HEDG Health, Econometrics and Data Group

THE UNIVERSITY of York

HEDG Working Paper 08/02

Decomposing Bodymass Index gaps between Mediterranean countries: A Counterfactual Quantile Regression Analysis*

Joan Costa-Font
Daniele Fabbri
Joan Gil

February 2008

ISSN 1751-1976

http://www.york.ac.uk/res/herc/research/hedg/wp.htm

DECOMPOSING BODY MASS INDEX GAPS BETWEEN MEDITERRANEAN COUNTRIES:

A COUNTERFACTUAL QUANTILE REGRESSION ANALYSIS*

JOAN COSTA-FONT [A, C, D, E] **DANIELE FABBRI** [B] **JOAN GIL** [C, D, E]

A. LSE HEALTH & EUROPEAN INSTITUTE, LONDON SCHOOL OF ECONOMICS

B. DIPARTIMENTO DI SCIENZE ECONOMICHE, UNIVERSITÀ DI BOLOGNA

C. DEPARTAMENT DE TEORIA ECONÒMICA, UNIVERSITAT DE BARCELONA

D. CAEPS, UNIVERSITAT DE BARCELONA

E. FEDEA, MADRID

This version: February, 2008

(*) **ACKNOWLEDGEMENTS**: We are grateful to Wiji Arulampalam and Mark L. Bryan for providing us with their STATA code to simulate distributions from quantile regressions. We thank participants at the XXVII Jornadas de Economía de la Salud, the IHEA 6th World Congress and the FEDEA research workshop for their comments and suggestions. Joan Costa-Font and Joan Gil acknowledge the financial support of the Spanish Ministry of Education (Project SEJ2005-06270) and the Government of Catalonia (Project 2005 SGR-460 and the XREEP-Xarxa de Referència en Economia i Polítiques Públiques) and Joan Costa-Font acknowledges the funds from the Fundació Ramon Llull.

CONTACT ADDRESS: European Institute, Cowdray House, Houghton Street, London.WC2A 2AE. Tel: +44 207 955 6484 Fax: +44 207 955 6803.

E-mail: J.Costa-Font@lse.ac.uk, d.fabbri@unibo.it, joangil@ub.edu.

ABSTRACT

Wide cross-country variation in obesity rates have been reported within European Union member states. However, health production determinants for these differences have been largely overlooked in the health economics literature. In this paper we propose a methodology for conducting standardized cross-country comparisons in BMI. The method we adopt is based on the estimation of the marginal density function of BMI in a given country implied by different counterfactual distributions of all the covariates included within a quantile regression framework. We apply our method to the analysis of the variation in BMI distribution in Spain with respect to Italy in the year 2003. Our findings suggest that Spain-to-Italy BMI gaps are largely explained by cross-country variation in the returns to each health input. Therefore, there appear to be differences in the country-specific behavioural responses to the caloric (im)balance.

KEYWORDS: BMI, country weight gap, quantile regression, counterfactual decomposition, Mediterranean countries, Italy, Spain.

JEL CLASSIFICATION: I18, J15, J16

1 INTRODUCTION

The expansion of overweight and obesity has reached alarming levels in the western world (WHO, 2003, Silvertonen *et al.*, 2004). Obesity and overweight currently affect about two thirds of the US population, stemming from a rapid expansion in the last two decades (Ruhm, 2007). Similar patterns are found in the EU, where the prevalence of obesity has tripled in the last twenty years (Branca *et al.*, 2007); 27% of European men and 38% of women are now considered obese. Such a fast process raises real concern, since obese individuals are more likely to suffer from chronic conditions including hypertension, diabetes, osteoporosis and heart disease, which represent a significant burden for the healthcare system and may have a devastating effect on individuals' quality of life, even if they may not increase mortality (Gregg and Guralnik, 2007). The European Commission's Green Paper on promoting healthy diets and physical activity, published in 2005, estimated that obesity accounts for as much as 7% of total healthcare costs in the EU.

Within Europe it has been reported that obesity rates vary widely from country to country (see Sanz-de-Galdeano 2005). This large variation is likely to be reflected in differences in the impact on the healthcare burden of particular diseases in different countries. Developing cross-country comparisons in obesity might shed some light on the underlying causes of the obesity epidemic. In particular, it can provide suggestive evidence of the role played by contextual or environmental effects that are not fully measured by surveys, taking into account differences in cultural attitudes towards food, and individuals' self-image which may directly or indirectly shape health production. Despite their obvious importance, few cross-country comparisons of this kind have been carried out. Notable exceptions are the papers by Sanz-de-Galdeano (2005) and Michaud et al. (2007). In this paper we propose a methodology for conducting standardised cross-country comparisons in BMI. Specifically, we focus on two important issues which have not been properly addressed before in this literature.

The first is the need to set a cut-off point for defining obesity and overweight. When setting a cut-off point, some information is inevitably lost, and the validity of the comparison is compromised. Contoyannis and Wildman (2007) report that two very different BMI distributions may produce similar obesity rates. For instance, in one country a large share of the population might be concentrated on or around the overweight or obesity threshold, making obesity prevalence sensitive to such cut-off points. Recent evidence also indicates that cross-country differences are larger in the right tail of the BMI distribution and that differences are particularly marked between genders (Ruhm, 2007). For these reasons, comparisons of obesity across countries (or across time[1]) should take the entire BMI distribution into consideration.

To our knowledge, only Contoyannis and Wildman (2007) have applied this approach. In their paper, changes in the BMI distributions of England and Canada between 1994 and 2001 are analysed, taking advantage of polarization and inequality measures that break down distributional changes in location and in its shape. However, their approach does not account for variation in covariates affecting the BMI distribution, such as education, income or lifestyle. This severely limits the chances of gaining insight into the reasons behind cross-country variation in BMI. To be able to do so, we need a method that can disentangle the effect of variation in underlying factors (for example, education, income and lifestyle) from the effects of variation in the returns to those factors. Moreover, it has been shown that these factors affect BMI to different degrees along its different quantiles (see Kan and Tsai, 2004, and Ruhm, 2007). Therefore a second methodological issue in cross-country comparisons of BMI distribution is the opportunity to adjust for covariates related to the variable of interest within a "full distribution" analysis. Our proposal is to adopt a method that extends the traditional Oaxaca decomposition of effects on the mean to the entire distribution of the variable of interest.

Our method is based on the estimation of the marginal density function of BMI in a given country implied by different counterfactual distributions of all the covariates included. This methodology is applied to the analysis of the variation in the BMI distribution in Spain with respect to Italy in the year 2003. For instance, we will estimate the BMI density that would have prevailed in Spain if all the covariates had been distributed in the same way as in Italy. By comparing this with the actual marginal distribution in Spain, we are able to disentangle the contribution of cross-country variations in the distribution of underlying covariates underpinning the Spain-to-Italy BMI gaps observed. We then estimate the BMI density that would have

[1] Flegal (2006) reports that BMI distributions in the USA appear to have right shifted and progressively become more skewed. Using US data, Freedman *et al.* (2000) found that the 10^{th} percentile of adult BMI rose $0.6kg/m^2$ between 1990-2000, whilst this effect increased up to 1.2 at the median and 3.2 at the 95% percentile.

prevailed in Spain if BMI had been determined according to returns to covariates estimated in Italy. By comparing it with the actual marginal distribution in Spain, we will be able to disentangle the contribution of cross-country variations in the returns to covariates to the BMI gaps observed between Spain and Italy.

The counterfactual nature of the exercise requires an estimation of the BMI distribution that is conditional on the variables of interest. We accomplish this first step by means of quantile regressions (Koenker and Bassett, 1978; Garcia *et al.*, 2001), that is, by estimating models for the quantiles of the conditional (log)BMI distribution. Unlike simple ordinary least-squares methods that focus on the impact of covariates upon the mean, we model the impact of covariates upon different quantiles of the conditional distribution. As a result, we take into account the case of extreme obesity and provide a more general picture of the effect of covariates on BMI. This will provide us with an understanding of factors exerting idiosyncratic effects on the body mass in the two countries examined. However, following Machado and Mata (2005), the model we use is not merely conditional: indeed, a conditional distribution does not reflect the variability of the covariates in the two populations under scrutiny. The second stage of our analysis is thus to marginalize the conditional distribution estimated in the previous step using different scenarios for the distribution of the populations' attributes.

We should stress that a cross-country analysis of Spain vs. Italy is interesting in itself. According to Sanz-de-Galdeano (2005) Spain is one of the EU's most obese member states, while Italy is the least. This large difference exists in spite of the fact that they are Mediterranean countries with very similar income per-capita levels, socio-economic characteristics, diet and smoking habits. The gaps observed must therefore be due to "behavioural" factors rather than to "unfavourable distribution in underlying determinants". Moreover, in this paper we stress the importance of analysing the entire distribution of BMI. The Spain-to-Italy BMI gap is positive and significant throughout the entire distribution, widening clearly towards the upper tail, especially in the female population. Specific features of BMI distribution across Spain and Italy will be further discussed below.

Our paper presents some new findings. First, in breaking down the gaps between Spain and Italy BMI, we find that they are to a large extent explained by differences in education and lifestyles. According to our counterfactual decomposition analysis, the Spain-to-Italy BMI differentials observed appear to be explained by cross-country differences in the returns to each health input. Cross-country differences in the distribution of health production inputs seem to play a minor role. Hence, country-specific behavioural (or environmental) responses to caloric (in) balance appear to account for these differences in weight and obesity.

The paper is organized as follows. In the next section, we describe the conceptual background to our analysis. In section 3, we present an overview of the econometric approach we adopt, in particular dealing with decomposition methodologies and counterfactual analysis. Section 4 describes the data and the empirical specification used. Section 5 presents the results from the empirical application, and section 6 concludes.

2 BACKGROUND AND GENERAL MOTIVATION

2.1 CROSS-COUNTRY COMPARISONS OF BMI

Cross-country comparisons of BMI provide suggestive evidence that may help to understand the sources of the obesity epidemic. An epidemic is expected to have a concomitant effect across countries, but cross-country differences may highlight variations that are useful to explore the determinants of difference in body mass. Indeed, whilst traditional health production approaches suggest that health status can be modified by individual combinations of inputs, including health depreciation and direct investments in health and knowledge, the country-specific production function incorporates the effect of environmental or behavioural (cultural) responses to changes on similar health inputs.

Cross-country comparison examines whether differences in health outcomes and, in our case in body mass, result from individual factors affecting health outcomes or from country-specific effects that determine the energy balance. Short-term fluctuations in calorie intake are likely to be dealt with effectively by an individual's metabolism, which is elastic up to a certain level of daily variation. However, when the excess calorie gain is longer-lasting, the calorie imbalance manifests itself in weight gain. For this reason, the "energy accounting" approach used by Cutler *et al.*, (2003) to explain the growth of obesity in the U.S. is an appropriate conceptual framework for multivariate regression analysis of obesity as a function of individual characteristics. Alternatively, one could conceive of BMI as a health outcome, the result of choices made in a health production model (see for example Lakdawalla and Philipson, 2002).

2.2 METHODOLOGICAL AND MEASUREMENT ISSUES

Country-specific BMI data provides a snapshot that reflects the steady-state or cumulative adiposity resulting from past and current health-related behaviours. The steady state BMI of individual *i* is determined by:

$$BMI_i = x_i \beta + \gamma f_i + \delta e_i + \varepsilon_i \qquad (1)$$

where f_i measures food consumption, e_i is physical activity or exercise, and x_i is a vector of individual characteristics. Finally, ε_i is a measure of unobservables. We assume a steady-state interpretation of equation (1), that is, body weight has stabilized in the population we are looking at, and also that health behaviours have been stable for some time. Under these conditions, health behaviours should correlate with body fat, provided that these behaviours actually impact long-term imbalances in energy intake and expenditure (see Michaud *et al.*, 2007). Given that body weight also affects demand for energy intake and expenditure, some of this relationship is unlikely to be causal, whilst cultural and country-specific effects are likely to remain as unobserved heterogeneity in the "equilibrium conditions".

Studies examining the determinants of obesity suffer from the well-known measurement error due to the misreporting of weight and height. This measurement error is exacerbated by the loss of information caused by the dichotomization of the BMI whenever cut-off points are used as approximate measures of obesity. In this context, working with the entire BMI distribution represents a major improvement. Moreover, the use of quantile regression provides additional advantages as it may account for individual heterogeneity (Jones and López, 2003). This approach is followed by Kan and Tsai (2004) to explore the relationship between knowledge of health risks and body mass at each quantile of the distribution. Similarly, Flegal (2006), and more recently Ruhm (2007), examine the whole BMI distribution for the US. In a similar vein, other studies analyse cross-country differentials in BMI. Sanz-de-Galdeano (2005) compare differences in adult BMI distribution by gender in several EU countries between 1998 and 2001. Michaud *et al.* (2007) compare the BMI distribution between some EU countries and the US for adults aged 50 and above. Finally, Contoyannis and Wildman (2007) concentrates on BMI distributions of England and Canada.

To date, no study has developed counterfactual decompositions of cross-country differences in BMI. Especially relevant are decompositions based on quantile regression, which account for individual heterogeneity in BMI distribution. Blinder (1973) and Oaxaca (1973) used traditional methods which break down differences in a relevant variable (in our case, BMI) between two groups into two additive elements: one attributed to the existence of differences in observable characteristics between the two groups (e.g. Italy vs. Spain) and the other attributed to differences in the returns to those characteristics. In decomposing weight gaps, some studies address differences in the outcome distribution using quantile regression analysis (Albrecht *et al.,* 2003) or, alternatively, expand the methodology to dichotomous outcomes (Fairlie, 2005). More recent applications include censored outcomes (Bauer and Sinning, 2005) and count data (Bauer *et al.*, 2006).

In this paper we apply a decomposition methodology to the entire BMI distribution using quantile regression. Decomposition approaches have to address a major issue, namely that differences in the variable of interest (BMI in our case) may result from the combined effect of differences in health production inputs in combination with differences in the returns to such inputs. To disentangle these different components, the labour economics literature has developed methods of counterfactual decomposition for continuously distributed variables (e.g. wages and BMI). Counterfactual decomposition analysis, adjusted for quantile regression, examines whether the BMI gap is attributed to country-specific differences in characteristics or to cross-country variation in the returns to those characteristics.

2.3 RELEVANCE OF THE ITALY VS. SPAIN GAP

We examine cross-country differences in BMI between two countries, Italy and Spain, which are exposed to similar conditions (e.g., Mediterranean diet) but which exhibit marked differences in other dimensions (e.g., behaviour and use of health inputs). A cross-country comparison between Italy and Spain is relevant given the marked differences in the prevalence of obesity in the two countries: currently Spain has one of Europe's highest mean BMI levels in both men and women, while Italy ranks among the lowest. Whilst obesity and overweight affect 58.9% of men and 47% of women in Spain, these figures fall to 50.5% and 36% respectively in Italy (Sanz-de-Galdeano, 2005).

[Insert Table 1 about here]

Table 1 provides some insights into the BMI distribution in both countries. Specifically, it shows average BMI and BMI levels in different deciles of the distribution. Given that anthropometric measures are gender-specific, we split the sample in men and women. The table also shows total and gender-specific BMI gaps between countries. **Table 1** suggests that body mass gaps are always statistically significant and, interestingly, they increase along the BMI scale, so that in absolute values the gap is roughly four times higher at percentile 95 (1.98) than at percentile 5 (0.52). These differentials apply to both genders, though overall the gap is higher for women[2]. A comparison of the arithmetical mean and the median levels of BMI shows that the distributions are right skewed with long right-hand tails. Another important advantage of examining the whole distribution is that by examining the tails we can estimate where the obesity cut-off point lies. In the Spanish BMI the cut-off point for obesity lies close to percentile 85, whilst in Italy it is somewhere between percentiles 90-95. Even more striking is the finding that these differentials increase at the top of the BMI distribution. Therefore, an approach that takes into account the whole distribution of the BMI, rather than focusing on particular extremes resulting from specific cut-off points, seems to be most useful.

Figure 1 reports the distribution of BMI for Italian and Spanish men and women. Interestingly, for both genders we find that the Spanish distribution is relatively more skewed towards the right: that is, there is a higher percentage of people in the healthy BMI intervals in Italy than in Spain. Possibly the distribution is slightly more skewed among women than among men. **Figure 2** reports changes across age groups; interestingly, in agreement with previous studies (Baum II and Ruhm, 2007) there is an age effect which here appears to be the result of a transition from healthy BMI levels in younger ages to unhealthy BMI levels at older ages. This is the case both for Italy and Spain. However in Spain the process is more accentuated and the difference between middle age and old age is much more significant.

[Insert Figures 1 and 2 about here]

---

[2] When the relative gap is computed, we find that it doubles both for women (Spanish women are 3.6% heavier at percentile 5 and 8.1% heavier at percentile 95) and for men (1% heavier at percentile 5 and 4.3% heavier at percentile 95).

On the basis of these findings, it seems that an approach that takes into account the whole distribution of the body mass to analyse existing weight to height gaps in Italy and Spain is the most suitable. We discuss the matter further in the following section.

3 METHODS

3.1 *THE QUANTILE REGRESSION FRAMEWORK*

The quantile regression (QR) model, first introduced by Koenker and Bassett (1978), specifies the conditional quantile as a linear function of observed covariates. Following Buchinsky (1998), let $Q_\theta(w|X)$ for $\theta \in (0, 1)$ denote the θth conditional quantile of the distribution of (log)BMI (*w*), given a vector, X, of *k* covariates. These conditional quantiles are expressed as:

$$Q_\theta(w|X) = X'\beta(\theta) \qquad (2)$$

where $\beta(\theta)$ is a vector of coefficients, that is, the QR coefficients.[3] Following Koenker and Bassett (1978), $\beta(\theta)$ can be estimated by minimizing the following objective function with respect to β

$$n^{-1}\left[\sum_{i:w_i \geq X_i'\beta}^{n} \rho\left(w_i - X_i'\beta\right) + \sum_{i:w_i < X_i'\beta}^{n} (1-\rho)\left(w_i - X_i'\beta\right)\right] \qquad (3)$$

Although equation (3) is not differentiable and so gradient optimization methods are not applicable, linear programming methods can be used to efficiently compute $\beta(\theta)$ (Koenker and Hallock, 2001) and consistent estimates of the covariance matrix can be obtained by using bootstrap techniques.

In the empirical analysis of BMI determinants, adopting a QR framework allows us to ascertain whether the impact of health inputs on body fat generation changes across the distribution of BMI – in particular at the tails of the BMI distribution. As a result, we can take

---

[3] The θth conditional quantile of the error term is zero, that is $Q_\theta(\varepsilon|X) = 0$.

into account the existence of unobserved heterogeneity by allowing health inputs to have different returns at different points of the BMI scale. If differences are found, then we can control efficiently for important factors behind the body mass gap between Spanish and Italian populations (for both sexes or for individuals with different educational attainment). Another major advantage is that the evaluation of the effects at the extremes of the distribution contains important information for identifying the centile at which the population begins to become obese.

3.2 COUNTERFACTUAL GAPS ESTIMATION

After applying the QR estimation to the body mass equations, the following step is to calculate cross-country (log)BMI gaps (also by gender and age). Nonetheless, instead of using the traditional Blinder-Oaxaca approach where the observed gap is evaluated and decomposed in the mean distribution of characteristics, we extend these gap decomposition elaborations by taking into account the entire (log)BMI distribution which is accomplished by means of QR. However, mirroring other applications in the labour economics literature (for instance, Machado and Mata, 2005, Arulampalam *et al.*, 2005 or de la Rica *et al.*, 2005) we are interested in the computation of the *counterfactual* gaps, which are measured as the difference in (log)BMI that Italians would face at the θth quantile if their distribution of characteristics were the same as that of Spaniards.

To compute quantile counterfactual distributions, we follow the bootstrap procedure proposed by Machado and Mata (2005)[4]. The cornerstone of the method is the estimation of a marginal density function of (log)BMI that is consistent with the estimated conditional distribution defined by (2) as well as the distribution of covariates. Their approach proceeds as follows:

Step 1: Generate a random sample of size *m*=5,000 from a uniform distribution *U[0,1]*: $\theta_1......\theta_m$. These numbers are the quantiles to be estimated.

[4] Machado and Mata (2005) describes alternative methods proposed by the literature to compute counterfactual densities.

Step 2: For each θ from step 1, estimate the conditional quantile $Q_\theta(w|X)$ yielding *m* estimates of the QR coefficients, that is, $\beta_S(\theta)$ and $\beta_I(\theta)$ obtained using the Spanish and the Italian datasets respectively.

Step 3: Generate a random sample of 5,000 individuals (with replacement) from the Spanish dataset (denoted by $\{X_i^*(S)\}$ $i=1....m$) and use their characteristics to predict the (log)BMI (*w*) using the estimated coefficients ($\beta_S(\theta)$ and $\beta_I(\theta)$) from step 2,

$$\{w_i^*(j) \equiv X_i^*(S)' \beta_j(\theta)\} \quad i=1....m \;\; and \quad j=S,I \tag{4}$$

deriving two sets of predicted (log) BMI covering the entire distribution.

Step 4: Using the distributions calculated in step 3, the estimated cross-country (log)BMI gap is then computed as the difference between the predicted BMI at each quantile using the Spanish and Italian datasets or differences in densities as: $f(w(S)) - f(w(I))$, where $f(w(\cdot))$ denotes an estimator of the marginal density function of *w* (the log of BMI) based on the observed sample $X(j)$, *j=S,I.*

3.3 *THE DECOMPOSITION PROCEDURE*

The last step is the decomposition of the estimated cross-country differences in the (log)BMI distribution in two parts or counterfactual densities. Denote by $f^*(w(\cdot))$ an estimator of the marginal density of *w* based on the generated sample $X_i^*(S)$. Hence the overall gap is approached as

$$f(w(S)) - f(w(I)) = [f^*(w(S)) - f^*(w(I); X(S))] + [f^*(w(I); X(S)) - f^*(w(I))] + residual \tag{5}$$

where $f^*(w(I); X(S))$ is the counterfactual marginal density of (log)BMI that would result for Italians if the distribution of their characteristics were the same as that of the Spanish population. According to Machado and Mata (2005) the first term of the right-hand side measures the "coefficient effect" or the contribution attributable to differences in the QR coefficients (a "returns" effect), whereas the second term of the right-hand side reflects the "covariate effect" or the contribution to the total gap due to differences in the distribution of covariates plus a residual aimed at measuring differences unaccounted for by the estimated model. This allows us to disentangle whether body mass differences result a) from differences in the returns to health inputs which can be associated with behavioural or environmental factors that are associated with country specific lifestyles, or b) from differences in the concentration of certain health inputs in Spain as compared to Italy.

This distinction is important, as it suggests that interventions should focus on changing the way people use health inputs rather than on the differences in their availability. Therefore, disentangling the effect of input availability or differences in available returns becomes the integral part of the empirical analysis.

4 DATA AND MODEL SPECIFICATION

4.1 DATA AND SAMPLE

As in previous studies in the literature (Contoyannis and Wildman, 2007; Ruhm, 2007), we use cross-sectional data from representative surveys, in this case from two countries which are geographically close to each other, Italy and Spain.

The databases best suited to the study are the national health surveys in each country. The data used for Spain were taken from the 2003 edition of the Spanish National Health Survey (SNHS), a biannual, cross-sectional nationwide representative survey which gathers information on aspects such as the population's perceptions of their state of health, primary and specialized health care utilization, consumption of medicines, perceived mortality, lifestyles, conducts related to risk factors, anthropometrical characteristics, preventive practices, and socioeconomic characteristics. The original sample contained 21,650 adults aged 16-99 from all Spanish regions. After removing the data relating to some subjects aged under of 18, in order to allow comparison with the Italian database – and some missing data on weight and height information, the estimated sample contained 20,787 individuals.

The Italian data are from the 2003 edition of the National Survey on Daily Life ("Indagine sugli Aspetti della Vita Quotidiana"), a survey that compiled multipurpose individual data, including data on health conditions, healthcare access, dietary habits and body weight and height. The original sample contains information on 20,547 complete households comprising 44,384 adult individuals (aged 18 or above). After deleting certain missing data, the final sample included 40,545 individuals. Both surveys are nationally and regionally representative and use very similar sampling procedures.[5] The wording of the two questionnaires is surprisingly similar and the information compiled can be easily compared. Our meticulous work of homogenization of variables and definitions could thus be applied. After this, we appended the two samples obtaining a joint dataset of 61,332 observations of individuals aged 18 and over. Finally we omitted observations for individuals older than 85 (770 in Italy and 442 in Spain) ending up with 60,120 observations: 39,775 (66%) from Italians and 20,345 (34%) from Spaniards.

4.2 DEFINITION OF VARIABLES AND MODEL SPECIFICATION

Following common practice, as our dependent variable we used a measure of body weight based on Quetelet's index, namely the individual's body mass index (BMI). This is defined as weight in kilograms divided by the square of height in meters (kg/m^2). According to the World Health Organization (1997) classification a BMI $\geq$ 25 kg/m^2 is defined as overweight and a BMI of $\geq$30 kg/m^2 as obese. To compute this indicator, however, self-reported data on height and weight were used for each respondent. The reason for using this index is that BMI correlates highly with body fat, though there are differences in relation to age and gender. Women concentrate more adiposity than men and, on average, older people may have more body fat than

[5] The SNHS-2003, for instance, follows a stratified multi-stage sampling procedure in which the primary strata are the Autonomous Communities and sub-strata are then defined according to population size in particular areas. Within the sub-strata, municipalities and sections (primary and secondary sampling units respectively) are selected using a proportional random sampling scheme. Finally, individuals are randomly selected from the sections. The Italian National Survey on Daily Life follows a two-stage sampling procedure, with municipalities as primary sampling units and households as secondary sampling units. Municipalities are stratified by population size. Municipalities with a population above a certain threshold are always included, whereas the smaller ones are selected at random.

younger adults with the same BMI (Gallagher *et al.*, 1996).[6] An additional issue worth noting, and always present in nationwide health surveys, is that self-reported data on height and weight include a certain degree of error. Although correction procedures have been proposed in the literature (Cawley, 2004, Chou *et al.*, 2004 and Burkhauser and Cawley, 2008), we are unable to explore the implications of (partially) correcting our BMI indicator, as parallel clinical measurements of the anthropometric variables were not available. However, given that our main concern is to analyse BMI differentials, we conclude that reporting errors of this kind are not a major issue if they are similar across countries.

[Insert Table 2 about here]

In our specification we included some of the most relevant variables usually found in models for BMI equilibrium relation. In agreement with the discussion above and the evidence on the determinants of body weight, our econometric specifications considered the following set of exogenous covariates for each country (see **Table 2**) all of which were proven to be highly significant in explaining the BMI: i) the age and age square of each respondent at the date of the interview (following Kan and Tsai, 2004). ii) for knowledge and socio-economic effects, we used three categories for educational attainment and one dummy variable measuring whether the individual was currently working, iii) for marital status, a dummy indicated whether the individual was married or not, iv) for lifestyle, five dummies were considered, for smoking habits, physical activity exerted at work, breakfast and frequency of meat consumption, v) a dummy variable for private medical insurance and vi) regionally aggregated data to account for regional specific factors or heterogeneity affecting the distribution of the BMI; an aggregate level, per capita GDP, was included. The latter is taken as a measure of an environmental factor which proxies aggregate affluence. This variable partially corrects the fact that our database does not control for individual/familiar income. Income is not recorded in the Italian survey, thus

[6] Unfortunately, BMI does not take into consideration body composition (adiposity vs. lean weight) or body fat distribution. This means it may fail to predict obesity among very muscular individuals and the elderly.

making our decision necessary[7]. In our view, this limitation leads to a minor loss in the empirical soundness of our specification. However, income is not unambiguously relevant here.[8] In the US, evidence suggests that the "socio-economic gradient" is mainly explained by ethnicity and education and very little is propagated through household income (Baum II and Ruhm, 2007)[9].

5 RESULTS

We estimated a model for equation 1 for each country and gender, given that health production functions are likely to differ between genders. Since quantile regression is sensitive to outliers, we checked the magnitude of this potential problem. To this end we adopted the Hadi (1992) procedure to detect multiple outliers in multivariate data. We ran the procedure separately on each sub-sample defined by country and gender to identify outliers in the multivariate distribution of individuals' height, weight and age. For Spain we identified 13 females and 11 males, and for Italy 7 and 4 respectively. We ran the full set of quantile regressions excluding these outliers. The results obtained were very similar to those obtained by ignoring the presence of outliers. Here we present the results of the estimation conducted on the sample selected according to the criteria described above and without the outliers.

[7] Of course, income can be imputed for the Italian sub-sample by way of standard matching techniques. However our two definitions of income would not be homogenous.

[8] Kan and Tsai (2004) found that EDUCATION was statistically significant for all percentiles, while INCOME and INCOMESQ were not. Sanz-de-Galdeano (2005) found that the income effect was relevant for females but not for males. She also notes that reverse causality may lead to an overestimation of this coefficient. Michaud *et al.* (2007) found that WEALTH and EDUCATION exert consistent effects on obesity; while INCOME has no effect on (European) males, it is "negative" for European females while it is "positive" for US males and females.

[9] Moreover the evidence for Spain indicates that less that 7% of income-related inequalities in obesity are explained by pure income effects (Costa-Font and Gil, 2008).

5.1 *OLS AND QR ANALYSIS*

As our first step, we analyze the QR results displayed in **Figure 3**. The plots show the coefficient estimates, $\beta_i(\theta), i = 1,...,k$ for $\theta \in (0,1)$, and the associated confidence bands (represented by the dots). For each variable, the plots provide information on the coefficient estimates for Spanish females (first column), Italian females (second column), Spanish males (third column) and Italian males (fourth column). The dots represent the 95% (heterogeneity-consistent) confidence interval for the regression deciles, obtained by the method in Hendricks and Koenker (1992). For comparison purposes, the coefficients estimated by mean regression (OLS) are reported as a solid horizontal line (for details on the OLS regression model, see **Table 3**). The information in this figure can be summarized to reflect the impact of each covariate upon BMI inequality. Indeed, as the dependent variable is in logs, the difference in the estimated coefficients at two different quantiles provides a measure of the impact of that covariate upon the (log of the) ratio between BMIs at these quantiles. Similarly, **Figures 4** and **5** indicate that whilst in Italy BMI returns to age have significantly different coefficients among age groups, in Spain coefficients are roughly equivalent, suggesting that significant individual heterogeneity should be taken into account in Italy. **Figure 5** suggests that returns to age seem to converge in both countries.

[Insert Table 3 and Figure 3, 4 and 5]

The results indicate that individual age exerts a quadratic effect on BMI. This quadratic relation implies that individuals' body mass increases as they age until they reach a peak, which differs according to population and quantile (see **Figure 4**), but the median is around 70 years for females in both countries and around 60 for males. This gender gap age effect remains constant in the middle of the BMI distribution, but falls off in the upper tail in both countries. As in other studies, we find that education is the most significant variable in explaining obesity (Chou *et al.*, 2002 and Kan and Tsai, 2004). The plots corresponding to the high education variable (EDU_HIGH) in **Figure 3** show that individuals holding a university degree have a lower body mass than individuals with a medium level of education (coefficients are negative). Individuals with a low level of education (EDU_LOW) have a higher body mass than individuals with a medium level of education (the coefficients are positive). Note that in both countries this "low-education gap" is larger for females than for males. The "high-education gap" decreases as we move up through the BMI distribution, while the “low-education gap” increases. These effects imply that the BMI distribution for the more educated is less dispersed than t for the medium educated, and the BMI distribution for the less educated is more dispersed still. So our results suggest that increasing the educational level exerts a protective effect (which is larger in females), and reduces BMI inequality.

The results on smoking behaviour are puzzling. According to the literature being an active smoker (CURR_SMOKE) correlates negatively with body mass, while being ex-smoker (PAST_SMOKE) increases it. Indeed, our evidence suggests that being an ex-smoker exerts a positive and almost constant impact on body weight for males, in both countries. Therefore increasing the share of ex-smokers among males (in both countries) produces a positive scale impact on the BMI distribution without affecting BMI inequality. Among Italian females, being a ex-smoker exerts a significant and increasingly positive effect only at the upper tail of the distribution. The impact on Spanish females is, on the other hand, negative and constant, despite being significantly different from zero only in the middle of the BMI distribution. Being an active smoker exerts a negative and almost constant effect on the female population; this effect is larger and significant in Spain. This implies that increasing the share of females’ active smokers produces a negative, purely scale, effect on the BMI distribution. In the male population, being an active smoker is never a significant determinant of BMI for Italians, while it changes from negative to zero for Spaniards along the BMI distribution.

Professional status appears to be a significant covariate: for females being EMPLOYED present lower mean BMI, whilst for men the effect is the opposite. In a way this result might appear counterintuitive if other factors are not taken into account. It may be due to the fact that in southern Mediterranean countries men have not taken on the role traditionally played by women in promoting healthy eating, so for men shorter cooking time translates into a higher probability of eating less healthy food. This result is not found among women. However, when looking at the entire BMI distribution our data also reveal that being EMPLOYED exerts a negative impact on body weight for men at the top of the BMI scale, especially in the Spanish population. Furthermore, being MARRIED exerts a positive impact on BMI which fades away in the upper tails, while being INSURED presents hardly any correlation with BMI. Diet habit dummies, BREAKFAST and NEVE_MEAT, are barely significant. Having breakfast vs. not having breakfast proves to be protective against obesity among males.

Regional GDP per capita (GDP_pc) exerts a negative impact on body mass, suggesting that people in affluent regions are leaner and less likely to be obese. This effect is more marked in the higher quantiles of male BMI distribution than in the lower ones, implying that regional prosperity reduces BMI dispersion for males. On the other hand, for the female populations, the negative impact of regional GDP is less important at the tails, so that increasing regional prosperity reduces BMI more in the middle of the distribution than in the tails, thus increasing inequality.

5.2 COUNTERFACTUAL DECOMPOSITION

The next step is to examine cross-country gaps between Italy and Spain in the (log)BMI. As we showed in the descriptive analysis, Spain is more (positively) right skewed than Italy. The Spain-to-Italy gap is positive and significant throughout the distribution and widens clearly towards the upper tail, especially in the case of females. Possible explanations include a right location shift and a more dispersed BMI distribution in Spanish male and female populations. In this section we explore the possible causes of these two features.

[Insert Table 4 and Figures 6a, 6b and 7]

In order to decompose the cross-country gaps in the (log)BMI distribution into gaps attributable to differences in the coefficients (returns to those attributes) and gaps due to differences in the covariates (individual attributes), we follow the procedures described earlier (with the number of replications set at 200). The results are summarized in **Figure 6a** for females and **Figure 6b** for males. In the first panel of both figures we plot the estimated Spain-to-Italy gap for each quantile of the log(BMI) distribution attributable to differences in the QR coefficients (setting the distribution of covariates to that of Spain). In the second panel we plot the estimated Spain-to-Italy gap attributable to differences in the distribution of covariates (assuming that the BMI equilibrium relation is the one prevailing in Italy). In both cases the estimates are plotted along with the 95% confidence band around them and the log(BMI) gap observed. Moreover, we add a horizontal line representing the counterfactual coefficient and covariate effects estimated from the mean regression models.

Our evidence clearly suggests that differences in the distributions of covariates across Spanish and Italian females are responsible for a slight rightward shift in the BMI distribution of the Spanish subjects. In other words, only a small part of the observed gap is due to a relatively more unfavourable distribution of characteristics among Spanish females. However, a larger and increasing part of the cross-country gap is due to the different returns to BMI determinants (the coefficients effect). According to our estimates, the relative importance of this effect is twice that of covariates at the lower tail of BMI distribution, and four times as high at the upper tail (see **Table 4**). If we look at males' distribution the effect arising through the difference in covariates becomes irrelevant and the whole difference, in scale and shape, across the two BMI distributions arises due to the coefficient effect.

Finally, we performed the same set of counterfactual decomposition analyses of the Spain-to-Italy log(BMI) gaps on age groups for both males and females: "the young" (aged 18 to 39 years), "the middle aged" (aged 40 to 59) and "the elderly" (aged 60 to 75). In "the young" log(BMI) Spain-to-Italy the raw gap is rapidly increasing for both sexes; in "the middle aged" it is increasing for females but remains almost constant among males; in "the elderly" it stays relatively constant for both sexes. Among the young and the middle aged the coefficient effects (the counterfactual distribution of log(BMI) gaps due to the difference in returns to characteristics) is positive and increasing, being positive and constant among the elderly. The covariate effects (the counterfactual distribution of log(BMI) gaps due to the difference in the distribution of covariates) is positive and almost constant among females, while it is not different from zero for males, with the notable exception of the young group.

6 CONCLUSION

This paper has addressed the question of cross-country comparison of body mass gaps in gender and age groups in two Mediterranean countries that are subject to different environmental conditions. We applied quantile regression methods and decomposed cross-country BMI gaps using counterfactual decomposition techniques. Surprisingly, even though the two countries examined, Italy and Spain, are presumably subject to similar patterns of health-related behaviour, the Spain-to-Italy BMI gap is large and increases at the upper end of the distribution, especially in females. Moreover, the decomposition analysis suggests that a large, increasing portion of the cross-country gap is due to different returns to BMI determinants or the "coefficient effect". In particular, our estimate indicates that this effect is twice that of covariates at the lower tail of BMI distribution, rising to four times as high at the upper tail. This is the case for the young and middle age group. Summarizing, Spaniards are on average heavier due to worse returns to environmental or behavioural factors rather than due to greater use of unhealthy health inputs.

Among the factors that explain the concentration of BMI at the upper tail of the distribution the role of education stands out, as some other studies have found (Baum II, and Ruhm, 2007). Indeed, education arguably confers both ability and informational effects that may influence the extent to which individuals exercise and feed themselves to maintain the caloric balance responsible for adiposity. Another important contribution is the heterogeneity of age effects on BMI, suggesting that generation-specific behavioural factors may underpin feeding patterns. Finally, smoking does not appear to explain differences in body mass; contrary to the suggestions of some authors, differences in exposure to smoking due to measures such as bans – which are stricter in Italy than in Spain – are not likely to explain differences in obesity. This finding questions some evidence that suggests that obesity could be the result of smoking bans, as it does not seem to be the underlying variable responsible for BMI gaps.

Our interpretation of these results is backed by some anecdotal evidence that is worth mentioning. Some Eurobarometer data in 2005 suggests that Italy comes first among people worried about putting on weight (62%) and Spain ranks second (50%) of EU-15. Interestingly, another Eurobarometer survey (2006) asked whether their body weight was too high; 35% of Italians and only 29% of Spaniards respond affirmatively, despite the fact that Italians are on average less obese.

Our results have certain policy implications. Cross-country differences in body mass seem to be mainly due to country-specific responses to health inputs rather than to health inputs distribution. Consequently, a similar effect on the availability of health inputs such as a tax or a barrier on the consumption of healthy food will lead to marked differences across countries as a result of the different environmental, behavioural or cultural responses, even in countries with similar access to healthy (Mediterranean) diets.

7 REFERENCES

Albrecht J, Bjorklund A, Vroman S. 2003. Is there a glass ceiling in Sweden?, *Journal of Labor Economics* **21**(1): 145-177.

Arulampalam W, Booth A L, Bryan M L. 2005. Is there a glass ceiling over Europe?. Exploring the gender pay gap across the wages distribution, ISER Working Paper 2005-25. Colchester: University of Essex.

Bauer T, Göhlmann S, Sinning M. 2006. Gender differences in smoking behavior. IZA Discussion Paper 2259. Bonn, Germany: Institute for the Study of Labor (IZA).

Bauer T, Sinning M. 2005. Blinder-Oaxaca decomposition for Tobit models. IZA Discussion Paper 1795. Bonn, Germany: Institute for the Study of Labor (IZA).

Baum II C L, Ruhm C J. 2007. Age, socioeconomic status and obesity growth. NBER Working Paper Series, 13289.

Blinder A S. 1973. Wage discrimination: reduced form and structural variables. *Journal of Human Resources* **8**: 436-455.

Branca F, Nikogosian H, Lobstein T. 2007. The challenge of obesity in the WHO European Region and the strategies for response. World Health Organization 2007.

Buchinsky M. 1997. The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics* **13**(1): 1–30.

Buchinsky M. 1998. Recent advances in quantile regression models. *Journal of Human Resources* **33**: 88-126.

Burkhauser R V, Cawley J. 2008. Beyond the BMI: the value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics* **17** (forth.).

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

Chou S Y, Grossman M., Saffer H. 2004. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics* **23**: 565-587.

Cohen D A, Finch B K, Bower A, Sastry N. 2006. Collective efficiency and obesity: the potential influence of social factors on health. *Social Science and Medicine* **62**: 769-778.

Commission of the European Communities. 2005. Green Paper "Promoting healthy diets and physical activity: a European dimension for the prevention of overweight, obesity and chronic diseases", COM(2005) 637 final. Available at: http://ec.europa.eu/health/ph_determinants/life_style/nutrition/documents/nutrition_gp_en.pdf

Contoyannis P, Wildman J. 2007. Using relative distributions to investigate the body mass index in England and Canada. *Health Economics* **16**: 929-944.

Costa-Font J, Gil J. 2008. What lies behind socio-economic inequalities in obesity in Spain? A decomposition approach. *Food Policy* **33**: 61-73.

Cutler D, Glaeser E, Shapiro J. 2003. Why have Americans become more obese? *Journal of Economic Perspectives* **17**(3): 93–118.

de la Rica S, Dolado J J, Llorens V. 2005. Ceiling and floors: gender wage gaps by education in Spain, IZA DP No. 1483, Bonn, Germany.

Fairlie R W. 2005. An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement* **30**(4): 305-316.

Finkelstein E, Ruhm C, Kosa K. 2005. Economic causes and consequences of obesity. *Annual Review of Public Health* **26**: 239–257.

Flegal, K M. 2006. Body mass index of healthy men compared with healthy women in the United States. *International Journal of Obesity* **30**: 374-379.

Freedman D S, Kahn L K, Serdula M K, Galuska D A, Doetz, W H. 2000. Trends and correlates of class 3 obesity in the United States from 1900 through 2000. *Journal of the American Medical Association* **288**: 1758-1762.

French S A, Story M, Jeffery R W. 2001. Environmental influences on eating and physical activity. *Annual Review of Public Health* **22**: 309-335.

Gallagher D, Visser M, Sepulveda D, Pierson R N, Harris T, Heymsfield S B. 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.

Garcia J, Hernández P J, López, A. 2001. How wide is the gap? An investigation of gender wage differences using quantile regression. *Empirical Economics* **26**: 149.167.

Gregg E W, Guralnik J M. 2007. Is disability obesity's price of longevity?. *Journal of the American Medical Association* **298**(17): 2066-2067.

Hadi A S. 1992. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)* **54**(3): 761-771.

Hendricks W, Koenker R. 1992. Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association* **87**(417): 58-68.

Jones A M, López Nicolás A. 2002. The importance of individual heterogeneity in the decomposition of measures of socioeconomic inequality in health: an approach based on quantile regression. UPF Economics and Business Working Paper No. 626.

Kan K, Tsai W D. 2004. Obesity and risk knowledge. *Journal of Health Economics* **23**: 907-934

Koenker R, Basset G. 1978. Regression quantiles. *Econometrica* **46**(1): 33-50.

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

Lakdawalla D, Philipson T. 2002. The growth of obesity and technological change: a theoretical and empirical examination. NBER Working Paper Series 8946.

Machado J A F, Mata J. 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics* **20**: 445-465.

Michaud P C, Van Soest A, Andreyeva T. 2007. Cross-country variation in obesity patterns among older Americans and Europeans. *Forum for Health Economics & Policy* 10(2) (Aging and Medical Care Costs), Article 8. The Berkeley Electronic Press.

Nayga R 2001. Effect of schooling on obesity: Is health knowledge a moderating factor?, *Education Economics* **2**: 129-137.

Oaxaca R. 1973. Male-female wage differentials in urban labor markets. *International Economic Review* **14**, 693-709.

Pagano R, La Vecchia C. 1994. Overweight and obesity in Italy, 1990-91. *International Journal of Obesity Related Metabolic Disorders* **18**(10): 665-669.

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

Ruhm C J. 2000. Are recessions good for your health?. *Quarterly Journal of Economics* **115**(2): 617-650.

Ruhm C J. 2007. Current and future prevalence of obesity and severe obesity in the United States. *Forum for Health Economics & Policy* **10**(2) (Obesity), Article 6. The Berkeley Electronic Press

Sanz de Galdeano A. 2005. The obesity epidemic in Europe. IZA Discussion Paper, 1814.

Silvertonen K, Sans S, Tolonen H, Monterde D, Kuulasmaa K, Kasteloot H, et al. 2004. Trends in obesity and energy supply in the WHO MONICA Project. *International Journal of Obesity and Related Metabolic Disorders* **28**(5) 710-718.

World Health Organization-WHO. 1997. Obesity: Preventing and managing the global epidemic. Report of a WHO consultation on obesity. Geneva, Switzerland: World Health Organization.

World Health Organization-WHO. 2003. Global Strategy on Diet, Physical Activity and Health. Geneva, Switzerland: World Health Organization.

TABLES AND FIGURES

Table 1: BMI distribution across Italy and Spain

	MEAN	Q5	Q10	Q25	Q50	Q75	Q90	Q95
ALL								
ITALY	24.7	19.0	20.1	22.0	24.3	26.9	29.4	31.3
SPAIN	25.7	19.6	20.7	22.8	25.3	28.1	31.2	33.2
DIFFERENCE	1.05	0.52	0.64	0.73	1.01	1.23	1.79	1.98
t-value	29.9	14.9	12.9	18.8	17.2	16.8	37.2	27.5
RELATIVE DIFFERENCE	0.043	0.028	0.032	0.033	0.041	0.046	0.061	0.063
FEMALES								
ITALY	23.9	18.4	19.3	20.9	23.4	26.2	29.4	31.3
SPAIN	25.3	19.0	20.0	21.9	24.6	27.9	31.3	33.8
DIFFERENCE	1.38	0.65	0.69	1.02	1.23	1.74	1.87	2.53
t-value	26.3	14.2	9.5	14.2	18.2	18.5	28.1	16.0
RELATIVE DIFFERENCE	0.058	0.036	0.036	0.049	0.052	0.067	0.063	0.081
MALES								
ITALY	25.5	20.7	21.6	23.2	25.1	27.5	29.7	31.4
SPAIN	26.2	20.9	22.0	23.8	26.0	28.4	31.0	32.7
DIFFERENCE	0.75	0.20	0.43	0.61	0.85	0.91	1.27	1.33
t-value	17.0	2.0	7.9	8.1	14.7	18.1	10.5	8.2
RELATIVE DIFFERENCE	0.030	0.010	0.020	0.026	0.034	0.033	0.043	0.043

Note: t-values for the differences are from simultaneous quantile regressions with bootstrapped (1000 replications) standard errors.

Figure 1: Gender specific BMI fraction histograms

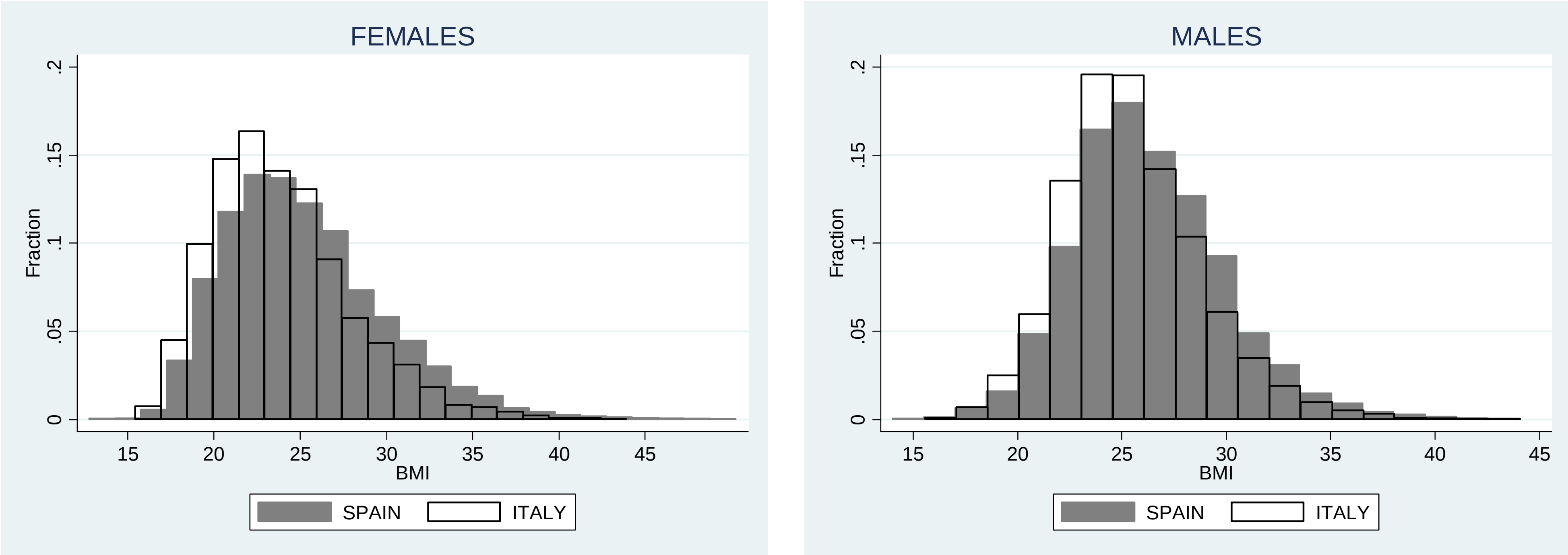


Figure 1: Gender specific BMI fraction histograms by age groups

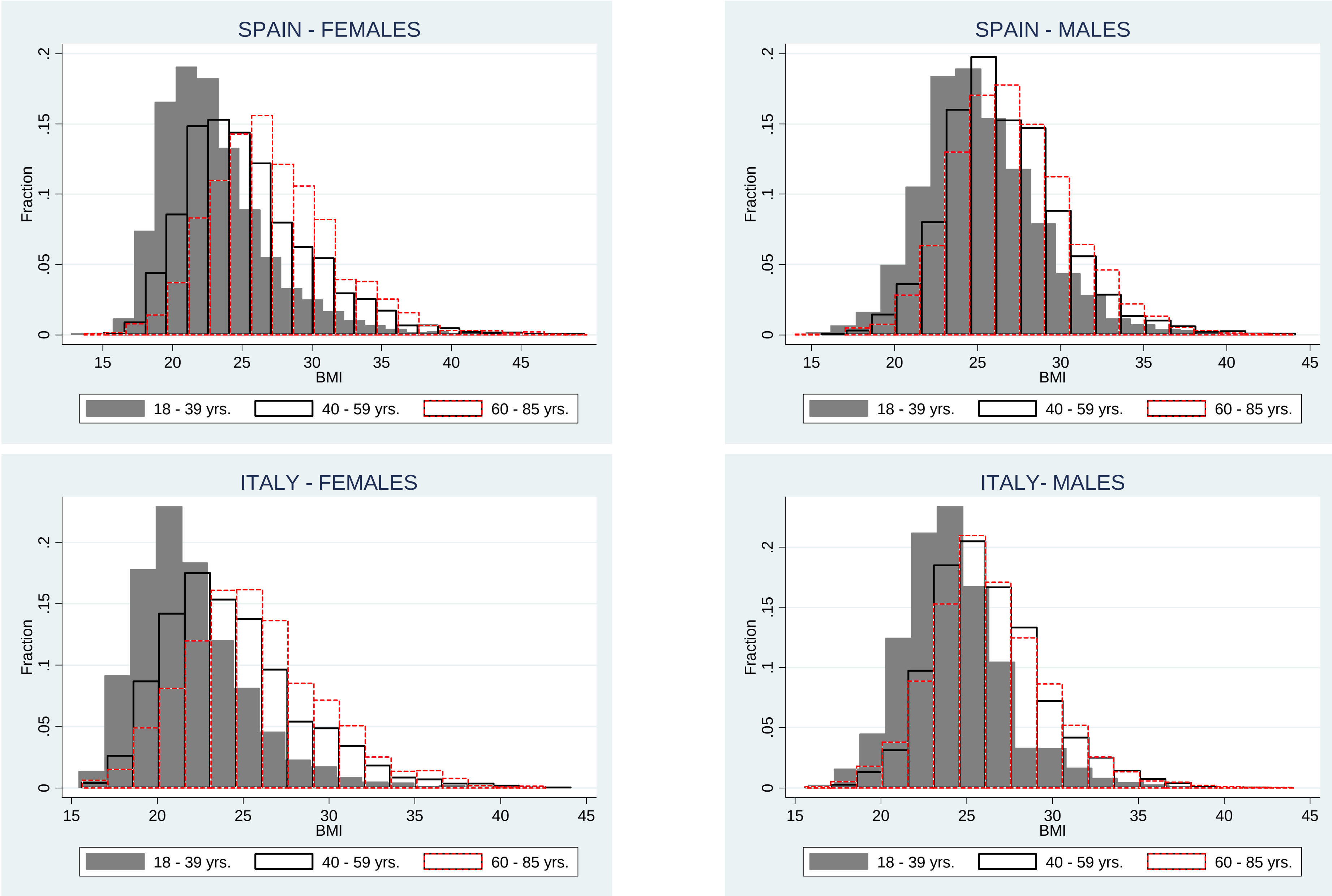


Table 2. Variable definitions, sample means and standard deviations in parentheses

VARIABLE	DESCRIPTION	SPAIN Female	SPAIN Male	ITALY Female	ITALY Male
	N° of observations	11025	9296	20572	19192
Log BMI	Log of Body Mass Index	3.21	3.26	3.16	3.23
		(0.18)	(0.14)	(0.16)	(0.13)
AGE	Age of the interviewed individual	51.27	48.02	48.58	46.98
		(18.53)	(17.49)	(17.78)	(17.12)
AGE_SQ	Square of age /100	29.71	26.12	26.76	25.00
		(19.34)	(17.80)	(18.05)	(16.93)
EDU_HIGH	=1 if university education; =0 otherwise	0.15	0.14	0.08	0.08
		(0.35)	(0.35)	(0.27)	(0.27)
EDU_LOW	=1 if primary or lower education; =0 otherwise	0.53	0.46	0.33	0.24
		(0.35)	(0.50)	(0.47)	(0.43)
MARRIED	=1 if married; =0 otherwise	0.64	0.76	0.59	0.62
		(0.48)	(0.43)	(0.49)	(0.48)
EMPLOYED	=1 if employed; 0 otherwise	0.35	0.61	0.35	0.60
		(0.48)	(0.49)	(0.48)	(0.49)
WORK_HARD	=1 if the employed has a "hard work"*; =0 otherwise	0.06	0.21	0.07	0.18
		(0.23)	(0.41)	(0.25)	(0.39)
CURR_SMOKE	=1 if current smoker; =0 otherwise	0.22	0.37	0.17	0.32
		(0.41)	(0.48)	(0.38)	(0.47)
PAST_SMOKE	=1 if quitted smoking; =0 otherwise	0.11	0.28	0.15	0.31
		(0.31)	(0.45)	(0.36)	(0.46)
INSURED	=1 if owner of a private health insurance; 0 otherwise	0.10	0.09	0.14	0.22
		(0.30)	(0.29)	(0.35)	(0.41)
BREAKFAST	=1 if habitual breakfast; =0 otherwise	0.97	0.92	0.94	0.90
		(0.17)	(0.26)	(0.24)	(0.30)
NEV_MEAT	=1 if eats meat less than once a week or never; =0 otherwise	0.06	0.03	0.08	0.07
		(0.23)	(0.17)	(0.27)	(0.25)
GDP_PC	Regional GDP** per capita Purchasing Power Standard / 1000 EU€	20.42	20.45	22.02	22.06
		(3.76)	(3.77)	(5.64)	(5.65)

Note: The reported means refer to each country sub-sample of adults aged 18-85 (Italy N=39,764; Spain N=20,321).

* "Hard work" is a work that implies considerable physical exertion;

** Regional GDP is defined for 20 Italian regions and 18 Spanish Autonomous Communities. This macro regional variable comes from EUROSTAT database.

Source: “Encuesta Nacional de Salud 2003” (MSC) for Spain and “Indagine sugli Aspetti della Vita Quotidiana 2003” (ISTAT) for Italy.

Table 3. The OLS determinants of Log(BMI)

	SPAIN Females	SPAIN Males	ITALY Females	ITALY Males
AGE	0.0116***	0.0104***	0.0112***	0.0099***
	0.0005	*0.0005*	*0.0004*	*0.0004*
AGE_SQ	-0.0089***	-0.0087***	-0.0086***	-0.0084***
	0.0005	*0.0005*	*0.0004*	*0.0004*
EDU_HIGH	-0.0419***	-0.0263***	-0.0406***	-0.0220***
	0.0043	*0.0039*	*0.0036*	*0.0030*
EDU_LOW	0.0501***	0.0081**	0.0492***	0.0172***
	0.0042	*0.0034*	*0.0030*	*0.0026*
MARRIED	0.0068**	0.0084***	0.0102***	0.0158***
	0.0034	*0.0032*	*0.0025*	*0.0023*
EMPLOYED	-0.0131***	0.0091**	-0.0186***	0.0075***
	0.0039	*0.0041*	*0.0026*	*0.0024*
WORK_HARD	0.0047	-0.0063*	0.0157***	0.0025
	0.0066	*0.0036*	*0.0043*	*0.0024*
CURR_SMOKE	-0.0338***	-0.0135***	-0.0123***	-0.0031
	0.0040	*0.0032*	*0.0028*	*0.0021*
PAST_SMOKE	-0.0174***	0.0146***	0.0113***	0.0140***
	0.0049	*0.0034*	*0.0030*	*0.0022*
INSURED	-0.0248***	0.0087*	-0.0095***	0.0018
	0.0048	*0.0046*	*0.0030*	*0.0021*
BREAKFAST	-0.0329***	-0.0260***	-0.0082**	-0.0220***
	0.0091	*0.0054*	*0.0043*	*0.0030*
NEV_MEAT	-0.0024	-0.0194**	-0.0060	-0.0062*
	0.0067	*0.0085*	*0.0037*	*0.0032*
GDP_PC	-0.0030***	-0.0018***	-0.0021***	-0.0014***
	0.0004	*0.0004*	*0.0002*	*0.0002*
CONSTANT	2.9701***	3.0359***	2.8920***	3.0068***
	0.0166	*0.0142*	*0.0098*	*0.0084*
R^2	0.2325	0.1190	0.2144	0.1477

Note: Standard errors (in *italics*) are computed using the Huber/White variance estimator.
***, **. * denotes significance levels at 1%, 5% and 10% respectively.

Figure 3: Log(BMI) returns to characteristics

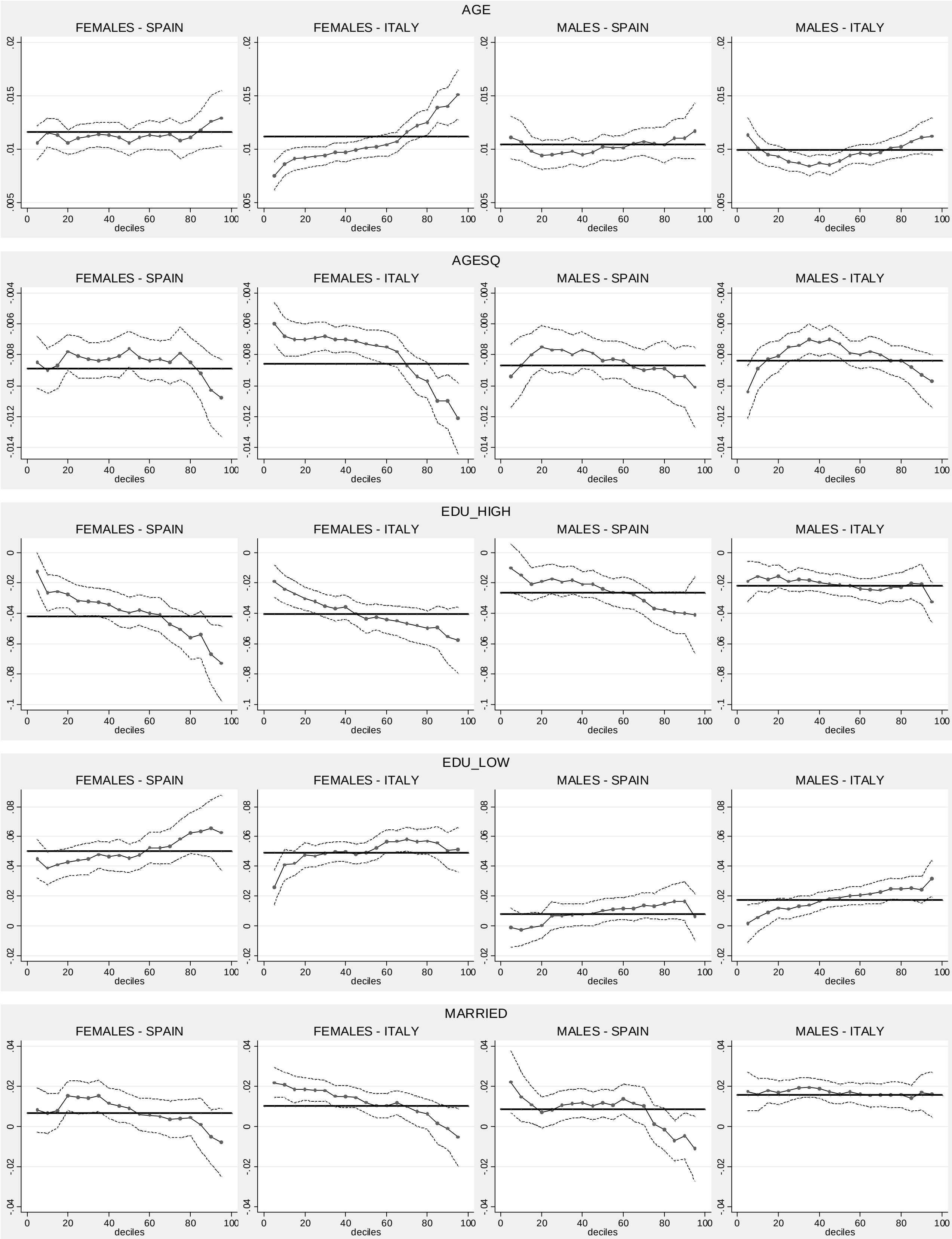


Figure 3: Log(BMI) returns to characteristics: continued

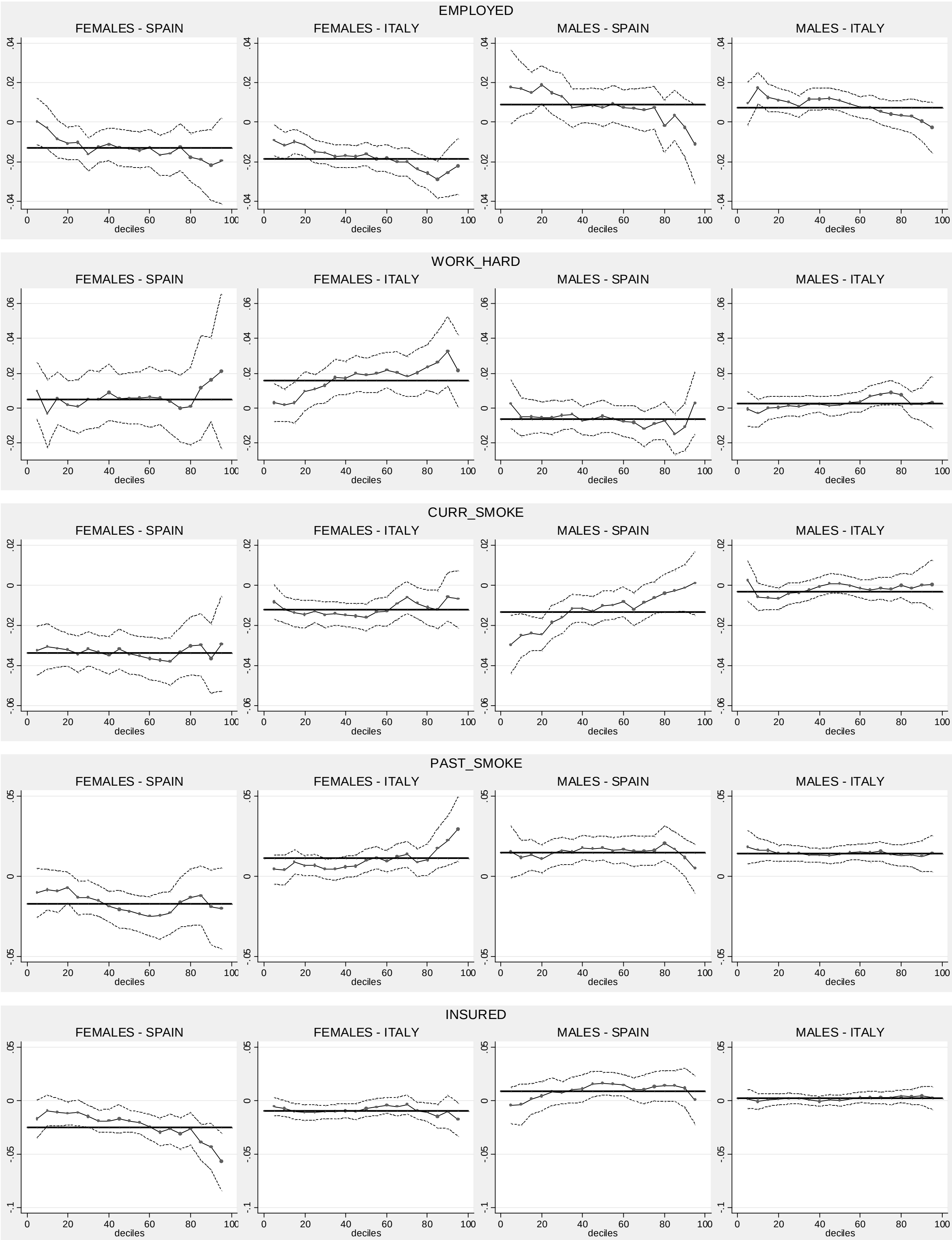


Figure 3: Log(BMI) returns to characteristics: continued

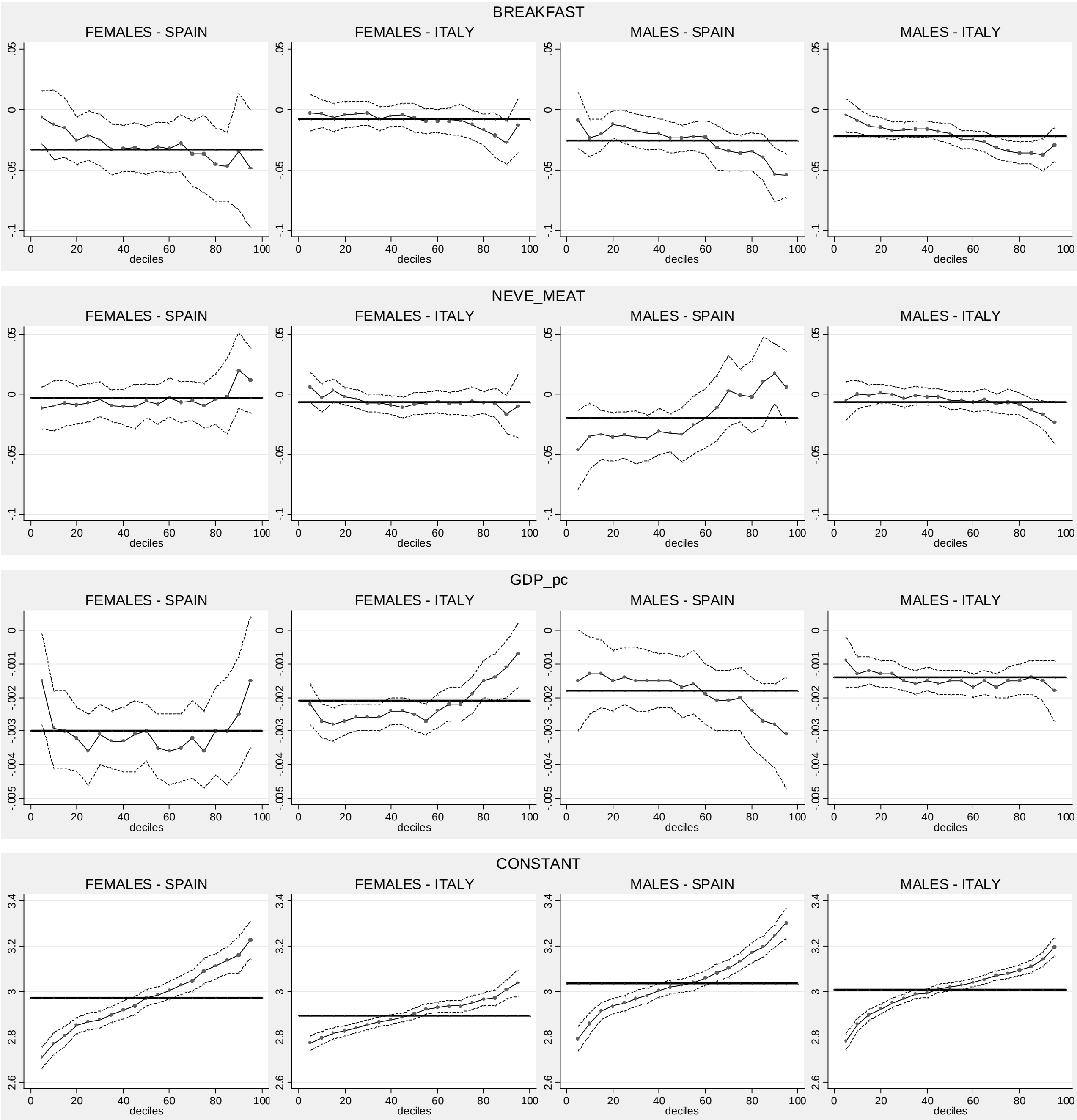


Figure 4: BMI returns to AGE at the mean and the tails

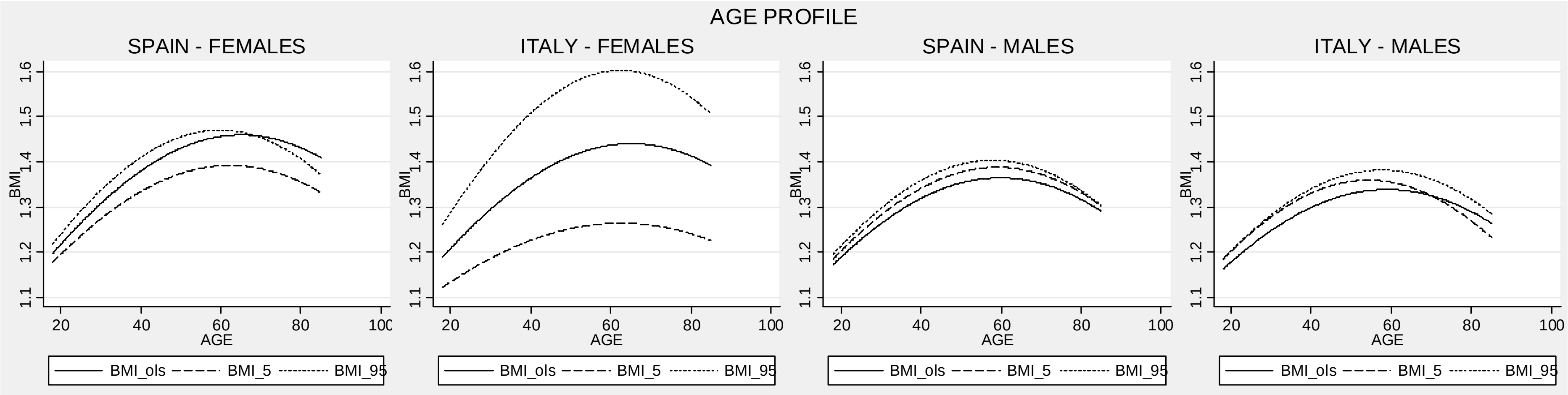


Figure 5: "Tipping AGES" at the BMI quantiles

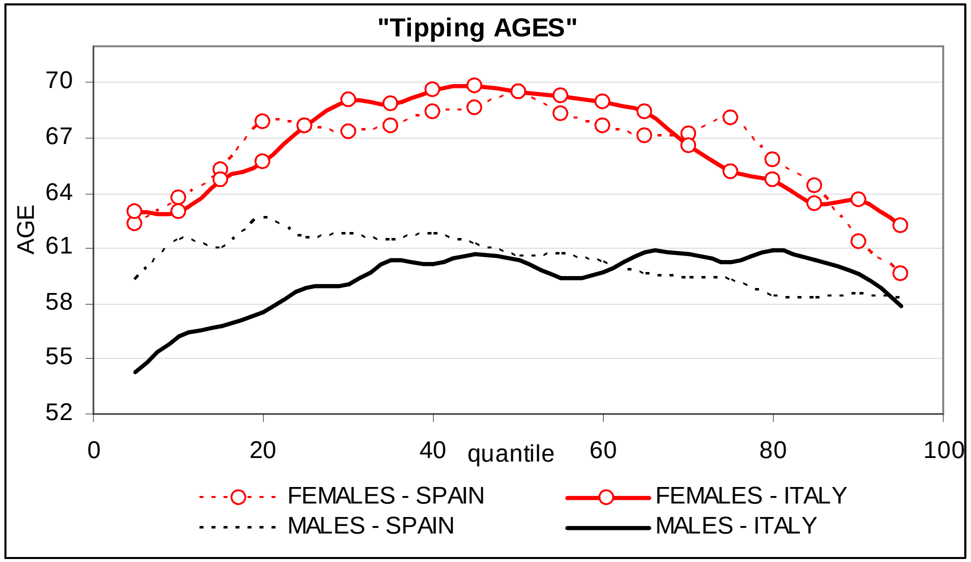


Figure 6a: SPAIN-to-ITALY log(BMI) GAP counterfactual decomposition analysis: FEMALES

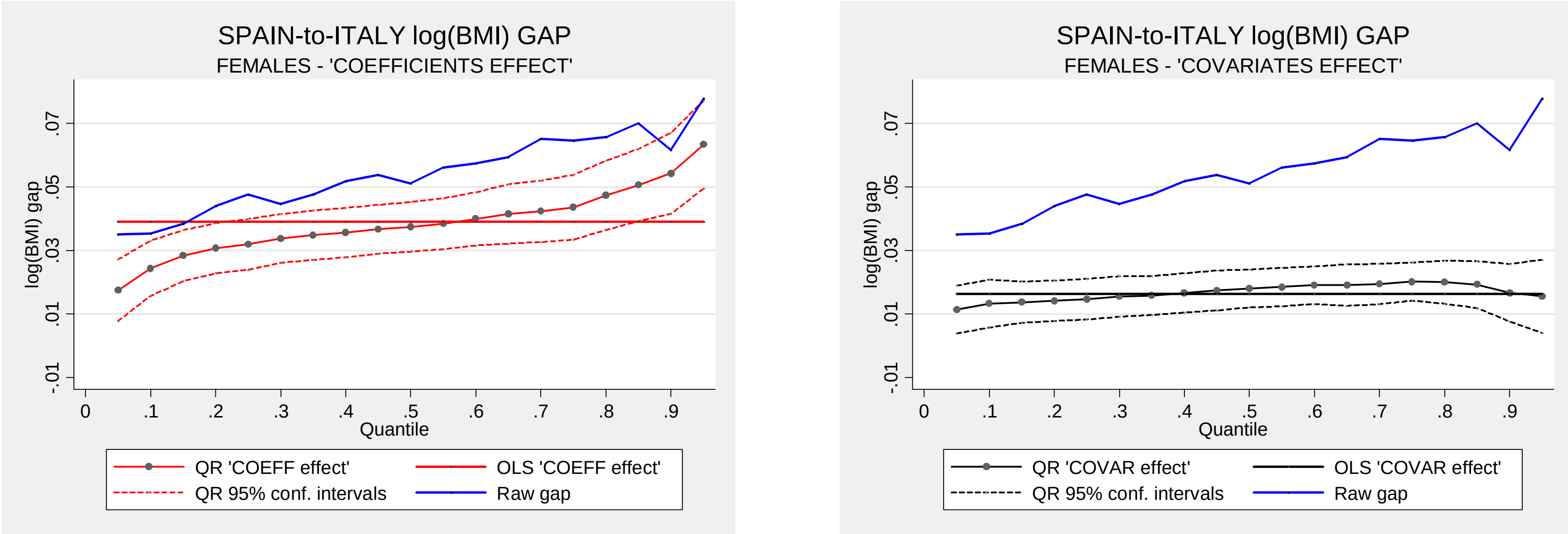


Figure 6b: SPAIN-to-ITALY log(BMI) GAP counterfactual decomposition analysis: MALES

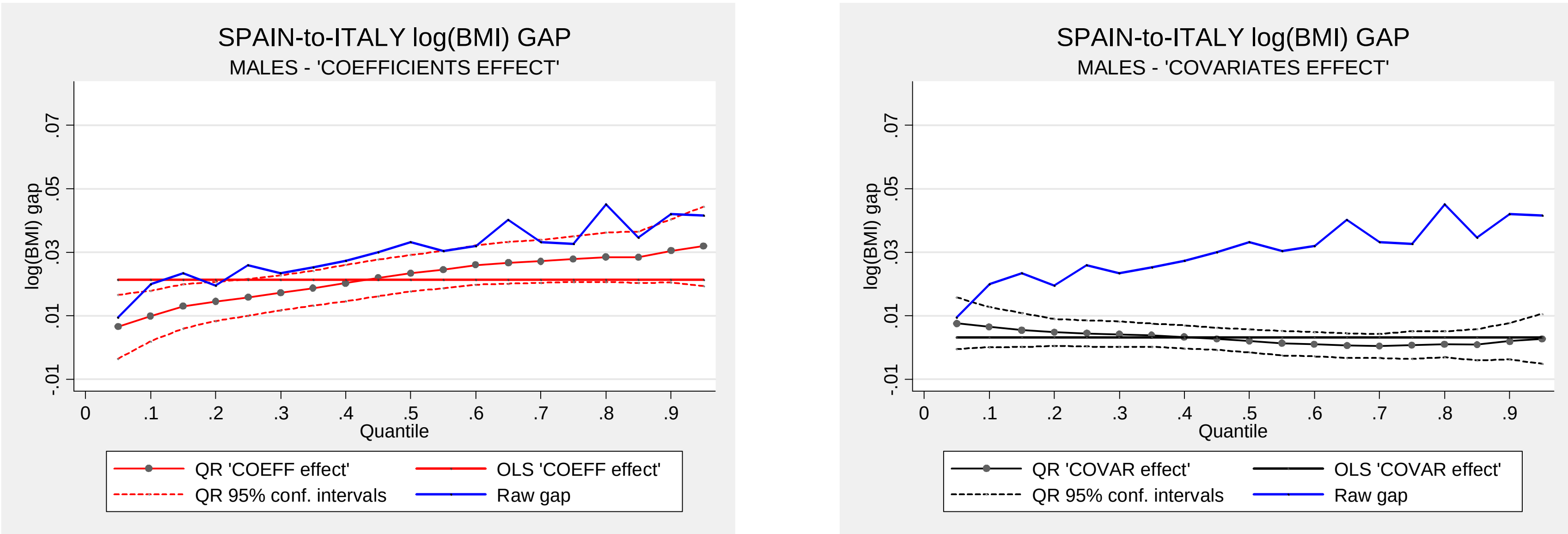
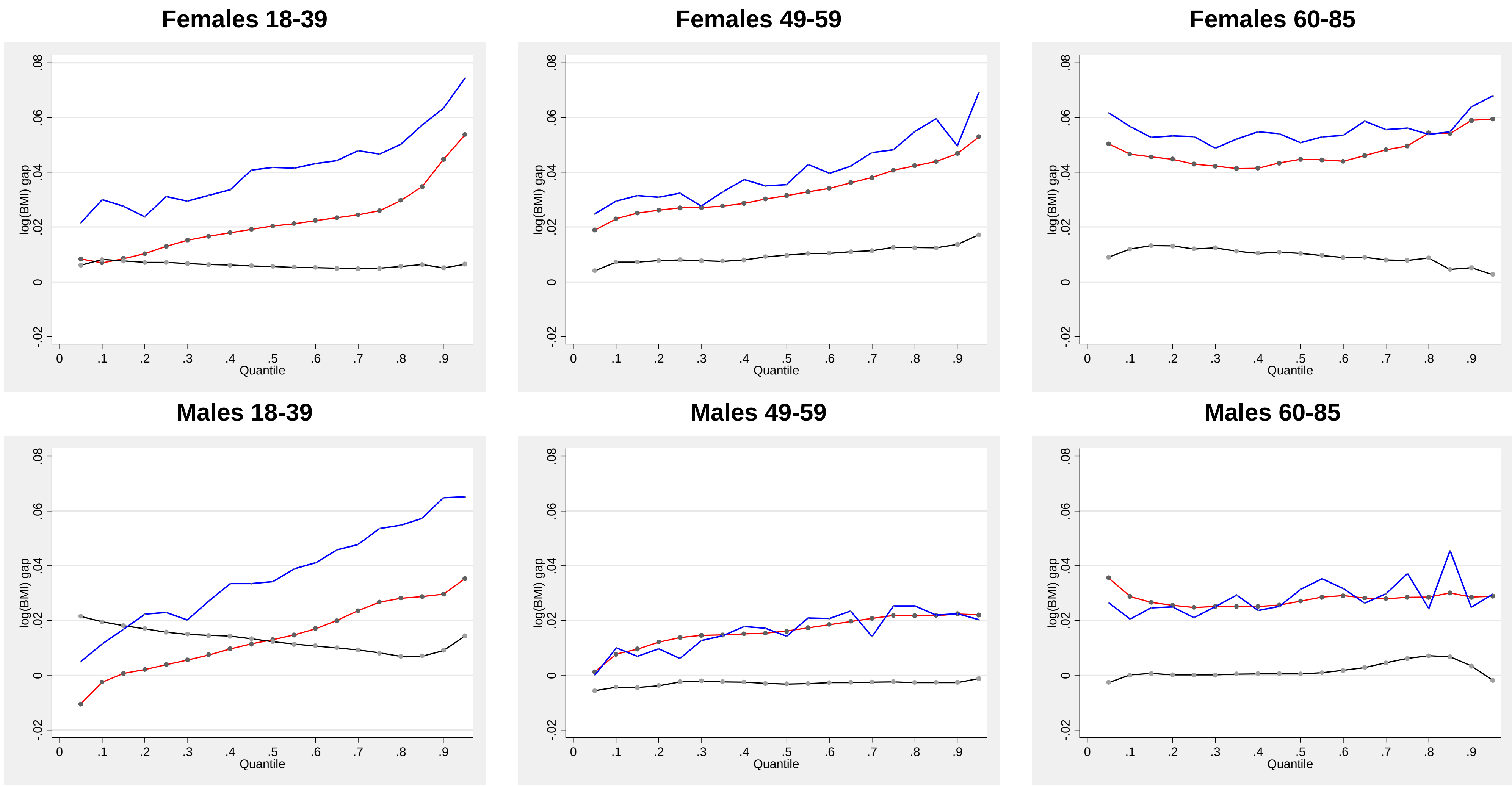


Table 4: SPAIN-to-ITALY log(BMI) GAP counterfactual decomposition

	RAW GAP	% explained by COEFFICIENTS	% explained by COVARIATES	RESIDUAL
		FEMALES		
MEAN	0.054	72.79	30.41	-3.21
Q_5	0.035	49.84	32.53	17.62
Q_{10}	0.035	69.08	37.62	-6.70
Q_{25}	0.048	67.06	30.80	2.14
Q_{50}	0.051	73.32	35.33	-8.66
Q_{75}	0.064	67.62	31.30	1.08
Q_{90}	0.062	88.19	27.12	-15.32
Q_{95}	0.078	81.37	20.04	-1.42
		MALES		
MEAN	0.028	75.905	11.326	12.77
Q_5	0.009	69.35	80.60	-49.95
Q_{10}	0.020	49.95	32.59	17.45
Q_{25}	0.026	60.77	17.16	22.07
Q_{50}	0.033	70.40	6.35	23.25
Q_{75}	0.033	85.42	2.44	12.14
Q_{90}	0.042	72.56	4.79	22.65
Q_{95}	0.042	76.66	6.61	16.73

Figure 7: SPAIN-to-ITALY log(BMI) GAP counterfactual decomposition analysis: by AGE CLASS and GENDER



APPENDIX

Table A1. Quantile Regression estimates of (Log)BMI: SPAIN - Females

	OLS	Q5	Q10	Q25	Q50	Q75	Q90	Q95
AGE	0.0116***	0.0106***	0.0115***	0.0110***	0.0106***	0.0108***	0.0126***	0.0129***
	0.0005	0.0008	0.0007	0.0007	0.0006	0.0008	0.0013	0.0013
AGE_SQ	-0.0089***	-0.0085***	-0.0090***	-0.0081***	-0.0076***	-0.0079***	-0.0103***	-0.0108***
	0.0005	0.0009	0.0007	0.0007	0.0006	0.0009	0.0012	0.0013
EDU_HIGH	-0.0419***	-0.0125**	-0.0267***	-0.0319***	-0.0397***	-0.0506***	-0.0670***	-0.0729***
	0.0043	0.0062	0.0061	0.0052	0.0052	0.0063	0.0100	0.0126
EDU_LOW	0.0501***	0.0451***	0.0388***	0.0441***	0.0454***	0.0584***	0.0654***	0.0626***
	0.0042	0.0066	0.0057	0.0051	0.0048	0.0065	0.0098	0.0128
MARRIED	0.0068**	0.0082	0.0065	0.0145***	0.0089**	0.0038	-0.0052	-0.0079
	0.0034	0.0057	0.0051	0.0042	0.0037	0.0048	0.0069	0.0086
EMPLOYED	-0.0131***	0.0004	-0.0028	-0.0102**	-0.0133***	-0.0125**	-0.0217**	-0.0195*
	0.0039	0.0060	0.0055	0.0044	0.0046	0.0061	0.0091	0.0110
WORK_HARD	0.0047	0.0097	-0.0031	0.0009	0.0056	-0.0002	0.0162	0.0212
	0.0066	0.0083	0.0099	0.0078	0.0075	0.0097	0.0123	0.0225
CURR_SMOKE	-0.0338***	-0.0325***	-0.0306***	-0.0343***	-0.0343***	-0.0334***	-0.0365***	-0.0293**
	0.0040	0.0062	0.0058	0.0046	0.0051	0.0064	0.0089	0.0120
PAST_SMOKE	-0.0174***	-0.0105	-0.0084	-0.0134**	-0.0218***	-0.0163**	-0.0193	-0.0202
	0.0049	0.0078	0.0064	0.0054	0.0056	0.0079	0.0118	0.0129
INSURED	-0.0248***	-0.0169*	-0.0093	-0.0114*	-0.0192***	-0.0309***	-0.0432***	-0.0567***
	0.0048	0.0090	0.0073	0.0062	0.0052	0.0074	0.0110	0.0137
BREAKFAST	-0.0329***	-0.0067	-0.0128	-0.0217**	-0.0338***	-0.0369**	-0.0347	-0.0489**
	0.0091	0.0112	0.0146	0.0105	0.0101	0.0163	0.0246	0.0248
NEV_MEAT	-0.0024	-0.0113	-0.0096	-0.0068	-0.0053	-0.0091	0.0202	0.0121
	0.0067	0.0088	0.0106	0.0082	0.0073	0.0096	0.0161	0.0139
GDP_PC	-0.0030***	-0.0015**	-0.0029***	-0.0036***	-0.0030***	-0.0036***	-0.0025***	-0.0015
	0.0004	0.0007	0.0006	0.0005	0.0004	0.0006	0.0008	0.0010
CONSTANT	2.9701***	2.7096***	2.7696***	2.8680***	2.9718***	3.0901***	3.1614***	3.2267***
	0.0166	0.0240	0.0243	0.0190	0.0185	0.0288	0.0411	0.0417
Pseudo R2	0.2325	0.1036	0.1234	0.1461	0.1492	0.1288	0.1013	0.0902

Note: Quantile regression estimates with bootstrapped standard errors. Replications set to 1500. ***, **. * denotes significance levels at 1%, 5% and 10% respectively.

Table A2. Quantile Regression estimates of (Log)BMI: SPAIN - Males

	OLS	Q5	Q10	Q25	Q50	Q75	Q90	Q95
AGE	0.0104***	0.0111***	0.0107***	0.0095***	0.0102***	0.0105***	0.0110***	0.0117***
	0.0005	0.0010	0.0009	0.0007	0.0006	0.0007	0.0010	0.0013
AGE_SQ	-0.0087***	-0.0094***	-0.0087***	-0.0077***	-0.0084***	-0.0089***	-0.0094***	-0.0101***
	0.0005	0.0011	0.0010	0.0007	0.0006	0.0008	0.0010	0.0013
EDU_HIGH	-0.0263***	-0.0103	-0.0149**	-0.0174***	-0.0238***	-0.0368***	-0.0399***	-0.0410***
	0.0039	0.0081	0.0070	0.0050	0.0044	0.0051	0.0070	0.0128
EDU_LOW	0.0081**	-0.0012	-0.0026	0.0068	0.0103***	0.0133***	0.0165**	0.0062
	0.0034	0.0067	0.0054	0.0048	0.0040	0.0044	0.0066	0.0080
MARRIED	0.0084***	0.0221***	0.0147**	0.0083**	0.0116***	0.0011	-0.0048	-0.0111
	0.0032	0.0079	0.0062	0.0039	0.0036	0.0048	0.0058	0.0083
EMPLOYED	0.0091**	0.0178*	0.0168**	0.0150***	0.0073	0.0073	-0.0027	-0.0109
	0.0041	0.0094	0.0068	0.0055	0.0047	0.0055	0.0075	0.0101
WORK_HARD	-0.0063*	0.0024	-0.0052	-0.0054	-0.0047	-0.0089*	-0.0109	0.0029
	0.0036	0.0071	0.0056	0.0050	0.0047	0.0047	0.0069	0.0092
CURR_SMOKE	-0.0135***	-0.0296***	-0.0250***	-0.0184***	-0.0102***	-0.0063	-0.0014	0.0010
	0.0032	0.0074	0.0055	0.0043	0.0038	0.0040	0.0059	0.0080
PAST_SMOKE	0.0146***	0.0152*	0.0116**	0.0143***	0.0175***	0.0160***	0.0116*	0.0048
	0.0034	0.0082	0.0055	0.0044	0.0038	0.0046	0.0060	0.0078
INSURED	0.0087*	-0.0046	-0.0038	0.0083	0.0159***	0.0132*	0.0119	0.0008
	0.0046	0.0086	0.0099	0.0065	0.0055	0.0069	0.0093	0.0116
BREAKFAST	-0.0260***	-0.0091	-0.0238***	-0.0144**	-0.0238***	-0.0360***	-0.0541***	-0.0547***
	0.0054	0.0118	0.0078	0.0070	0.0054	0.0075	0.0113	0.0093
NEV_MEAT	-0.0194**	-0.0462***	-0.0349**	-0.0337***	-0.0334***	-0.0006	0.0175	0.0064
	0.0085	0.0170	0.0141	0.0098	0.0116	0.0113	0.0127	0.0155
GDP_PC	-0.0018***	-0.0015**	-0.0013**	-0.0014***	-0.0017***	-0.0020***	-0.0028***	-0.0031***
	0.0004	0.0008	0.0006	0.0005	0.0005	0.0005	0.0006	0.0009
CONSTANT	3.0359***	2.7910***	2.8577***	2.9476***	3.0255***	3.1316***	3.2450***	3.3003***
	0.0142	0.0284	0.0247	0.0174	0.0154	0.0195	0.0257	0.0345
Pseudo R2	0.1190	0.0810	0.0787	0.0728	0.0679	0.0593	0.0535	0.0469

Note: Quantile regression estimates with bootstrapped standard errors. Replications set to 1500. ***, **. * denotes significance levels at 1%, 5% and 10% respectively.

Table A3. Quantile Regression estimates of (Log)BMI: ITALY - Females

	OLS	Q5	Q10	Q25	Q50	Q75	Q90	Q95
AGE	0.0112***	0.0075***	0.0086***	0.0093***	0.0101***	0.0122***	0.0140***	0.0151***
	0.0004	0.0007	0.0006	0.0005	0.0004	0.0006	0.0009	0.0012
AGE_SQ	-0.0086***	-0.0060***	-0.0068***	-0.0069***	-0.0073***	-0.0094***	-0.0110***	-0.0121***
	0.0004	0.0007	0.0006	0.0005	0.0005	0.0006	0.0009	0.0012
EDU_HIGH	-0.0406***	-0.0191***	-0.0242***	-0.0324***	-0.0437***	-0.0481***	-0.0553***	-0.0575***
	0.0036	0.0055	0.0047	0.0038	0.0048	0.0059	0.0091	0.0111
EDU_LOW	0.0492***	0.0260***	0.0410***	0.0468***	0.0492***	0.0565***	0.0505***	0.0512***
	0.0030	0.0059	0.0052	0.0037	0.0035	0.0042	0.0062	0.0076
MARRIED	0.0102***	0.0217***	0.0207***	0.0181***	0.0119***	0.0075*	-0.0011	-0.0053
	0.0025	0.0038	0.0032	0.0028	0.0028	0.0038	0.0054	0.0073
EMPLOYED	-0.0186***	-0.0092**	-0.0117***	-0.0148***	-0.0160***	-0.0236***	-0.0254***	-0.0222***
	0.0026	0.0040	0.0034	0.0029	0.0030	0.0040	0.0062	0.0072
WORK_HARD	0.0157***	0.0030	0.0017	0.0107**	0.0189***	0.0202***	0.0324***	0.0217**
	0.0043	0.0055	0.0047	0.0043	0.0050	0.0070	0.0101	0.0106
CURR_SMOKE	-0.0123***	-0.0084*	-0.0122***	-0.0131***	-0.0160***	-0.0089**	-0.0058	-0.0069
	0.0028	0.0044	0.0034	0.0028	0.0034	0.0040	0.0062	0.0072
PAST_SMOKE	0.0113***	0.0042	0.0037	0.0069**	0.0098***	0.0086*	0.0222***	0.0294***
	0.0030	0.0046	0.0048	0.0034	0.0037	0.0044	0.0079	0.0102
INSURED	-0.0095***	-0.0057	-0.0073*	-0.0111***	-0.0072*	-0.0093**	-0.0107	-0.0175**
	0.0030	0.0043	0.0038	0.0036	0.0037	0.0040	0.0079	0.0079
BREAKFAST	-0.0082**	-0.0029	-0.0035	-0.0037	-0.0071	-0.0127**	-0.0277***	-0.0133
	0.0043	0.0077	0.0059	0.0052	0.0061	0.0062	0.0092	0.0114
NEV_MEAT	-0.0060	0.0061	-0.0027	-0.0036	-0.0076	-0.0058	-0.0163**	-0.0099
	0.0037	0.0063	0.0061	0.0039	0.0047	0.0062	0.0081	0.0135
GDP_PC	-0.0021***	-0.0022***	-0.0027***	-0.0026***	-0.0025***	-0.0019***	-0.0011***	-0.0007
	0.0002	0.0003	0.0003	0.0002	0.0002	0.0003	0.0004	0.0005
CONSTANT	2.8920***	2.7720***	2.7959***	2.8393***	2.9020***	2.9502***	3.0088***	3.0373***
	0.0098	0.0168	0.0143	0.0111	0.0123	0.0157	0.0208	0.0291
Pseudo R2	0.2144	0.0788	0.0993	0.1272	0.1390	0.1255	0.1012	0.0871

Note: Quantile regression estimates with bootstrapped standard errors. Replications set to 1500. ***, **. * denotes significance levels at 1%, 5% and 10% respectively.

Table A4. Quantile Regression estimates of (Log)BMI: ITALY - Males

	OLS	Q5	Q10	Q25	Q50	Q75	Q90	Q95
AGE	0.0099***	0.0113***	0.0101***	0.0088***	0.0089***	0.0101***	0.0111***	0.0112***
	0.0004	0.0008	0.0006	0.0005	0.0004	0.0005	0.0007	0.0009
AGE_SQ	-0.0084***	-0.0104***	-0.0089***	-0.0075***	-0.0073***	-0.0084***	-0.0093***	-0.0097***
	0.0004	0.0009	0.0007	0.0005	0.0004	0.0005	0.0008	0.0009
EDU_HIGH	-0.0220***	-0.0190***	-0.0159***	-0.0191***	-0.0215***	-0.0231***	-0.0208***	-0.0325***
	0.0030	0.0068	0.0048	0.0032	0.0037	0.0044	0.0067	0.0067
EDU_LOW	0.0172***	0.0017	0.0056	0.0113***	0.0188***	0.0250***	0.0242***	0.0319***
	0.0026	0.0065	0.0048	0.0035	0.0029	0.0036	0.0045	0.0062
MARRIED	0.0158***	0.0174***	0.0159***	0.0178***	0.0160***	0.0157***	0.0170***	0.0160***
	0.0023	0.0049	0.0042	0.0027	0.0025	0.0033	0.0045	0.0057
EMPLOYED	0.0075***	0.0095*	0.0173***	0.0101***	0.0110***	0.0042	0.0007	-0.0026
	0.0024	0.0056	0.0040	0.0030	0.0027	0.0034	0.0051	0.0064
WORK_HARD	0.0025	-0.0005	-0.0030	0.0012	0.0017	0.0089**	0.0023	0.0034
	0.0024	0.0051	0.0041	0.0029	0.0030	0.0035	0.0048	0.0075
CURR_SMOKE	-0.0031	0.0022	-0.0058*	-0.0041	0.0007	-0.0020	0.0001	0.0004
	0.0021	0.0051	0.0035	0.0027	0.0024	0.0031	0.0044	0.0062
PAST_SMOKE	0.0140***	0.0181***	0.0163***	0.0144***	0.0137***	0.0135***	0.0124**	0.0142**
	0.0022	0.0053	0.0038	0.0025	0.0027	0.0032	0.0049	0.0057
INSURED	0.0018	0.0016	-0.0009	0.0018	-0.0002	0.0025	0.0043	0.0025
	0.0021	0.0046	0.0037	0.0026	0.0026	0.0032	0.0044	0.0054
BREAKFAST	-0.0220***	-0.0048	-0.0092*	-0.0177***	-0.0202***	-0.0344***	-0.0376***	-0.0294***
	0.0030	0.0071	0.0052	0.0038	0.0042	0.0043	0.0069	0.0072
NEV_MEAT	-0.0062*	-0.0052	0.0002	0.0000	-0.0051	-0.0061	-0.0167***	-0.0232***
	0.0032	0.0083	0.0059	0.0038	0.0038	0.0054	0.0059	0.0089
GDP_PC	-0.0014***	-0.0009**	-0.0013***	-0.0013***	-0.0015***	-0.0015***	-0.0015***	-0.0018***
	0.0002	0.0004	0.0002	0.0002	0.0002	0.0002	0.0003	0.0004
CONSTANT	3.0068***	2.7799***	2.8537***	2.9484***	3.0197***	3.0795***	3.1408***	3.1966***
	0.0084	0.0189	0.0145	0.0110	0.0090	0.0113	0.0164	0.0203
Pseudo R2	0.1477	0.0787	0.0825	0.0863	0.0874	0.0830	0.0717	0.0602

Note: Quantile regression estimates with bootstrapped standard errors. Replications set to 1500. ***, **. * denotes significance levels at 1%, 5% and 10% respectively.