563	170.719
	(32.749)	(35.488)	(52.891)	(38.538)	(46.623)	(51.289)	(48.307)	(44.591)	(44.860)	(91.005)	(87.869)	(105.107)
R^2	0.09	0.07	0.09	0.04	0.07	0.04	0.06	0.05	0.04	0.05	0.03	0.04

*Note: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** ≤ 0.01, ** ≤ 0.05, * ≤ 0.1. The description of variables are reported in Table 3.1.*

Table 3.5: BMI OLS and quantile regressions: model (2)

	OLS		**Q 0.1**		**Q 0.25**		**Q 0.5**		**Q 0.75**		**Q 0.9**	
Variable	M	F	M	F	M	F	M	F	M	F	M	F
Phys Act	-0.646***	-1.124***	-0.009	-0.211**	-0.139	-0.546***	-0.486***	-0.844***	-0.929***	-1.386***	-1.444***	-2.233***
	(0.112)	(0.118)	(0.127)	(0.093)	(0.093)	(0.047)	(0.119)	(0.103)	(0.119)	(0.188)	(0.233)	(0.222)
Strenuousness	-0.076	-0.298.	0.05	0.082	0.109	-0.078	0.147	-0.23	-0.133	-0.188	-0.17	-0.549
	(0.132)	(0.169)	(0.115)	(0.119)	(0.107)	(0.182)	(0.117)	(0.212)	(0.141)	(0.18)	(0.268)	(0.381)
$\text{Price}_{F\&V}$	0.434***	0.609***	0.36***	0.302**	0.301**	0.444***	0.437***	0.409**	0.421**	0.959***	0.598**	0.721**
	(0.147)	(0.165)	(0.138)	(0.127)	(0.123)	(0.139)	(0.142)	(0.169)	(0.179)	(0.273)	(0.299)	(0.299)
Price_{TA}	-0.282	-1.15***	-0.643*	-0.726**	-0.256	-1.208**	-0.414	-1.057**	-0.553	-0.913	0.191	-1.734*
	(0.28)	(0.307)	(0.363)	(0.358)	(0.352)	(0.477)	(0.464)	(0.468)	(0.567)	(0.775)	(0.822)	(0.959)
Rest/FF	5.434*	5.619*	6.569**	5.056	3.569	3.062	5.295**	3.717	6.459**	5.228*	4.604	15.215*
	(3.101)	(3.323)	(2.944)	(3.22)	(2.294)	(3.442)	(2.508)	(2.467)	(3.133)	(2.697)	(5.567)	(6.228)
Rest/FF^2	-2.810**	-2.694*	-3.524**	-2.118	-1.975**	-1.364	-2.619**	-1.669*	-3.341***	-2.591**	-2.593	-7.582***
	(1.348)	(1.403)	(1.377)	(1.355)	(0.980)	(1.428)	(1.102)	(1.007)	(1.294)	(1.116)	(2.538)	(2.564)
N_Cigarettes	-0.053***	-0.046***	-0.059***	-0.058***	-0.054***	-0.054***	-0.058***	-0.029***	-0.048***	-0.037***	-0.039**	-0.026
	(0.009)	(0.010)	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)	(0.010)	(0.011)	(0.015)	(0.022)
Work_Mother	0.238**	0.377***	0.147	0.328***	0.304***	0.360***	0.299***	0.494***	0.282*	0.542***	0.206	0.338*
	(0.118)	(0.118)	(0.123)	(0.089)	(0.085)	(0.087)	(0.101)	(0.098)	(0.153)	(0.142)	(0.211)	(0.196)
Black	0.649	0.545	0.434	-0.408	0.839*	-0.402	0.508	0.782**	0.437	0.523	0.167	4.098***
	(0.583)	(0.744)	(0.799)	(0.413)	(0.497)	(0.817)	(0.499)	(0.386)	(0.619)	(0.762)	(1.289)	(1.153)
Age	0.322***	0.299***	0.223***	0.182***	0.245***	0.222***	0.282***	0.262***	0.340***	0.362***	0.419***	0.439***
	(0.022)	(0.023)	(0.018)	(0.017)	(0.019)	(0.016)	(0.017)	(0.020)	(0.020)	(0.023)	(0.031)	(0.055)
Age^2	-0.003***	-0.003***	-0.002***	-0.002***	-0.002***	-0.002***	-0.003***	-0.002***	-0.003***	-0.003***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Net_Income	0.514	-1.221**	0.866	0.375	0.831	-0.738.	0.741	-1.46***	0.331	-1.716**	-0.422	-2.306***
	(0.552)	(0.499)	(0.666)	(0.272)	(0.578)	(0.434)	(0.571)	(0.466)	(0.812)	(0.837)	(1.131)	(0.767)
Net_Income^2	-0.405	0.123	-0.460	-0.230	-0.255	0.204	-0.393	0.231	-0.512	0.381	-0.009	0.214
	(0.319)	(0.179)	(0.438)	(0.187)	(0.439)	(0.218)	(0.323)	(0.247)	(0.570)	(0.489)	(0.571)	(0.234)
Couple	0.517**	0.457.	0.921***	0.442**	0.823***	0.574***	0.679***	0.569***	0.399*	0.391	0.322	0.221
	(0.232)	(0.235)	(0.192)	(0.221)	(0.158)	(0.113)	(0.200)	(0.147)	(0.240)	(0.239)	(0.453)	(0.485)
Married	0.604***	0.425*	1.288***	0.432**	1.089***	0.383**	0.890***	0.344	0.457**	0.535**	0.051	0.695
	(0.226)	(0.240)	(0.158)	(0.180)	(0.162)	(0.182)	(0.155)	(0.224)	(0.214)	(0.244)	(0.335)	(0.463)
Divorced	0.271	0.456	1.046***	-0.190	0.690**	0.157	0.746***	0.412	0.011	0.687*	0.036	1.020*
	(0.366)	(0.344)	(0.244)	(0.216)	(0.272)	(0.334)	(0.282)	(0.268)	(0.316)	(0.366)	(0.699)	(0.600)
Separated	-0.387	0.163	0.154	0.172	0.084	0.090	0.202	-0.139	-0.493	0.845	-1.546	0.833
	(0.493)	(0.490)	(0.486)	(0.380)	(0.369)	(0.456)	(0.649)	(0.581)	(0.380)	(0.578)	(1.374)	(0.816)
Widowed	0.796**	0.648.	1.557***	0.225	1.530***	0.295	0.967***	0.540	0.725.	0.680**	0.190	1.614***
	(0.331)	(0.332)	(0.355)	(0.265)	(0.271)	(0.239)	(0.311)	(0.415)	(0.384)	(0.326)	(0.443)	(0.607)
Degree	-1.107***	-1.775***	-0.702***	-0.923***	-0.565***	-1.239***	-0.877***	-1.713***	-1.101***	-2.164***	-2.235***	-2.614***
	(0.228)	(0.233)	(0.177)	(0.229)	(0.183)	(0.158)	(0.188)	(0.176)	(0.229)	(0.217)	(0.361)	(0.437)
Diploma	-0.526***	-1.065***	-0.108	-0.572***	-0.160	-0.904***	-0.358***	-1.046***	-0.579***	-1.137***	-1.157***	-1.031***
	(0.183)	(0.189)	(0.133)	(0.125)	(0.146)	(0.141)	(0.129)	(0.124)	(0.210)	(0.237)	(0.246)	(0.348)
Alevel	-0.581**	-0.700***	-0.099	-0.548***	-0.303*	-0.930***	-0.587***	-1.043***	-1.070***	-0.761**	-1.309***	-0.290
	(0.233)	(0.259)	(0.156)	(0.165)	(0.166)	(0.198)	(0.187)	(0.149)	(0.233)	(0.337)	(0.385)	(0.782)
Olevel	-0.412**	-0.739***	0.326**	-0.503***	0.164	-0.732***	-0.254*	-0.687***	-0.604***	-1.000***	-1.488***	-0.595*
	(0.204)	(0.216)	(0.152)	(0.179)	(0.139)	(0.220)	(0.146)	(0.235)	(0.210)	(0.284)	(0.296)	(0.342)
D_{2004}	-0.860***	-1.364***	-0.731***	-0.372**	-0.623***	-0.949***	-0.846***	-1.113***	-0.882***	-1.893***	-1.346***	-1.868***
	(0.186)	(0.205)	(0.197)	(0.150)	(0.193)	(0.196)	(0.219)	(0.279)	(0.275)	(0.321)	(0.394)	(0.504)
Cons.	21.179	119.175***	68.399	83.632	26.112	137.320**	37.310	122.377**	58.419	64.796	-45.313	185.415
	(32.350)	(35.188)	(47.490)	(50.991)	(45.333)	(58.274)	(58.865)	(58.542)	(72.175)	(105.989)	(109.700)	(129.646)
R^2	0.09	0.08	0.09	0.04	0.07	0.05	0.06	0.05	0.04	0.06	0.03	0.06

*Note: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** ≤ 0.01, ** ≤ 0.05, * ≤ 0.1. The description of variables are reported in Table 3.1.*

Table 3.6: BMI effect of a 1% increase of selected variables: model(1)

	OLS		**Q 0.1**		**Q 0.25**		**Q 0.5**		**Q 0.75**		**Q 0.9**	
Variable	M	F	M	F	M	F	M	F	M	F	M	F
Job_hours (x100)	0.042***	-0.016	0.039***	0.031*	0.045***	-0.001	0.048***	-0.02	0.041***	-0.022	0.035**	-0.022
	(0.012)	(0.021)	(0.011)	(0.017)	(0.009)	(0.017)	(0.01)	(0.018)	(0.012)	(0.024)	(0.014)	(0.034)
$\text{Price}_{F\&V}$	0.016***	0.024***	0.015**	0.016***	0.014***	0.017***	0.016***	0.021***	0.013***	0.03***	0.019***	0.029***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)	(0.011)
Price_{TA}	-0.007	-0.042***	-0.019	-0.034	-0.014	-0.044***	-0.016	-0.037**	-0.013	-0.038**	0.01	-0.055
	(0.011)	(0.012)	(0.02)	(0.023)	(0.012)	(0.016)	(0.014)	(0.015)	(0.016)	(0.019)	(0.026)	(0.039)
Rest/FF	0.196*	0.181	0.317**	0.223**	0.14	0.156	0.201***	0.164	0.214*	0.179*	0.149	0.547***
	(0.119)	(0.13)	(0.154)	(0.108)	(0.104)	(0.101)	(0.071)	(0.103)	(0.119)	(0.106)	(0.182)	(0.196)
N_Cigarettes	-0.002***	-0.001***	-0.003***	-0.003***	-0.002***	-0.002***	-0.002***	-0.001**	-0.001***	-0.001**	-0.001*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Net_Income	-0.004	-0.051***	0.017	0.008	0.018	-0.032**	0.006	-0.057***	-0.024	-0.044*	-0.014	-0.081**
	(0.021)	(0.02)	(0.028)	(0.015)	(0.019)	(0.014)	(0.021)	(0.014)	(0.029)	(0.025)	(0.042)	(0.038)

*Note: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

APPENDIX 3.A.1

Table 3.A.1: BMI OLS and quantile regressions: model(1), people over 50

	OLS		**Q 0.1**		**Q 0.25**		**Q 0.5**		**Q 0.75**		**Q 0.9**	
Variable	M	F	M	F	M	F	M	F	M	F	M	F
Job_hours	0.001	-0.028***	0.004	-0.005	0.006	-0.013	0.010**	-0.02	0.002	-0.034**	-0.001	-0.081***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)	(0.007)	(0.006)	(0.011)
Strenuousness	-0.303	-0.518	0.077	-0.066	-0.02	-0.380	-0.096	-0.712**	-0.666**	-0.607	-0.788.	-0.085
	(0.250)	(0.319)	(0.271)	(0.368)	(0.220)	(0.285)	(0.149)	(0.318)	(0.286)	(0.512)	(0.413)	(0.736)
$Price_{F\&V}$	0.370*	0.711***	0.185	0.434	0.254	0.611**	0.604***	0.826***	0.621.	1.096***	0.480**	0.638
	(0.222)	(0.253)	(0.250)	(0.319)	(0.174)	(0.241)	(0.178)	(0.232)	(0.323)	(0.331)	(0.235)	(0.400)
$Price_{TA}$	-0.410	-1.422***	-1.262	-1.121*	-1.043	-1.503**	-0.610	-1.315*	-0.319	-1.425**	0.428	-1.452
	(0.456)	(0.475)	(0.789)	(0.629)	(0.660)	(0.740)	(0.698)	(0.710)	(0.785)	(0.718)	(1.512)	(1.326)
Rest/FF	-0.724	7.223	6.622	2.959	4.902	6.868*	-1.088	6.550	-6.530	5.137	-7.888	16.919**
	(5.145)	(5.289)	(6.471)	(5.104)	(4.633)	(3.997)	(4.759)	(5.707)	(5.270)	(5.967)	(9.371)	(8.592)
$Rest/FF^2$	0.125	-3.344	-3.602	-0.234	-2.342	-2.582	0.636	-2.963	3.016	-2.791	2.854	-8.353**
	(2.337)	(2.250)	(3.159)	(2.019)	(2.540)	(1.625)	(2.260)	(2.421)	(2.297)	(2.374)	(3.988)	(3.963)
N_Cigarettes	-0.062***	-0.061***	-0.078***	-0.087***	-0.072***	-0.089***	-0.067***	-0.053***	-0.062***	-0.052***	-0.048	-0.011
	(0.014)	(0.017)	(0.013)	(0.019)	(0.016)	(0.013)	(0.010)	(0.012)	(0.018)	(0.014)	(0.037)	(0.023)
Work_Mother	0.142	0.948***	0.374	0.549.	0.293	0.95***	0.179	1.137***	0.244	1.002**	0.005	1.595**
	(0.209)	(0.284)	(0.247)	(0.322)	(0.197)	(0.262)	(0.19)	(0.343)	(0.221)	(0.436)	(0.360)	(0.765)
Black	-0.514	2.033*	-0.177	-0.320	-0.145	2.755**	-0.161	2.155**	-0.533	1.891	-0.863	2.630
	(1.036)	(1.101)	(1.537)	(2.386)	(1.493)	(1.376)	(0.739)	(0.959)	(1.069)	(1.670)	(1.505)	(1.672)
Age	0.175	0.546***	0.239.	0.272	0.245**	0.496***	0.347***	0.649***	0.265***	0.515***	0.208	0.471*
	(0.116)	(0.122)	(0.132)	(0.168)	(0.105)	(0.092)	(0.072)	(0.126)	(0.093)	(0.153)	(0.205)	(0.245)
Age^2	-0.002**	-0.005***	-0.002**	-0.002**	-0.002***	-0.004***	-0.003***	-0.005***	-0.003***	-0.004***	-0.003*	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Net_Income	0.147	-0.167	0.07	-0.171	1.000	-0.293	-0.347	-1.415	-0.896	0.329	-0.112	0.327
	(0.859)	(0.974)	(1.014)	(1.027)	(0.846)	(0.810)	(1.312)	(0.957)	(0.868)	(2.036)	(1.724)	(2.119)
Net_Income^2	-0.774	0.123	-0.144	0.655	-0.853	0.306	-0.245	0.737	-0.398	-0.187	-1.221	-0.593
	(0.473)	(0.447)	(0.592)	(0.437)	(0.630)	(0.587)	(0.840)	(0.664)	(0.634)	(1.589)	(0.885)	(1.423)
Couple	-0.828	-0.758	0.231	0.789	-0.389	0.625	0.001	-0.142	-1.557	-0.871	-1.607	-2.700**
	(0.653)	(0.676)	(0.644)	(0.675)	(0.430)	(0.652)	(0.730)	(0.742)	(1.065)	(0.966)	(1.060)	(1.162)
Married	-0.665	0.014	0.122	0.775	-0.445	0.495	0.218	0.046	-1.079*	0.363	-2.107***	-0.799
	(0.493)	(0.542)	(0.402)	(0.616)	(0.374)	(0.415)	(0.434)	(0.438)	(0.631)	(0.750)	(0.692)	(1.095)
Divorced	-1.195*	0.756	-0.346	0.003	-0.995.*	0.554	-0.024	0.881	-1.148	1.483***	-2.197*	1.236
	(0.630)	(0.640)	(0.385)	(0.605)	(0.564)	(0.516)	(0.665)	(0.643)	(0.794)	(0.706)	(1.146)	(1.310)
Separated	-2.531***	1.459	-1.140	0.474	-1.498***	1.556**	-0.842	1.306	-3.261***	2.977.	-5.991***	1.345
	(0.796)	(1.213)	(0.827)	(1.935)	(0.488)	(0.727)	(0.943)	(1.759)	(0.751)	(1.664)	(1.362)	(1.825)
Widowed	-0.580	0.356	0.789	0.633	-0.135	0.441	0.273	0.379	-0.982	0.765	-2.295**	0.173
	(0.547)	(0.568)	(0.499)	(0.656)	(0.569)	(0.515)	(0.511)	(0.457)	(0.697)	(0.655)	(0.900)	(0.943)
Degree	-1.549***	-1.468***	-1.413***	-1.32***	-1.077***	-1.676***	-0.989***	-1.128***	-1.826***	-1.967***	-2.927***	-1.038
	(0.356)	(0.422)	(0.396)	(0.426)	(0.405)	(0.346)	(0.237)	(0.281)	(0.295)	(0.336)	(0.532)	(0.704)
Diploma	-0.760***	-1.144***	-0.054	-0.855***	-0.349*	-1.167***	-0.449**	-1.113***	-1.034***	-1.117***	-1.535***	-0.935**
	(0.240)	(0.252)	(0.306)	(0.194)	(0.197)	(0.197)	(0.182)	(0.254)	(0.275)	(0.291)	(0.448)	(0.473)
Alevel	-0.710**	-0.806.	0.277	-0.843*	-0.315	-1.132***	-0.303	-0.873***	-1.443***	-0.313	-2.167***	-0.612
	(0.357)	(0.452)	(0.314)	(0.442)	(0.362)	(0.339)	(0.275)	(0.251)	(0.264)	(0.604)	(0.547)	(0.901)
Olevel	-0.396	-0.757**	0.161	-0.513*	-0.126	-0.732***	-0.265	-0.855***	-0.604.	-1.127**	-1.027*	-0.450
	(0.309)	(0.312)	(0.288)	(0.301)	(0.189)	(0.280)	(0.233)	(0.278)	(0.354)	(0.464)	(0.542)	(0.492)
D_{2004}	-0.883***	-1.493***	-0.784	-0.548	-0.853**	-1.215***	-1.241***	-1.608***	-1.032**	-2.066***	-0.943*	-2.095***
	(0.282)	(0.313)	(0.493)	(0.437)	(0.348)	(0.399)	(0.284)	(0.425)	(0.470)	(0.434)	(0.538)	(0.729)
Constant	51.371	137.002**	160.410*	121.731	129.775	151.024	51.106	110.190	24.778	114.255	-50.989	153.887
	(54.818)	(55.373)	(95.565)	(84.828)	(84.982)	(92.669)	(88.972)	(88.426)	(95.051)	(94.332)	(192.407)	(160.665)
R^2	0.06	0.05	0.04	0.03	0.03	0.03	0.04	0.03	0.04	0.03	0.06	0.04

*Note: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** ≤ 0.01, ** ≤ 0.05, * ≤ 0.1. The description of variables are reported in Table 3.1.*

APPENDIX 3.A.2

Table 3.A.2: BMI OLS and quantile regressions: model (2), people over 50

	OLS		Q 0.1		Q 0.25		Q 0.5		Q 0.75		Q 0.9	
Variable	M	F	M	F	M	F	M	F	M	F	M	F
Phys_Activity	-0.579***	-1.576***	-0.003	-0.573***	-0.027	-0.968***	-0.418**	-1.309***	-0.797***	-2.085***	-1.182***	-2.521***
	(0.145)	(0.123)	(0.154)	(0.087)	(0.078)	(0.039)	(0.121)	(0.114)	(0.123)	(0.178)	(0.254)	(0.232)
Strenuousness	-0.334	-0.595*	0.111	-0.443	0.078	-0.534**	0.074	-0.697**	-0.779**	-0.457	-0.484	-0.936
	(0.243)	(0.313)	(0.178)	(0.313)	(0.259)	(0.236)	(0.260)	(0.345)	(0.336)	(0.405)	(0.447)	(0.630)
$Price_{F\&V}$	0.340	0.598**	0.171	0.417*	0.234	0.441**	0.649***	0.776***	0.646***	0.773**	0.785**	0.341
	(0.222)	(0.247)	(0.238)	(0.239)	(0.200)	(0.184)	(0.195)	(0.192)	(0.240)	(0.350)	(0.376)	(0.380)
$Price_{TA}$	-0.451	-1.452***	-1.16***	-1.147**	-0.988*	-1.468**	-0.975	-1.759***	-0.590	-1.372	0.580	-2.088*
	(0.452)	(0.471)	(0.453)	(0.579)	(0.592)	(0.716)	(0.713)	(0.636)	(0.810)	(0.999)	(1.062)	(1.113)
Rest/FF	-0.062	9.369*	6.812	4.242	5.813	7.627**	0.368	9.120**	-5.635	7.853	-2.975	18.982***
	(5.098)	(5.254)	(6.374)	(4.955)	(4.664)	(3.369)	(3.251)	(4.305)	(5.243)	(6.309)	(9.506)	(7.151)
$Rest/FF^2$	-0.192	-4.176*	-3.748	-0.776	-2.729	-2.835**	0.028	-3.728**	2.422	-3.951	1.248	-8.864***
	(2.317)	(2.239)	(3.009)	(1.965)	(2.206)	(1.361)	(1.495)	(1.833)	(2.204)	(2.732)	(4.636)	(2.758)
N_Cigarettes	-0.065***	-0.076***	-0.079***	-0.101***	-0.071***	-0.101***	-0.066***	-0.064***	-0.067***	-0.069**	-0.038	-0.035
	(0.014)	(0.016)	(0.012)	(0.011)	(0.014)	(0.012)	(0.01)	(0.013)	(0.021)	(0.029)	(0.025)	(0.026)
Work_Mother	0.092	0.327	0.380.	0.260	0.336	0.561***	0.202	0.574***	0.109	0.362	-0.264	0.142
	(0.204)	(0.234)	(0.201)	(0.261)	(0.212)	(0.156)	(0.206)	(0.216)	(0.280)	(0.387)	(0.323)	(0.395)
Black	-0.486	2.243**	-0.075	-0.424	0.051	3.497**	-0.149	3.124***	-0.796	2.065**	-1.108	2.812
	(1.037)	(1.111)	(1.188)	(1.697)	(1.894)	(1.505)	(0.967)	(0.993)	(0.955)	(0.990)	(1.483)	(2.098)
Age	0.203.	0.703***	0.232	0.299**	0.223**	0.578***	0.341***	0.763***	0.275**	0.722***	0.315	0.826***
	(0.111)	(0.118)	(0.143)	(0.124)	(0.114)	(0.081)	(0.109)	(0.107)	(0.108)	(0.177)	(0.272)	(0.144)
Age^2	-0.002***	-0.006***	-0.002**	-0.003***	-0.002**	-0.005***	-0.003***	-0.006***	-0.003***	-0.006***	-0.003.	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Net_Income	0.147	-0.632	0.144	0.303	0.994	-0.477	0.060	-1.532	-0.876	-1.275	0.266	-0.448
	(0.882)	(0.944)	(0.632)	(1.179)	(0.941)	(1.015)	(0.872)	(1.359)	(1.450)	(1.829)	(1.624)	(2.471)
Net_Income^2	-0.725	0.315	-0.157	0.352	-0.827	0.144	-0.311	0.839	-0.280	0.399	-1.221	-0.266
	(0.488)	(0.459)	(0.344)	(0.931)	(0.707)	(0.997)	(0.542)	(0.892)	(1.367)	(0.801)	(1.034)	(1.396)
Couple	-0.859	-0.869	0.292	0.604	-0.484	0.549	-0.046	-0.281	-1.603	-1.322	-1.892	-3.008**
	(0.651)	(0.669)	(0.484)	(0.677)	(0.491)	(0.658)	(0.530)	(0.784)	(1.045)	(0.930)	(1.906)	(1.292)
Married	-0.667	-0.054	0.148	0.538	-0.512	0.307	0.143	0.133	-1.199*	-0.003	-1.994*	-1.491
	(0.491)	(0.534)	(0.471)	(0.540)	(0.330)	(0.424)	(0.436)	(0.447)	(0.654)	(0.615)	(1.045)	(1.246)
Divorced	-1.180*	0.695	-0.357	-0.125	-1.076***	0.266	-0.202	0.870	-1.449**	1.078	-1.529	0.460
	(0.632)	(0.632)	(0.525)	(0.699)	(0.398)	(0.423)	(0.544)	(0.641)	(0.658)	(0.749)	(1.341)	(1.551)
Separated	-2.584***	1.520	-1.158	0.696	-1.449**	1.885***	-1.091	1.627	-3.256***	2.285.	-5.679***	0.987
	(0.818)	(1.118)	(0.939)	(1.410)	(0.575)	(0.727)	(1.171)	(1.497)	(0.729)	(1.190)	(2.080)	(2.326)
Widowed	-0.529	0.372	0.825	0.595	-0.125	0.469	0.280	0.513	-0.996	0.252	-1.805	-0.401
	(0.547)	(0.560)	(0.588)	(0.59)	(0.300)	(0.457)	(0.470)	(0.469)	(0.718)	(0.581)	(1.286)	(1.264)
Degree	-1.543***	-1.370***	-1.406***	-1.246***	-1.047***	-1.754***	-1.026**	-1.395***	-1.753***	-1.600***	-2.801***	-2.208***
	(0.352)	(0.412)	(0.270)	(0.386)	(0.351)	(0.286)	(0.408)	(0.458)	(0.373)	(0.289)	(0.503)	(0.656)
Diploma	-0.762***	-0.987***	-0.054	-0.759***	-0.314	-1.098***	-0.439*	-0.977***	-1.001***	-0.925****	-1.603***	-0.930**
	(0.238)	(0.247)	(0.198)	(0.179)	(0.209)	(0.201)	(0.246)	(0.239)	(0.302)	(0.318)	(0.258)	(0.397)
Alevel	-0.701**	-0.697	0.268	-0.919*	-0.313	-1.144***	-0.340	-0.753	-1.480***	-0.094	-2.177***	-0.489
	(0.356)	(0.455)	(0.330)	(0.502)	(0.231)	(0.350)	(0.397)	(0.479)	(0.278)	(0.534)	(0.623)	(0.724)
Olevel	-0.398	-0.685**	0.199	-0.492	-0.093	-0.902***	-0.307	-0.645***	-0.635**	-1.155***	-0.932.	-0.860
	(0.307)	(0.308)	(0.318)	(0.308)	(0.232)	(0.197)	(0.314)	(0.233)	(0.324)	(0.361)	(0.477)	(0.633)
D_{2004}	-0.899***	-1.451***	-0.802**	-0.556**	-0.829**	-1.118***	-1.359***	-1.543***	-1.396***	-1.698***	-1.227**	-1.681***
	(0.281)	(0.308)	(0.330)	(0.277)	(0.327)	(0.278)	(0.381)	(0.332)	(0.412)	(0.545)	(0.546)	(0.611)
Cons.	57.969	143.655***	148.677**	125.599	124.757	156.885*	94.344	165.653**	58.121	124.540	-99.315	244.998*
	(54.418)	(54.803)	(63.649)	(82.289)	(77.051)	(90.381)	(83.130)	(78.065)	(98.150)	(123.968)	(131.879)	(135.358)
R^2	0.07	0.07	0.04	0.04	0.03	0.04	0.04	0.04	0.05	0.05	0.06	0.06

*Note: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1. The description of variables are reported in Table 3.1.*

Chapter 4

Smoking and Body Weight: Causal Estimates from the UK

4.1 Introduction

The purpose of this chapter is to estimate the influence of smoking on body weight in the UK using a longitudinal sample extracted from two waves of the British Household Panel Survey (BHPS), controlling for a large number of socio-demographic variables. We take advantage of the fact that we are able to observe a sample of the population of smokers in two periods, 2004 and 2006, and verify that some individuals experience transition from the status of smoker to that of non-smokers. Thus, the longitudinal structure of our data allows us to isolate (e.g., identify) the effect on body weight associated with reductions in nicotine consumption.

Our analysis differs from the current health economic literature in two important respects. First, we propose a very general framework that allows to test the influence of multi-treatments, including the effect of quitting and of reducing significantly cigarettes consumption, on weight. Second, we contribute to the literature focusing our attention on the effects of smoking on different quantiles of the BMI distribution, within the difference-in-differences (DID) framework, extending the quantile regression to identify these effects consistently on subgroups of the population belonging to different BMI categories, (i.e. overweight and obese individuals that are of particular interest for policy makers).

4.2 An econometric approach in evaluating smoking reductions on BMI

In this section, we argue that quantile regression can be used to consistently describe the effect of smoking on BMI. The specific motivation to adopt this empirical model is that overweight and obese people, are generally not considered properly when estimates are carried out at the average or at the median of the BMI distribution. This implies that the usual ATE estimator provides incorrect point estimates and that the QTE estimator should be used instead. We generalise our model to account also for the effect of reducing smoking, rather than focusing only on quitting, under a multi-treatment framework.

4.2.1 Preliminaries: standard regression and average treatment effect (ATE)

Our aim is to estimate the causal effect of smoking changes on BMI for a representative sample of the UK population. We start specifying a benchmark model where $BMI(i,t)$ is the body mass index of individual i at time t. Our sample is composed by individuals observed in a pre-treatment period $t = 0$, and in a post-treatment period $t = 1$. We denote with $D(i,t) = 1$ if an individual has been exposed to the treatment between $t = 0$ and $t = 1$ and with $D(i,t) = 0$ if he has not been exposed to the treatment (control group) in the period analysed. In this context we assume as treated those subjects which in $t = 1$ reduced to zero the number of cigarettes smoked. We define at this stage as control group those individuals who did not experience the treatment. We will define more precisely these groups in the next section.

Suppose that BMI is generated through a components of variance process:

$$BMI(i,t) = \delta(t) + \beta D(i,t) + \tau(i) + \nu(i,t) \tag{4.1}$$

where $\delta(t)$ is a time-specific component, β represents the effect of treatment, nicotine reduction experienced by quitters, $\tau(i)$ is an individual specific component, or unobserved characteristics, such as concerns about health, and $\nu(i,t)$ is an individual-transitory shock that has mean zero at each period and is possibly correlated in time.

A sufficient condition for identification is that selection for treatment does not depend on the individual-transitory shocks. By following Abadie (2005), we obtain this formulation:

$$BMI(i,t) = \mu + \eta D(i,1) + \delta t + \beta D(i,t) + \epsilon(i,t) \tag{4.2}$$

where $\mu = E[\tau(i)|D(i,1) = 0] + \delta(0)$, $\eta = E[\tau(i)|D(i,1) = 1] - E[\tau(i)|D(i,1) = 0]$ and $\delta = \delta(1) - \delta(0)$.

Notice that the model allows any kind of dependence between selection for treatment, $D(i,1) = 1$, and the individual-specific component, $\eta(i)$, and under the identifying condition discussed above, the DID estimate of β can be obtained using ordinary least squares on equation (4.2). If we consider that, although we can not observe $BMI(i,t)$ for $D(i,1) = 0$ and $D(i,1) = 1$ for the same individual simultaneously, we can estimate the treatment effect as the difference between BMI of treated and untreated individuals compared before and after the treatment took place. To see this, let us define $E[BMI(i,0)|D(i,t) = 1]$ and $E[BMI(i,0)|D(i,t) = 0]$ as the expected values of BMI of the control and treatment group in the pre treatment period, and $E[BMI(i,1)|D(i,t) = 1]$ and $E[BMI(i,1)|D(i,t) = 0]$, as the same quantities in the post treatment period, then the treatment effect can be estimated by:

$$E[BMI(i,1) - BMI(i,0)|D(i,t) = 1] - E[BMI(i,1) - BMI(i,0)|D(i,t) = 0] = \beta \tag{4.3}$$

from the above equation, β corresponds to the difference between the within-groups BMI variations in pre- and post-treatment periods, being netted out the effect of unobserved confounders. This estimation technique is known also as the difference-in-differences (DID) estimator.

It is also possible to control for differences between treatment and control groups by introducing covariates linearly in the model. In this way identification is achieved under the more general assumption that $P(D(i,1) = 1|X(i), \nu(i,t) = P(D(i,1) = 1|X(i))$. Or equivalently the selection for treatment has to be uncorrelated with the transitory shock conditioned on a set of observable variables. Equation (4.2) can be rewritten as:

$$BMI(i,t) = \mu + X(i)'\pi + D(i,1)\tau + \delta t + \beta D(i,t) + \epsilon(i,t) \tag{4.4}$$

where $X(i)$ is the vector of observed characteristics that are assumed to be uncorrelated with $\epsilon(i,t)$. This specification is particularly indicated when differences in the distribution of treated and untreated individuals are supposed to exist according to the covariates of interest. The effect of quitting smoking on BMI can be estimated by the following:

$$E[BMI(i,1) - BMI(i,0)|D(i,1) = 1, X] - E[BMI(i,1) - BMI(i,0)|D(i,1) = 0, X] = \beta. \tag{4.5}$$

Provided that D is not endogenous, the OLS estimate of β is the known conditional ATE.

4.2.2 Quantile regression and quantile treatment effect (QTE)

Consider the following synthetic representation of the linear regression specification of BMI presented in equation (4.2):

$$BMI(i,t) = f(D,t,X)\phi + \epsilon(i,t), \;\; i = 1,2,...n, \;\; t = 0,1 \tag{4.6}$$

where $f(\bullet)$ includes the already described variables related to treatment, time, unobserved specific individual characteristics - associated respectively with the parameters β, η and δ - and the covariates matrix X, associated with the vector of parameters π; and $\epsilon(i,t)$ is still the error term. The objective function for individual BMI, under OLS, is expressed as:

$$\sum_{i=1}^{n} \left(BMI(i,t) - f(\bullet)\phi\right)^2, \tag{4.7}$$

the estimated vector $\hat{\phi}$ is the solution to the minimization of the sum of squared residuals presented in equation (4.7). Equation (4.6) can be estimated also using a quantile regression, that is a direct generalization of the least absolute deviation (LAD) technique

proposed by Koenker and Hallock (2001), Koenker and Bassett (1978). The specification for the θ-th conditional quantile of BMI can be expressed as:

$$BMI^{\theta}(i,t) = f(\bullet)\phi^{\theta} + e^{\theta}(i,t), \;\; i = 1,2,...n, \;\; t = 0,1 \tag{4.8}$$

where θ is the quantile of interest chosen in the interval (0, 1), ϕ^{θ} is the vector of parameters associated with θ, and e^{θ} is the corresponding error term. The solution of equation (4.8) is obtained by minimizing the asymmetric weighted sum of absolute deviations at specific quantiles of the outcome variable:

$$\underset{\phi \in \Re^k}{min} \left[\sum_{i:BMI(i,t) \geq f(\bullet)\phi} \theta \left| BMI(i,t) - f(\bullet)\phi^{\theta} \right| + \sum_{i:BMI(i,t) \leq f(\bullet)\phi} (1-\theta) \left| BMI(i,t) - f(\bullet)\phi^{\theta} \right| \right], \tag{4.9}$$

If θ is equal to 0.5, equation (4.8) turns into the LAD estimator and $f(\bullet)\phi^{\theta}$ describes the effect of the variables of interest at the median of the BMI distribution. While, if θ is chose from the upper tail of BMI distribution, $f(\bullet)\phi^{\theta}$ will characterise the behaviour of overweight and obese individuals.

We present now the definition of quantile treatment effect (QTE), developed firstly in a static framework following Bhattacharya (1963), Doksum (1974), Lehmann (1974). To do this, let's consider F^i, with $i = 1, 0$, be the BMI empirical distributions for individuals belonging to treatment and control groups, respectively. Let $d(bmi)$ be the *''horizontal distance"* between F^1 and F^0 at a given quantile θ of the BMI distribution, then the difference between the quantiles of the outcome variable between treated and controls will be given as $QTE = F_{\theta}^{1}(bmi)^{-1} - F_{\theta}^{0}(bmi)^{-1}$.

We can also extend the previous definition to the case in which observations are recorded in different time periods. If the BMI distribution function varies over time, and is defined as $F_{\theta}^{i,t}$, where $t = 1, 0$, and we define the corresponding distribution functions for control and treated individuals as $F_{\theta}^{0,t}$ and $F_{\theta}^{1,t}$, respectively. Then, the QTE will be obtained for any given θ, as:

$$\beta(\theta) = [F_{\theta}^{1,0}(bmi)^{-1} - F_{\theta}^{0,0}(bmi)^{-1}] - [F_{\theta}^{1,1}(bmi)^{-1} - F_{\theta}^{0,1}(bmi)^{-1}]. \tag{4.10}$$

Thus we can express the estimate of the effect of smoking on different quantiles of the BMI distribution as follows:

$$\begin{aligned} &E[BMI^{\theta}(i,1) - BMI^{\theta}(i,0)|D(i,1) = 1, X]- \\ &E[BMI^{\theta}(i,1) - BMI^{\theta}(i,0)|D(i,1) = 0, X] = \beta^{\theta}. \end{aligned} \tag{4.11}$$

Here, we extend the models proposed to estimate the average and quantile treatment effects with a DID estimator under a multiple treatment framework. We justify this extension because BMI could be not only affected by quitting smoking but also by a significant reduction in the number of cigarettes smoked. To incorporate the multi-treatment effect equation (4.2) will be described now using the following expression:

$$BMI(i,t) = \mu + X(i)^{'}\pi + \sum_{j=1}^{n} D_j(i,1)\tau_j + \delta t + \sum_{j=1}^{n} D_j(i,t)\beta_j + \epsilon(i,t) \tag{4.12}$$

where D(j) are indicators related to the j-th treatment group, with j=1,...n, while the other parameters and variables are the same already described.

The average treatment effects in the case of multi-treatment will be given by:

$$\begin{aligned} &E[BMI(i,1) - BMI(i,0)|D^{*}(i,1) = 1, X]- \\ &E[BMI(i,1) - BMI(i,0)|D^{*}(i,1) = 0, X] = \sum_{j=1}^{n} \beta_j. \end{aligned} \tag{4.13}$$

where $D^{*} = 1$ is $\sum_{j=1}^{n} D(i,1) = 1$. The OLS estimates of $\sum_{j=1}^{n} \beta_j$ from equation (4.X) are the composite DID estimates of the ATE. Analogously, we derive the quantile regression estimates for β_j as:

$$\begin{aligned} &E[BMI^{\theta}(i,1) - BMI^{\theta}(i,0)|D^{*}(i,1) = 1, X]- \\ &E[BMI^{\theta}(i,1) - BMI^{\theta}(i,0)|D(i,1) = 0, X] = \beta^{\theta *} = \sum_{j=1}^{n} \beta_j^{\theta}. \end{aligned} \tag{4.14}$$

that are still DID estimates of the QTE.

4.3 Identification strategy, data and empirical models

4.3.1 Identification from the BHPS

The dataset used in this chapter was extracted from the BHPS, a multi-purpose survey which collects information at household and individual levels for a representative sample of the UK population. The original sample was drawn from 250 areas of England and was subsequently enlarged to include Scotland and Wales in 1999 and Northern Ireland in 2002. The dataset is formed by 18 waves: the first survey was conducted in 1980 but, for our purposes, we will use only the 14^{th} and 16^{th} wave, conducted respectively in 2004 and 2006, because two anthropometric characteristics of individuals (height and weight) were collected. We then derived the BMI for a balanced panel of the UK population (13,230 individuals for each year).

The DID estimates of the effects of quitting smoking on BMI are carried out using a wide range of control groups. These control groups are obtained from different partitions of our sample, obtained according to the nicotine consumption variation occurred in the period 2004-2006, for each individual. Formally we define:

D1. **Treated group of smokers who quit smoking ($\mathbf{TG}_Q$)**. Individuals who were smokers in 2004 and became non-smokers in 2006.

D2. **Treated group of smokers who reduces smoking ($\mathbf{TG}_R$)**. Individuals who were smokers in 2004 and reduced significantly (at least 50%) nicotine consumption in 2006.

D3. **Control group of smokers ($\mathbf{CG}_S$)**. Individuals who were smokers in 2004 and kept the same status in 2006.

Individual body weight variations of TG_Q are assumed to be affected, between 2004 and 2006, both by the effect of quitting smoking and a time effect, while weight changes of individuals who continue to smoke (CG_S) are supposed to be only affected by the time effect. In addition, in our specification also smoking reductions (TG_R), if significantly large[1], may produce a relevant effect on BMI.

Under these assumptions we compare the average weight gain of quitters (and in the multi-treatment model also of individuals reducing cigarettes consumption) to that of smokers, controlling for differences between these groups related to: age, sex and other individual characteristics. A problem with this approach is that the two groups are likely to differ in terms of unobservable characteristics - such as general health concerns or the discount factor associated to the future - that may be also related to BMI (for a discussion, see Baum (2009)).

Now, we extend the representativeness of the smokers control group by including in CG_S also non-smokers, forming a new control group composed by the entire population, with the exception of treated individuals, (CG_{ALL})[2]. CG_{ALL} is formed by individuals who kept constant nicotine assumption, whether they were smokers or non-smokers. Formally, we have:

D4. **Control group of smokers and non-smokers (CG_{ALL})**. Individuals who were smokers or non-smokers in 2004 and kept the same status in 2006.

The reason for using this new control group is that, our estimates using CG_S may be biased upward or downward by endogeneity. Quitters are known to be generally less concerned about their weight than individuals continuing to smoke, Meyer (1995), so that using CG_S as comparison group may bias upward ATEs and QTEs estimates respect to the true causal effect. Cawley et al. (2004) suggest that some people may use

[1]We considered significant a reduction of 50% of current cigarettes consusmption. We also carried out estimates with thresholds of 30% and 70% for cigarettes consumption redcutions. But the results were very similar to those obtained with the 50% threshold. All the tables related to these estimates are are available upon request by the authors.

[2]In principle, we exclude from this control group non-smoker individuals, irrespective if ex-smokers or not in 2004, that starts smoking in 2006. We can anticipate that the dimension within our sample results negligible for the estimates.

nicotine to control weight, and more specifically that some individuals may be afraid of the consequences of quitting (or reducing) smoking. This problem is commonly known as reverse causality bias. In order to further account for this issue we consider also a control group formed only by non-smokers who kept the same status during the observational period:

D5. **control group of non-smokers($\mathbf{CG_{NS}}$)**. Individuals who were non-smokers in 2004 and kept the same status in 2006.

Since CG_{NS} is composed by individuals who do not control their weight by smoking, we can assume that using this control group we are able to indirectly check for the dimension of the effect of reverse causality.

Quitters have also been shown to differ from smokers in relation to their concerns about health and especially in their attitude toward the future. Many studies have demonstrated how the principal motivation for quitting (or reducing) smoking is generally found in individual concerns about health (see, for a review, McCaul et al. (2006))[3]. This implies that some quitters, or individuals reducing smoking significantly, will be also more careful about their weight than individuals continuing to smoke. In other words there may be unobserved confounders affecting simultaneously BMI and smoking behaviours and biasing downward our estimates. For this reason we specify a control group, formed by individuals with the most similar unobservable characteristics to quitters but that formally are still classified as smokers. Thus, in line with the procedure proposed by Del Bono and Vuri (2011) we selected from BHPS a group of individuals classified as smokers in 2004 and 2006, but who are going to quit in 2008. We assume that the BMI variation between 2004 and 2006 of future quitters is the best comparison in terms of unobservable characteristics for TG_{Q}.

Formally, this control group is defined as:

[3]Bickel et al. (1999) propose as an explanation of downward bias in estimated weight gain the propensity of quitters to be more future-oriented on average. Third, quitters are less impulsive on average while Terracciano and Costa (2004) explains it from a behaviour less impulsive on average of quitters.

D6. **Control group of next period quitters**(CG_{Q08}). Individuals who are smokers in 2004 and 2006, but will quit smoking in 2008.

The same argumentation can be used to define a similar control group of individuals who reduced smoking consumption in 2008, defined formally as:

D7. **Control group of next period reducers** (CG_{R08}). Individuals who are smokers in 2004 and 2006, but consistently reduced smoking in 2008.

In the next sub-section, we will justify empirically the use of QTE estimator and describe the econometrics models used to estimate the effects of quitting or reducing smoking.

4.3.2 Empirical models

In this section, we analyse the key variables of our model. To emphasise differences between treatment and control groups, we investigate the BMI empirical distributions for individuals classified as "non-smokers" and "heavy smokers", where we consider as "heavy smokers" adults smoking more than 20 cigarettes per day.

The solid line in Figure 1 defines the BMI empirical distribution of non-smokers, while the dashed line is associated with heavy-smokers. The density functions are both skewed and non-normally distributed; moreover we can notice how they present a clearly distinct shape. We can notice that the non-smokers' empirical distribution is shifted to the right, implying that this group also have on average higher BMI with respect to heavy smokers. This result is consistent with the hypothesis that nicotine reductions may increase overweight and obesity rates, while the empirical density functions suggest to use estimation techniques that allow to measure heterogeneous effects on the entire BMI distribution.

Table 1 (panel a) shows that, the average BMI variation for TG_Q is about 0.85 points, and that smaller variations are found for CG_S (0.24 BMI points). The unconditional ATE

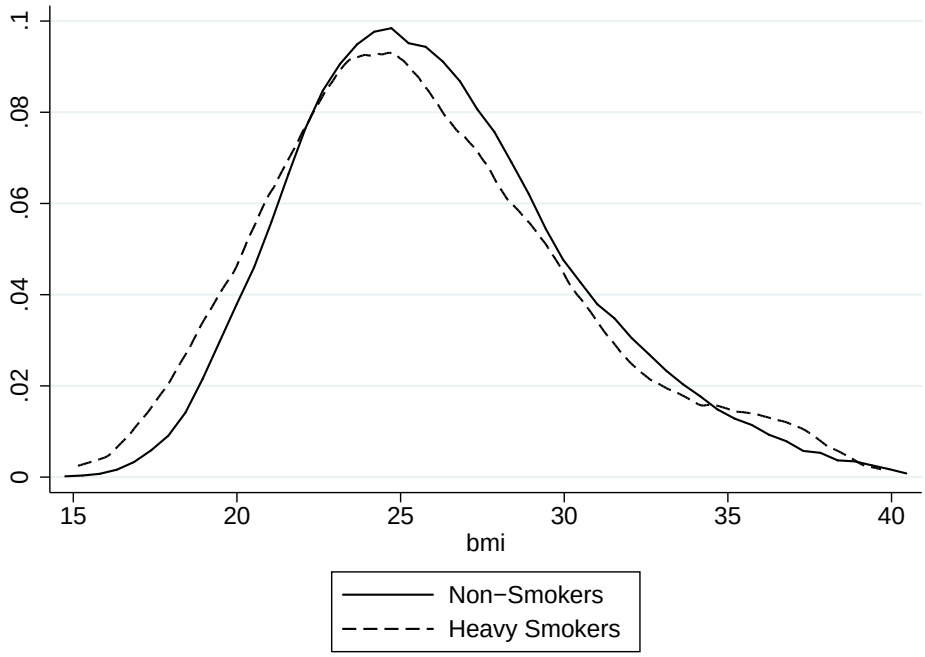


Figure 4.1: BMI kernel density estimate, by smoking status

is 0.61 points between the two groups. Also using other control groups, the estimates of the ATE is about 0.6 points. In panel (b) of the same table are listed the effects of smoking reductions (TG_R) on BMI. For this treatment group we estimate an ATE very close to zero (0.01) and the same result holds for all the other control groups.

Previously we checked for significant differences in BMI of smokers and non-smokers over the BMI empirical distribution. Now, we use BMI classes to evaluate the unconditional ATE for underweight, normalweight, overweight and obese individuals. From Table 2, it is possible to highlight three relevant results. First, the impact of quitting smoking on BMI is greater for obese individuals (BMI increases by 1.08 points), irrespective of the control group used. Second, smoking reductions have a positive, although small, impact on obese people BMI (from 0.25 and 0.29) compensated by a (small) negative effect on underweight and normalweight individuals. Note that, at this stage we can not give any indication about the significance of these effects that will be further discussed in the next section. Third, from the table clearly emerges the heterogeneous spontaneous dynamic of BMI in the obese group. BMI showed an increasing trend for each weight category, in line with UK statistics, unless for obese individuals, who are characterised by a negative

trend. This result suggests that people at higher health risk tend to reduce their BMI, irrespective if they are smokers, quitters or individuals reducing cigarettes consumption, but those belonging to TG_Q also tend to increase their weight for the effect of quitting smoking. Consequently their total variation, in terms of weight, results very close to zero. However, the effects of time trends and quitting smoking on BMI, obtained from the descriptive analysis, are not clear and precisely determined and need to be further investigated under an inferential framework.

Table 4.1: BMI absolute variation and ATE, by smoking status

	N. obs	**BMI**	**BMI**	**BMI absolute var**	**ATE**
		2004	**2006**	**2004/2006**	**2004/2006**
TG_Q	443	25.04	25.89	0.85	-
CG_S	2455	25.12	25.36	0.24	0.61
CG_{ALL}	10946	25.82	26.06	0.25	0.6
CG_{NS}	8491	26.02	26.27	0.25	0.6
CG_{Q08}	281	24.89	25.15	0.27	0.58
		2004	**2006**	**2004/2006**	**2004/2006**
TG_R	694	25.25	25.50	0.25	-
CG_S	1761	25.07	25.30	0.24	0.01
CG_{ALL}	10252	25.85	26.10	0.25	0.00
CG_{NS}	8491	26.02	26.27	0.25	0.00
CG_{Q08}	589	25.12	25.37	0.25	0.00

Notes: We indicate with TG_Q and TG_R the treated groups and with CG_S, CG_{ALL}, CG_{NS} and CG_{08}, the control groups already described in section 3

We now propose the following baseline model to estimate the effect of quitting smoking on BMI at a specific quantile θ of the BMI distribution.

$$BMI(i) = \mu^{\theta} + X(i)^{'}\pi^{\theta} + D(i)\tau^{\theta} + t(i)\delta^{\theta} + (D(i) \times t(i))\beta^{\theta} + \epsilon(i)^{\theta} \tag{4.15}$$

all the variables and related parameters have been already discussed in section 2.

The covariates matrix X includes: gender; health status (five modalities from excellent

Table 4.2: BMI absolute variation and ATE, by BMI classes and smoking status

	BMI Abs var				**ATE**			
					BMI			
	< 19	$19 - 25$	$25 - 30$	> 30	< 19	$19 - 25$	$25 - 30$	> 30
TG_Q	1.15	0.71	1.00	0.78	-	-	-	-
CG_S	0.35	0.41	0.20	-0.3	0.81	0.30	0.80	1.08
CG_{NS}	0.44	0.42	0.24	-0.19	0.71	0.29	0.76	0.97
CG_{ALL}	0.40	0.42	0.23	-0.21	0.75	0.29	0.76	0.99
CGQ08	0.47	0.35	0.03	-0.09	0.68	0.36	0.97	0.87
	< 19	$19 - 25$	$25 - 30$	> 30	< 19	$19 - 25$	$25 - 30$	> 30
TG_R	0.34	0.30	0.25	0.05	-	-	-	-
CG_S	0.35	0.45	0.18	-0.45	-0.01	-0.16	0.07	0.51
CG_{NS}	0.44	0.42	0.24	-0.19	-0.10	-0.12	0	0.24
CG_{ALL}	0.41	0.43	0.23	-0.22	-0.07	-0.13	0.01	0.28
CG_{Q08}	0.27	0.43	0.27	-0.09	0.07	-0.13	-0.02	0.14

Notes: We indicate with TG_Q and TG_R the treated groups and with CG_S, CG_{ALL}, CG_{NS} and CG_{08}, the control groups already described in section 3

to very poor); length of sickness (long term against short term); work conditions (hours work weekly; strenuousness of job; presence in the household of a working mother); race (black or white); age (18-30, 30-40, 40-50, 50-60, > 60); income (classes by quintile); marital status (married; couple not married; divorced; separated; widowed); education (degree; diploma; A_{level}; O_{level}); and three regional dummies (England, Wales, Scotland). The summary statistics of covariates for TG_Q and CG_S are listed in Appendix 4.A.1.

However, to further control the effect of omitted variables[4], we included in the covariates matrix also a dummy variable that measures alcohol consumption habits, that is assumed to be a proxy for the unobserved individual heterogeneity in addictive behaviours and takes value 1 if the respondent goes for a drink outside at least once a week (reference category: going for a drink outside rarely or not going). We decided to include this variable because alcohol consumption is known to be a complement to smoking (Aristei and Pieroni, 2010), but is also known to be positively correlated with body

[4]Although measurement error may be a concern here, because height and body weight were not directly measured.

weight.

We also explicit the quantile regression counterpart of equation (4.15) that admits multi-treatment effects:

$$BMI(i) = \mu^{\theta} + X(i)^{'}\pi^{\theta} + \sum_{j=1}^{2} D_j(i)\tau_j^{\theta} + \delta^{\theta}t(i) + \sum_{j=1}^{2}(D_j(i) \times t(i))\beta_j^{\theta} + \epsilon(i)^{\theta} \qquad (4.16)$$

D(j) represents the j-th treatment group, where j=1 corresponds to the treatment associated with quitting smoking and j=2 corresponds to the treatment associated with reducing cigarettes consumption more than 50%. The quantile regression estimates of $\beta(j)$ are still the DID estimates of the QTE for the $j-th$ treatment group.

Heterogeneous behaviours of individuals belonging to different BMI classes may also affect our estimates and mislead the interpretation of our results. As already discussed previously and clearly shown in Table 2, irrespectively from the control group used, the absolute BMI variation between 2004 and 2006, for individuals classified as obese in 2004, was negative. This result is in contrast with the increase generally showed by the other categories and by the full sample. This implies that, if not taken properly into account, the reverse trend in the spontaneous dynamic of BMI of the obese group may contamine negatively the effect of quitting smoking. We propose then to estimate our baseline model on subsamples of the population corresponding to the well known BMI classes of underweight, normalweight, overweight and obese individuals by specifying equation (4.16) for $h = 1, ..., 4$ categories of BMI. In the results section we will discuss only the estimates obtained for the obese group. Formally we can express the DID estimates of ATEs and QTEs for the h BMI classes as follows:

$$\begin{aligned} BMI(i,h) = \mu_h^{\theta} + X(i,h)^{'}\pi_h^{\theta} + D(i,h)\tau_h^{\theta} + t(i,h)\delta_h^{\theta} \\ + (D(i,h) \times t(i,h))\beta_h^{\theta} + \epsilon(i,h)^{\theta} \end{aligned} \qquad (4.17)$$

The same model can be easily written to incorporate multi-treatment effects, where also the treatment related to reducing cigarettes consumption is considered. Formally:

$$BMI(i,h) = \mu_h^\theta + X(i,h)^{'}\pi_h^\theta + \sum_{j=1}^{2} D_j(i,h)\tau_{j,h}^\theta + \delta^\theta t(i,h) + \sum_{j=1}^{2}(D_j(i,h) \times t(i,h))\beta_{j,h}^\theta + \epsilon(i,h)^\theta \tag{4.18}$$

Estimates will be carried out also in this case for each control group previously specified.

4.4 Results

In this section, we discuss the estimated ATE and QTE for the BMI specifications described by equations (4.15) and (4.16). As we already showed these models estimate the effect of quitting (TG_Q), equation (4.15) and quitting or reducing smoking (TG_Q and TG_R), equation (4.16) using as control group smokers (CG_S). Table 4.3 lists the estimated coefficients obtained by OLS and quantile regression for values of θ that correspond to the $10^{th}, 25^{th}, 50^{th}, 75^{th}$ and 90^{th} quantile of the BMI distribution. These estimates are in line with the unconditional ATEs obtained from Table 2, quitting smoking leads to a small but significant increase in terms of BMI in the time period considered. The estimated ATE is about 0.6 with standard error 0.099. This estimate is found to be robust with respect to a number of checks carried out using different control groups, that we are going to discuss in depth in the next sub-sections. We also estimate significant parameters associated to the BMI spontaneous dynamic. The estimated BMI trend is 0.22 (s.e. = 0.035), meaning that part of BMI variation (about 30%) can be attributed to external factors affecting BMI dynamic. The upper part of Table 4.2 lists the DID estimates of QTEs for TG_Q. We can notice that adding control variables do not affect estimates significantly, QTEs are estimated to be 0.48 (s.e.=0.27) at the 50^{th} quantile, 0.8 (s.e.=0.41) at the 75^{th}, while no significant effects are found at the other quantiles. From our estimates we cannot conclude in favour of a clear effect on obese individuals. Estimates at the 90^{th} quantile are not significant, but a significant effect is found at the 75^{th} quantile, which roughly corresponds to the lower bound of the obesity category, (BMI = 30). The

parameters associated with the BMI trend (δ) are significant, only above the median, where the estimated coefficient is 0.29 (s.e.=0.15), which increases to 0.31 (s.e=0.17) at the 75^{th} quantile and to 0.34 (s.e.=0.19) at the 90^{th}. Overall, QTE estimates exhibit a significant variability and in general they present an increasing dynamic over the BMI distribution. The lower part of Table 4.3 shows instead the parameters associated to the effects of both quitting and reducing smoking. Although the estimated β and δ are close to those listed in the upper part of Table 4.3, the effect of smoking reduction is found to be not significant both when covariates where included or excluded from the model at each quantile analysed. We also propose to estimate the presence of heterogeneous effects of quitting smoking among socio-economic groups, by specifying a full set of interactions between the treatment and the socio-economic characteristics of interest. The results are shown in Table 4.4 and from these estimates we can not find systematically significant differences among the groups considered, except for individuals aged 18-29, 30-39, couples or married individuals in some quantiles.

Table 4.3: Causal effect of quitting smoking on BMI, CG_S

						$\mathbf{TG}_Q$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.07	-0.04	0.61**	0.37	0.19	0.16	0.09	0.10	-0.41	-0.24	-0.33	-0.56
	(0.205)	(0.205)	(0.249)	(0.237)	(0.240)	(0.192)	(0.340)	(0.279)	(0.407)	(0.319)	(0.349)	(0.486)
β	0.60***	0.56***	-0.03	0.08	0.47	0.14	0.48	0.48*	0.78*	0.80*	0.72	0.83
	(0.099)	(0.103)	(0.430)	(0.355)	(0.449)	(0.307)	(0.467)	(0.275)	(0.426)	(0.416)	(0.609)	(0.621)
δ	0.25***	0.22***	0.01	-0.08	0.08	0.09	0.30**	0.29*	0.45**	0.31*	0.49**	0.34*
	(0.035)	(0.039)	(0.221)	(0.149)	(0.170)	(0.103)	(0.139)	(0.152)	(0.199)	(0.179)	(0.232)	(0.199)
Constant	25.11***	23.50***	19.94***	18.04***	22.08***	20.39***	24.51***	22.73***	27.67***	26.44***	30.75***	31.45***
	(0.087)	(0.547)	(0.141)	(0.467)	(0.125)	(0.610)	(0.102)	(0.607)	(0.121)	(0.836)	(0.160)	(1.080)
Observations	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700
R-squared	0.00	0.04										
Adj. R-squared	0.00	0.03										
Pseudo R-squared			0.00	0.02	0.00	0.02	0.00	0.04	0.00	0.04	0.00	0.02
						$\mathbf{TG}_Q + \mathbf{TG}_R$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.12	-0.08	0.51**	0.29	0.18	0.16	0.09	-0.06	-0.46	-0.30	-0.58	-0.55
	(0.215)	(0.216)	(0.240)	(0.270)	(0.239)	(0.254)	(0.292)	(0.250)	(0.429)	(0.342)	(0.393)	(0.546)
$\eta 1$	-0.14	-0.12	-0.16	0.21	-0.15	-0.15	-0.14	-0.16	-0.07	-0.22	-0.30	-0.29
	(0.193)	(0.192)	(0.231)	(0.210)	(0.230)	(0.215)	(0.217)	(0.287)	(0.298)	(0.296)	(0.464)	(0.532)
β	0.64***	0.62***	-0.02	0.35	0.49	0.38	0.48	0.67*	0.84	0.84*	0.81	0.63
	(0.102)	(0.106)	(0.377)	(0.382)	(0.389)	(0.364)	(0.404)	(0.354)	(0.529)	(0.508)	(0.707)	(0.715)
$\beta 1$	0.06	0.10	-0.02	0.02	0.12	0.22	0.16	0.20	0.11	0.29	-0.10	-0.10
	(0.079)	(0.083)	(0.407)	(0.311)	(0.310)	(0.297)	(0.324)	(0.385)	(0.407)	(0.406)	(0.643)	(0.631)
δ	0.23***	0.18***	0.02	-0.10	0.07	0.04	0.22	0.23	0.38**	0.33	0.40	0.41
	(0.042)	(0.046)	(0.236)	(0.196)	(0.197)	(0.158)	(0.184)	(0.178)	(0.195)	(0.237)	(0.330)	(0.321)
Constant	25.17***	23.17***	20.00***	16.88***	22.10***	19.30***	24.59***	22.46***	27.74***	25.91***	31.00***	31.58***
	(0.105)	(0.536)	(0.154)	(0.481)	(0.157)	(0.578)	(0.120)	(0.650)	(0.140)	(0.894)	(0.244)	(1.353)
Observations	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700	5,700
R-squared	0.00	0.04										
Adj. R-squared	0.00	0.03										
Pseudo R-squared			0.00	0.02	0.00	0.02	0.00	0.04	0.00	0.04	0.00	0.02

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 4.4: Interaction effect with socio demographics covariates on BMI, using CG_S

	CG_S					
Variables	**OLSc**	**Q10c**	**Q25c**	**Q50c**	**Q75c**	**Q90c**
Female	0.18	-0.16	0.11	1.04	0.12	-1.03
	(0.215)	(0.864)	(0.833)	(0.823)	(1.063)	(1.427)
Black	0.24	-1.08	0.46	0.11	-1.41	2.49
	(1.223)	(2.808)	(4.315)	(4.707)	(5.497)	(6.696)
Net Income Q20	-0.05	-0.22	-1.13	1.36	-0.51	0.41
	(0.816)	(1.361)	(1.268)	(1.630)	(1.246)	(2.232)
Net Income Q40	0.52	-0.19	-1.07	0.76	1.24	0.61
	(0.870)	(1.122)	(1.433)	(1.285)	(1.467)	(2.593)
Net Income Q60	1.07	1.38	0.52	2.19	1.33	1.92
	(0.810)	(1.284)	(1.447)	(1.557)	(1.428)	(2.076)
Net Income Q80	0.23	0.80	-0.78	0.08	-0.57	0.06
	(0.890)	(1.362)	(1.281)	(1.127)	(1.382)	(2.041)
Age 18 - 29	-0.87	-0.24	-2.26*	-0.74	1.16	-2.58
	(0.646)	(1.212)	(1.228)	(1.672)	(1.916)	(3.105)
Age 30 - 39	-0.94*	0.59	-1.94	-1.22	-1.56	-1.75
	(0.561)	(1.083)	(1.234)	(1.200)	(1.993)	(2.473)
Age	-0.78	-0.17	-1.19	-0.51	-1.82	1.64
	(0.506)	(1.461)	(1.051)	(1.123)	(2.000)	(2.593)
agegroupd4qs06	-0.17	-0.68	-1.36	0.64	-0.95	-0.98
	(0.667)	(1.509)	(1.355)	(1.297)	(2.148)	(2.661)
degreeqs06	-0.29	-1.20	-1.12	1.07	-0.84	0.57
	(0.445)	(1.490)	(1.171)	(1.241)	(1.575)	(2.427)
diplomaqs06	0.18	-0.40	-0.07	0.97	-0.57	-0.46
	(0.364)	(1.201)	(1.108)	(1.271)	(1.327)	(2.245)
alevelqs06	-0.31	-0.74	-0.73	0.98	-1.63	2.07
	(0.476)	(1.094)	(0.951)	(1.305)	(1.664)	(3.284)
olevelqs06	0.03	-0.74	0.19	0.70	0.64	0.01
	(0.372)	(1.462)	(1.447)	(1.319)	(1.619)	(2.075)
coupleqs06	1.33**	0.83	1.94	0.96	3.52***	-0.11
	(0.665)	(1.123)	(1.339)	(1.330)	(1.274)	(2.606)
marriedqs06	0.15	-0.12	-0.63	-0.03	2.74**	-0.14
	(0.571)	(1.062)	(1.269)	(1.360)	(1.279)	(2.619)
divorcedqs06	0.26	0.06	1.27	-0.77	3.42	-1.45
	(0.816)	(2.374)	(2.067)	(1.910)	(2.221)	(3.347)
widowedqs06	-0.66	0.32	-0.37	-1.72	1.54	-0.07
	(0.721)	(1.803)	(1.997)	(2.208)	(2.322)	(3.211)
Constant	24.74***	19.16***	21.63***	24.03***	27.02***	32.15***
	(0.663)	(0.659)	(0.739)	(0.846)	(1.013)	(1.434)
Observations	5,796	5,796	5,796	5,796	5,796	5,796
R-squared	0.05					
Adj. R-squared	0.03	.	.	.	.	.

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** ≤ 0.01, ** ≤ 0.05, * ≤ 0.1.*

4.4.1 Robustness by control group

In this section we compare the estimates obtained using CG_S with those obtained with the other control groups. Our baseline results, under CG_S, show that the coefficients associated to η, measure of the initial differences in BMI between treated and control groups, are never significant. This result is not surprising because, under this specification, we are comparing groups of homogeneous individuals (smokers and ex-smokers), which have similar initial BMI, as is also illustrated from Fig (1). Instead, the effect of quitting smoking, β, is significant both at the average and at some quantiles of the BMI distribution, the 50^{th} and the 75^{th} respectively.

If we compare these results with those obtained using CG_{ALL} and CG_{NS}, listed respectively in Tables 4.6 and 4.7, we find that, as before, BMI shows a positive and significant trend above the median; the effect of quitting is slightly smaller than that obtained from previous estimates, although these differences are very small if measured in terms of weight variations. The only significant difference between the estimates obtained from the control groups analysed, are related to the η coefficients, which, in the two latter specifications are negative and significant. This result is actually in line with findings from the medical literature, where according to studies proposed by French and Jeffery (1995), Grunberg (1985), Klesges et al. (1989) smokers show lower weight than non-smokers because nicotine consumption increases the metabolic rate and number of calories consumed by our organism during the day. In fact after smoking a cigarette, heart may beat 10-20 more times per minute, Dill et al. (1934), Glauser et al. (1970), and the amount of weight gained after quitting is usually close to the initial gap between smokers and non-smokers, at least in the short-term, Higgins (1967), Karvonen et al. (1959), Keys et al. (1966).

Table 4.7 shows the estimated coefficients obtained using CG_{08} as control group. In this case we can notice how the η parameters are not significant, (consistently with those estimated using CG_S), while the ATE (β) estimates are found to be higher when we control for unobservable characteristics, 0.66 BMI points (s.e. $= 0.135$). The same conclusion holds for QTEs, where we estimate significant coefficients at the median and at the 75^{th} quantile of the BMI distribution, respectively of 0.50 (s.e. $= 0.290$) and 0.87 (s.e. $= 0.513$). The control groups supposed to account for unobservable characteristics of smokers (CG_{08})

and reverse causality (CG_{ALL} and CG_{NS}) do not reveal an excessive distortion from the baseline estimates.

The coefficients associated with smoking reductions are not significant in all the specifications previously presented. All the other parameters are in line with those already obtained from the robustness analysis conducted in this section for the treated group of quitters.

Table 4.5: Causal effect of quitting smoking on BMI, CG_{ALL}

						TG_Q						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.78***	-0.66***	-0.29*	-0.47*	-0.51***	-0.42*	-0.71***	-0.48***	-1.14***	-0.64*	-1.21***	-1.30***
	(0.189)	(0.191)	(0.164)	(0.242)	(0.171)	(0.241)	(0.273)	(0.167)	(0.382)	(0.335)	(0.310)	(0.478)
β	0.60***	0.56***	-0.16	0.38	0.31	0.09	0.49	0.39*	1.01*	0.58*	0.84	0.83
	(0.094)	(0.098)	(0.261)	(0.351)	(0.280)	(0.381)	(0.380)	(0.230)	(0.569)	(0.347)	(0.811)	(0.583)
δ	0.25***	0.22***	0.14	0.07	0.24***	0.14***	0.30***	0.27***	0.22*	0.27***	0.37***	0.34**
	(0.016)	(0.018)	(0.089)	(0.054)	(0.037)	(0.052)	(0.063)	(0.061)	(0.111)	(0.080)	(0.141)	(0.134)
Constant	25.82***	24.15***	20.83***	18.37***	22.78***	20.85***	25.31***	23.76***	28.41***	27.01***	31.63***	30.72***
	(0.040)	(0.326)	(0.063)	(0.389)	(0.032)	(0.265)	(0.061)	(0.361)	(0.085)	(0.563)	(0.099)	(0.719)
Observations	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256
R-squared	0.00	0.07										
Adj. R-squared	0.00	0.06										
Pseudo R-squared			0.00	0.04	0.00	0.04	0.00	0.06	0.00	0.06	0.00	0.004
						TG_Q + TG_R						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.83***	-0.74***	-0.38**	-0.58*	-0.56**	-0.49**	-0.79***	-0.60***	-1.21***	-0.72**	-1.21***	-1.40***
	(0.190)	(0.191)	(0.185)	(0.297)	(0.224)	(0.240)	(0.272)	(0.162)	(0.436)	(0.309)	(0.312)	(0.449)
$\eta 1$	-0.87***	-0.91***	-1.08***	-0.81***	-0.88***	-0.91***	-0.94***	-0.92***	-0.81***	-0.94***	-0.94**	-0.82**
	(0.165)	(0.166)	(0.215)	(0.176)	(0.182)	(0.179)	(0.204)	(0.180)	(0.271)	(0.233)	(0.449)	(0.367)
β	0.60***	0.56***	-0.08	0.40	0.34	0.12	0.54*	0.42*	1.06*	0.60*	0.79	0.94
	(0.094)	(0.098)	(0.259)	(0.393)	(0.307)	(0.368)	(0.311)	(0.251)	(0.558)	(0.347)	(0.710)	(0.651)
$\beta 1$	0.04	0.08	-0.06	0.01	-0.03	0.08	0.11	0.20	0.28	0.15	-0.13	-0.15
	(0.068)	(0.071)	(0.336)	(0.273)	(0.242)	(0.227)	(0.316)	(0.303)	(0.333)	(0.357)	(0.566)	(0.503)
δ	0.25***	0.22***	0.06	0.07	0.22***	0.13**	0.24***	0.26***	0.17	0.28***	0.43***	0.35***
	(0.016)	(0.019)	(0.063)	(0.053)	(0.064)	(0.052)	(0.075)	(0.066)	(0.116)	(0.071)	(0.138)	(0.119)
Constant	25.88***	24.29***	20.92***	18.53***	22.83***	20.98***	25.38***	23.90***	28.48***	27.17***	31.63***	30.77***
	(0.041)	(0.326)	(0.053)	(0.317)	(0.058)	(0.410)	(0.066)	(0.313)	(0.061)	(0.431)	(0.106)	(0.920)
Observations	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256	22,256
R-squared	0.00	0.07										
Adj. R-squared	0.00	0.07										
Pseudo R-squared			0.00	0.04	0.00	0.05	0.00	0.07	0.00	0.07	0.00	0.004

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 4.6: Causal effect of quitting smoking on BMI, CG_{NS}

						$\mathbf{TG}_Q$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.98***	-0.96***	-0.54***	-0.71**	-0.75***	-0.74***	-0.95***	-0.80***	-1.27***	-0.97***	-1.32***	-1.41***
	(0.190)	(0.195)	(0.194)	(0.290)	(0.204)	(0.214)	(0.260)	(0.162)	(0.360)	(0.244)	(0.294)	(0.478)
β	0.60***	0.55***	-0.21	0.02	0.34	0.10	0.48	0.56*	1.00**	0.69**	0.82	0.81
	(0.094)	(0.100)	(0.334)	(0.435)	(0.371)	(0.397)	(0.363)	(0.314)	(0.459)	(0.343)	(0.728)	(0.638)
δ	0.25***	0.25***	0.19**	0.06	0.21**	0.20***	0.30***	0.25***	0.23*	0.24**	0.40***	0.31**
	(0.018)	(0.021)	(0.073)	(0.067)	(0.083)	(0.065)	(0.102)	(0.072)	(0.125)	(0.116)	(0.143)	(0.128)
Constant	26.03***	24.78***	21.08***	19.43***	23.03***	21.66***	25.54***	24.44***	28.53***	27.30***	31.74***	30.64***
	(0.045)	(0.381)	(0.047)	(0.327)	(0.045)	(0.349)	(0.051)	(0.333)	(0.088)	(0.514)	(0.128)	(1.051)
Observations	17,422	17,422	17,422	17,422	17,422	17,422	17,422	17,422	17,422	17,422	17,422	17,422
R-squared	0.00	0.08										
Adj. R-squared	0.00	0.08										
Pseudo R-squared			0.00	0.05	0.00	0.05	0.01	0.08	0.01	0.07	0.00	0.05
						$\mathbf{TG}_Q + \mathbf{TG}_R$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.98***	-0.96***	-0.54**	-0.74***	-0.75***	-0.75***	-0.95***	-0.82***	-1.27***	-0.91***	-1.32***	-1.49**
	(0.190)	(0.194)	(0.242)	(0.258)	(0.196)	(0.233)	(0.301)	(0.200)	(0.433)	(0.330)	(0.362)	(0.580)
$\eta 1$	-1.02***	-1.21***	-1.24***	-1.24***	-1.08***	-1.24***	-1.10***	-1.15***	-0.86***	-1.17***	-1.05***	-1.05**
	(0.166)	(0.169)	(0.164)	(0.170)	(0.192)	(0.216)	(0.207)	(0.159)	(0.295)	(0.271)	(0.376)	(0.417)
β	0.60***	0.55***	-0.21	0.04	0.34	0.09	0.48	0.52*	1.00*	0.64*	0.82	0.83
	(0.094)	(0.099)	(0.344)	(0.350)	(0.304)	(0.328)	(0.362)	(0.305)	(0.512)	(0.368)	(0.724)	(0.650)
$\beta 1$	0.04	0.07	-0.19	0.04	-0.03	0.09	0.05	0.12	0.22	0.25	-0.10	-0.12
	(0.069)	(0.073)	(0.342)	(0.272)	(0.256)	(0.287)	(0.314)	(0.259)	(0.442)	(0.432)	(0.602)	(0.437)
δ	0.25***	0.25***	0.19**	0.08	0.21**	0.18***	0.30***	0.27***	0.23**	0.28***	0.40***	0.32**
	(0.018)	(0.021)	(0.088)	(0.062)	(0.087)	(0.062)	(0.098)	(0.072)	(0.106)	(0.105)	(0.118)	(0.141)
Constant	26.03***	24.87***	21.08***	19.18***	23.03***	21.69***	25.54***	24.50***	28.53***	27.59***	31.74***	31.37***
	(0.045)	(0.361)	(0.055)	(0.377)	(0.039)	(0.380)	(0.036)	(0.314)	(0.080)	(0.453)	(0.113)	(0.771)
Observations	18,802	18,802	18,802	18,802	18,802	18,802	18,802	18,802	18,802	18,802	18,802	18,802
R-squared	0.01	0.08										
Adj. R-squared	0.01	0.08										
Pseudo R-squared			0.00	0.05	0.00	0.05	0.01	0.08	0.01	0.07	0.00	0.05

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 4.7: Causal effect of quitting smoking on BMI, CG_{08}

						$\mathbf{TG}_Q$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	0.18	0.16	0.70***	0.53	0.51	0.18	-0.09	0.28	-0.01	-0.10	-0.28	0.14
	(0.263)	(0.263)	(0.246)	(0.351)	(0.366)	(0.222)	(0.399)	(0.279)	(0.432)	(0.352)	(0.658)	(0.645)
β	0.70***	0.66***	0.01	-0.17	0.39	0.20	0.67	0.50*	0.89*	0.87*	1.18	0.63
	(0.126)	(0.135)	(0.481)	(0.475)	(0.515)	(0.355)	(0.530)	(0.290)	(0.536)	(0.513)	(1.063)	(0.853)
δ	0.15*	0.09	-0.04	-0.09	0.16	-0.10	0.12	0.03	0.34	0.15	0.03	0.28
	(0.086)	(0.097)	(0.403)	(0.350)	(0.357)	(0.236)	(0.354)	(0.247)	(0.393)	(0.349)	(0.755)	(0.510)
Constant	24.86***	23.35***	19.84***	18.38***	21.77***	21.42***	24.68***	23.15***	27.28***	25.15***	30.70***	27.32***
	(0.187)	(1.037)	(0.205)	(1.193)	(0.308)	(1.440)	(0.261)	(1.336)	(0.240)	(1.320)	(0.560)	(3.005)
Observations	1,770	1,770	1,770	1,770	1,770	1,770	1,770	1,770	1,770	1,770	1,770	1,770
R-squared	0.01	0.08										
Adj. R-squared	0.01	0.06										
Pseudo R-squared			0.00	0.03	0.00	0.04	0.01	0.06	0.01	0.06	0.00	0.03
						$\mathbf{TG}_Q$ + $\mathbf{TG}_R$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.25	-0.16	0.63***	0.43	0.14	0.05	-0.26	-0.08	-0.69*	-0.23	-0.75*	-0.89*
	(0.230)	(0.230)	(0.232)	(0.383)	(0.235)	(0.267)	(0.270)	(0.251)	(0.406)	(0.310)	(0.418)	(0.524)
$\eta 1$	-0.28	-0.21	-0.07	0.08	-0.18	-0.25	-0.41*	-0.33	-0.28	-0.21	-0.48	-0.42
	(0.211)	(0.210)	(0.255)	(0.212)	(0.220)	(0.250)	(0.248)	(0.320)	(0.276)	(0.333)	(0.475)	(0.554)
β	0.66***	0.61***	-0.20	-0.02	0.34	0.12	0.58*	0.62*	1.02**	0.92*	0.75	0.54
	(0.107)	(0.112)	(0.382)	(0.418)	(0.417)	(0.398)	(0.305)	(0.306)	(0.491)	(0.515)	(0.871)	(0.629)
$\beta 1$	0.10	0.12	-0.17	-0.09	-0.03	0.16	0.15	0.37	0.24	0.07	-0.16	-0.37
	(0.085)	(0.091)	(0.443)	(0.282)	(0.338)	(0.316)	(0.376)	(0.442)	(0.362)	(0.500)	(0.666)	(0.730)
δ	0.19***	0.16***	0.17	0.06	0.22	0.14	0.20	0.09	0.21	0.28	0.46	0.65
	(0.054)	(0.059)	(0.260)	(0.223)	(0.271)	(0.217)	(0.252)	(0.284)	(0.237)	(0.353)	(0.431)	(0.442)
Constant	25.29***	24.13***	19.91***	18.41***	22.13***	20.83***	24.86***	23.43***	27.96***	27.04***	31.17***	32.99***
	(0.137)	(0.659)	(0.193)	(0.501)	(0.207)	(0.938)	(0.169)	(0.920)	(0.174)	(0.947)	(0.235)	(1.298)
Observations	4,304	4,304	4,304	4,304	4,304	4,304	4,304	4,304	4,304	4,304	4,304	4,304
R-squared	0.00	0.05										
Adj. R-squared	0.00	0.04										
Pseudo R-squared			0.00	0.02	0.00	0.02	0.01	0.04	0.01	0.04	0.00	0.02

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

4.4.2 Investigating smoking effects on BMI of obese people

OLS and quantile regressions estimates of the effect of quitting smoking on BMI in the obese sample are showed in Tab 4.9, 4.10, 4.11 and 4.12. These estimates were obtained respectively with the usual treatment and control groups. The ATE for obese individuals is estimated to be about 0.96 (s.e. = 0.390), while QTEs as previously discussed result significant only at the bottom of the BMI distribution. The estimated QTEs are respectively 1.71 (s.e. = 0.600) and 1.02 (s.e. = 0.610) at the 10^{th} and 25^{th} quantile, these values, as already discussed are counterbalanced by an average decrease of BMI for this category, measured by δ, which however is not strong enough to overcome the positive variation induced by quitting smoking. Under CG_{ALL} and CG_{NS} we find similar results in terms of estimated ATEs and QTEs, with the exception that under these control groups we estimate a more persistent significance across quantiles, extended now also at the median, where the QTE is estimated to be 0.98 (s.e. = 0.439) for CG_{ALL} and 1.01 (s.e. = 0.440) for CG_{NS}. All the other coefficients are close to those estimated using CG_S and the differences found are negligible. Finally, Table 4.12 lists the estimates obtained with CG_{08} as control group. we can notice that, under this specification the ATE, is not significant and the 10^{th} quantile is the only significant QTE with estimated parameter close to that of CG_S, 1.74 (s.e. = 1.016).

Also for the obese sub-sample no significant effects related to smoking reductions are found from our estimates, and we can conclude also in this case that quitting is responsible for the entire variation in BMI.

Table 4.8: Causal effect of quitting smoking on BMI, obese individuals CG_S

						$\mathbf{TG}_Q$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.48	-0.42	0.10	-0.22	-0.22	-0.18	-0.18	-0.56	-1.67**	-1.31	-0.96	-0.94
	(0.363)	(0.400)	(0.159)	(0.308)	(0.216)	(0.359)	(0.451)	(0.477)	(0.715)	(1.043)	(1.285)	(1.136)
β	1.04***	0.96**	1.44**	1.71***	1.05*	1.02*	0.80	0.97	0.82	0.38	2.24	1.93
	(0.365)	(0.390)	(0.631)	(0.600)	(0.632)	(0.61)	(0.527)	(0.648)	(1.039)	(1.258)	(1.995)	(1.824)
δ	-0.26*	-0.25*	-1.35***	-1.35***	-0.42*	-0.51**	0.09	-0.06	0.46	0.23	0.10	0.10
	(0.133)	(0.139)	(0.299)	(0.278)	(0.246)	(0.223)	(0.289)	(0.337)	(0.441)	(0.429)	(0.527)	(0.499)
Constant	32.94***	33.52***	30.18***	28.21***	30.83***	30.57***	32.19***	33.62***	34.57***	36.55***	37.03***	39.70***
	(0.138)	(0.924)	(0.080)	(1.249)	(0.120)	(1.104)	(0.146)	(1.399)	(0.284)	(1.677)	(0.363)	(1.144)
Observations	752	752	752	752	752	752	752	752	752	752	752	752
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.02	0.00	0.03	0.00	0.05	0.01	0.05	0.00	0.03
						$\mathbf{TG}_Q$ + $\mathbf{TG}_R$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.42	-0.40	0.15	-0.18	-0.12	-0.29	-0.07	-0.45	-1.67*	-1.29	-0.96	-1.22
	(0.378)	(0.416)	(0.135)	(0.293)	(0.208)	(0.352)	(0.472)	(0.476)	(0.872)	(0.789)	(1.189)	(1.107)
$\eta 1$	0.21	0.23	0.29	-0.02	0.28	0.05	0.06	0.64	0.15	0.16	0.00	0.55
	(0.314)	(0.334)	(0.204)	(0.237)	(0.224)	(0.338)	(0.624)	(0.483)	(0.727)	(0.670)	(0.600)	(0.652)
β	1.05***	0.96**	1.72**	1.66***	1.09*	1.20*	0.72	1.04	0.82	0.34	2.20	1.99
	(0.380)	(0.404)	(0.697)	(0.617)	(0.659)	(0.659)	(0.601)	(0.683)	(1.335)	(1.246)	(1.973)	(1.624)
$\beta 1$	-0.15	-0.18	0.56	0.16	0.10	0.07	0.14	-0.56	-0.22	-0.29	-0.83	-1.18
	(0.305)	(0.318)	(0.551)	(0.573)	(0.541)	(0.548)	(0.797)	(0.708)	(1.026)	(0.938)	(0.888)	(0.817)
δ	-0.28*	-0.27*	-1.63***	-1.34***	-0.46*	-0.67**	0.06	0.03	0.46	0.32	0.13	0.31
	(0.154)	(0.160)	(0.271)	(0.265)	(0.267)	(0.304)	(0.330)	(0.422)	(0.542)	(0.467)	(0.485)	(0.467)
Constant	32.90***	33.66***	30.13***	28.33***	30.73***	30.53***	32.18***	33.50***	34.57***	36.10***	37.03***	39.30***
	(0.162)	(0.914)	(0.076)	(1.101)	(0.119)	(1.002)	(0.156)	(1.339)	(0.328)	(1.803)	(0.360)	(1.215)
Observations	752	752	752	752	752	752	752	752	752	752	752	752
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.03	0.00	0.03	0.01	0.05	0.01	0.05	0.00	0.02

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 4.9: Causal effect of quitting smoking on BMI, obese individuals CG_{ALL}

						$\mathbf{TG}_Q$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.35	-0.41	-0.00	-0.17	-0.39**	-0.40*	-0.23	-0.70*	-1.27*	-0.82	-0.11	-0.02
	(0.339)	(0.353)	(0.140)	(0.161)	(0.162)	(0.244)	(0.373)	(0.356)	(0.687)	(0.509)	(1.287)	(1.027)
β	0.97***	0.89***	1.36**	1.49**	1.01*	0.89*	0.84**	0.98**	0.83	0.47	1.79	0.61
	(0.343)	(0.347)	(0.616)	(0.604)	(0.592)	(0.514)	(0.410)	(0.439)	(1.309)	(0.922)	(2.200)	(1.700)
δ	-0.19***	-0.17***	-1.27***	-1.34***	-0.38***	-0.45***	0.04	0.16	0.44***	0.40***	0.55*	0.37*
	(0.051)	(0.053)	(0.130)	(0.111)	(0.096)	(0.080)	(0.062)	(0.107)	(0.162)	(0.147)	(0.314)	(0.209)
Constant	32.81***	33.03***	30.29***	29.50***	31.00***	30.82***	32.23***	32.61***	34.17***	35.12***	36.19***	37.59***
	(0.055)	(0.465)	(0.038)	(0.454)	(0.026)	(0.418)	(0.032)	(0.656)	(0.066)	(0.768)	(0.067)	(0.855)
Observations	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.02	0.00	0.03	0.00	0.05	0.01	0.05	0.00	0.03
						$\mathbf{TG}_Q + \mathbf{TG}_R$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.34	-0.40	-0.00	-0.15	-0.39**	-0.40	-0.23	-0.68*	-1.25*	-0.85	-0.11	0.02
	(0.339)	(0.353)	(0.124)	(0.205)	(0.178)	(0.273)	(0.357)	(0.388)	(0.726)	(0.826)	(1.064)	(1.110)
$\eta 1$	0.22	0.10	0.13	-0.12	0.01	0.07	0.01	0.04	0.36	0.18	0.39	0.41
	(0.267)	(0.270)	(0.217)	(0.157)	(0.193)	(0.243)	(0.584)	(0.508)	(0.464)	(0.638)	(0.460)	(0.500)
β	0.97***	0.89**	1.37**	1.45***	1.01*	0.85*	0.84**	0.96*	0.82	0.54	1.79	0.50
	(0.343)	(0.347)	(0.653)	(0.477)	(0.605)	(0.509)	(0.423)	(0.523)	(1.175)	(1.201)	(1.933)	(1.547)
$\beta 1$	-0.15	-0.15	0.21	0.62	0.02	-0.18	0.16	-0.10	-0.01	-0.61	-0.80	-0.81
	(0.264)	(0.264)	(0.488)	(0.432)	(0.437)	(0.421)	(0.649)	(0.590)	(0.630)	(0.758)	(0.692)	(0.670)
δ	-0.19***	-0.17***	-1.28***	-1.35***	-0.38***	-0.43***	0.04	0.16	0.46**	0.42**	0.55*	0.42**
	(0.052)	(0.054)	(0.125)	(0.126)	(0.110)	(0.087)	(0.089)	(0.113)	(0.200)	(0.179)	(0.295)	(0.210)
Constant	32.80***	33.02***	30.29***	29.48***	31.00***	30.82***	32.23***	32.60***	34.15***	35.29***	36.19***	37.71***
	(0.056)	(0.470)	(0.034)	(0.530)	(0.040)	(0.396)	(0.042)	(0.686)	(0.074)	(0.701)	(0.073)	(0.865)
Observations	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638	3,638
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.03	0.00	0.03	0.01	0.05	0.01	0.05	0.00	0.02

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 4.10: Causal effect of quitting smoking on BMI, obese individuals CG_{NS}

						$\mathbf{TG}_Q$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.32	-0.42	-0.01	-0.14	-0.40**	-0.46	-0.23	-0.77**	-1.25	-0.83	-0.08	0.63
	(0.339)	(0.352)	(0.116)	(0.214)	(0.165)	(0.287)	(0.357)	(0.376)	(0.804)	(0.575)	(1.127)	(0.767)
β	0.96***	0.90***	1.30*	1.12*	0.97*	0.93*	0.84*	1.01**	0.90	0.58	2.05	0.47
	(0.344)	(0.348)	(0.709)	(0.617)	(0.587)	(0.517)	(0.444)	(0.440)	(1.379)	(0.977)	(2.114)	(1.692)
δ	-0.18***	-0.17***	-1.20***	-1.28***	-0.34***	-0.34***	0.04	0.11	0.38*	0.41**	0.29	0.48*
	(0.055)	(0.058)	(0.121)	(0.131)	(0.093)	(0.107)	(0.123)	(0.128)	(0.198)	(0.188)	(0.282)	(0.258)
Constant	32.78***	33.22***	30.30***	29.86***	31.01***	30.89***	32.24***	33.13***	34.15***	35.43***	36.16***	37.23***
	(0.059)	(0.539)	(0.044)	(0.566)	(0.042)	(0.577)	(0.039)	(0.853)	(0.062)	(0.809)	(0.060)	(0.986)
Observations	2,994	2,994	2,994	2,994	2,994	2,994	2,994	2,994	2,994	2,994	2,994	2,994
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.02	0.00	0.03	0.00	0.05	0.01	0.05	0.00	0.03
						$\mathbf{TG}_Q$ + $\mathbf{TG}_R$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.32	-0.41	-0.01	-0.24	-0.40**	-0.31	-0.23	-0.73*	-1.25*	-0.86	-0.08	0.47
	(0.340)	(0.354)	(0.105)	(0.190)	(0.181)	(0.302)	(0.403)	(0.424)	(0.724)	(0.877)	(1.290)	(1.193)
$\eta 1$	0.24	0.07	0.12	-0.01	0.00	0.10	-0.00	-0.15	0.36	0.29	0.42	0.31
	(0.268)	(0.270)	(0.189)	(0.155)	(0.220)	(0.225)	(0.707)	(0.331)	(0.653)	(0.451)	(0.507)	(0.502)
β	0.96***	0.89**	1.30**	1.30**	0.97*	0.86*	0.84*	0.91*	0.90	0.75	2.05	0.53
	(0.344)	(0.349)	(0.575)	(0.532)	(0.547)	-0.518	(0.485)	(0.500)	(0.982)	(1.101)	(1.992)	(1.575)
$\beta 1$	-0.16	-0.14	0.13	0.41	-0.02	-0.07	0.16	0.09	0.07	-0.50	-0.54	-0.63
	(0.265)	(0.265)	(0.455)	(0.393)	(0.464)	(0.427)	(0.826)	(0.632)	(0.854)	(0.779)	(0.768)	(0.732)
δ	-0.18***	-0.17***	-1.20***	-1.29***	-0.34***	-0.35***	0.04	0.12	0.38	0.36**	0.29	0.50*
	(0.055)	(0.058)	(0.155)	(0.126)	(0.083)	(0.108)	(0.102)	(0.120)	(0.240)	(0.178)	(0.271)	(0.259)
Constant	32.78***	33.26***	30.30***	30.02***	31.01***	30.95***	32.24***	33.51***	34.15***	35.22***	36.16***	37.43***
	(0.059)	(0.497)	(0.040)	(0.498)	(0.035)	(0.493)	(0.053)	(0.700)	(0.089)	(0.842)	(0.072)	(1.065)
Observations	3,160	3,160	3,160	3,160	3,160	3,160	3,160	3,160	3,160	3,160	3,160	3,160
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.03	0.00	0.03	0.01	0.05	0.01	0.05	0.00	0.02

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

Table 4.11: Causal effect of quitting smoking on BMI, obese individuals CG_{08}

						$\mathbf{TG}_Q$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	0.12	0.69	0.15	0.63	-0.53*	0.48	0.05	0.24	-0.74	0.56	0.95	1.39
	(0.421)	(0.503)	(0.307)	(0.742)	(0.283)	(0.638)	(0.537)	(0.583)	(0.806)	(0.848)	(1.448)	(1.121)
β	0.90*	0.60	2.10**	1.74*	1.73*	0.86	0.62	0.81	0.93	0.91	1.13	0.00
	(0.487)	(0.575)	(1.065)	(1.016)	(0.990)	(0.915)	(0.690)	(0.825)	(1.482)	(1.124)	(1.913)	(1.273)
δ	-0.12	0.12	-2.01***	-1.32*	-1.10	-0.54	0.26	0.25	0.34	0.17	1.21	1.23
	(0.346)	(0.417)	(0.762)	(0.784)	(0.696)	(0.741)	(0.453)	(0.706)	(0.863)	(0.981)	(0.914)	(1.209)
Constant	32.34***	34.20***	30.13***	28.79***	31.14***	30.59***	31.96***	34.55***	33.64***	35.72***	35.12***	40.06***
	(0.250)	(2.182)	(0.273)	(3.024)	(0.192)	(3.575)	(0.265)	(3.153)	(0.552)	(3.483)	(0.626)	(3.905)
Observations	210	210	210	210	210	210	210	210	210	210	210	210
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.02	0.00	0.03	0.00	0.05	0.01	0.05	0.00	0.03
						$\mathbf{TG}_Q + \mathbf{TG}_R$						
Variables	**OLS**	**OLSc**	**Q10**	**Q10c**	**Q25**	**Q25c**	**Q50**	**Q50c**	**Q75**	**Q75c**	**Q90**	**Q90c**
η	-0.56	-0.30	0.15	-0.31	-0.39**	-0.20	-0.18	-0.30	-1.77	-1.49*	-0.96	-0.82
	(0.397)	(0.439)	(0.205)	(0.408)	(0.182)	(0.320)	(0.430)	(0.484)	(1.118)	(0.773)	(1.385)	(0.991)
$\eta 1$	0.00	0.07	0.28	-0.17	0.01	0.02	0.05	0.28	-0.16	0.17	-0.45	0.40
	(0.337)	(0.366)	(0.266)	(0.330)	(0.251)	(0.330)	(0.704)	(0.464)	(0.955)	(0.681)	(0.641)	(0.713)
β	1.00**	0.79*	1.40**	1.36**	1.34**	0.87	0.46	-0.04	0.76	0.29	2.17	1.75
	(0.396)	(0.435)	(0.699)	(0.685)	(0.684)	(0.745)	(0.782)	(0.756)	(1.800)	(1.180)	(2.101)	(1.621)
$\beta 1$	-0.13	-0.20	0.23	0.17	0.35	-0.12	-0.22	-1.06	-0.06	-0.59	-0.42	-0.87
	(0.329)	(0.341)	(0.527)	(0.555)	(0.627)	(0.451)	(1.031)	(0.818)	(1.274)	(0.947)	(0.949)	(0.844)
δ	-0.22	-0.13	-1.30***	-1.20***	-0.71*	-0.35	0.43	0.73	0.51	0.49	0.17	0.07
	(0.201)	(0.212)	(0.432)	(0.363)	(0.411)	(0.290)	(0.598)	(0.531)	(0.865)	(0.575)	(0.492)	(0.519)
Constant	33.02***	34.27***	30.14***	30.82***	31.00***	31.69***	32.19***	34.33***	34.67***	37.37***	37.03***	39.06***
	(0.212)	(1.081)	(0.164)	(1.584)	(0.130)	(0.968)	(0.226)	(1.373)	(0.583)	(1.538)	(0.389)	(1.709)
Observations	568	568	568	568	568	568	568	568	568	568	568	568
R-squared	0.01	0.05										
Adj. R-squared	0.00	0.01										
Pseudo R-squared			0.00	0.03	0.00	0.03	0.01	0.05	0.01	0.05	0.00	0.02

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** $\leq$ 0.01, ** $\leq$ 0.05, * $\leq$ 0.1.*

4.4.3 IV estimates

In this section we present a robustness analysis of our estimates using IV. Moreover we adopted the technique proposed by Chernozhukov and Hansen (2008) to obtain IVQR estimates for each quantile of the BMI distribution. This sensitivity analysis is meant to check whether the DID approach was successful in eliminating the distortion due to unobserved confounders which may affect simultaneously BMI and smoking behaviours. We propose this analysis because the decision to quit smoking could be endogenously determined or affected by sample selection.

The instrumental variable used in our analysis is the percentage of smokers in the same socio-economic group of respondent. This variable is a measure of the effect of social interactions on smoking. Social interactions are believed to affect many socio-economic outcomes, because the utility that an individual receives from consuming a given good is increased by other individuals in the same group consuming the same good. The main idea is to verify if the average behaviour, related to smoking, in a group affects the behaviour of the individual that belongs to that group. For this reason we calculated the percentages of smokers in each region according to age, income level, education, type of occupation and marital status. This variable is expected to negatively affect the decision of quitting, because the higher is the percentage of smokers in a given socio-economic group the lower will be the probability of quitting, and viceversa the lower is this percentage the higher will be the probability to quit, Aristei and Pieroni (2010).

IV estimates can not be directly estimated from equation (4.15), because the decision to quit smoking is expressed as interaction between two variables, the treatment and time dummies. So, in order to perform the standard IV procedure at the average of BMI distribution we specified equation (4.15) in the differences. Since our interest is focused on the entire BMI distribution we estimated the IV model also for selected quantiles, using the Instrumental Variable Quantile Regression (IVQR). This estimation technique, proposed by Chernozhukov and Hansen (2008) allows to obtain IV estimates for quantile regressions using a numerical method analogous to two stages least squares, but computationally easy and fast[5]. However, in our case it is not possible to perform IVQR on the differences of

[5]For a more detailed description of the Instrumental Variable Quantile Regression (IVQR) technique, see Chernozhukov and Hansen (2008).

BMI because the analogy between equation (4.15) and its counterpart in the differences no longer applies when we use a quantile estimator. In fact in this case one would estimate the effect of quitting smoking on the BMI variations distribution, rather than on the BMI distribution. Since our work is interested in the analogy between quantiles and BMI classes we are forced to use the specification from equation (4.15) and we consequently need to modify lightly the IVQR routine described below.

The endogenous variable in our model is the decision to quit smoking between 2004 and 2006, and the parameter associated, as already discussed, is β. We are interested in obtaining IVQR estimates of β for selected quantiles of the BMI distribution, from now on β_{IVQR}. In order to perform this task we proceeded to: 1) estimate the first step regression between quitting smoking as a function of the the instrument $Z(i)$ and the covariance matrix $X(i)$, 2) obtain the fitted values from the first step regression and 3) substitute $D(i,t)$ with the fitted values obtained in the previous step, and interact them with the year dummy. From now on the original procedure proposed by Chernozhukov and Hansen (2008) was used to estimate β_{IVQR} and the results are reported in Table (4.13) using CG_S. We do not report the same results also for CG_{08} because the numerosity of this control group is not sufficient to obtain consistent parameters estimates.

The estimated ATE of 0.70 (s.e. = 0.13), listed in Table 4.13, is higher than that obtained using OLS in the baseline model, 0.56 (s.e. = 0.103), if we look at QTEs instead we can see how also in this case the values are found to be different, not in terms of significance, in fact β_{IVQR} are significantly different from zero at the median and at the 75^{th} quantile, but in terms of magnitude of their effects, β_{IVQR} takes respectively the values of 0.76 (s.e. = 0.301) and 1.15 (s.e. = 0.456) at the median and 75^{th} quantile, while standard QR coefficients are 0.48 (s.e. = 0.275) and 0.80 (s.e. = 0.416) at the same quantiles of the BMI distribution. However these effects are not very high in terms of weight.

In the obese subsample, we find that QTEs parameters are significant only at the bottom of the BMI distribution, and also in this case they are higher than those obtained from standard QR estimates. The ATE is not significant and this result is very similar to that obtained using CG_{08} as control group. The effect of quitting smoking is estimated to increase BMI by 1.93 points at the 10^{th}, 1.81 at the 25^{th}, and only 0.74 at the median of its

distribution for the obese group. The values estimated via standard QR were 1.71 at the 10^{th}, 1.02 at the 25^{th}, and the effect at the median was non significant. These estimates suggest that endogeneity is responsible for the distortion in our baseline estimates, but again this distortion seems to be not very large in terms of weight gains.

We propose also a set of tests, reported in Chernozhukov and Hansen (2008) for the following hypothesis: 1) the treatment has no impact on the outcomes ($\beta_{IVQR}(\tau) = 0$), 2) the treatment has a constant effect on the outcomes, 3) stochastic dominance of treatments, ($\beta_{IVQR}(\tau) \geq 0$) and 4) exogeneity, or Hausman test between IVQR and QR estimates ($\beta_{IVQR}(\tau) = \beta_{QR}(\tau)$).

The results are reported in Table 4.14 for the baseline model and for the baseline model on obese individuals. We can see that the test of no causal effect and constant causal effect are strongly rejected in both samples, while the hypothesis of stocastic dominance is accepted in both cases, finally the test of exogeneity or equality among IVQR and QR estimates rejects the null hypothesis and confirms the presence of endogeneity in our estimates.

Table 4.12: Causal effect of quitting smoking on BMI, IV estimates

			CG$_S$			
Variables	**OLS**	**Q10**	**Q25**	**Q50**	**Q75**	**Q90**
η	-	0.32***	0.09	-0.41*	-0.66***	-0.60*
	-	(0.128)	(0.244)	(0.215)	(0.232)	(0.361)
β	0.70***	0.28	0.47	0.76**	1.15***	0.58
	(0.132)	(0.212)	(0.325)	(0.301)	(0.415)	(0.369)
δ	0.23***	-0.12	0.04	0.28**	0.27**	0.37**
	(0.039)	(0.131)	(0.112)	(0.124)	(0.135)	(0.145)
			CG$_S$, Obese			
Variables	**OLS**	**Q10**	**Q25**	**Q50**	**Q75**	**Q90**
η	-	-0.32***	0.41***	-0.04	0.08	0.57**
	-	(0.05)	(0.07)	(0.11)	(0.21)	(0.25)
β	0.65	1.93***	1.81***	0.74***	0.00	0.34
	(0.53)	(0.13)	(0.14)	(0.17)	(0.26)	(0.37)
δ	-0.24	-1.38***	-0.59***	0.05	0.51***	-0.06
	(0.15)	(0.27)	(0.24)	(0.06)	(0.06)	(0.05)

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** ≤ 0.01, ** ≤ 0.05, * ≤ 0.1.*

Table 4.13: Process tests for the BMI equation. Subsample size = $5n^{2/5}$

		CG$_S$	
Null Hypothesis	**Kolmogorov–Smirnov statistic**	**90% Critical value**	**95% Critical value**
No effect. $\beta_{IVQR}(\tau) = 0$	4.33	1.61	1.71
Costant effect $\beta_{IVQR}(\tau) = \alpha$	4.59	3.56	3.92
Dominance $\beta_{IVQR}(\tau) \geq 0$	0.00	1.26	1.48
Exogeneity $\beta_{IVQR}(\tau) = \beta_{QR}(\tau)$	4.82	2.94	3.18
		CG$_S$, Obese	
Null Hypothesis	**Kolmogorov–Smirnov statistic**	**90% Critical value**	**95% Critical value**
No effect. $\beta_{IVQR}(\tau) = 0$	4.00	1.31	1.50
Costant effect $\beta_{IVQR}(\tau) = \alpha$	2.75	2.35	2.69
Dominance $\beta_{IVQR}(\tau) \geq 0$	0.00	0.76	0.98
Exogeneity $\beta_{IVQR}(\tau) = \beta_{QR}(\tau)$	4.87	2.60	3.30

*Notes: Standard errors are shown in round brackets and significant levels are reported with the following notation: p-value *** ≤ 0.01, ** ≤ 0.05, * ≤ 0.1.*

4.5 A primer conclusion: the effects on body weight

In this section we use the previously etimates ATEs and QTEs, to calculate the effect of quitting smoking on weight rather than on BMI. Since BMI is expressed as the ratio between weight and squared height, the variation in terms of weight can be easily obtained multiplying the coefficients, estimated through OLS and quantile regressions, by the average of the squared height for each sample analysed. The results of this analysis are summarised in Figure 4.2 and 4.3, for the entire sample and the obese group respectively. The effect of quitting smoking was estimated, as already discussed, considering a two years time period. Now, for simplicity, we will calculate the effect on one year only.

Figure 4.2 shows the estimated effects of quitting smoking on the entire sample, for each control group analysed. As we see from the figure the average weight variation attributed to quitting smoking is about 0.8 kg per year under the control groups CG_S, CG_{ALL} and CG_{NS}, while under CG_{08} is about 1 kg. Similarly the QTEs are 0.7 kg under CG_S and CG_{08}, and about 0.8 kg and 0.6 kg under CG_{NS} and CG_{ALL} at the median. At the 75^{th} quantile the weigth gain is about 1 kg, and ranges from 0.9 kg under CG_{ALL} to 1.2 kg under CG_S. Finally, the effect on weight under CG_{08} is slightly higher and around 1.4 kg.

The estimated effects on weigth using IV are in line with our baseline results and the differences found are almost negligible. The ATE is estimated to be 1.1 kg, the same value is obtained at the median of the distribution, while at the 75^{th} quantile the effect is significantly higher, if compared to previous results, and is about 1.8 kg.

Figure 4.3 shows the weigth variations calculated for obese individuals, also in this case we do not find very different results across control groups. We can notice how the estimated ATE using CG_S is of 1.5 kg, and that a very close result is obtained under CG_{NS} and CG_{ALL}. The QTE at the 10^{th} quantile ranges from almost 2 kg under CG_{NS} to about 2.5 kg under CG_S and CG_{ALL}, while at the 25^{th} quantile is slightly slower and around 1.5 kg. A significant effect is estimated also at the median of the BMI distribution but only under CG_{ALL} and CG_{NS} with value similar to that obtained at the 25^{th} quantile. Finally, both ATEs under CG_{08} and by IVQR under CG_S are non-significant. QTEs at the 10^{th} quantile is about 2.5 kg for each specification, while at the 25^{th} quantile and at

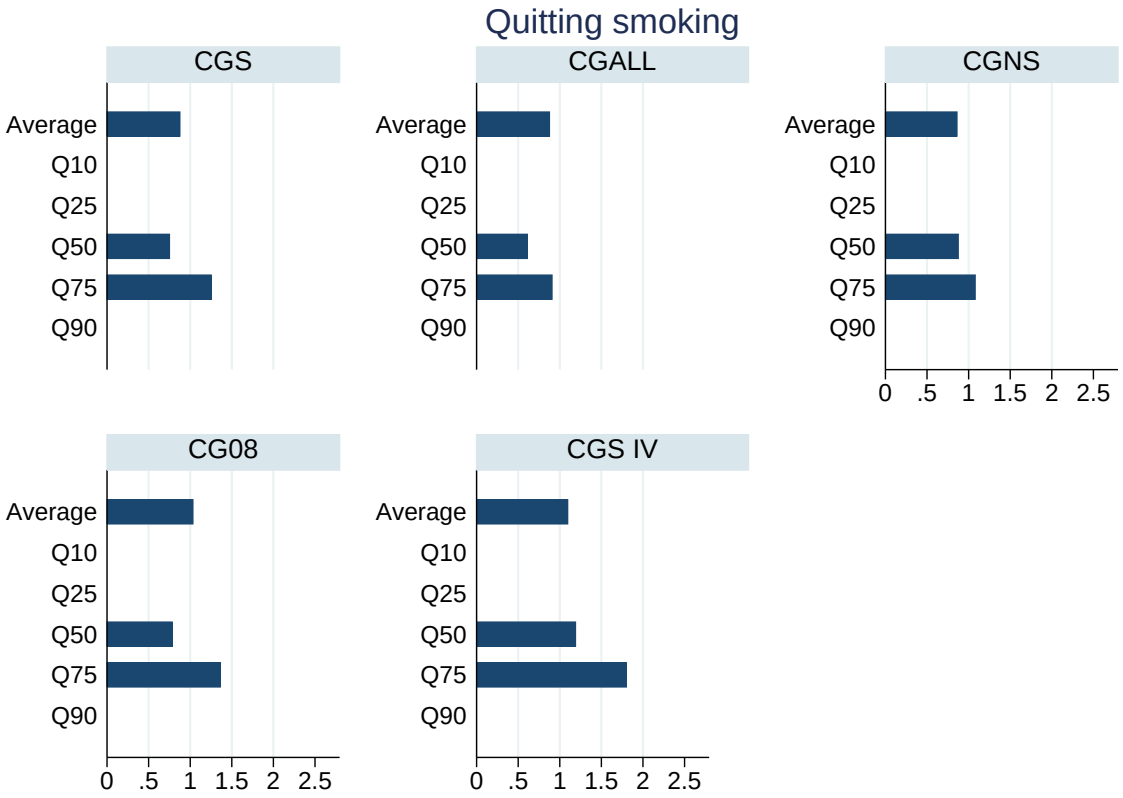


Figure 4.2: Quitting smoking effect on weight (kg), by control group

the median is signicant only for the IVQR specification.

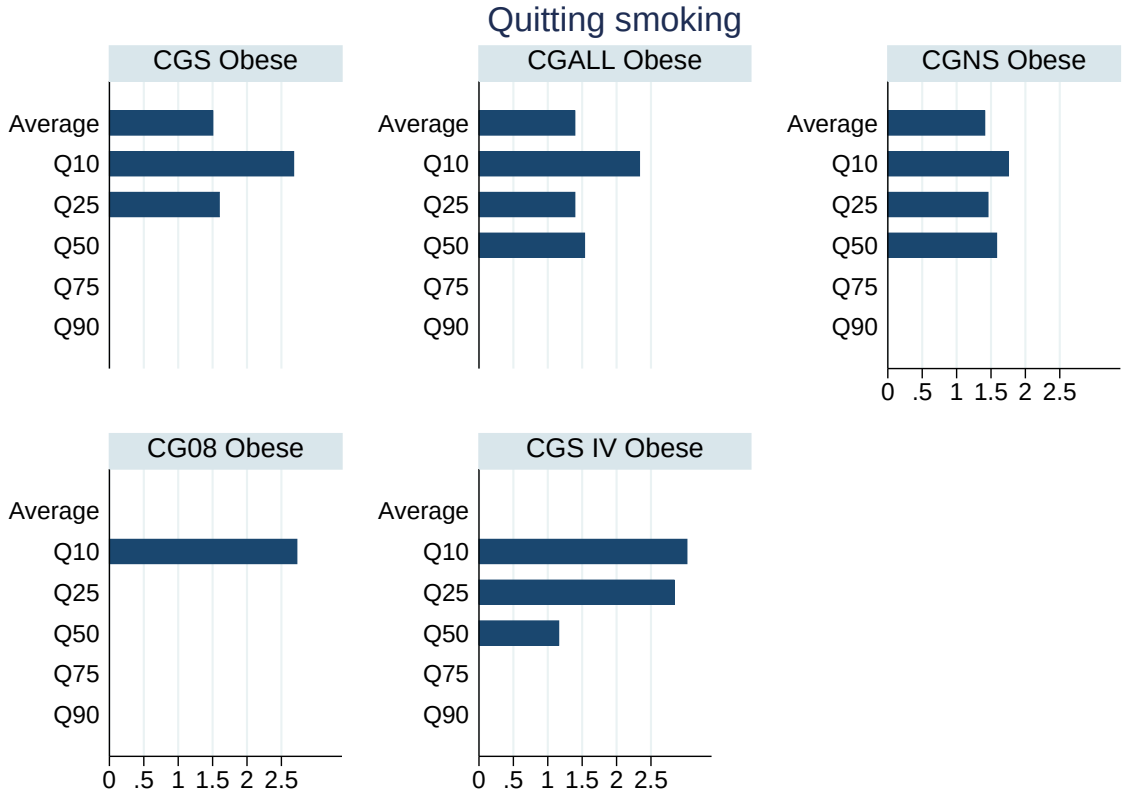


Figure 4.3: Quitting smoking effect (kg), IV estimates by control group

APPENDIX 4.A.1

Table 4.A.1: Descriptive statistics of covariates

	Respondent quitted smoking between 2004 and 2006 (TG_Q)				**Respondent continued smoking between 2004 and 2006 (CG_S)**			
	2004		**2006**		**2004**		**2006**	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Male	0.49	0.5	0.49	0.5	0.46	0.5	0.46	0.5
Female	0.51	0.5	0.51	0.5	0.54	0.5	0.54	0.5
Goes for a drink at least once a week	0.37	0.48	0.29	0.46	0.34	0.47	0.33	0.47
Goes for a drink at least once a month	0.23	0.42	0.28	0.45	0.24	0.43	0.22	0.41
Goes for a drink several times a year	0.16	0.37	0.18	0.38	0.17	0.38	0.18	0.39
Goes for a drink once a year or less	0.05	0.22	0.04	0.2	0.07	0.25	0.06	0.24
Goes for a drink never/almost never	0.19	0.39	0.21	0.41	0.18	0.38	0.21	0.41
Health status is excellent	0.22	0.41	0.21	0.41	0.17	0.38	0.19	0.39
Health status is good	0.44	0.5	0.43	0.5	0.45	0.5	0.44	0.5
Health status is fair	0.23	0.42	0.23	0.42	0.25	0.43	0.24	0.43
Health status is poor	0.09	0.29	0.11	0.31	0.1	0.3	0.1	0.3
Health status is very poor	0.02	0.13	0.02	0.15	0.03	0.17	0.03	0.18
Non-long term sick	0.94	0.24	0.95	0.22	0.91	0.28	0.91	0.28
Long term sick	0.06	0.24	0.05	0.22	0.09	0.28	0.09	0.28
Hours worked weekly (including overtime)	39.59	11.43	38.17	12.36	37.91	13.61	37.64	12.99
Strenuous job	0.24	0.43	0.24	0.43	0.33	0.47	0.31	0.46
Non-strenuous job	0.76	0.43	0.76	0.43	0.67	0.47	0.69	0.46
Working mother	0.41	0.49	0.37	0.48	0.35	0.48	0.34	0.47
Non-working mother	0.59	0.49	0.63	0.48	0.65	0.48	0.66	0.47
White	0.99	0.11	0.99	0.11	0.99	0.08	0.99	0.08
Black	0.01	0.11	0.01	0.11	0.01	0.08	0.01	0.08
Aged 18-30	0.26	0.44	0.21	0.41	0.22	0.42	0.18	0.39
Aged 30-40	0.25	0.43	0.25	0.43	0.24	0.42	0.23	0.42
Aged 40-50	0.15	0.35	0.18	0.38	0.21	0.41	0.23	0.42
Aged 50-60	0.12	0.33	0.12	0.33	0.18	0.38	0.18	0.39
Aged 60+	0.22	0.41	0.24	0.43	0.15	0.36	0.18	0.39
1st quintile of net income	0.23	0.42	0.2	0.4	0.28	0.45	0.26	0.44
2nd quintile of net income	0.18	0.39	0.21	0.41	0.23	0.42	0.23	0.42
3rd quintile of net income	0.23	0.42	0.21	0.41	0.21	0.41	0.2	0.4
4th quintile of net income	0.2	0.4	0.2	0.4	0.16	0.36	0.17	0.38
5th quintile of net income	0.15	0.36	0.19	0.39	0.12	0.32	0.14	0.35
Couple	0.18	0.39	0.18	0.39	0.2	0.4	0.19	0.39
Married	0.46	0.5	0.5	0.5	0.43	0.5	0.44	0.5
Divorced	0.06	0.24	0.08	0.27	0.09	0.29	0.1	0.3
Separated	0.03	0.16	0.02	0.13	0.03	0.18	0.03	0.17
Widowed	0.05	0.21	0.05	0.22	0.05	0.21	0.05	0.23
Never Married	0.22	0.42	0.17	0.38	0.2	0.4	0.19	0.39
Degree	0.13	0.34	0.14	0.35	0.06	0.24	0.07	0.25
Diploma	0.26	0.44	0.29	0.45	0.25	0.43	0.28	0.45
Alevel	0.11	0.31	0.1	0.3	0.12	0.32	0.11	0.31
Olevel	0.17	0.38	0.17	0.37	0.19	0.39	0.18	0.38
Loweducated	0.1	0.3	0.09	0.29	0.1	0.29	0.09	0.28
No qualificication	0.2	0.4	0.19	0.39	0.26	0.44	0.25	0.43
Still in school	0	0	0	0	0.01	0.08	0.01	0.07
England	0.52	0.5	0.54	0.5	0.43	0.49	0.44	0.5
Wales	0.14	0.35	0.14	0.35	0.17	0.38	0.17	0.38
Scotland	0.17	0.37	0.16	0.37	0.19	0.39	0.19	0.39
Northern Ireland	0.15	0.36	0.15	0.36	0.19	0.4	0.19	0.4

Conclusions

The results shown in previous chapters of this thesis showed how, cross-price elasticities estimated by demand systems provide a consistent framework to evaluate substitution effects among goods in Italy. In particular, we presented estimates of the long-run substitution effects of the categories of unhealthy and healthy foods for Italy. By showing the close pattern linking unhealthy foods with more energy-dense foods, our findings suggest that the largest rise in healthy food prices, versus unhealthy ones, have favoured the consumption of high-calorie foods. This result matches the increase in body weight recently recorded in Italy.

One peculiarity of our results is that, as relative healthy food prices rise, individuals who are male, below the relative poverty line, and those with lower education tend to substitute energy-light foods for more energy-dense. The changes in relative prices, which have caused healthy foods to become 10 basis points more expensive in Italy in ten years, have generated a mechanism of substitution towards foods which are cheaper and high in calories, affecting mainly some disadvantaged groups.

While for the UK, we showed how different effects of socio-economic causes on individual body weight have important implications as regards whether the UK government should recommend policies for adult obesity reduction. Our OLS regression results support literature findings regarding the significant determinants of obesity. However, quantile regressions reveal the sensitivity of these determinants to BMI distributions. While significant in the OLS case, a lack of physical exercise consistently increased BMI only at higher levels of the quantile distribution, reinforcing findings that proper physical exercise can reduce the phenomenon of obesity. From evidence that there are remarkable differences in gender behaviour, the most significant revelation regards the relative prices of food. The effect of higher prices for healthy foods such as fruit and vegetables in increasing body weight is significant in all the quantiles of the specifications proposed, and stronger in obese people, although its influence in reducing body weight appears to be quite small. Also, the effect on BMI of a reduction in the relative prices of take-away restaurants is significant and increasing for women but is not statistically significant for men. Moreover, our results reveal that obesity is lower for men who spend less extra-time

at work. However, the increased density of restaurants and fast-food shops, while non-significant in the OLS case, does affect the calorie intake and consumption of overweight people, with a remarkable effect for women over 90^{th} quantile. A key implication of our findings is that obesity control policies are unlikely to succeed equally across gender at different BMI levels.

Learning about the effects of smoking habits on weight gains requires a setting which generates an exogenous variation in future outcomes. The 2005 Italian Clean Indoor Air Law generates such a variation and is the background for our research design. With this law, individuals of the same age, but born in adjacent birth-cohorts face different smoking restrictions in public places.

With this approach, we find that smokers significantly respond to the smoking ban, giving rise to sudden negative changes in cigarette consumption. We attempt to estimate the magnitudes of smoking habit changes on weight with a simplified structural model. We estimate this impact over an ex-post period of three years (2005-2007) by an IV approach, which we compared with an RD design able to assess the effects of nicotine reductions on BMI in a one year post-period. As a general result, the baseline model predicts a large impact in terms of BMI changes with the shortest horizons of the RD sample. Estimates for employed individuals and for men and women, show how the effect of the smoking ban on men and employees has very large adverse effects on smoking habits, but their weight gains are smaller than those at the average of the population whereas women tend to gain weight largely in response to smaller reductions in cigarette consumption.

We also conclude that a decreasing smoking participation rate or cigarette consumption has limited effects on weight gains in both overweight and obese people. Our results show that anti-smoking policies, generally favored by society, may not play a role in increasing the cost of overweight and obesity, when we look towards the future.

In the UK quitting smoking is found to increase obesity and BMI in general. Individuals, regardless from the BMI class they belong, show a significant increase in their BMI in correspondence of the average variation of their category. Consequently quitting smoking increases the risk of overweight and obesity especially for those individuals who were normalweight or overweight. In fact, keeping all other variables constant, quitters already at the threshold of normalweight or overweight classes will become respectively

overweight or obese in the future, while, obese tend to remain trapped in the same BMI class even in the occurence of a weight loss. The misleading non-significant global effect obtained for obese individuals from the whole sample was due to the fact that in the estimation process we were ignoring individual specific trends, which are found to differ greatly among individuals belonging to different BMI classes. The effect of reducing smoking, rather than quitting, was found to have no significant influence on BMI. In conclusion, the estimated weight gains attributable to smoking reductions are not so large and do not support, at least in the long run, the hypothesis that quitting smoking favors weight increases dangerous for health and costly for the society.

Bibliography

Abadie, A. (2005), 'Semiparametric difference-in-differences estimators', *Review of Economic Studies* **72**(1), 1–19.

Adda, J., Berlinski, S. and Machin, S. (2006), 'Short-run economic effects of the scottish smoking ban', *International Journal of Epidemiology* **36**(1), 149–154.

Allen, R. G. D. (1938), *Mathematical analysis for economists*, Macmillan London.

Angrist, J. D. (1995), 'The economic returns to schooling in the west bank and gaza strip', *American Economic Review* **85**(5), 1065–87.

Anonymous (2005), 'Ireland's smoking ban is an admirable achievement', *Lancet* (365), 1282.

Aristei, D. and Pieroni, L. (2010), 'Habits, complementarities and heterogeneity in alcohol and tobacco demand: A multivariate dynamic model', *Oxford Bulletin of Economics and Statistics* **72**(4), 428–457.

Atella, V., Pace, N. and Vuri, D. (2008), 'Are employers discriminating with respect to weight?: European evidence using quantile regression', *Economics & Human Biology* **6**(3), 305–329.

Attfield, C. (1997), 'Estimating a cointegrated demand system', *European Economic Review* **41**, 61–73.

Attfield, C. (2004), 'Stochastic trends demographics and demand systems', *University of Bristol mimeographs* .

Auld, M. C. and Powell, L. M. (2009), 'Economics of food energy density and adolescent body weight', *Economica* **76**(304), 719–740.

Baker, M., Gruber, J. and Milligan, K. (2008), 'Universal child care, maternal labor supply, and family well-being', *Journal of Political Economy* **116**(4), 709–745.

Ball, K. and Crawford, D. (2005), 'Socioeconomic status and weight change in adults: a review', *Social Science Medicine* **60**, 1987–2010.

Barefoot, J., Heitmann, B., Helms, M., Williams, R., Surwit, R. and Siegler, I. (1998), 'Symptoms of depression and changes in body weight from adolescence to mid-life.', *International Journal of Obesity Related Metabolic Disorders.* **22**(7), 688–694.

Barnett, W. A. and Serletis, A. (2008), 'Consumer preferences and demand systems', *Journal of Econometrics* **147**(2), 210–224.

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

Becker, G. S. and Murphy, K. M. (1988), 'A theory of rational addiction', *Journal of Political Economy* **96**(4), 675–700.

Bhattacharya, P. (1963), 'On an analog of regression analysis', *Annal of Mathematical Statistics* **34**, 1459–1473.

Bickel, W. K., Odum, A. L. and Madden, G. J. (1999), 'Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers', *Psychopharmacology* **146**(1), 447–54.

Blanchflower, D. G., Oswald, A. J. and Landeghem, B. V. (2009), 'Imitative obesity and relative utility', *Journal of the European Economic Association* **7**(2-3), 528–538.

Bleich, S., Cutler, D., Murray, C. and Adams, A. (2007), Why is the developed world obese?, NBER Working Papers 12954, National Bureau of Economic Research, Inc.

Brunello, G., Fort, M., Schneeweis, N. and Winter-Ebmer, R. (2011), The causal effect of education on health: What is the role of health behaviors?, Economics working papers 2011-17, Department of Economics, Johannes Kepler University Linz, Austria.

Brunello, G., Michaud, P.-C. and de Galdeano, A. S. (2009), 'The rise of obesity in europe: an economic perspective', *Economic Policy* **24**, 551–596.

Burkhauser, R. V. and Cawley, J. (2008), 'Beyond bmi: The value of more accurate measures of fatness and obesity in social science research', *Journal of Health Economics* **27**(2), 519–529.

Cash, S. B., Sunding, D. L. and Zilberman, D. (2004), Fat taxes and thin subsidies: Prices, diet, and health outcomes, Technical report.

Cawley, J., Markowitz, S. and Tauras, J. (2004), 'Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation', *Journal of Health Economics* **23**(2), 293–311.

Chaloupka, F. (1992), 'Clean indoor air laws, addiction and cigarette smoking', *Applied Economics* **24**, 193–205.

Chaloupka, F. and Saffer, H. (1992), 'Clean indoor air laws and the demand for cigarettes', *Contemporary Economic Policy* **10**, 72–83.

Chernozhukov, V. and Hansen, C. (2008), 'Instrumental variable quantile regression: A robust inference approach', *Journal of Econometrics* **142**(1), 379–398.
URL: *http://ideas.repec.org/a/eee/econom/v142y2008i1p379-398.html*

Chesher, A. (1998), 'Individual demands from household aggregates: time and age variation in the composition of diet', *Journal of Applied Econometrics* **13**(5), 505–524.

Chou, S.-Y., Grossman, M. and Saffer, H. (2004), 'An economic analysis of adult obesity: results from the behavioral risk factor surveillance system', *Journal of Health Economics* **23**(3), 565–587.

Chouinard, H. H., Davis, D. E., LaFrance, J. T. and Perloff, J. M. (2007), 'Fat taxes: Big money for small change', *Forum for Health Economics & Policy* **10**(2), 2.

Cinciripini, P., Wetter, D., Fouladi, R., Blalock, J., Carter, B., Cinciripini, L. and Baile, W. (2003), 'The effects of depressed mood on smoking cessation: mediation by postcessation self-efficacy.', *Journal of Consulting and Clinical Psychology.* **71**(2), 292–301.

Conforti, P., Pierani, P. and Rizzi, P. (2001), Food and nutrient demands in italy actual behaviour and forecast through a multistage quadratic system with heterogeneous preferences, Working Paper 303, Univerity of Siena.

Coondoo, D., Majumder, A. and Ray, R. (2004), 'A method of calculating regional consumer price differentials with illustrative evidence from india', *Review of Income and Wealth* **50**(1), 51–68.

Costa, D. L. and Steckel, R. H. (1995), Long-term trends in health, welfare, and economic growth in the united states, NBER Historical Working Papers 0076, National Bureau of Economic Research, Inc.

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

Cranfield, J. A. L. and Pellow, S. (2004), 'The role of global vs. local negativity in functional form selection: an application to canadian consumer demands', *Economic Modelling* **21**(2), 345–360.

Currie, J., Vigna, S. D., Moretti, E. and Pathania, V. (2010), 'The effect of fast food restaurants on obesity and weight gain', *American Economic Journal: Economic Policy* **2**(3), 32–63.

Cutler, D., Glaeser, E. and Shapiro, J. (2003), Why have americans become more obese?, Technical report.

de Saint Pol, T. (2009), 'Evolution of obesity by social status in france, 1981-2003', *Economics & Human Biology* **7**(3), 398–404.

Deaton, A. (1985), 'Panel data from time series of cross-sections', *Journal of Econometrics* **30**(1-2), 109–126.

Deaton, A. (1986), 'Demand analysis in z griliches and md intriligator', *Handbook of Econometrics* **3**, 1767–1839.

Deaton, A. and Paxson, C. (1999), Mortality, education, income, and inequality among american cohorts, NBER Working Papers 7140, National Bureau of Economic Research, Inc.

Deaton, A. S. and Muellbauer, J. (1980), 'An almost ideal demand system', *American Economic Review* **70**(3), 312–26.

Del Bono, E. and Vuri, D. (2011), 'Job mobility and the gender wage gap in italy', *Labour Economics* **18**(1), 130–142.

Department of Health (2006), Forecasting obesity to 2010, Technical report.

Dill, D., Edwards, H. and Forbes, W. (1934), 'Tobacco smoking in relation to blood sugar, blood lactic acid and metabolism', *Am J Physiol* **109**, 118–22.

Doksum, K. (1974), 'Empirical probability plots and statistical inference for nonlinear models in the twosample case', *Annals of Statistics* **2**, 267–277.

Drewnowski, A. and Darmon, N. (2005), 'Food choices and diet costs: an economic analysis', *Journal of Nutrition* **135**, 900–904.

Drewnowski, A., Darmon, N. and Briend, A. (2004), 'Replacing fats and sweets with vegetables and fruits – a question of cost', *American Journal of Public Health* **94**(9), 1555–1559.

Edgerton, D. L. (1997), 'Weak separability and the estimation of elasticities in multistage demand systems', *American Journal of Agricultural Economics* **79**(1), 62–79.

Eisenberg, D. and Quinn, B. (2006), 'Estimating the causal effect of smoking cessation on weight gain: An instrumental variables approach.', *Health Services Research* **41**(6), 2255–2266.

Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica* **64**(4), 813–36.

Eriksen, M. and Chaloupka, F. (2007), 'The economic impact of clean indoor air laws', *A Cancer Journal for clinicians* **57**(6), 367–378.

Etilé, F. (2007), 'Social norms, ideal body weight and food attitudes', *Health Economics* **16**(9), 945–966.

Evans, W. N., Farrelly, M. C. and Montgomery, E. (1999), 'Do workplace smoking bans reduce smoking?', *American Economic Review* **89**(4), 728–747.

Fang, H., Ali, M. M. and Rizzo, J. A. (2009), 'Does smoking affect body weight and obesity in china?', *Economics & Human Biology* **7**(3), 334–350.

Fichtenberg, C. and Glantz, S. (2004), 'Effect of smoke-free workplaces on smoking behavior: Systematic review', *British Medical Journal* (325), 188–194.

Flegal, K. (2007), 'The effects of changes in smoking prevalence on obesity prevalence in the united states', *American Journal of Public Health* **97**, 1510–1514.

Fong, G., Hyland, A., Borland, R., Hammond, D., Hastings, G., McNeill, A., Anderson, S., Cummings, K., Allwright, S., Mulcahy, M., Howell, F., Clancy, L., Thompson, M., Connolly, G. and Driezen, P. (2006), 'Reductions in tobacco smoke pollution and increases in support for smoke-free public places following the implementation of comprehensive smoke-free workplace legislation in the republic of ireland: findings from the itc ireland/uk survey', *Tobacco Control* **15**, 51–58.

French, S. and Jeffery, R. (1995), 'Weight concerns and smoking: a literature review', *Annals of Behavioral Medicine* **17**, 234–244.

Gallagher, D., Visser, M. and Sepulveda, D. (1996), 'How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?', *American Journal of Epidemiology* **143**(3), 228–239.

Gallus, S., Pacifici, R. and Colombo, P. e. a. (2006), 'Prevalence of smoking and attitude towards smoking regulation in italy, 2004', *Eur. J. Cancer Prev.* **15**, 77–81.

Garn, S., Leonard, W. and Hawthorne, V. (1986), 'Three limitations of the body mass index', *American Journal of Clinical Nutrition* **44**, 996–997.

Görg, H. and Strobl, E. (2002), 'Multinational companies and entrant start-up size: Evidence from quantile regressions', *Review of Industrial Organization* **20**(1), 15–31.

Gelbach, J. B., Klick, J. and Stratmann, T. (2009), Cheap donuts and expensive broccoli: The effect of relative prices on obesity, Working paper.

Gil, J. and Mora, T. (2011), 'The determinants of misreporting weight and height: The role of social norms', *Economics & Human Biology* **9**(1), 78–91.

Glauser, S., Glauser, E., Reidenberg, M., Rusy, B. and Tallarida, R. (1970), 'Metabolic changes associated with the cessation of cigarette smoking', *Arch Environ Health* **20**(1), 377–81.

Gorman, W. (1995), *Separability and Aggregation*, C. Blackorby and A. Shorrocks. Editors Collected Works of W.M. Gorman Vol 1 Clarendon Press Oxford.

Gruber, J. and Frakes, M. (2006), 'Does falling smoking lead to rising obesity?', *Journal of Health Economics* **25**(2), 183–197.

Grunberg, N. (1985), 'Nicotine, cigarette smoking, and body weight', *British Journal of Addiction* **80**, 369–377.

Guthrie, J. F., Andrews, M. S., Frazao, E., Leibtag, E., Lin, B.-H., Mancino, L., Nord, M., Prell, M. A., Smallwood, D. M., Variyam, J. N. and Ver Ploeg, M. (2007), Can food stamps do more to improve food choices? an economic perspective, Economic Information Bulletin 59417, United States Department of Agriculture, Economic Research Service.

Hahn, J., Todd, P. and Van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression-discontinuity design', *Econometrica* **69**(1), 201–09.

Hartog, J., Pereira, P. and Vieira, J. (2001), 'Changing returns to education in portugal during the 1980s and early 1990s: Ols and quantile regression estimators', *Applied Economics* **33**(8), 1021–1037.

Havnes, T. and Mogstad, M. (2010), Is universal child care leveling the playing field? evidence from non-linear difference-in-differences, IZA Discussion Papers 4978, Institute for the Study of Labor (IZA).

Hicks, J. R. (1986), *A Revision of Demand Theory*, number 9780198285502 *in* 'OUP Catalogue', Oxford University Press.

Higgins, M.W.and Kjelsberg, M. (1967), 'Characteristics of smokers and nonsmokers in tecumseh, michigan', *American Journal of Epidemiology* **86**, 60–77.

House of Commons Health Committee (2004), Obesity, Technical report.

Imbens, G. W. and Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–75.

Imbens, G. W. and Rubin, D. B. (1997), 'Estimating outcome distributions for compliers in instrumental variables models', *Review of Economic Studies* **64**(4), 555–74.

ISTAT (2007), Annual multipurpose survey of households "aspetti della vita quotidiana", Technical report.

Jacobs, D. and Gottenberg, S. (1981), 'Smoking and weight: The minnesota lipid research clinic', *American Journal of Public Health* (71), 391–396.

Jacobs, M., Knapp, P., Anderson, L., Karnsh, N., Meissner, R. and Richman, S. (1965), 'Relationship of oral frustration factors with heavy cigarette smoking in male college students.', *J Nerv Ment Dis* **141**, 161–171.

Jeffery, R. (1996), 'Bias in reported body weight as a function of education, occupation, health and weight concern', *Addictive behaviours* **21**, 217–222.

Johansen, S. (1995), 'Identifying restrictions of linear equations with applications to simultaneous equations and cointegration', *Journal of Econometrics* **69**(1), 111–132.

Juster, H., Loomis, B., Hinman, T. and et al. (2007), 'Declines in hospital admissions for acute myocardial infarction in new york state after implementation of a comprehensive smoking ban.', *Am J Public Health* **97**(11), 2035–2039.

Kapteyn, A., Alessie, R. and Lusardi, A. (2005), 'Explaining the wealth holdings of different cohorts: Productivity growth and social security', *European Economic Review* **49**(5), 1361–1391.

Karvonen, M., Keys, A., Orma, E., Fidanzo, F. and Brozek, J. (1959), 'Cigarette smoking, serum cholesterol, blood pressure and body fatness: observations in finland.', *Lancet* **1**, 492–494.

Keys, A., Aravanis, C. and Blackburn, H. (1966), 'Epidemiologic studies related to coronary heart disease: characteristics of men aged 40-59 in seven countries', *Acta Med Scand Suppl* **460**, 1–75.

Klesges, R., Meyers, A., Klesges, L. and LaVasque, M. (1989), 'Smoking, body weight, and their e?ects on smoking behavior: a comprehensive review of the literature', *Psychology Bulletin* **106**, 204–230.

Koenker, R. and Hallock, K. F. (2001), 'Quantile regression', *Journal of Economic Perspectives* **15**(4), 143–156.

Koenker, R. W. and Bassett, Gilbert, J. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Komlos, J. and Baur, M. (2004), 'From the tallest to (one of) the fattest: the enigmatic fate of the american population in the 20th century', *Economics & Human Biology* **2**(1), 57–74.

Kriegbaum, M., Larsen, A., Christensen, U., Lund, R. and Osler, M. (2010), 'Reduced probability of smoking cessation in men with increasing number of job losses and partnership breakdowns.', *J Epidemiol Community Health* (55), 271–278.

Kuchler, F., Golan, E., Variyam, J. and Crutchfield, S. (2005), 'Obesity policy and the law of unintended consequences', *Amber Waves* **3**, 26–33.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root?', *Journal of Econometrics* **54**(1-3), 159–178.

Lakdawalla, D. and Philipson, T. (2002), The growth of obesity and technological change: A theoretical and empirical examination, NBER Working Papers 8946, National Bureau of Economic Research, Inc.

Lehmann, E. (1974), 'Nonparametrics: Statistical methods based on ranks', *Holden-Day Inc., San Francisco* .

Lewbel, A. (1999), 'Consumer demand systems and household equivalence scales', *Handbook of Applied Econometrics* pp. 167–196.

Lewbel, A. and Ng, S. (2005), 'Demand systems with nonstationary prices', *The Review of Economics and Statistics* **87**(3), 479–494.

Lin, B.-H. and Guthrie, J. F. (2007), Can food stamps do more to improve food choices? an economic perspective-how do low-income households respond to food prices, Economic Information Bulletin 59432, United States Department of Agriculture, Economic Research Service.

Liu, F., Zhang, N., Cheng, K.-W. and Wang, H. (2010), 'Reduced smoking and rising obesity: Does smoking ban in the workplace matter?', *Economics Letters* **108**(3), 249–252.

Loureiro, M. L. and Nayga, Rodolfo M., J. (2005), Obesity rates in oecd countries: An international perspective, Technical report.

Luk, R., Ferrence, R. and Gmel, G. (2011), 'The economic impact of a smoke-free bylaw on restaurant and bar sales in ottawa, canada', *Addiction* (101), 738–745.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag Berlin.

Machin, S., Marie, O. and Vujic, S. (2011), 'The crime reducing effect of education', *Economic Journal* **121**(552), 463–484.

McCarthy, H., Cole, T., Fry, T., J., S.A. and Prentice, A. (2006), 'Body fat reference curves for children', *International Journal of Obesity* **30**, 598–602.

McCaul, K. D., Hockemeyer, J. R., Johnson, R. J., Zetocha, K., Quinlan, K. and Glasgow, R. E. (2006), 'Motivation to quit using cigarettes: A review', *Addictive Behaviours* **31**(1), 42–56.

Meyer, B. D. (1995), 'Natural and quasi-experiments in economics', *Journal of Business Economic Statistics* **13**, 151–161.

Michaud, P.-C., van Soest, A. and Andreyeva, T. (2007), Cross-country variation in obesity patterns among older americans and europeans, Social and Economic Dimensions of an Aging Population Research Papers 185, McMaster University.

Muellbauer, J. (1976), 'Community preferences and the representative consumer', *Econometrica* **44**(5), 979–99.

Ng, S. (1995), 'Testing for homogeneity in demand systems when the regressors are nonstationary', *Journal of Applied Econometrics* **10**(2), 147–63.

Nonnemaker, J., Finkelstein, E., Engelen, M., Hoerger, T. and Farrelly, M. (2009), 'Have efforts to reduce smoking really contributed to the obesity epidemic?', *Economic Inquiry* **47**(2), 366–376.

Oreopoulos, P. (2006), 'Estimating average and local average treatment effects of education when compulsory schooling laws really matter', *American Economic Review* **96**(1), 152–175.

Pesaran, M. H. (1997), 'The role of economic theory in modelling the long run', *Economic Journal* **107**(440), 178–91.

Pesaran, M. H. and Shin, Y. (2002), 'Long-run structural modelling', *Econometric Reviews* **21**(1), 49–87.

Philipson, T. (2001), 'The world-wide growth in obesity: an economic research agenda', *Health Economics* **10**(1), 1–7.

Pieroni, L., Lanari, D. and Salmasi, L. (2011), 'Food prices and overweight patterns in italy', *The European Journal of Health Economics* pp. 1–19. 10.1007/s10198-011-0350-y.

Powell, L. and Chaloupka, F. (2009), 'Food prices and obesity: Evidence and policy implications for taxes and subsidies', *The Milbank Quarterly* **87**(1), 229–257.

Prentice, A. and Jebb, S. (2001), 'Beyond body mass index', *Obesity Reviews* **2**(3), 141–147.

Rashad, I., Grossman, M. and 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.

Saikkonen, P. and Lütkepohl, H. (2000), 'Trend adjustment prior to testing for the cointegrating rank of a var process', *Journal of Time Series Analysis* **21**, 435–456.

Sargent, R., Shepard, R. and Glantz, S. (2004), 'Reduced incidence of admissions for myocardial infarction associated with public smoking ban: before and after study.', *BMJ* **328**, 977.

Schlosser, E. (2001), *Fast food nation*, New York time Bestsellers.

Schroeter, C., Lusk, J. and Tyner, W. (2008), 'Determining the impact of food price and income changes on body weight', *Journal of Health Economics* **27**(1), 45–68.

Schunck, R. and Rogge, B. (2010), 'Unemployment and its association with health-relevant actions: investigating the role of time perspective with german census data.', *International Journal of Public Health* (55), 271–278.

Scollo, M., Lal, A. and Glantz, S. (2003), 'Review of the quality of studies on the economic effects of smoke-free policies on the hospitality industry', *Tobacco Control* **12**, 13–20.

Service, N. H. (2009), Statistics on obesity, physical activity and diet: England, february 2009., Technical report.

Smalley, K., Knerr, A., Kendrick, Z., Colliver, J. and Owen, O. (1990), 'Reassessment of body mass indices', *American Journal of Clinical Nutrition* **52**, 405–408.

Tauras, J. A. (2005), 'An empirical analysis of adult cigarette demand', *Eastern Economic Journal* **31**(3), 361–375.

Trochim, W. (1984), 'Research design for program evaluation.', *Beverly Hills: Sage Publications.* .

Trochim, W. (2001), 'Regression discontinuity design.', *International Encyclopedia of the Social and Behavioral Sciences* **19**, 12940–12945.

Trochim, W. and Campbell, D. (1960), 'Regression-discontinuity analysis: An alternative to the ex-post facto design.', *Journal of Educational Psychology* **51**, 309–317.

US Department of Health and Human Services (1972), 'The health consequences of smoking: A report of the surgeon general', *DHHS Publication CDC* (72-7516).

US Department of Health and Human Services (2006), '2006 surgeon general's report on the health consequences of involuntary exposure to tobacco smoke', *DHHS Publication CDC* (90-8416).

Viscusi, W. K. and Hersch, J. (2001), 'Cigarette smokers as job risk takers', *The Review of Economics and Statistics* **83**(2), 269–280.

Wasserman, J., Manning, W., Newhouse, J. and Winkler, J. (1991), 'The effect of excise taxes and regulations on cigarette smoking', *Journal of Health Economics* **10**, 43–64.

Wurtman, J. (1993), 'Depression and weight gain: the serotonin connection.', *International Journal of Obesity Related Metabolic Disorders.* **29**(2-3), 183–192.

Yurekli, A. and Zhang, P. (2000), 'The impact of clean indoor air laws and cigarette smuggling on demand for cigarettes: an empirical model', *Health Economics* **9**, 159–170.

Yusuf, S., Hawken, S. and Ounpuu, S. (2005), 'Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study', *Lancet* **366**, 1640–1649.

Zagorsky, J. L. (2005), 'Health and wealth: The late-20th century obesity epidemic in the u.s', *Economics & Human Biology* **3**(2), 296–313.

Zheng, X. and Zhen, C. (2008), 'Healthy food, unhealthy food and obesity', *Economics Letters* **100**(2), 300–303.