



Decomposing body mass index gaps between Mediterranean countries: A counterfactual quantile regression analysis

Joan Costa-Font^a, Daniele Fabbri^{b,*}, Joan Gil^c

^a Department of Social Policy & European Institute London School of Economics, UK

^b Department of Economics, University of Bologna, Piazza Scaravilli, 2 - 40126 Bologna, Italy

^c Department of Economic Theory, Universitat de Barcelona, Spain

ARTICLE INFO

Article history:

Received 27 November 2008

Received in revised form 27 August 2009

Accepted 28 August 2009

JEL classification:

I18

J15

J16

Keywords:

BMI

Country weight gap

Quantile regression

Counterfactual decomposition

Mediterranean countries

Italy

Spain

ABSTRACT

Wide cross-country variation in obesity rates has been reported between European Union member states. Although the existing cross-country differences have not been analyzed in depth, they contain important information on health production determinants. In this paper we apply a methodology for conducting standardized cross-country comparisons of body mass index (BMI). We draw on estimations of the marginal density function of BMI for Italy and Spain in 2003, two countries with similar GDP and socio-economic conditions. We produce different counterfactual distribution estimates using covariates (health production inputs) specified in a quantile regression. Our findings suggest that Spain-to-Italy BMI gaps among females are largely explained by cross-country variation in the returns to each covariate, especially for younger women. We find that adverse underlying determinants do not explain the gap observed in particular between younger Spanish females and their Italian counterfactuals; behavioural differences appear to be the key. We tentatively conclude that Spanish policy on obesity should target mainly younger females.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The prevalence of overweight and obesity is rapidly expanding in the western world (WHO, 2003). It is estimated that obesity and overweight currently affect about two-thirds of the US population, stemming from a rapid expansion in the last two decades (Ruhm, 2007). Although in Europe the prevalence of obesity is on average less than in the U.S. (Brunello et al., 2008), it is rising at a faster rate, having tripled in the last 20 years (Branca et al., 2007). This situation poses important cost containment issues to the health system and calls into question the efficiency of individual lifestyles in producing health.

Indeed, the European Commission's Green Paper on promoting healthy diets and physical activity (Commission of the European Communities, 2005), estimated that obesity already accounts for 7% of total healthcare costs in the EU.

An issue that has not been deeply explored is the fact that obesity rates vary widely within Europe (Sanz-de-Galdeano, 2005; Branca et al., 2007). For example, female obesity rates in Finland and Spain (around 13%) are almost double those in Italy (7.4%). Similarly, among men, Spain (13.9%) and Finland (12.4%) again have the highest obesity rates, whereas Ireland (8.4%) and Italy (9%) have the lowest (Sanz-de-Galdeano, 2005). The reasons for this large variation, which is likely to be reflected into differences in the healthcare burden of obesity-related diseases, are not well understood (Costa-Font and Gil, 2004, 2005). Developing appropriate cross-country comparisons in obesity may shed some light on the underlying causes of

* Corresponding author. Tel.: +39 051 2098669; fax: +39 051 2098040.
E-mail address: d.fabbri@unibo.it (D. Fabbri).

the obesity epidemic, especially the role of economic and social incentives that affect people's lifestyles. Few studies of this kind have been carried out to date. Exceptions are the papers by Contoyannis and Wildman (2007), Sanz-de-Galdeano (2005) and Michaud et al. (2007).

This paper contributes to the literature on cross-country comparisons in obesity on two specific fronts. First, we endorse an empirical strategy that is not sensitive to the definition of cut-off points applied to the BMI distribution. Obesity and overweight are typically examined as binary variables. This implies that some information is inevitably lost, and the validity of the comparison compromised. Two very different BMI distributions may produce similar obesity rates: for instance, in one country a large share of the population may be concentrated around the overweight or obesity threshold, making obesity prevalence sensitive to the definition of cut-off points (Contoyannis and Wildman, 2007). Therefore, cross-country (or across time¹) comparisons of obesity should take the entire BMI distribution into consideration.² To our knowledge, only Contoyannis and Wildman (2007) have addressed this issue within a country-to-country comparison. They analyze changes in the BMI distributions of England and Canada between 1994 and 2001, taking advantage of polarization and inequality measures that break down distributional changes in location and in shape.

A second methodological issue that we explore is that of accounting for (cross-country) differences in the distribution of relevant determinants (e.g., age, education or employment status) and for differences in the returns³ to these determinants. Contoyannis and Wildman (2007) for instance do not account for cross-country variation in covariates affecting the BMI distribution, such as education, income or lifestyle, which might play an important role in determining the shape of the BMI distribution. Evidence in Sanz-de-Galdeano (2005) and Michaud et al. (2007) suggests that cross-country differences in the returns to relevant determinants of obesity can be non-negligible even between similar countries. Moreover, these factors have been shown to exert a heterogeneous influence on individuals' BMI at different points of its distribution (Kan and Tsai, 2004; Ruhm, 2007).

Our empirical strategy extends the Oaxaca decomposition (Oaxaca, 1973) of effects on mean values of a variable of interest to its entire distribution. The method estimates the marginal density function of BMI in a given country implied by different counterfactual distributions of all the covariates included. This counterfactual decomposition approach has been developed by Machado and Mata (2005) and repeatedly applied to the analysis of changes in the wage schedule

across time and space (Garcia et al., 2001; Albrecht et al., 2003; Melly, 2005; Arulampalam et al., 2007). The methodology is applied here to an analysis of the variation in the BMI distribution in Spain with respect to Italy in the year 2003. We estimate the counterfactual 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 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 prevailed in Spain if BMI values were the same function of variables as in Italy. By comparing this last counterfactual distribution with the actual marginal distribution in Spain, we are 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 Basset, 1978), namely, by estimating models for the quantiles of the conditional BMI distribution (measured in logarithms⁴). Unlike 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 or underweight and provide a more general picture of the effect of covariates on BMI. 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 step of our analysis is thus to estimate the marginal density function of (log)BMI that is consistent with the conditional distribution estimated in the previous step, using the same distribution of the population's attributes, specifically that of Spain.

The interest of examining Spain and Italy lies in the fact that both are Mediterranean countries with similar income per capita levels, socio-economic characteristics, and smoking habits. According to Sanz-de-Galdeano (2005), Spain is one of the EU's most obese member states, while Italy is the least. 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. We would therefore expect the gaps observed to be due more to differences in the way underlying determinants affect BMI in the two countries rather than to differences in the distribution of these determinants.

2. Data and preliminary evidence on the Spain vs. Italy cross-country case in obesity

The Italian data are from the 2003 edition of the National Survey on Daily Life ("Indagine sugli Aspetti della

¹ Flegal (2006) reports that BMI distributions in the USA appear to have shifted to the right and progressively become more skewed, and Freedman et al. (2000) found that the 10th percentile of adult BMI rose 0.6 kg/m² between 1990 and 2000, while this effect increased up to 1.2 at the median and 3.2 at the 95th percentile.

² Recent evidence also indicates that cross-country differences are larger in the right tail of the BMI distribution (Ruhm, 2007; Komlos et al., 2009).

³ By "differences in the returns to a given determinant" we mean the differences in estimated coefficients across the two countries.

⁴ This is done just for the sake of exposition. Modeling BMI in logarithms, the difference in the estimated coefficients at two different quantiles provides a measure of the impact of each given covariate upon the (log of the) ratio between BMIs at these quantiles.

Table 1
BMI distribution across Italy and Spain.

	Mean	Q5	Q10	Q25	Q50	Q75	Q90	Q95
All								
Italy	24.4	19.0	19.9	21.8	24.1	26.5	29.4	31.1
Spain	25.3	19.3	20.3	22.3	24.8	27.7	30.7	32.7
Difference	0.87	0.33	0.39	0.54	0.73	1.15	1.31	1.55
<i>t</i> -Statistic	18.9	4.1	6.1	7.5	14.1	16.8	9.6	13.3
Relative difference	0.035	0.017	0.020	0.025	0.030	0.043	0.045	0.050
Females								
Italy	23.4	18.4	19.1	20.6	22.7	25.4	28.7	31.1
Spain	24.5	18.7	19.6	21.3	23.7	26.9	30.5	33.2
Difference	1.14	0.38	0.46	0.73	1.01	1.46	1.80	2.05
<i>t</i> -Statistic	16.7	8.6	9.4	11.3	10.7	14.6	8.1	8.2
Relative difference	0.049	0.021	0.024	0.036	0.045	0.058	0.063	0.066
Males								
Italy	25.4	20.8	21.6	23.1	25.0	27.3	29.6	31.2
Spain	26.0	20.8	21.8	23.5	25.7	28.1	30.8	32.4
Difference	0.61	−0.01	0.24	0.38	0.71	0.77	1.22	1.20
<i>t</i> -Statistic	10.8	−0.2	2.6	6.8	10.7	5.9	6.4	7.6
Relative difference	0.024	−0.001	0.011	0.016	0.028	0.028	0.041	0.038

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

Vita Quotidiana”), that collects individual data, including data on health conditions and dietary habits. The original sample contains information on 20,547 complete households comprising 44,384 adult individuals (aged 18 or above). The data used for Spain are from the 2003 edition of the Spanish National Health Survey (“Encuesta Nacional de Salud”), a biannual, cross-sectional nationwide representative survey which gathers information on the population’s health and healthcare use. The original sample contained 21,650 adults aged 16–99. Both surveys are nationally and regionally representative. They use similar sampling procedures, and the wording of the two questionnaires is comparable. In order to have for Italy a sample with a similar structure to the Spanish one (i.e., one respondent selected at random for each sampled family) we resized the Italian dataset to 20,547 individuals, one for each sampled household. After removing the data relating to subjects aged below 18 and above 65, and those with missing data on weight and height,⁵ our final sample comprises 15,147 Spaniards and 14,522 Italians.⁶

We use the individual’s body mass index⁷ (BMI), computed using self-reported data on weight and height, as our dependent variable.

We report average BMI and BMI levels in different deciles of the distribution (Table 1). We split the sample by gender and report also gender-specific BMI gaps between countries. A comparison of the mean and the median levels of BMI shows that the distributions are right-skewed with long right-hand tails. In Spain the cut-off point for obesity lies close to the 85th percentile, while in Italy it is between the 90th and 95th percentiles. Body mass index gaps are always positive and statistically significant and increase along the BMI distribution, so that the gap is almost five times higher at the 95th percentile (1.55) than at 5th percentile (0.33). These differentials apply to both genders, though overall the gap is higher for women.⁸ The Italian BMI distribution is more positively skewed compared to the Spanish distribution for both genders (skewness is .28 in Italy vs. .15 in Spain for men and .54 vs. .51 respectively for women), which means that there is a higher percentage of people in the healthy BMI interval in Italy than in Spain. The distribution is more positively skewed among women than among men. In accordance with previous studies (Baum II and Ruhm, 2007), we notice an age effect which appears to be the result of a transition from healthy BMI levels in younger ages to unhealthy BMI levels at older ages (Fig. 1). This is the case for both Italy and Spain. However, in Spain the process is more pronounced and the differences are larger.

Italy and Spain are *prima facie* very similar. An important empirical question is therefore what drives these differences in the BMI distribution across the two countries. To explore this issue we need a method to decompose the gaps in BMI at different points of the distribution into components attributable to (cross-country) differences in the distribution of relevant determinants vis-à-vis (cross-country) differences in

⁵ Missing data on BMI for Spain were 271 out of 15,418, less than 1.8%. No missing data on BMI was present in the Italian sample.

⁶ We exclude the elderly population from the sample to reduce composition bias arising through larger mortality among the more obese. Moreover, as we use self-reported data, this selection helps in reducing the well-known misreporting problems affecting self-reported weight and height at old ages (Boström and Diderichsen, 1997; Kuczmarski et al., 2001; Gil and Mora, 2009). However, we recognize that a different cross-country pattern in self-reporting anthropometric data could bias our results. Hence, we hypothesize that this self-reporting pattern is similar conditional on our set of covariates.

⁷ The BMI is defined as weight in kilograms divided by the square of height in meters (kg/m²). According to the World Health Organization (1997) an individual with a BMI ≥ 25 kg/m² is considered overweight and one with a BMI ≥ 30 kg/m² is obese.

⁸ We find that it doubles both for women (Spanish women are 2.1% heavier at percentile 5 and 6.6% heavier at percentile 95) and for men (roughly the same at percentile 5 and 3.8% heavier at percentile 95).

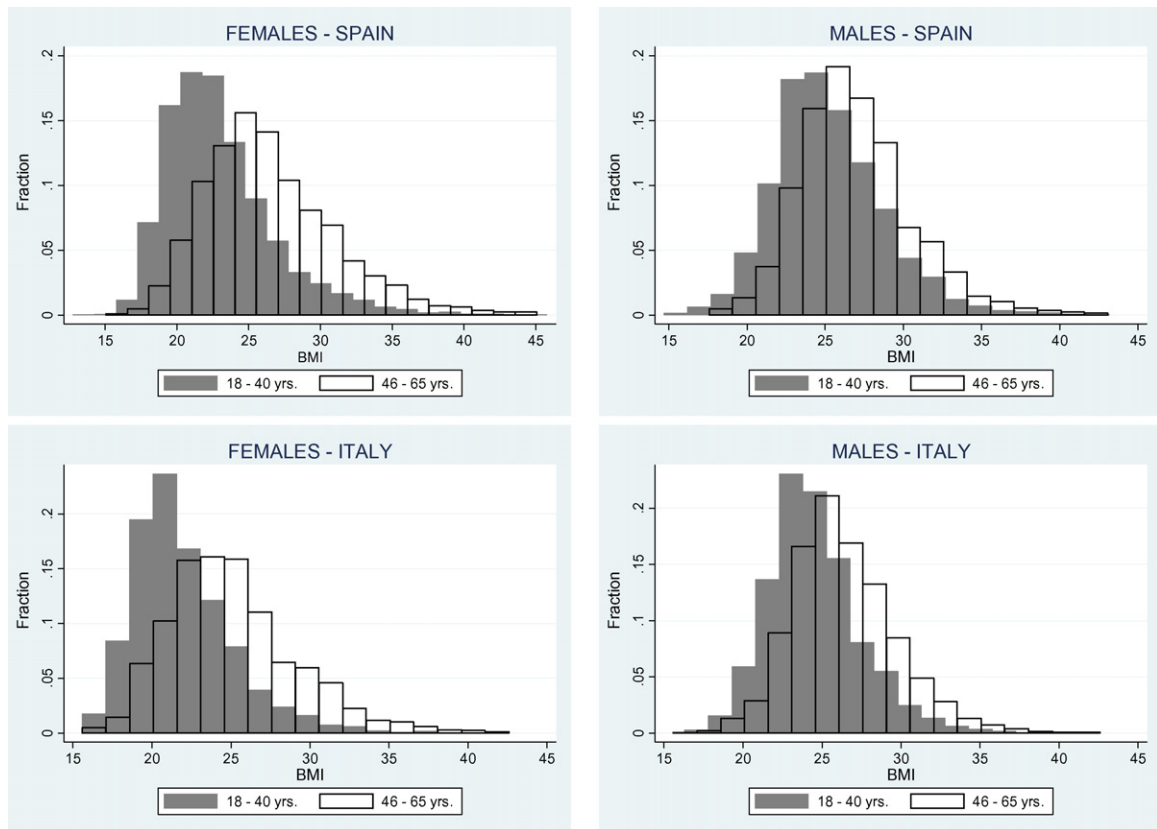


Fig. 1. Gender-specific BMI fraction histograms by age groups.

the returns to relevant health production determinants. The latter would tell us what would be the partial effect on BMI of a change in one specific variable, i.e. maintaining the rest constant.

3. Background

At a given point in time, individual BMI provides a snapshot that reflects the accumulated difference between an individual's calorie intake and calorie expenditure that he experienced in his life. Therefore current individual BMI should be properly modeled within a dynamic framework where the impact of current and past permanent shocks in calorie imbalance is disentangled in face of individual unobserved heterogeneity (mainly reflecting the influence of genetics and the social environment). Cutler et al. (2003) adopted such a dynamic approach at a macro-level. In the empirical literature it has become common practice to estimate models for the conditional mean of adiposity, either measured as a dichotomous (obesity) or continuous (BMI) variable, by regressing it on factors such as age, education, income, food consumption, physical activity and lifestyles, measured at the same point in time. In the emerging framework the respondent i 's measure of adiposity is determined by:

$$w_i = X' \beta + \varepsilon_i \quad (1)$$

where X is a vector of health production determinants and ε_i is a measure of unobservables. A plausible interpretation of this relationship rests on a steady-state view of the energy-accounting equations: provided that body weight has “settled” or stabilized in the population, an equation such as (1) will reflect “equilibrium conditions” in the relationship between underlying determinants and adiposity.⁹

Within the framework of Eq. (1) a standard decomposition to deal with cross-country comparison of adiposity prevalence is proposed by Oaxaca (1973) and Blinder (1973) as follows:

$$\bar{w}_S - \bar{w}_I = \sum_{j=1}^K \bar{X}_I^{(j)'} (\hat{\beta}_S^{(j)} - \hat{\beta}_I^{(j)}) + \sum_{j=1}^K (\bar{X}_S^{(j)'} - \bar{X}_I^{(j)'}) \hat{\beta}_S^{(j)} + \text{residual} \quad (2)$$

where \bar{w}_c is the mean adiposity in country $c = S$ (Spain) or I (Italy), $\hat{\beta}_c^{(j)}$ is a column vector of estimated coefficients for the set of regressors comprising the j th among K variables, and $\bar{X}_c^{(j)'}$ is a row vector of regressors' means for the set of regressors comprising the j th among K variables. Therefore the Blinder–Oaxaca decomposition allows us to

⁹ According to Michaud et al. (2007) these circumstances are plausibly met for the adult to elderly population, i.e. those aged 50 or above.

disentangle the cross-country gap in adiposity, measured at the mean of each country's distribution, in two parts: one that is attributable to cross-country differences in the "mean" returns to the country's levels of underlying factors and another one attributable to differences in the average distribution of underlying determinants.

However, this approach is not well suited to the study of obesity because adiposity, unlike wages in the labour market, is not a monotonic indicator of "performance". The more an individual earns in the labour market, the better off he is. However, an increase in adiposity reduces health hazards in the case of underweight individuals, but increases them among individuals already overweight. Decomposing the "country shift" in the distribution around the conditional mean does not shed light on what happens in the interesting parts of the distribution, the very right and the very left tails. Therefore, in this context, working with the entire BMI distribution represents a clear advantage.¹⁰ 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.

We apply a decomposition methodology to the entire BMI distribution using quantile regression. We detail the method in the following section.

4. Methods

The quantile regression (QR) model (Koenker and Basset, 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 θ th conditional quantile of the distribution of $(\log)\text{BMI}(w)$, given a vector, X , of k covariates. These conditional quantiles are expressed as:

$$Q_\theta(w|X) = X'\beta(\theta) \quad (3)$$

where $\beta(\theta)$ is a vector of coefficients, that is, the QR coefficients.¹¹ Following Koenker and Basset (1978), $\beta(\theta)$ can be estimated by minimizing the following objective function with respect to β

$$n^{-1} \left[\sum_{i:w_i \geq X_i'\beta} \theta(w_i - X_i'\beta) + \sum_{i:w_i < X_i'\beta} (\theta - 1)(w_i - X_i'\beta) \right] \quad (4)$$

Although Eq. (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 enables us to ascertain whether the impact of covariates on adiposity changes across the

distribution of BMI – in particular at the tails of the distribution. As a result we can allow for a different impact of the same covariate at the lower tail of the distribution (the underweight) and at the upper one (the obese).

4.1. Counterfactual densities estimation

After obtaining the QR estimates of the BMI equations for each country, the following step is to compute quantile counterfactual distributions, i.e. those levels of $(\log)\text{BMI}$ that Italians (Spaniards) would face at the θ th quantile if the distribution of their characteristics were the same as that of Spaniards (Italians). To do this, we use the bootstrap procedure proposed by Machado and Mata (2005).¹² The cornerstone of the method is the estimation of a marginal density function of $(\log)\text{BMI}$ that is consistent with the estimated conditional distribution defined by (3) as well as the distribution of covariates. Their approach proceeds as follows:

Step 1: Generate a random sample of size m from a uniform distribution $U[0,1]$: $\theta_1, \dots, \theta_m$. These numbers are the quantiles to be estimated.¹³

Step 2: For the Spanish and Italian dataset and for each value of θ from step 1, estimate the conditional quantile $Q_\theta(w|X)$, where w denotes $(\log)\text{BMI}$, yielding m estimates of the QR coefficients for each dataset, that is, $\beta_S(\theta)$ and $\beta_I(\theta)$ for Spain and Italy, respectively.

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

$$\{w_i^*(j) \equiv X_i^*(S)'\beta_j(\theta)\} \quad i = 1, \dots, m \quad \text{and} \quad j = S, I \quad (5)$$

deriving two sets of predicted $(\log)\text{BMI}$ covering the entire distribution.

Machado and Mata (2005) do not provide any explicit argument for setting the size of the random sample, m . Since each of the 99 quantiles can be sampled on average $m/99$ times, setting $m = 4500$, as they do, produces on average about 45.5 predicted values of $\log(\text{BMI})$ for each quantile. We set a slightly higher value, i.e. $m = 5000$.¹⁴

4.2. The decomposition procedure

We denote by $f(w(c))$ an estimator of the marginal density of w in country c based on the observed sample and by $f^*(w(c))$ an estimator of the marginal density of w in country c based on the generated sample $\{w_i^*(c)\}$.

¹⁰ Working with the entire distribution of BMI reduces the impact of the well-known measurement error due to the misreporting of weight and height whenever cut-off points are used as approximate measures of obesity.

¹¹ The θ th conditional quantile of the error term is zero, that is $Q_\theta(\varepsilon|X) = 0$.

¹² Machado and Mata (2005) describe alternative methods to compute counterfactual densities.

¹³ At this step we "built" the list of quantiles to be estimated. This list is made of numbers between .01 and .99 with replications. Since we can estimate 99 quantile regressions and we generate a sample of size m each quantile should be resampled $m/99$ times on average.

¹⁴ In their paper Machado and Mata (2005) work with a sample size of about 5000 obs. and they set $m = 4500$. Our sample size is (for each gender-country) more than 7100 so our choice seems fair.

Table 2

Variable definitions, sample means and standard deviations in parentheses.

Variable	Description	Females		Males	
		Spain	Italy	Spain	Italy
		7813 ^a	7367 ^a	7312 ^a	7148 ^a
Log BMI	Log of body mass index	3.18 (0.17)	3.14 (0.16)	3.25 (0.14)	3.23 (0.12)
Age	Age of the interviewed individual	41.85 (12.98)	42.11 (12.80)	41.08 (12.50)	42.14 (12.56)
Age squared	Square of age/100	19.20 (11.07)	19.37 (11.00)	18.44 (10.53)	19.33 (10.77)
Years of education	Years spent in full time education	9.56 (4.54)	10.16 (4.02)	9.62 (4.28)	10.25 (3.81)
Married	=1 if married; =0 otherwise	0.61 (0.49)	0.58 (0.49)	0.60 (0.49)	0.56 (0.50)
Employed	=1 if employed; 0 otherwise	0.49 (0.50)	0.47 (0.50)	0.77 (0.42)	0.74 (0.44)
Smoker	=1 if current smoker; =0 otherwise	0.30 (0.46)	0.22 (0.42)	0.43 (0.49)	0.37 (0.48)
GDP	Regional GDP ^b per capita purchasing power standard/1000 EU€	20.54 (3.83)	22.19 (5.66)	20.52 (3.82)	22.27 (5.65)

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

^a No. of observations.^b Regional GDP for the 20 Italian regions and 18 Spanish autonomous communities comes from EUROSTAT database.

Counterfactual densities are denoted as $f^*(w(I); X_i^*(S))$ for the density that would result in country I (Italy) if all covariates were distributed as in country S (Spain). Hence the overall gap is approached as follows:

$$f(w(S)) - f(w(I)) = [f^*(w(S)) - f^*(w(I); X(S))] + [f^*(w(I); X(S)) - f^*(w(I))] + \text{residual} \quad (6)$$

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), and 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 cross-country BMI differences result from differences in the returns to covariates, which can be associated with behavioural or environmental factors related to country-specific lifestyles, or from differences in the distribution of common characteristics or covariates in Spain as compared to Italy.¹⁵

5. Model specification

In agreement with the discussion above and the literature on the determinants of body weight, our econometric specification contains the following set of covariates (Table 2).

- i) Following Kan and Tsai (2004) we enter the age and age squared of each respondent at the date of the interview to capture non-linear effects across age groups.

- ii) To control for differences due to education, we enter the number of years individuals spent in full time education.¹⁶ It should be noted that education can be endogenous to BMI (Cutler and Lleras-Muney, 2006). Evidence on causal effects of education on BMI is still inconclusive (Arendt, 2005). We should therefore be cautious in attaching a causal interpretation to our estimates.
- iii) We include a dummy variable for the individual being currently employed. We would expect workers to be leaner as more calories are spent at working. However reverse causality might occur, as the obese may be discriminated against in the labour market. Therefore our estimates of the partial effect of employment status on BMI could be overstated.¹⁷
- iv) We control for the individual being married as, among other explanations, it reduces competition for mating and hence relaxes the incentives to "keep in good shape". On the other hand being married might be a source of stability and maintaining a more intense social life.
- v) We include a dummy variable for being an active smoker. The negative association between (falling) smoking rates and (increasing) obesity has been noted in the literature. However it should be noted that smoking can be endogenous to BMI. Evidence on causal effects of education on BMI is still inconclusive (Chou et al., 2004; Gruber and Frakes, 2006). We are therefore cautious in attaching a causal interpretation to our estimates.
- vi) To account for regional environmental factors behind individual BMI, we include regional GDP per capita to proxy for aggregate affluence in BMI production. This

¹⁵ An issue emerges here pertaining to the order we follow to decompose the cross-country gaps. In the paper we compare the Spanish distribution of BMI with the one that would have emerged if Spain were "producing" BMI according to its Italian "counterfactual". This choice is driven by policy relevance. We could reverse this order and compare the Italian distribution with its Spanish "counterfactual". Results might differ in general. However in our case they prove to be very similar indeed. Italy-to-Spain decomposition analysis is available upon request.

¹⁶ We opt for this regressor instead of including dummies for educational attainments. These might be more prone to measurement error given different cross-country classifications.

¹⁷ Morris (2006, 2007) addresses the issue of causality between obesity and employability. Assuming BMI to be exogenous he finds that it exerts a positive and significant effect on occupational attainment in males and a negative and significant effect in females. Turning to IV estimates however no significant effect is found and no endogeneity problems emerges with respect to BMI.

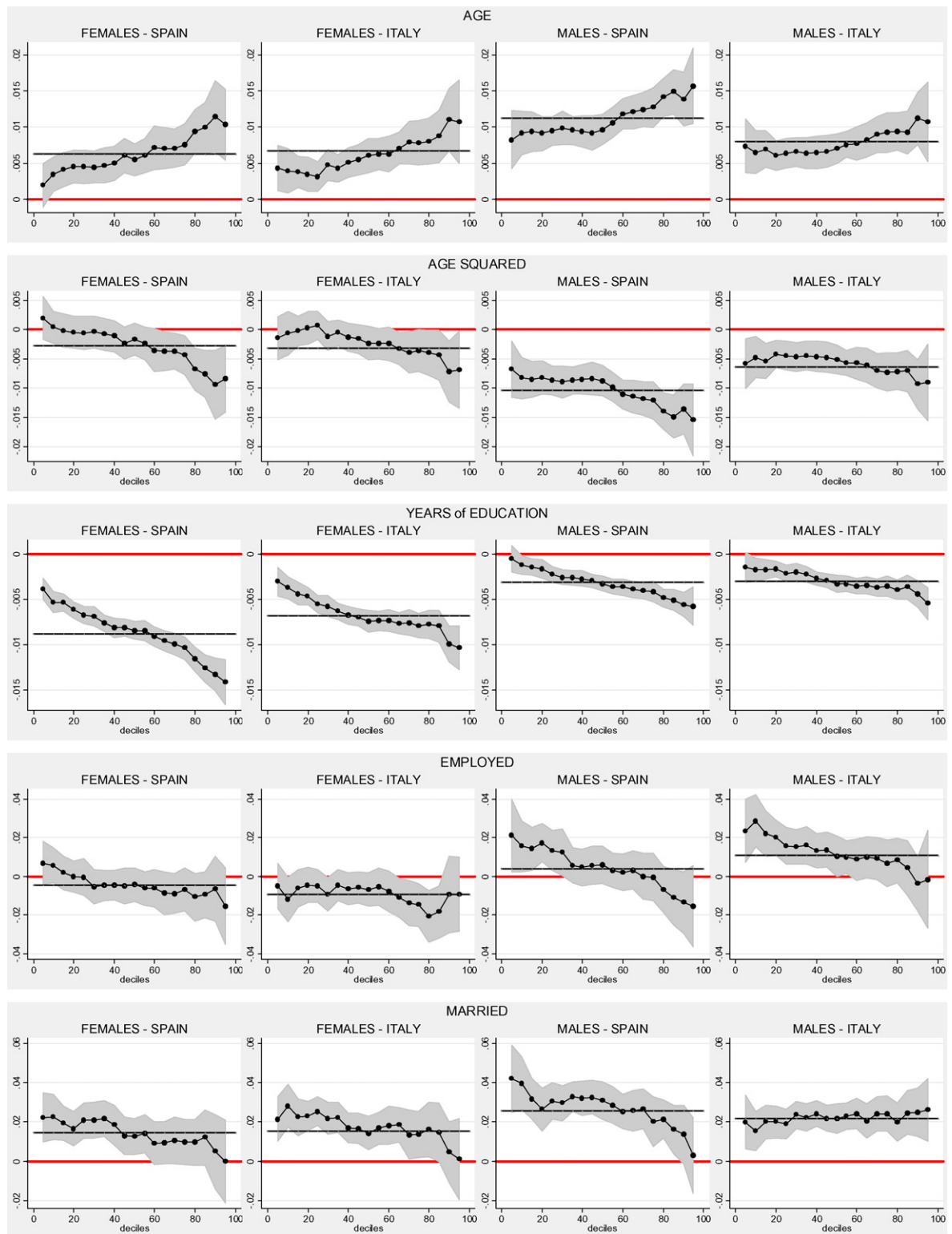


Fig. 2. Log(BMI) returns to characteristics.

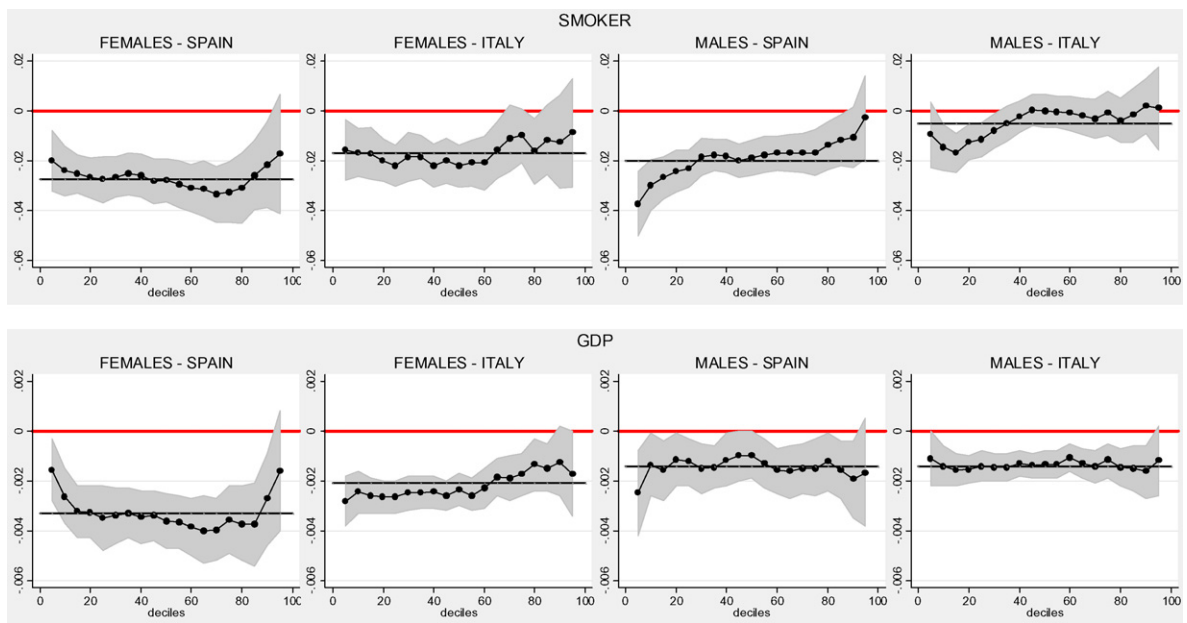


Fig. 2. (Continued).

variable partially corrects the fact that we could not include individual/family income in our specification, since income is not recorded in the Italian survey.¹⁸ Moreover the evidence for Spain indicates that less than 7% of income-related inequalities in obesity are explained by pure income effects (Costa-Font and Gil, 2008).¹⁹

6. Results

We conduct our analysis by gender. Although quantile regression is robust to the presence of outliers, we check for the magnitude of this potential problem at the extreme quantiles by adopting the procedure to detect multiple outliers in multivariate data proposed by Hadi (1992). 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 12 females and 10 males, and for

Italy 4 females and 3 males. We ran the full set of quantile regressions excluding these outliers. The results obtained were very similar to those obtained ignoring the presence of outliers. Here we present the estimates on the sample reduced for excluding the outliers.

6.1. OLS and QR analysis

As our first step, we analyze the quantile regression results (Fig. 2). The plots show the coefficient estimates, $\beta_i(\theta)$, $i = 1, \dots, k$ for $\theta = .05, \dots, .95$, and the associated confidence bands (the shaded area). 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 shaded areas represent the 95% confidence intervals for the regression deciles.²⁰ 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.

The results indicate that age exerts a quadratic effect on BMI. This quadratic relation implies that individuals' BMI value increases as they age until they reach a peak which differs according to population and quantile (Fig. 3). This gender age effect remains constant in the middle of the

¹⁸ Income is not unambiguously relevant here. 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.

¹⁹ We report below a robustness check on the Spanish sample for the relevance of excluding income from our specification. Results suggest that we cannot exclude income to exert an independent effect in the case of females. Anyway excluding income does not seem to have a major impact on partial effects of included regressors. Baum II and Ruhm (2007) provide evidence suggesting that in the US the "socio-economic gradient" in obesity is mainly explained by ethnicity and education and very little is propagated through household income.

²⁰ Bootstrap sample of size 200 was used for the calculation of the standard errors.

Table 3
The OLS determinants of log(BMI).

	Females		Males	
	Spain	Italy	Spain	Italy
Age	0.0062*** <i>0.0010</i>	0.0067*** <i>0.0010</i>	0.0112*** <i>0.0009</i>	0.0080*** <i>0.0009</i>
Age squared	−0.0027** <i>0.0012</i>	−0.0031*** <i>0.0011</i>	−0.0104*** <i>0.0011</i>	−0.0063*** <i>0.0011</i>
Years of education	−0.0088*** <i>0.0004</i>	−0.0068*** <i>0.0005</i>	−0.0031*** <i>0.0004</i>	−0.0030*** <i>0.0004</i>
Employed	−0.0044 <i>0.0038</i>	−0.0095** <i>0.0038</i>	0.0039 <i>0.0043</i>	0.0107*** <i>0.0040</i>
Married	0.0145*** <i>0.0042</i>	0.0155*** <i>0.0038</i>	0.0259*** <i>0.0038</i>	0.0216*** <i>0.0032</i>
Smoker	−0.0274*** <i>0.0039</i>	−0.0170*** <i>0.0040</i>	−0.0200*** <i>0.0031</i>	−0.0051* <i>0.0029</i>
GDP	−0.0033*** <i>0.0005</i>	−0.0021*** <i>0.0003</i>	−0.0014*** <i>0.0004</i>	−0.0014*** <i>0.0002</i>
Constant	3.1311*** <i>0.0210</i>	3.0304*** <i>0.0192</i>	3.0322*** <i>0.0185</i>	3.0531*** <i>0.0173</i>
R ²	0.2323	0.2156	0.1299	0.1386
No. of observations	7813	7367	7312	7148

Note: Standard errors (in italics) are computed using the Huber/White variance estimator.

An *F*-test of the hypotheses that the parameters are the same in the two countries clearly rejects the null for both males and females.

*** Significance levels at 1%.

** Significance levels at 5%.

* Significance levels at 10%.

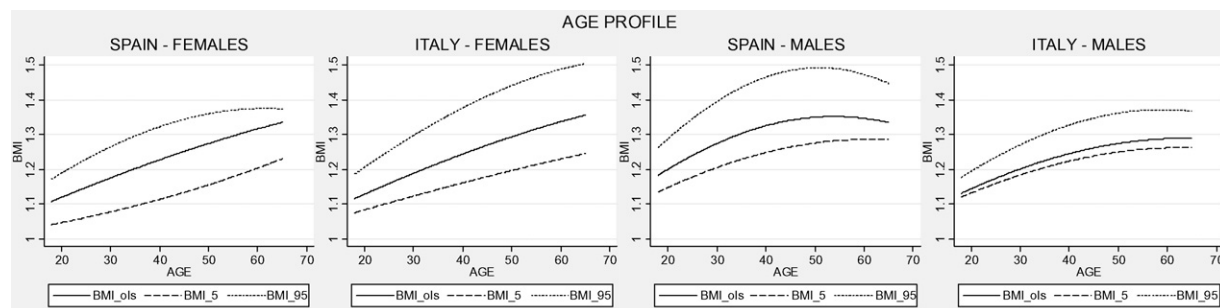


Fig. 3. BMI returns to AGE at the mean and the tails.

BMI distribution, but falls off in the upper tail in both countries. As in other studies, we find that education is one of the most significant variables in explaining obesity (Chou et al., 2004; Kan and Tsai, 2004).

The plot corresponding to YEARS OF EDUCATION shows that individuals with more years of completed education have a lower BMI value than individuals with a shorter education career (coefficients are negative). Note that in both countries the impact of this covariate is larger for females than for males. A feature concealed by the OLS estimates but identified by the quantile analysis is that the “protective effect” of years of education intensifies as we move up through the BMI distribution, mainly among women. This effect implies that samples with more educated individuals show a lower BMI dispersion than samples of less educated people. So our quantile regression

estimates suggest that increasing the educational level is not only associated with a reduction (which is larger in females) in overall obesity, but also leads to a reduction in BMI inequality.

Working status²¹ is a significant correlate for mean BMI only for the Italian population: EMPLOYED Italian females have a lower mean BMI, while for men the effect is the opposite. Our estimates reveal that in the male population being EMPLOYED is associated with higher BMI at the very lower tail of the BMI scale.

According to the OLS estimates being MARRIED produces a positive, proportional scale effect on BMI.

²¹ Since we do not control for income, EMPLOYED is probably acting as a proxy for income.

The quantile analysis reveals that this effect is actually constant among Italian males but fades away along the distribution of BMI for the other populations. Namely, being married is protective against underweight with respect to the middle values of BMI while it does not affect BMI at the very upper tails.

According to some literature being an active smoker (SMOKER) is negatively correlated with BMI. We find evidence of this pattern of correlation in both countries. In particular 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 hardly a significant determinant of BMI for Italians, while it changes from negative to zero for Spaniards along the BMI distribution.

Finally, our OLS estimates suggest that regional GDP per capita (GDP) exerts a small, negative impact on BMI values, suggesting that people in affluent regions are leaner and less likely to be obese. The analysis of this effect at the quantiles shows that the impact stays constant along the whole distribution with no exception.

6.2. Counterfactual decompositions

Our next step is to examine cross-country gaps in BMI (henceforth, always measured in logs) between Italy and Spain. A natural way to consider this comparison is as an aid to identify how in Spain, where obesity is a major health policy issue, the distribution of the variables we are concerned with differs with respect to an otherwise very similar country where obesity is less of a problem.

As we showed in the descriptive analysis the Spain-to-Italy gap is positive and significant throughout the distribution and widens towards the upper tail, especially in the case of females. This can be ascribed to two underlying features. Spaniards could have BMI values that are proportionally larger than their Italian counterfactuals, so that Spanish BMI distributions mimic those of Italians, apart from “right location shifts”. At the same time, Spaniards' BMI values may be more dispersed than those of Italians so that the upper tail of the distribution is “thicker”. In terms of policy the first feature points towards the need for a national strategy aimed at reducing the overall gap in adiposity as a way of containing the proportionally higher implied costs incurred by the healthcare system and the economy as a whole. On the other hand, the second raises the need for a more focussed strategy for reducing within-country inequality in health, as proxied by obesity and severe obesity, like for instance nutritional information campaigns targeted on the already obese.

The aim of a decomposition analysis is to ascribe these two features to “effects” arising through (cross-country) differences in the distribution of underlying determinants of BMI vis-à-vis (cross-country) differences in the returns to each determinant. In the traditional Blinder–Oaxaca linear decomposition, the first set of effects is termed the

“endowments effect” and the second the “coefficients effect”.²²

In order to decompose the cross-country gaps in the BMI distribution into shares 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. To calculate standard errors we bootstrap this procedure 200 times. The results are summarized in Fig. 4a for females and Fig. 4b for males. In the first panel of both figures we plot the estimated Spain-to-Italy gap for each quantile of the 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. In both cases the estimates are plotted along with the 95% confidence band around them and the BMI gap observed. Moreover, we add as a reference a horizontal line representing the coefficient and covariate effects estimated from the mean regression models, i.e. the Blinder–Oaxaca linear decomposition. Details on the linear and quantile decompositions are also presented in separate tables (Tables 4 and 5). A linear decomposition conducted using the Blinder–Oaxaca method is helpful only for interpreting any factors behind the existence of a locational shift in the distribution of interest. However it is silent on mean-preserving changes in the dispersion, such as a polarization.

According to the linear decomposition analysis, differences in the distribution of covariates across Spanish and Italian females are responsible for a small rightward locational shift in the BMI distribution of the Spanish population. In other words, a small part of the observed difference (16% of it) is due to the relatively unfavourable distribution of characteristics among Spanish females. A larger part of the gap (83%) is due to the different returns to BMI determinants (the coefficients effect). Therefore the locational shift due to coefficients is five times as large as the part due to covariates. By taking advantage of the decomposition analysis along the entire distribution, we see that, for women, differences in the underlying distribution of covariates reflect a small but significant part of the Spain-to-Italy gap at the upper quantiles (Table 5). Nonetheless differences in the returns to covariates explain the main part of it. Moreover, as we move towards the upper tail of the distribution the part of this gap explained by differences in the returns to covariates becomes larger and larger.

The Blinder–Oaxaca decomposition suggests that only a negligible part of the locational shift in the males' distribution can be ascribed to a difference in the underlying distribution of covariates, while most of the locational shift arises through difference in their returns. The quantile decomposition does not add much to this pattern emerging from the linear decomposition. According to our OLS estimates, the extent of variation we are able

²² In the literature on gender gap in labour economics remuneration this last is usually interpreted as capturing the “discrimination effect” suffered by women, net of differences in individual productivity.

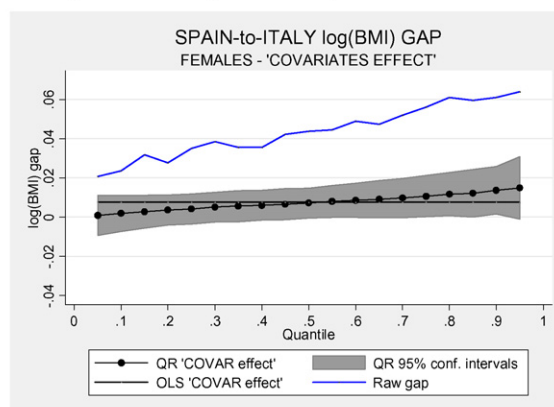
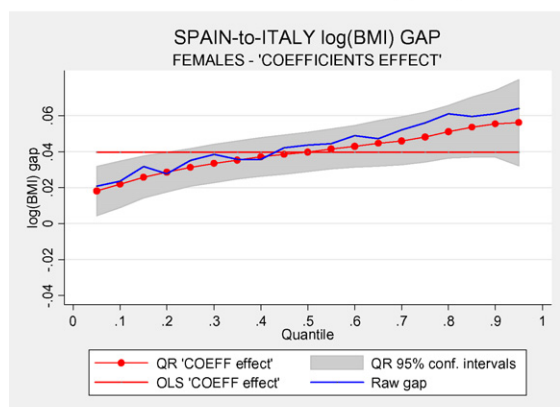
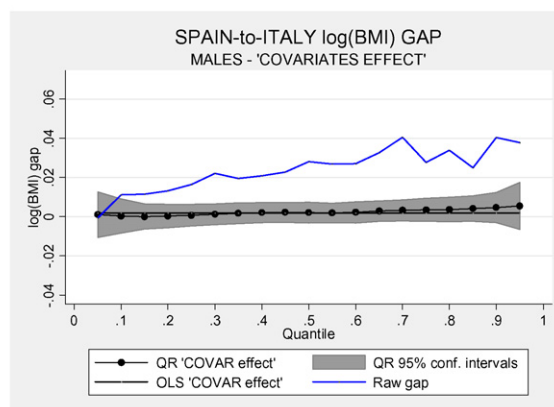
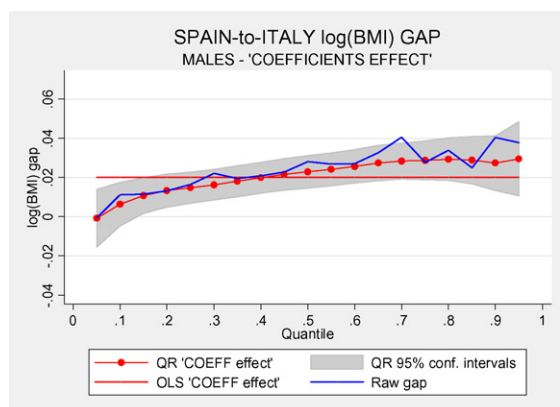
a: SPAIN-to-ITALY log(BMI) GAP counterfactual quantile decomposition analysis: FEMALES**b: SPAIN-to-ITALY log(BMI) GAP counterfactual quantile decomposition analysis: MALES**

Fig. 4. (a) Spain-to-Italy log(BMI) gap counterfactual quantile decomposition analysis: females. (b) Spain-to-Italy log(BMI) gap counterfactual quantile decomposition analysis: males.

Table 4
Spain-to-Italy log(BMI) gap linear decomposition (Blinder–Oaxaca).

Linear decomposition	Estimate	95% confidence interval	
Females			
Coefficients	3.98%	0.035	0.045
Covariates	0.76%	0.006	0.009
Residual	0.01%		
Log BMI raw gap	4.75%		
Males			
Coefficients	2.01%	0.016	0.024
Covariates	0.19%	0.000	0.003
Residual	0.18%		
Log BMI raw gap	2.38%		

Note: Terms in bold type are significant at 5% significance level.

to explain with our regressors is rather limited in the case of males, making it rather inevitable that the coefficients will explain much of the difference in that case.

We tentatively conclude that Spanish women have higher BMI value than Italians, both on average and particularly in the upper tail of the distribution, mainly because of more adverse country-specific “equilibrium

conditions” in the relation between underlying determinants and adiposity. Therefore, if we take Italy as the most appropriate “yardstick” for Spain, our exercise suggests that in Spain females’ BMI value is larger and has a thicker upper tail as the result of individual behaviours that deviate from those of their Italian counterfactuals. Results for males prove to be rather inconclusive given the poor explanatory value of the regressors we include in our specification.

Finally, we consider the robustness of our conclusions about females by performing the same set of counterfactual decomposition analysis on two distinct age groups: “the young” (aged 18–40 years) and “the middle-aged” (aged 46–65) (Fig. 5). For the sake of completeness we report also the results for males but abstain from commenting on them. The Spain-to-Italy BMI raw gap is always positive and increases quite rapidly, suggesting that BMI values are both higher and with thicker upper tail in Spain than in Italy both among “the young” and “the middle-aged”. For young females, the differences in the returns to common underlying determinants explain most of the rightward shift and thicker upper tail observed in the Spanish distribution. For middle-aged women, the differences in the distribution of underlying determinants

Table 5

Spain-to-Italy log(BMI) gap counterfactual quantile decomposition.

Quantile	BMI raw gap	Coefficients	95% confidence interval		Covariates	95% confidence interval	
Females							
10	2.4%	2.2% [93%]	0.009	0.035	0.2% [8%]	−0.0074	0.0112
25	3.5%	3.1% [89%]	0.021	0.042	0.4% [12%]	−0.0035	0.0119
50	4.4%	4.0% [91%]	0.029	0.051	0.7% [16%]	−0.0004	0.0147
75	5.6%	4.8% [86%]	0.034	0.062	1.1% [19%]	0.0002	0.0212
90	6.1%	5.5% [91%]	0.037	0.074	1.4% [22%]	0.0015	0.0258
Males							
10	1.1%	0.6% [57%]	−0.005	0.018	0.0% [3%]	−0.008	0.009
25	1.6%	1.5% [90%]	0.007	0.023	0.1% [5%]	−0.005	0.006
50	2.8%	2.3% [81%]	0.015	0.031	0.2% [7%]	−0.003	0.007
75	2.8%	2.9% [104%]	0.019	0.039	0.3% [13%]	−0.002	0.009
90	4.0%	2.7% [68%]	0.013	0.041	0.5% [11%]	−0.003	0.012

Note: In squared brackets we report the share of the total gap attributable to each component. Terms in bold type are significant at 5% significance level.

capture a non-negligible part of the rightward shift in the distribution; nonetheless gap in the very upper tail proves to be mostly captured by a coefficient effect.

We are aware that some potentially relevant determinants of cross-country differences in obesity are omitted in our analysis. As a general strategy we opted for a conservative choice of regressors in order to limit endo-

geneity problems. For instance we decided to ignore the role played by cross-country differences in food consumption habit and differences in household incomes or assets. Concerning the latter we included education and employment which may be partly correlated with income or household assets. However income could have an independent impact on the partial effect of education and employ-

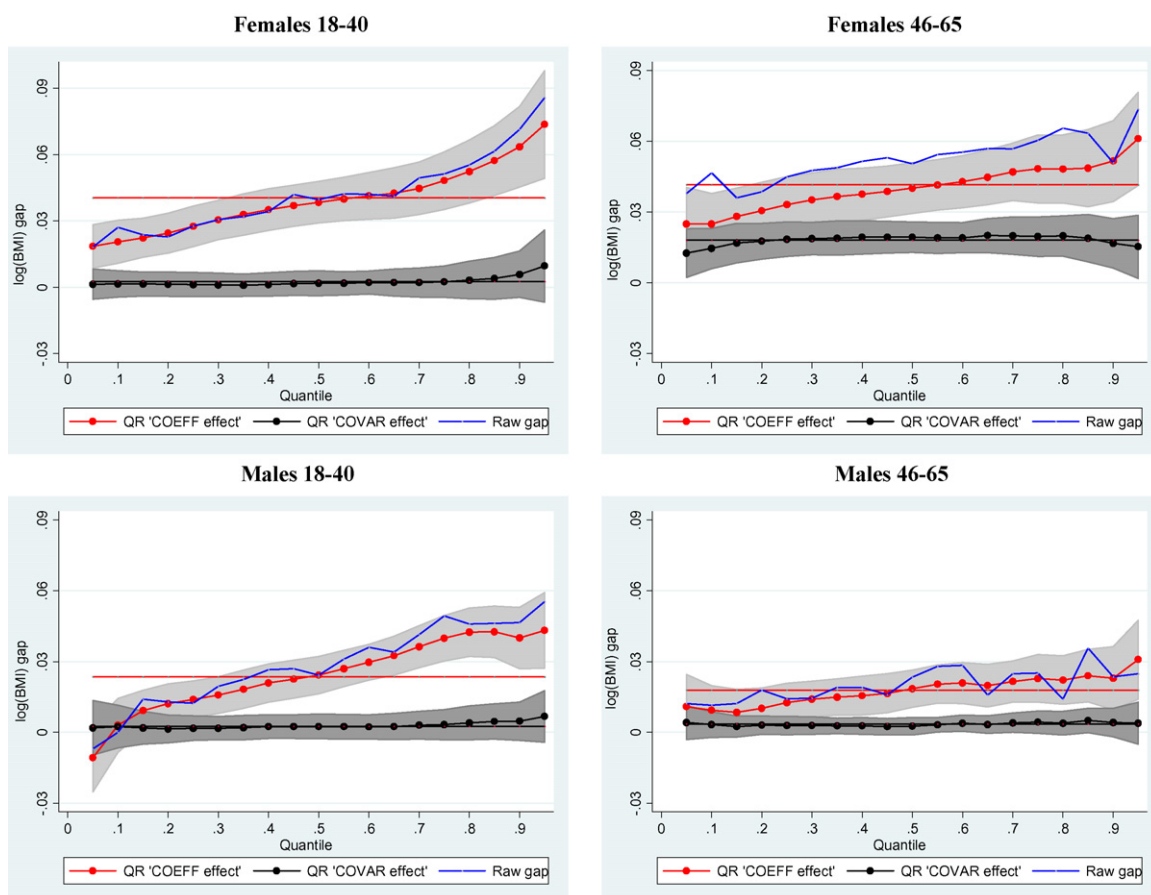


Fig. 5. Spain-to-Italy log(BMI) gap counterfactual decomposition analysis: by age class and gender.

ment so that its exclusion might significantly change our results. Since this piece of information is only available for Spain we cannot check the robustness of our decomposition analysis to the exclusion of income. As an indirect sensitivity analysis we examine its impact on individual BMI in the Spanish sample. See [Appendix A](#). The evidence ([Table A1](#)) suggests that in the Spanish sample we cannot exclude that income exerts an independent effect (in the case of females). Anyway excluding income does not seem to have a major impact on partial effects of included regressors.

As a final robustness check we replicate our decomposition analysis including a set of food consumption habit dummies. We add controls for the individual being a daily consumer of beer or wine, having breakfast daily, seldom consume meat and follow a “Mediterranean diet”.²³ These are clearly endogenous. Their impact both on the linear and the counterfactual quantile decomposition is close to negligible ([Tables A2 and A3](#)).

7. Conclusions

This paper has addressed the issue of how to conduct cross-country comparisons in BMI. Using a quantile regression approach we propose a counterfactual decomposition of BMI differentials evaluated at different quantiles. We apply this strategy to a representative sample of Italy and Spain, two Mediterranean countries with similar income per capita levels, cultural values and socio-economic characteristics. Descriptive evidence suggests that BMI values are proportionally higher in Spaniards than in Italians and are shifted towards the upper tail of the distribution, meaning that Spanish BMI distributions are more dispersed than Italians. We apply both Blinder–Oaxaca linear decomposition analysis and counterfactual quantile decomposition according to [Machado and Mata \(2005\)](#) to assess the role of differences in the effects of underlying determinants on BMI in the two countries vis-à-vis the cross-country differences in the distribution of underlying determinants. Our analysis shows that the counterfactual quantile decomposition analysis can provide insights which otherwise would be concealed if we were to rely on the standard linear decomposition.

Some limitations of our study should also be considered. Some relevant determinants of cross-country differences in obesity are omitted in our analysis. For instance we could not control for differences in household incomes or assets since this piece of information was only available for Spain. Moreover we opted for a conservative choice of regressors in order to limit endogeneity problems. For this reason we decided to ignore also the role played by cross-country differences in food consumption habits. Our robustness checks suggest that at least for these regressors their exclusion does not seem to affect our results. A further limitation concerns causality. Given that we do not

have access to longitudinal data and natural experiments, we cannot claim causality for the partial effects we estimate. Finally, as long as measured anthropometric information is not available in our datasets we are unable to assess the impact on our results of a potential differential cross-country pattern in self-reporting weight and height.

Given the limited explanatory value of our regressors for males' BMI we mainly devote the concluding comments to the results pertaining to the female populations. Compared with Italians, Spanish women have a rightward shifted and more dispersed distribution of BMI towards the upper tail. The Blinder–Oaxaca linear decomposition analysis suggests that differences in the distribution of covariates are responsible for only a small part of the rightward locational shift in the BMI distribution of the Spanish population, while the part due to coefficients is five times as large. Counterfactual quantile decomposition analysis shows that differences in the underlying distribution in covariates reflects a small part of this dispersion towards the upper tail compared with the larger role played by the differences in the returns to underlying determinants. We performed the same decomposition analyses in two distinct age groups: “young” (aged 18–40 years) and “middle-aged” (aged 46–65). For young females, the differences in the returns to common underlying determinants explain most of the rightward shift and thicker tails observed in the Spanish distribution of BMI. For middle-aged women, the differences in the distribution of underlying determinants appear to capture a non-negligible part of the rightward shift in the distribution; nonetheless higher dispersion towards the very upper tail proves to be mostly captured by a coefficient effect. Therefore, as Italy is assumed to be the most appropriate benchmark for Spain, our exercise suggests that BMI values in Spanish females, in particular those in the younger age group, are higher and more dispersed towards the upper tail as the result of individual behaviours that deviate from their Italian counterfactuals. Younger Spanish females appear to be in greater need of prevention interventions addressing these behaviours than middle-aged ones. Unfavourable underlying determinants explain a non-negligible part of the gap for the latter.

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 XREEPXarxa de Referència en Economia i Polítiques Públiques) and Joan Costa-Font acknowledges the funds from the Fundació Ramon Llull. Comments and remarks from the editor and four anonymous referee were valuable in helping us to improve the paper. Usual disclaimers apply.

²³ We consider the individual to follow a Mediterranean diet if he declares to jointly consume meat less frequently than daily, milk and cheese more frequently than weekly and vegetables, fruit, pasta and bread daily.

Table A1

The OLS determinants of log(BMI) in Spain.

Variables	Females	Females	Males	Males
Age	0.0062*** [0.0010]	0.0062*** [0.0010]	0.0112*** [0.0009]	0.0111*** [0.0009]
Age squared	−0.0027** [0.0012]	−0.0026** [0.0012]	−0.0104*** [0.0011]	−0.0104*** [0.0011]
Years of education	−0.0088*** [0.0004]	−0.0083*** [0.0004]	−0.0031*** [0.0004]	−0.0033*** [0.0004]
Employed	−0.0044 [0.0038]	−0.0020 [0.0039]	0.0039 [0.0043]	0.0030 [0.0043]
Married	0.0145*** [0.0042]	0.0142*** [0.0042]	0.0259*** [0.0038]	0.0272*** [0.0039]
Smoker	−0.0274*** [0.0039]	−0.0272*** [0.0039]	−0.0200*** [0.0031]	−0.0200*** [0.0031]
GDP	−0.0033*** [0.0005]	−0.0029*** [0.0005]	−0.0014*** [0.0004]	−0.0015*** [0.0004]
Income per capita		−0.0021*** [0.00052]		0.0007 [0.0005]
Missing income		−0.0189*** [0.0051]		0.0054 [0.0043]
Observations	7813	7813	7312	7312
R ²	0.232	0.234	0.130	0.130

Note: Standard errors (in italics) are computed using the Huber/White variance estimator.

** Significance levels at 5%.

*** Significance levels at 1%.

Appendix A

In the Spanish survey household income is collected in bracket form. We impute to each respondent the interval mean for each income bracket and convert it into per capita values by dividing it for the number of family components. Values are then denominated in hundreds of EU€. Individuals with missing value on income per capita are assigned a “missing” dummy variable. These comprise true missing and individuals that refused to report on family income.

Results are provided below. The general message is that concerning the Spanish sample we cannot exclude that income exerts an independent effect (at least in the case of females). Anyway excluding income does not seem to have a major impact on partial effects of included regressors.

Appendix B

See Tables A2 and A3.

Table A2

Spain-to-Italy log(BMI) gap linear decomposition (Blinder–Oaxaca) when food consumption habits are accounted for.

Linear decomposition	Estimate	95% confidence interval	
Females			
Coefficients	3.81%	0.033	0.043
Covariates	0.94%	0.006	0.013
Residual	0.002%		
Log BMI raw gap	4.75%		
Males			
Coefficients	2.01%	0.016	0.025
Covariates	0.44%	0.001	0.007
Residual	−0.06%		
Log BMI raw gap	2.38%		

Note: Terms in bold type are significant at 5% significance level.

This decomposition is performed including a set of food consumption habit variables in the specification. Namely four dummies for daily consumption of beer or wine, habitual breakfast in the morning, eating meat less than once a week or never, following a “Mediterranean diet”.

Table A3

Spain-to-Italy log(BMI) gap counterfactual quantile decomposition when food consumption habits are accounted for.

Quantile	BMI raw gap	Coefficients	95% confidence interval		Covariates	95% confidence interval	
Females							
10	2.4%	2.1% [91%]	0.008	0.034	0.3% [12%]	−0.0064	0.0122
25	3.5%	3.1% [87%]	0.021	0.041	0.6% [16%]	−0.0022	0.0133
50	4.4%	3.9% [89%]	0.028	0.050	0.9% [21%]	0.0016	0.0169
75	5.6%	4.6% [83%]	0.033	0.060	1.3% [24%]	0.0031	0.0235
90	6.1%	5.2% [85%]	0.033	0.070	1.6% [27%]	0.0036	0.0291

Table A3 (Continued)

Table 12 (continued)							
Quantile	BMI raw gap	Coefficients	95% confidence interval		Covariates	95% confidence interval	
Males							
10	1.1%	0.8% [74%]	−0.003	0.019	0.1% [14%]	−0.007	0.010
25	1.6%	1.5% [93%]	0.007	0.023	0.2% [13%]	−0.003	0.008
50	2.8%	2.3% [81%]	0.014	0.031	0.4% [14%]	−0.001	0.009
75	2.8%	2.8% [102%]	0.018	0.038	0.5% [18%]	−0.001	0.011
90	4.0%	2.6% [63%]	0.011	0.040	0.7% [18%]	−0.001	0.015

Note: In squared brackets we report the share of the total gap attributable to each component. Terms in bold type are significant at 5% significance level. This decomposition is performed including a set of food consumption habit variables in the specification. Namely four dummies for daily consumption of beer or wine, habitual breakfast in the morning, eating meat less than once a week or never, following a “Mediterranean diet”.

References

- Albrecht, J., Bjorklund, A., Vroman, S., 2003. Is there a glass ceiling in Sweden? *Journal of Labor Economics* 21 (1), 145–177.
- Arendt, J.N., 2005. Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review* 24, 149–160.
- Arulampalam, W., Booth, A.L., Bryan, M.L., 2007. Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *Industrial and Labor Relations Review* 60 (2), 163–186.
- Baum II, C.L., Ruhm, C.J., 2007. Age, socioeconomic status and obesity growth. NBER Working Paper Series, n° 13289.
- Blinder, A.S., 1973. Wage discrimination: reduced form and structural variables. *Journal of Human Resources* 8, 436–455.
- Boström, G., Diderichsen, F., 1997. Socio-economic differentials in misclassification of height, weight and body mass index based on questionnaire data. *International Journal of Epidemiology* 26 (4), 860–866.
- Branca, F., Nikogosian, H., Lobstein, T., 2007. The Challenge of Obesity in the WHO European Region and the Strategies for Response. World Health Organization.
- Brunello, G., Michaud, P.-C., Sanz-de-Galdeano, A., 2008. The rise in obesity across the Atlantic: an economic perspective. IZA Discussion Paper, n° 3529.
- Buchinsky, M., 1998. Recent advances in quantile regression models. *Journal of Human Resources* 33, 88–126.
- 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.
- 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., 2004. Social interactions and the contemporaneous determinants of individuals' weight. *Applied Economics* 36 (20), 2253–2263.
- Costa-Font, J., Gil, J., 2005. Obesity and the incidence of chronic diseases in Spain: a seemingly unrelated probit approach. *Economics and Human Biology* 3, 188–214.
- 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.M., Glaeser, E.L., Shapiro, J.M., 2003. Why have Americans become more obese? *Journal of Economic Perspectives* 17 (3), 93–118.
- Cutler, D.M., Lleras-Muney, A., 2006. Education and health: evaluating theories and evidence. NBER Working Paper Series, n° 12352.
- 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.
- García, 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.
- Gil, J., Mora, T., 2009. The determinants of misreporting weight and height: the role of social norms. FEDEA Working Paper no. 2009-01.
- Gruber, J., Frakes, M., 2006. Does falling smoking lead to rising obesity? *Journal of Health Economics* 25, 183–197.
- Hadi, A.S., 1992. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)* 54 (3), 761–771.
- 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.
- Komlos, J., Breitfelder, A., Sunder, M., 2009. The transition to post-industrial BMI values among US children. *American Journal of Human Biology* 21 (2), 151–160.
- Kuczmarski, M.F., Kuczmarski, R.J., Najjar, M., 2001. Effects of age on validity of self-reported height, weight and body mass index: findings from the third national Health and Nutrition Examination Survey, 1988–1994. *Journal of the American Dietetic Association* 101 (1), 28–34.
- Machado, J.A.F., Mata, J., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics* 20, 445–465.
- Melly, B., 2005. Public–private sector wage differentials in Germany: evidence from quantile regression. *Empirical Economics* 30 (2), 505–520.
- 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.
- Morris, S., 2006. Body mass index and occupational attainment. *Journal of Health Economics* 25, 347–364.
- Morris, S., 2007. The impact of obesity on employment. *Labour Economics* 14, 413–433.
- Oaxaca, R., 1973. Male–female wage differentials in urban labor markets. *International Economic Review* 14, 693–709.
- 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, n° 1814.
- World Health Organization, 1997. Obesity: Preventing and Managing the Global Epidemic. World Health Organization, Geneva, Switzerland.
- World Health Organization, 2003. Global Strategy on Diet, Physical Activity and Health. World Health Organization, Geneva, Switzerland.