A Profile of Obesity in Ireland, 2002-2007

David Madden, University College Dublin

A Profile of Obesity in Ireland, 2002-2007

David Madden

(University College Dublin)

March 2010

Abstract: Using the nationally representative Slan dataset we take a number of approaches to profile the change in obesity in Ireland over the 2002-2007 period. There is no evidence of either first or second order stochastic dominance between the two years. There is evidence that obesity and overweight are relatively more concentrated amongst males, the old and those with lower educational achievement. While obesity rose slightly over the period this was due to a rise in the average level of body mass index rather than a change in the shape of the distribution. Finally a semi-parametric decomposition of the change in the distribution over time indicates that the change in obesity arose not because of changes in population characteristics but rather the in the impact of these characteristics on body mass index.

Keywords: Obesity, Body Mass Index, Decomposition.

JEL Codes: I12, I31, I32.

Corresponding Author: David Madden,
          School of Economics,
          University College Dublin,
          Belfield,
          Dublin 4,
          Ireland.

Phone: 353-1-7168396
Fax: 353-1-2830068
E-Mail: david.madden@ucd.ie

A Profile of Obesity in Ireland, 2002-2007

Introduction

This paper provides a profile of the change in obesity in Ireland between 2002 and 2007, using the Slan surveys for those years. There are two dimensions to the profile provided. First of all we analyse the total distribution of body mass index (BMI) for both years and use stochastic dominance techniques to examine changes across the complete distribution of BMI. Secondly we employ a series of decompositions to obesity and the change in obesity over time. Obesity typically does not fall randomly across the population. It may differ by age, gender and education. We carry out decompositions of obesity measures along these three dimensions. We also employ decomposition techniques to examine the change over time. The Shapley decomposition breaks down the change in obesity into that part accounted for by a change in the *average* level of BMI and that part accounted for by a change in the *distribution* of BMI. Finally, we estimate a reduced form model of BMI and apply the semi-parametric decomposition approach of Dinardo, Fortin and Lemieux, (henceforth DFL, 1996) to examine the role of various factors in the change in obesity observed over the 2002-2007 period. This approach enables us to construct counterfactuals of the total distribution of BMI under different scenarios.

The remainder of the paper proceeds as follows. In the next section we make some general observations concerning the measurement of obesity and we motivate the stochastic dominance approach for examining the change in obesity between 2002 and 2003. We then discuss our data and present the stochastic dominance results. The next section carries out decompositions by age, gender and education while we

then explain and present results for the Shapley decomposition. We then present results for the DFL decomposition while the final section presents concluding comments.

Measuring Obesity

Obesity is clearly one of the most pressing health issues in the developed (and increasingly in the developing) world. For example, a recent edition of the *New England Journal of Medicine* is devoted to the topic with one article suggesting that:

> "Unless effective population-level interventions to reduce obesity are developed, the steady rise in life expectancy observed in the modern era may soon come to an end and the youth of today may, on average, live less healthy and possibly even shorter lives than their parents" (Olshansky et al., NEJM, 2005).

Meanwhile an editorial in *The Lancet* stated:

> "Excess bodyweight is one of the most blatantly visible, yet most neglected, risk factors contributing to the overall burden of disease worldwide. At least 1.1 billion adults and 10% of children are now overweight or obese, leading to decreased life expectancy due to cardiovascular disease, type 2 diabetes and some types of cancer" (The Lancet, 2006).

Given this extent of concern regarding obesity, it is important that we measure it accurately. The most common measure of obesity used is derived from body mass index (BMI). BMI is obtained by dividing weight (in kilos) by height (in metres) squared. The World Health Organisation defines overweight and obesity with respect to BMI as follows:

Range of BMI	Weight Definition
<20	Underweight
20-24.9	Normal Weight
25-29.9	Overweight
30-39.9	Obese
≥40	Severely Obese

Thus obesity is defined as a value of BMI greater than or equal to 30.

Note that there is criticism of BMI as a measure of obesity with some authors suggesting that other measures such as total body fat, percent body fat and waist circumference are superior measures of fatness (see Cawley and Burkhauser, 2006). Since much of the contribution of this paper is primarily methodological, while acknowledging the importance of this issue, we still feel it is useful to apply our approach to BMI as the likelihood is that it will remain the most commonly used indicator of obesity for the foreseeable future. Also, the approaches to analyzing obesity which we suggest here could in principle be applied to measures such as total body fat etc.

Most analysis of obesity usually proceeds by calculating the fraction of the population with BMI above the obesity threshold and proceeding from there. There are a number

of problems with this approach, however. First of all, using the simple fraction of the population with BMI above a particular threshold ignores much of the available information. It is a crude aggregate measure which is insensitive to how far above the threshold obese people are and is also insensitive to the distribution of BMI above the obesity threshold. These issues, and suggestions to overcome them, have been discussed in Jolliffe (2004) and Madden (2006) and are briefly reviewed again in section 4 below.

A second problem with the WHO approach to measuring obesity is the potential sensitivity of the measure to the choice of 30 as the obesity threshold. An individual with a BMI of 29.9 may be observationally indistinguishable from one with a BMI of 30. Yet one will count as obese and the other will not. Thus a comparison of obesity between two populations may be sensitive to the (arbitrary) choice of BMI threshold. Ideally, we would like to be able to compare obesity between two populations in a manner which is not sensitive to choice of BMI threshold. This is where stochastic dominance can help.

Probably the main application of stochastic dominance in economics is in relation to assets with monetary payoffs where it is used to rank the payoff distributions of assets in terms of their level of return and the dispersion of the return i.e. the level of risk attached to the asset. It can also been used in poverty and income distribution analysis and it is extremely useful when making non-parametric comparisons between distributions of continuous variables such as BMI.

Suppose we have two distributions of BMI with cumulative density functions (CDF) $F(BMI)$ and $G(BMI)$ respectively. Then CDF $F(BMI)$ first-order stochastically dominates $G(BMI)$ if and only if, for all monotone non-increasing functions $\alpha(BMI)$:

$$\int \alpha(BMI)dF(BMI) \leq \int \alpha(BMI)dG(BMI)$$

where the integral is taken over the whole range of BMI. Thus the average value of α is no greater in distribution F than in distribution G, as long as the valuation function is such that lower is better i.e. it is monotone non-increasing.[1] In this sense distribution F stochastically dominates distribution G. An equivalent way of expressing this is to say that for all BMI,

$$G(BMI) \leq F(BMI)$$

so that the CDF of distribution G is no greater than that of distribution F i.e. distribution F always has more mass in the lower part of the distribution.

In figure 1 we show two cumulative distributions of BMI, *F(BMI)* and *G(BMI)*. Here *F* stochastically dominates *G* in the sense referred to above. The relevance of this in the context of obesity can be shown as follows: suppose we decide there is a critical level of BMI, BMI*, which is the level denoting obesity. In figure 1, the value of the CDF at BMI* for distribution *F* is higher than that for distribution *G*. Thus the fraction of the population in distribution *F* with BMI less than the critical threshold is *lower* than that in distribution *G*. In other words the fraction with BMI in excess of BMI*, i.e. the fraction which is obese, is higher in distribution *G* than in distribution *F*. And because we have stochastic dominance in the sense that $G(BMI) \leq F(BMI)$

---

[1] In the case of BMI we can regard the α function as being similar to a health utility function which is non-increasing in BMI.

this will be true no matter what critical value of BMI is chosen to indicate obesity. Obesity is higher in *G* regardless of the chosen threshold.

Figure 2 shows an instance where stochastic dominance is not observed. At BMI* we have *F(BMI*)>G(BMI*)*, but at BMI** we have *F(BMI**)<G(BMI**)*. Thus the ranking of the distributions in terms of obesity is sensitive to the choice of threshold. In this instance there are two choices open to the analyst. One is to put a restriction on the range of BMI over which we look for stochastic dominance. Thus suppose we decide that a value of BMI_L is a reasonable lower bound for the obesity threshold (obviously there is no upper bound). Then we can search whether stochastic dominance is observed for values of BMI in excess of BMI_L.

An alternative is to place restrictions on the function α(BMI). This leads us on to the second type of stochastic dominance known as second-order stochastic dominance. We say that distribution $F(BMI)$ second-order stochastically dominates distribution $G(BMI)$ if and only if, for all monotone non-increasing and *convex* functions $\alpha(BMI)$ the previous inequality holds i.e.

$$\int \alpha(BMI)dF(BMI) \leq \int \alpha(BMI)dG(BMI)$$

Once again second-order stochastic dominance can be expressed equivalently as

$$D_G(BMI) = \int_{BMI_L}^{BMI^U} G(t)dt \leq \int_{BMI_L}^{BMI^U} F(t)dt = D_F(BMI)$$

so that second-order stochastic dominance is checked, not by comparing the CDFs themselves, but by comparing the integrals below them.[2] When $\alpha(BMI)$ is convex,

---

[2] Note that first order dominance necessarily implies second order dominance.

this implies that second-order dominance holds for obesity measures which are sensitive to the depth of obesity. We can then employ the Second-Order Dominance Condition above. Note the limits of the integrals. Because we are not worried about BMI levels below the lower bound, BMI_L, we only calculate the integrals for those whose BMI exceeds this lower bound up to the highest observed value of BMI, BMI^U.

Since first-order stochastic dominance implies second-order stochastic dominance, this check should only apply when first order dominance is not observed. It is possible to justify an obesity measure which is sensitive to the depth of obesity by noting that for many of the conditions listed above for which obesity is a contributory factor, the risk ratio is increasing with BMI.

We now discuss our data and present stochastic dominance results for 2002 and 2007.

Data and Stochastic Dominance Results

Our data comes from the Survey of Lifestyle, Attitudes and Nutrition in Ireland, usually known as the Slán (the Irish word for "safe") survey. The Slan surveys were carried out in 1998, 2002 and 2007. For our purposes in this paper, the correspondence between the questions asked in 2002 and 2007 is closest and so it is these two years which form the basis of our study. The Slan surveys are comprehensive, nationally representative surveys with sample sizes in 2002 and 2007 of 5992 and 10364 respectively. It is worth pointing out that Slan 2007 was a face-to-face interview in the respondent's house, while Slan 2002 was a self-completed postal survey. Both approaches have their advantages and disadvantages: while interviewers

can prompt and provide help to respondents in a face-to-face situation, the presence of the interviewer may affect the response to some questions. In the case of the self-reported survey there is always the danger than some respondents may not fully understand the question. Morgan et al (2008) provide greater detail.

Before examining the data for stochastic dominance, we first present summary statistics for BMI for the two years in question. Note we trim the data of the top and bottom 0.5% by BMI for fear of very large and very small values reflecting measurement error. Table 1 provides some information on BMI for 2002 and 2007. We can see that mean and median BMI have both increased slightly (by less than one per cent). The overweight rate (percentage of the sample with BMI over 25) has increased by about two per cent while the obesity rate (percentage of the sample with BMI over 30) has increased by less than one per cent.

Figure 2 presents kernel densities for BMI in 2002 and 2007, while figure 3 shows the difference in the densities. The distributions are quite close. 2002 shows somewhat more mass around the median value of about 25, while 2007 shows more mass in the 26-27 and 34-35 region.

In terms of eye-balling the CDFs there is one slight transformation which we have to carry out. Because our concern is with the proportion of the population *in excess* of a BMI threshold (rather than the proportion below an income threshold as would be the case in a poverty study) we need to examine the complementary cumulative distribution function (which we can then transform into an equivalent cumulative distribution function).

Thus for any threshold value of BMI, say BMI*, the cumulative distribution of body mass index gives the probability that the random variable (in this case BMI) takes on a value less than or equal to BMI* i.e. *F(BMI*)=P(BMI≤BMI*).* In examining obesity we are concerned with the complementary CDF i.e. $F_C(BMI^*) = P(BMI \geq BMI^*)$. However since it must be the case that $P(BMI \geq BMI^*) = P(k - BMI \leq k - BMI^*)$, for any constant, *k*, then analysing the complementary CDF for *BMI* is equivalent to analysing the CDF for *k-BMI*, where it is convenient to set $k = BMI_{max}$, the maximum value of BMI observed in either 2002 or 2007.[3]

As eye-balling the transformed CDFs can be quite difficult, instead we show the difference between them for a range of *k-BMI* from 0 to 25. This corresponds to a range of actual BMI from the maximum value observed, just less than 45, to 20, and it seems reasonable to suggest that any threshold of BMI for overweight or obesity is certain to lie within this interval. Figure 4 shows this difference, along with the 95% confidence interval. There is a clear crossing (at a value corresponding to BMI of about 37) and so first order dominance does not hold.[4] Obesity (and overweight) rates were generally lower in 2002 for mild and intermediate degrees of overweight and obesity, but they were *higher* for more severe degrees of obesity (i.e. BMI in excess of about 37).

---

[3] This transformation is necessary as we are drawing these curves using the DASP package of Arrar and Duclos which does not construct curves for complementary CDFs. I am very grateful for their permission to use the package.

[4] Figure 4 shows the crossing at a value of "transformed" BMI of about 7. Since *k* is set at a value just above 44, this implies that since *k-BMI*=7, then BMI must be about 37.

What about second-order stochastic dominance? Figure 5 shows the area under the CDF for *k-BMI* and once again we do not observe dominance with a crossing at a level of BMI corresponding to about 35.

Since we do not observe either first or second order dominance between 2002 and 2007 it is not possible to make unambiguous comments regarding the change in obesity in Ireland over this period. Any comparison of the two years must necessarily involve the use of a particular obesity index and a particular BMI threshold. However the use of specific indices does permit the decomposition of such indices in useful ways, and this is the subject matter of the next section.

Decomposition of Obesity

Obesity typically does not fall randomly across the population. Rates of obesity can differ according to a number of factors. The three factors which we examine here are age, gender and education. We apply decompositions for two different measures of obesity. The first we label BMI0 and it is the standard measure which simply gives the proportion of the population with BMI below a certain critical level (in this case we give results for two critical levels, 25 and 30). BMI1 is a measure which takes account of the depth of obesity. Suppose the critical threshold level of BMI above which people are obese is given by BMI^*, then the BMI gap for individual i will be given by $BMI_i - BMI^*$ (note the gap is only measured for people who are above the threshold). The obesity gap measure, which we label BMI1, is then given by the sum of these gaps expressed as a percentage of total BMI in the community. Thus

$$BMI1 = \frac{\sum\limits_{BMI_i > BMI^*} (BMI_i - BMI^*)}{n.\mu_{BMI}}$$

where μ_{BMI} is average BMI for the community (see Jolliffe, 2004 and Madden, 2006).

First of all we note that taking the population as a whole, tables 1 and 2 show slight increases in both BMI0 and BMI1 between 2002 and 2007 for both the obesity and overweight thresholds. However, the results broken down by age and gender are perhaps of more interest.

In terms of the breakdown by age and gender we partition the population into four groups, taking 45 as the threshold age between young and old. In terms of education we also use four groups: those with primary school education or less (left school at around age 12), those with intermediate second level schooling (left school around 15-16), those with complete secondary school education (left school at 17-18) and those with third level education.

Table 2 gives the breakdown of obesity and overweight by age and gender. There is a fairly clear age and gender dimension to both. Males have a higher rate of obesity and overweight than females, and so too do older people relative to younger. Hence the obesity rate for older males is more than 30% higher than for the population as a whole. Compared to younger females, older males have almost twice the rate of obesity. Interestingly, the age gradient becomes less severe for males in 2007, compared to 2002, yet it becomes more pronounced for females.

What about measures that take account of the depth of obesity? For males, the age gradient here is much less severe in the case of BMI in excess of 30. There is also no clear gender gradient to be observed either for this measure in the case where the BMI threshold is 30, though such a gradient is observed for the threshold of 25.

We now turn to the gradient by education. Here we see that when using a threshold of 30 there is a clear gradient in BMI0 for 2002, but this gradient is much less sharp in 2007. This reduction in the gradient between 2002 and 2007 can also be observed when using a threshold of 25. For the most part the reduction in the gradient is caused by an increase in BMI amongst those with 3rd level education.

When using measures which take account of the depth of obesity there is no real reduction in the gradient for the threshold of 30, though there is some reduction for the lower threshold of 25. Once again, these changes mostly arise from an increase in BMI amongst those with 3rd level education.

The above decompositions are useful in that they show how obesity can differ according to factors such as age, gender and education. However, these factors in themselves can be highly correlated and such correlations should be taken into account. This is carried out via regression analysis in section 6. Before that however in section 5 we break down the change in obesity between 2002 and 2007 into that part accounted for by changes in average BMI and that part accounted for by changes in the shape of the distribution of BMI.

The Shapley Decomposition Over Time

Changes in obesity over time can arise for one of either two reasons. Either the average level of BMI in the population rises or else, while average BMI remains unaffected, the distribution of BMI changes, with greater numbers of people above the critical threshold. Figure 6 illustrates both these cases. In the left hand diagram the probability density function for BMI shifts to the right, increasing the mean and consequently increasing the mass of the distribution to the right of the critical obesity threshold. In the right hand diagram, the mean remains the same. However, there is a mean-preserving spread in the distribution and so once again the mass of the distribution to the right of the critical obesity threshold increases. In most cases, there will be changes in both the mean and the distribution of BMI and it can be useful to decompose the total change in BMI into changes arising from the mean and changes arising from the distribution.

As we have observed already, the measurement of obesity has much in common with the measurement of poverty, in that both issues are concerned with critical thresholds. Given this similarity it is hardly surprising that the decomposition issue has been addressed in the poverty literature. Suppose we characterise our measure of obesity as $O = O(\mu, L, BMI^*)$ where μ is the average level of BMI, L is the Lorenz curve for the distribution of BMI and BMI^* is the critical obesity threshold (note that the cumulative distribution function for BMI will be completely characterised by its mean and Lorenz curve).

If subscripts "0" and "1" refer to the two time periods in question, then the change in obesity over time $O_1 - O_0$ can be written as

$$O_1 - O_0 = F_1(BMI^*) - F_0(BMI^*) = O_1(\mu_1, L_1, BMI^*) - O_0(\mu_0, L_0, BMI^*)$$

where F_i is the cumulative distribution function for period "i". This can then be decomposed into growth and redistribution effects denoted by $O(\mu_1, L_0, BMI^*) - O(\mu_0, L_0, BMI^*)$ and $O(\mu_1, L_1, BMI^*) - O(\mu_1, L_0, BMI^*)$ respectively.

However, as is the case with any path dependence type problem, the choice of which configuration to use as the base period is arbitrary. Here in the growth part we calculate the marginal effect of the change in mean BMI with the distribution held constant at the *initial* configuration. However, we calculate the marginal impact of redistribution holding mean BMI constant at the *final* configuration. We could just vas easily have carried out a decomposition with the base periods changed and there is no logical reason for preferring one configuration over another. Following the approach outlined in Shorrocks and Kolenikov and Shorrocks we take the average of the two effects respectively thus giving a growth effect of

$$\frac{1}{2}[O(\mu_1, L_0, BMI^*) - O(\mu_0, L_0, BMI^*)] + \frac{1}{2}[O(\mu_1, L_1, BMI^*) - O(\mu_0, L_1, BMI^*)]$$

and a redistribution effect of

$$\frac{1}{2}[O(\mu_0, L_1, BMI^*) - O(\mu_0, L_0, BMI^*)] + \frac{1}{2}[O(\mu_1, L_1, BMI^*) - O(\mu_1, L_0, BMI^*)].$$

As explained in Shorrocks (1999) these two expressions are the growth and distribution components for a two-way Shapley decomposition of the change in

obesity. The Shapley decomposition arises from the classic co-operative game theory problem of dividing a pie fairly. The solution is that each player is assigned his marginal contribution averaged over all possible coalitions of agents. Shorrocks' interpretation was to consider the various n factors which contribute together to determine the value of an indicator such as obesity or poverty and then assign to each factor the average marginal contributions taken over the n! possible ways in which the factors may be removed in sequence. The decomposition is always exact as the factors are treated symmetrically.

Figure 7 shows a two-way Shapley decomposition for the change in obesity into a growth and distribution component. Since we have two factors (n=2, growth and distribution) we have 2!=2 possible routes. If we also allowed the obesity threshold to change then we would have n=3 and 3!=6 possible routes.

Table 5 presents results for the Shapley decomposition of obesity for the two measures of obesity, BMI0 and BMI1.

Dealing first of all with the case of obesity (BMI>30), the actual change in BMI0 over the period in question was small, an increase of less than one per cent. This change was down solely to a change in the average level of BMI. In fact, had average BMI remained constant then changes in the distribution of BMI would in fact have led to a fall in obesity. Similarly BMI1, which takes account of the depth of obesity, also showed a very small increase and once again all of this was accounted for by changes in the average level of BMI.

For the case of overweight (BMI>25) the change between 2002 and 2007 was greater in magnitude than for obesity. Once again, the bulk of the change was accounted for by a change in the average level of BMI. Three quarters of the change in BMI0 was accounted for by the average level of BMI, while in the case of BMI1 the increase in the average level accounted for 98 per cent of the change.

Are there any policy implications to be drawn from these decompositions? Most of the change in obesity is accounted for by a change in the average level of BMI, as opposed to a change in BMI amongst the overweight/obese. This suggests that policies to combat obesity should perhaps be targeted at a general audience, rather than at the more specific group towards the right hand tail of the distribution. But it is also worth remembering that the overall change in obesity rates is quite modest.

The Dinardo-Fortin-Lemieux Decomposition

The decompositions in sections 4 and 5 looked at the change in obesity according to the dimensions of age, gender and education and also in terms of a shift in the average level of BMI versus a change in the shape of the distribution. Section 6 looks at the decomposition of the change in the distribution of BMI due to certain sets of factors. Following the seminal paper by DFL we employ a sequential counterfactual approach to analysing the effect of each set of factors. This involves a re-weighting of the sampling weights for each individual (for gender, socio-economic characteristics and self-assessed health) and a rescaling of BMI (reflecting changes in the "returns" to BMI accounted for by various factors).

17

This approach is similar in many ways to the classic Blinder-Oaxaca (BO) decomposition (Blinder, 1973 and Oaxaca, 1973), except that while the BO decomposition provides a decomposition evaluated at the mean (which is arguably of relatively little interest in an obesity study), the DFL procedure provides the decomposition across the whole of the distribution. In our application we are looking at the decomposition of the change in the distribution of BMI over time (as opposed to decomposition by, say, gender). The approach involves the sequential construction of counterfactual distributions for changes in various sets of factors. The set of factors we choose to focus on are the socio-demographic attributes of individuals, their self-assessed health and the "returns", in terms of obesity, to these factors.

Suppose that we are decomposing the change in BMI between two time periods (2002 and 2007 which we label period "0" and "1" respectively). We estimate the probability density of BMI in period t, $f_t(y)$, using kernel density methods. Thus if $(y_{t1}, \ldots, y_{tN})$ is a random sample of N observations with sampling weights $(\theta_{t1}, \ldots, \theta_{tN})$ and $\sum \theta_{ti} = 1$, the kernel density estimate of $f_t(y)$ is

$$\hat{f}_t(y) = \sum_{i=1}^{N} \frac{\theta_{ti}}{b} . K\left(\frac{y - y_{ti}}{b}\right)$$

where K is the kernel function and b is the bandwidth. We use the kernel density facility in STATA with the Epanechnikov kernel and the bandwidth optimally chosen. We have already seen the kernel density of BMI in 2002 and 2007 in figure 2 while figure 3 shows the difference between the kernel densities (the density for BMI in

18

2002 minus the density for BMI in 2007).   We now explain how we go about constructing the sequence of counterfactuals.

The sequence of counterfactuals involves either a re-weighting of sample weights for each individual (reflecting changes in socio-demographic attributes and health outcomes) or a re-scaling of BMI (reflecting different "returns" to BMI). The first counterfactual we examine is that of socio-demographic attributes. If we believe that BMI is correlated with certain socio-demographic attributes, then it seems reasonable that changes in such socio-economic attributes could lead to changes in observed BMI, without any change in underlying "behaviour". Thus we estimate how certain socio-demographic attributes vary between 2002 and 2007. The attributes chosen are age, gender, education, marital status, principal economic status, and smoking status. The estimated relationship is then used to adjust the 2002 sampling weights to reflect the change in attributes between 2002 and 2007. This will give greater weight to those with attributes more similar to 2007 attributes and less weight to those household with attributes which are less similar.

Thus the density of BMI in period t is expressed as the integral of the density of BMI conditional on household demographic attributes, x, and self-assessed health outcomes, h:

$$f_t(y) = \int_{(h,x)\in\Omega_{(e,x)}} f_t(y; x, h) dF_t(h, x)$$

where $\Omega_{(h,x)}$ is the domain of individual health outcomes and socio-demographic attributes and $F_t(h, x)$ is the joint distribution of (h, x) for an individual in period t.

the sequence we are following first analyses the effects of changes in socio-demographic attributes and then changes in health, conditional on socio-demographic attributes. Thus we re-write the above expression as

$$f_t(y) = \int_{x \in \Omega_x} \int_{h \in \Omega_{h|x}} f_t(y; x, h).dF_t(h|x)dF_t(x).$$

Thus we construct a counterfactual density allowing the distribution of socio-demographic attributes to be as observed in 2007 but holding the conditional distribution of health outcomes and density of BMI to be as in 2002.

$$f_0^X(y) = \int_{x \in \Omega_x} \int_{h \in \Omega_{h|x}} f_0(y; x, h).dF_0(h|x)dF_1(x)$$

$$= \int_{x \in \Omega_x} \int_{h \in \Omega_{h|x}} f_0(y; x, h).dF_0(h|x)\psi_x(x)dF_0(x)$$

where we have replaced $dF_1(x)$ with $\psi_x(x)dF_0(x)$ and $\psi_x(x) = \dfrac{dF_1(x)}{dF_0(x)}$ is a re-weighting function which rescales the period 0 density of attributes to obtain the period 1 density. Using Bayes' rule, this function can be expressed as

$$\psi_x(x) = \frac{P(t=1, x)}{P(t=0, x)} = \frac{P(t=1|x)}{P(t=0|x)} \cdot \frac{P(t=0)}{P(t=1)}$$

and $P(t=1|x)$ is the conditional probability that a household with attributes x is observed in period 1, while $P(t=1)$ is the unconditional probability that the household is observed in period 1.

Thus in order to obtain estimates for the re-weighting function, we first pool observations for period 0 and period 1 and then estimate the probability that individual i is observed in period 1, given attributes x using a logit model for the binary dependent variable.[5] Estimates from this model can then be used to predict, for each individual observed in period 0, the relative probability that it would be observed in period 1 versus period 0 i.e. $\dfrac{P(t=1|x)}{P(t=0|x)}$. We can then adjust this by the ratio of the unconditional probabilities, $\dfrac{P(t=0)}{P(t=1)}$, to obtain the estimated re-weight for this individual, $\hat{\psi}_x(x_0)$. The counterfactual density for BMI which takes account of the changes in attributes is then given by

$$\hat{f}_0^X(y) = \sum_{i=1}^{N} \frac{\hat{\psi}_x(x_0)\theta_{0i}}{b} . K\left(\frac{y - y_{0i}}{b}\right)$$

and the estimated marginal effect of the change in the distribution which is explained by socio-demographic attributes is

$$\Delta \hat{f}^X(y) = \hat{f}_0^X(y) - f_0(y).$$

Figures 8a and 8b provide the counterfactual distribution and the change in distribution for socio-demographic attributes. From figure 8a it is clear that the two densities are very close together. However an examination of figure 8b shows that if the only factor that had changed between 2002 and 2007 was the change in socio-demographic attributes then the density in the low 20s would have increased, while that in the low to mid 30s would have decreased. Figure 3 however shows that in

---

[5] Results of this estimation and others used in the calculation of these counterfactuals are available from the author on request.

actual fact, the outcome was close to, though not exactly, the mirror image, with decreases in the low 20s, and increases in the late 20s and mid to late 30s.

This is confirmed by the results in table 7, where the measures of obesity and overweight for 2002(1) i.e. the first counterfactual, all show small decreases, compared to the actual measures in 2002 e.g. if the only factor that had changed between 2002 and 2007 was the change in socio-demographic attributes, then the rate of obesity would have fallen from 0.133 to 0.127.

We next construct a counterfactual density which permits the distribution of self-assessed health outcomes to be as in 2007 but the distribution of BMI conditional upon health to be as in 2002. Thus we would have

$$f_0^{HX}(y) = \int_{x\in\Omega_x} \int_{h\in\Omega_{h|x}} f_0(y;x,h).dF_1(h|x)dF_1(x)$$

$$= \int_{x\in\Omega_x} \int_{h\in\Omega_{h|x}} f_0(y;x,h).\psi_{h|x}(h,x)dF_0(h|x).\psi_x(x)dF_0(x)$$

where we have replaced $dF_1(h|x)$ with $\psi_{h|x}(h,x)dF_0(h|x)$ and $\psi_{h|x}(h,x) = \dfrac{dF_1(h|x)}{dF_0(h|x)}$

is a re-weighting function which rescales the period 0 density of health outcomes conditional on socio-demographic attributes to obtain the period 1 density. Given that health is self-assessed on a five-point scale there are five potential outcomes and we define $h_m = 1$ if the individual chooses health state m and $h_m = 0$ otherwise. Thus the re-weighting function is

$$\psi_{h|x}(h,x) = \sum_{m=1}^{5} h_m \cdot \frac{P_1(h_m = 1|x)}{P_0(h_m = 1|x)}$$

where $P_i(h_m = 1|x)$ is the probability of health state m in period i, given attributes x.

Thus to estimate the re-weighting function we estimate for 2002 and 2007 ordered logit models for health states conditional on attributes and obtain predicted probabilities for each outcome, $\hat{P}_t(h_m = 1|x_{t,i})$. For each household observed in 2002 we can use the predicted value from this model to predict the relative probability of health outcome h_m in 2007 compared to 2002 and calculate

$$\psi_{h|x}(h_{0,i}, x_{0,i}) = \sum_{m=1}^{5} h_{m0i} \cdot \frac{P_1(h_m = 1|x_{0i})}{P_0(h_m = 1|x_{0i})} \cdot$$

Given this value for the estimated re-weight, the counterfactual for individual i which takes account of changes in health states is estimated by

$$\hat{f}_0^{H,X}(y) = \sum_{i=1}^{N} \frac{\hat{\psi}_{h|x}(h_{0i}, x_{0i}) \cdot \hat{\psi}_x(x_0)\theta_{0i}}{b} \cdot K\left(\frac{y - y_{0i}}{b}\right).$$

The marginal effect of changes in health states is then given by:

$$\Delta\hat{f}^{H,X}(y) = \hat{f}_0^{H,X}(y) - \hat{f}_0^{X}(y).$$

The results of this counterfactual can be seen by examining figures 9a and 9b and looking at the results for 2002(2) in table 7. These show a further slight fall in all obesity measures, except for BMI1 when the threshold is 30, where there is a slight increase. However note that even though BMI1 increases compared to 2002 (1), it is still lower than for the original 2002 distribution.

The final factor which we wish to take account of is the change in the BMI "return" to attributes and health states. We construct a counterfactual density allowing the BMI returns to attributes to be as in 2007, by adjusting each individual's 2002 BMI by the predicted change, given their attributes.

Thus we regress BMI on a vector of attributes for 2002 and 2007, $\hat{y}_{ti} = X'_{ti}\beta_t + \varepsilon_{ti}$. We then compute the predicted change in returns $\Delta\hat{y}_i = X'_{0i}(\hat{\beta}_1 - \hat{\beta}_0)$ and $\hat{y}^R_{0i} = \hat{y}_{0i} + \Delta\hat{y}_i$ and obtain the counterfactual density which takes account of these changes

$$\hat{f}_0^{R,H,X}(y) = \sum_{i=1}^{N} \frac{\hat{\psi}_{h|x}(h_{0i}, x_{0i}).\hat{\psi}_x(x_0)\theta_{0i}}{b} . K\left(\frac{y - \hat{y}^R_{0i}}{b}\right)$$

and the marginal effect is given by

$$\Delta\hat{f}^{R,H,X}(y) = \hat{f}_0^{R,H,X}(y) - \hat{f}_0^{H,X}(y).$$

Then the total change in BMI between 2002 and 2007 can be decomposed into explained and unexplained components:

$$\hat{f}_1(y) - \hat{f}_0(y) = (\hat{f}_0^{R,H,X}(y) - \hat{f}_0(y)) + (\hat{f}_1(y) - \hat{f}_0^{R,H,X}(y))$$

where $\hat{f}_0^{R,H,X}(y) - \hat{f}_0(y)$ represents the total change in the distribution explained by the sets of factors we have examined here and $\hat{f}_1(y) - \hat{f}_0^{R,H,X}(y)$ represents the change in the distribution which is unexplained.

The result of this counterfactual is more dramatic than the preceding ones. All measures of obesity and overweight show quite a large increase compared to the previous counterfactual, 2002 (2), and in most cases they are above the measures for 2002 and 2007, the exception being BMI1 with a threshold of 30. This is reflected in figures 10a and 10b, where distribution for 2002 lies below that of 2002 (3) for most values of BMI, except those in excess of about 33. Similarly, figures 11a and 11b show the comparison between 2002 (3) and 2007, and for the most part the distribution for 2002 (3) lies above that of 2007, once again the exception being values of BMI between about 33 and 37.

Conclusions

This paper has provided an overview of the change in obesity (as measured by BMI) in Ireland from 2002 to 2007. We found that neither first nor second order stochastic dominance held, thus it is not possible to make unambiguous statements regarding the change in obesity i.e. statements which are independent of the chosen BMI threshold for obesity. When the conventional thresholds of 30 and 25 are adopted then there is a small increase in both obesity and overweight between 2002 and 2007.

Analysis of obesity by age, gender and education level reveals quite appreciable differences across these categories. Being male, being older and having lower education achievement are all associated with higher rates of obesity and overweight. There is some evidence that the age gradient for males falls somewhat over the period under review, while that age gradient for females increase. The education gradient appears to diminish between 2002 and 2007 and this is mainly arises from an increase in obesity amongst those with third level education.

A Shapley decomposition of the change in obesity over time also reveals that all of the increase is accounted for by an increase in the average level of BMI, as opposed to a change in the distribution of BMI towards the right-hand tail. When looking at overweight, the change in the distribution makes some contribution, but the bulk of the change once again arises from an increase in average BMI.

Finally, the DFL decomposition across the whole of the distribution suggests that none of the (admittedly quite small) change in obesity is accounted for by changes in the characteristics of the population. Rather it is changes in the effect of these characteristics on BMI (the "returns" to the characteristics) which is the main driving force.

References

Blinder, A. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources*. 8: 436-455.

Cawley, J., and R. Burkhauser. 2006. "Beyond BMI: The Value of More Accurate Measures of Fatness and Obesity in Social Science Research." *National Bureau of Economic Research*. Working Paper No. 12291.

"Curbing the Obesity Epidemic". Editorial. *The Lancet*, 367: 1549.

DiNardo, J., N. Fortin and T. Lemieux. 1996. "Labour Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, 64: 1001-1044.

Jolliffe, D. 2004. "Continuous and Robust Measures of the Overweight Epidemic: 1971-2000." *Demography,* 41, 303-314.

Kalenikov. S., and A. Shorrocks. 2005. **"**A Decomposition Analysis of Regional Poverty in Russia." *Review of Development Economics*. 9: 25-46.

Madden, D. 2006. "Body Mass Index and the Measurement of Obesity.", *UCD Centre for Economic Research* Working Paper, WP06/27.

Morgan, K., H. McGee, D. Watson, I. Perry, M. Barry, E. Shelley, J. Harrington, M. Molcho, R. Layte, N. Tully, E. van Lente, M. Ward, J. Lutomski, R. Conroy, R. Brugha. 2008. *SLAN 2007: Survey of Lifestyle, Attitudes and Nutrition in Ireland.* Main Report, Dublin. Department of Health and Children.

Oaxaca, R. 1973. "Male-Female Wage Differentials in Urban Labour Markets." *International Economic Review*. 14: 693-709.

Olshansky, S.J., D. Passaro, R. Hershow, J. Layden, B. Carnes, J. Brody, L. Hayflick, R. Butler, D. Allinson, and D. Ludwig. 2005. "A Potential Decline in Life Expectancy in the United States in the 21st Century." *New England Journal of Medicine*, 352: 1138-1145.

Shorrocks, A. 1999. *Decomposition Procedures for Distributional Analysis: a Unified Framework Based on the Shapley Value*, manuscript, University of Essex.

Table 1: BMI Summary Statistics, 2002 and 2007

Year	Mean	Median	% above 25	% above 30
2002 (N=5481)	25.39	24.82	47.84	13.32
2007 (N=9646)	25.56	25.04	50.05	14.08

Table 2: Relative Obesity (BMI>30) by Age and Gender

	2002				2007			
	BMI0		**BMI1**		**BMI0**		**BMI1**	
Group	**Rate**	**Relative Share**	**Rate**	**Relative Share**	**Rate**	**Relative Share**	**Rate**	**Relative Share**
Young M	0.128	*0.962*	0.012	*0.882*	0.128	*0.910*	0.013	*0.090*
Old M	0.181	*1.356*	0.015	*1.088*	0.184	*1.309*	0.016	*1.094*
Young F	0.107	*0.802*	0.013	*0.919*	0.105	*0.744*	0.012	*0.079*
Old F	0.133	*1.001*	0.015	*1.103*	0.154	*1.093*	0.018	*1.223*
Pop	**0.133**	**1.000**	**0.014**	**1.000**	**0.141**	**1.000**	**0.015**	**1.000**

Table 3: Relative Overweight (BMI>25) by Age and Gender

	2002				2007			
	BMI0		**BMI1**		**BMI0**		**BMI1**	
Group	**Rate**	**Relative Share**	**Rate**	**Relative Share**	**Rate**	**Relative Share**	**Rate**	**Relative Share**
Young M	0.546	*1.141*	0.767	*1.053*	0.532	*1.063*	0.076	*0.992*
Old M	0.648	*1.356*	0.096	*1.321*	0.665	*1.328*	0.099	*1.290*
Young F	0.343	*0.716*	0.056	*0.767*	0.354	*0.707*	0.055	*0.723*
Old F	0.464	*0.971*	0.073	*1.000*	0.500	*0.999*	0.082	*1.066*
Pop	**0.478**	**1.000**	**0.073**	**1.000**	**0.501**	**1.000**	**0.077**	**1.000**

Table 4: Relative Obesity (BMI>30) by Education

Group	2002 BMI0 Rate	2002 BMI0 Relative Share	2002 BMI1 Rate	2002 BMI1 Relative Share	2007 BMI0 Rate	2007 BMI0 Relative Share	2007 BMI1 Rate	2007 BMI1 Relative Share
Primary	0.191	*1.429*	0.018	*1.303*	0.172	*1.220*	0.020	*1.351*
Inter	0.175	*1.310*	0.018	*1.343*	0.176	*1.254*	0.019	*1.289*
Leaving	0.131	*0.982*	0.015	*1.106*	0.133	*0.944*	0.013	*0.881*
3rd Level	0.082	*0.613*	0.008	*0.597*	0.112	*0.800*	0.011	*0.761*
Pop	**0.133**	**1.000**	**0.014**	**1.000**	**0.141**	**1.000**	**0.015**	**1.000**

Table 5: Relative Overweight (BMI>25) by Education

Group	2002 BMI0 Rate	2002 BMI0 Relative Share	2002 BMI1 Rate	2002 BMI1 Relative Share	2007 BMI0 Rate	2007 BMI0 Relative Share	2007 BMI1 Rate	2007 BMI1 Relative Share
Primary	0.570	*1.193*	0.095	*1.305*	0.578	*1.154*	0.094	*1.226*
Inter	0.561	*1.175*	0.090	*1.240*	0.556	*1.110*	0.090	*1.176*
Leaving	0.465	*0.974*	0.074	*1.012*	0.472	*0.942*	0.071	*0.930*
3rd Level	0.390	*0.816*	0.052	*0.713*	0.454	*0.907*	0.065	*0.847*
Pop	**0.478**	**1.000**	**0.073**	**1.000**	**0.501**	**1.000**	**0.077**	**1.000**

Table 6: Shapley Decomposition for Change in Obesity 2002-2007

Overweight, BMI>25				
Measure	**2002**	**2007**	**Growth (%)**	**Distribution(%)**
BMI0	0.478	0.500	76.4	23.6
BMI1	0.073	0.077	98.2	1.8
Obese, BMI>30				
Measure	**2002**	**2007**	**Growth (%)**	**Distribution(%)**
BMI0	0.133	0.141	135.5	-35.5
BMI1	0.014	0.015	100.3	-0.3

Table 7: Sequential Obesity and Overweight

Overweight, BMI>25					
Measure	**2002**	**2002 (1)**	**2002 (2)**	**2002 (3)**	**2007**
BMI0	0.478	0.473	0.472	0.517	0.500
BMI1	0.073	0.071	0.0705	0.078	0.077
Obese, BMI>30					
Measure	**2002**	**2002 (1)**	**2002 (2)**	**2002 (3)**	**2007**
BMI0	0.133	0.127	0.126	0.141	0.141
BMI1	0.014	0.012	0.013	0.014	0.015

Figure 1: Stochastic Dominance Observed
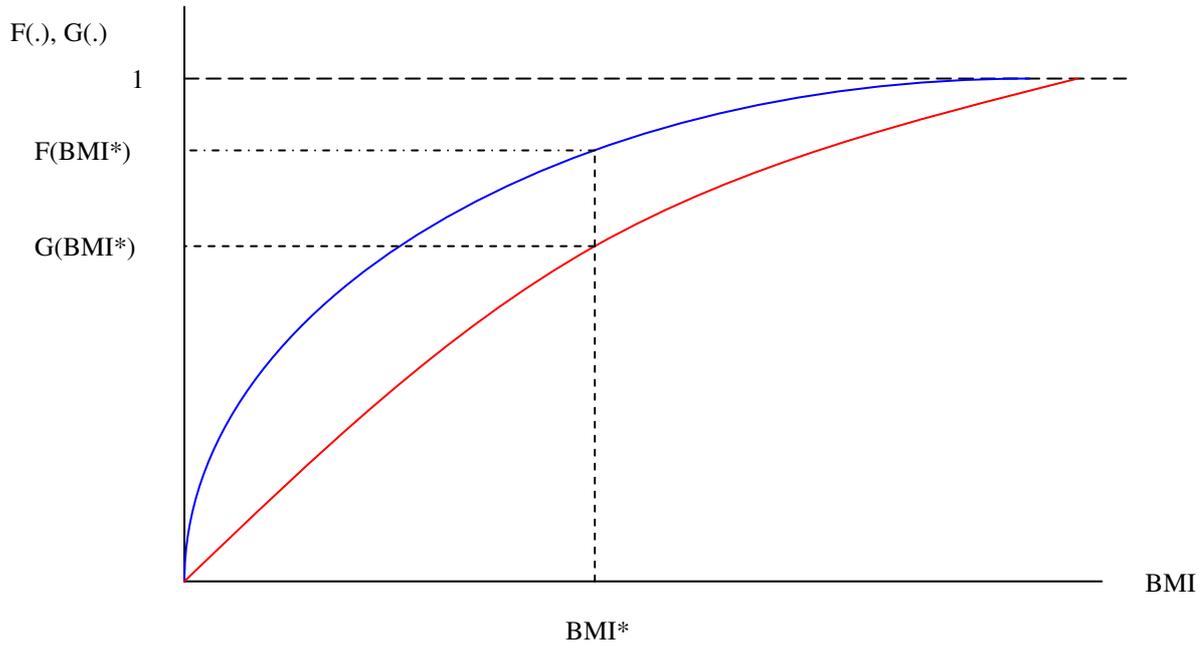
Figure 2: Stochastic Dominance Not Observed
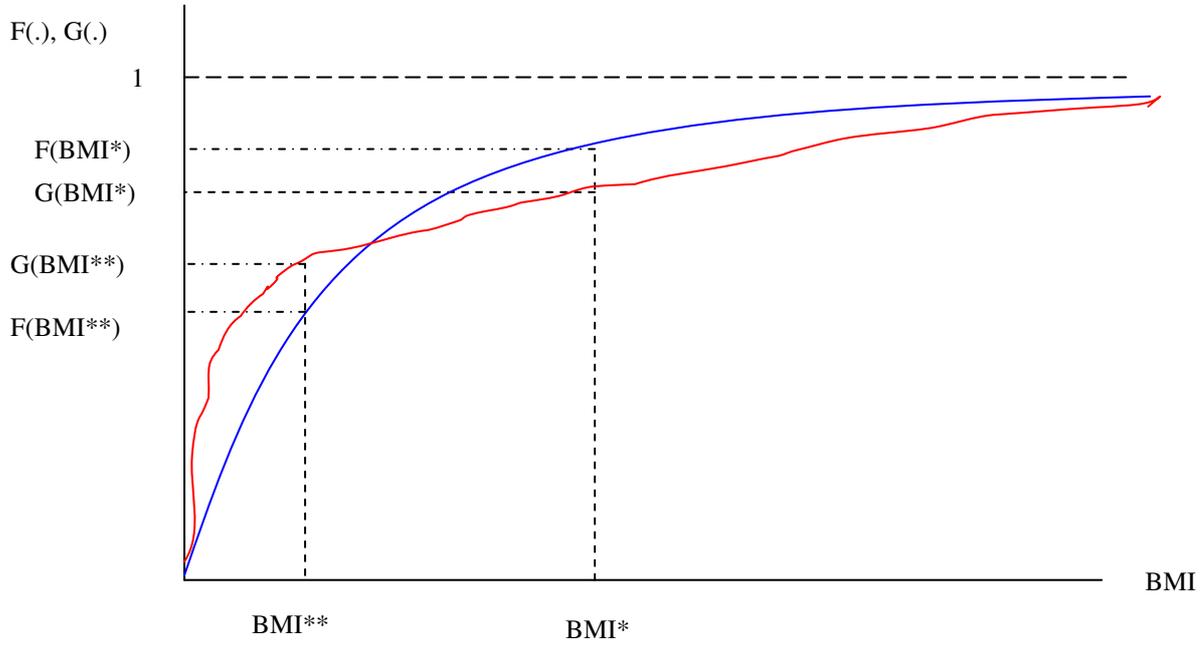
Figure 2: Kernel Densities, BMI, 2002 and 2007
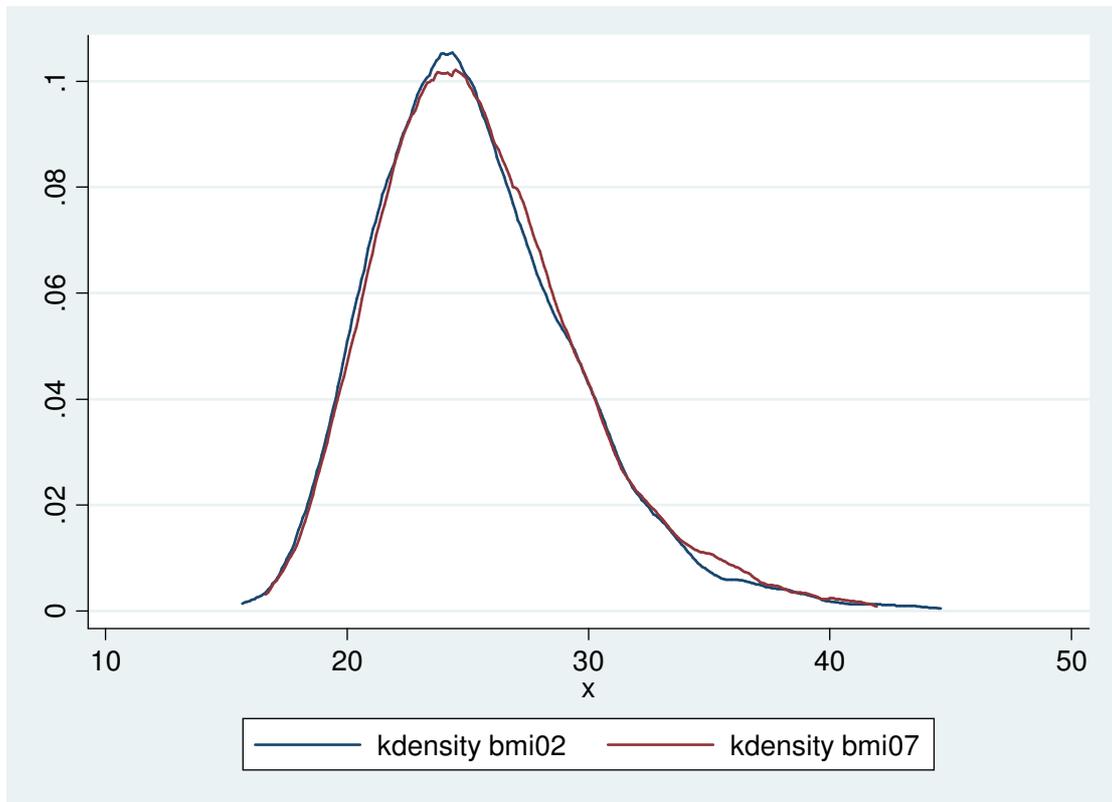
Figure 3: Change in Kernel Distribution between 2002 and 2007
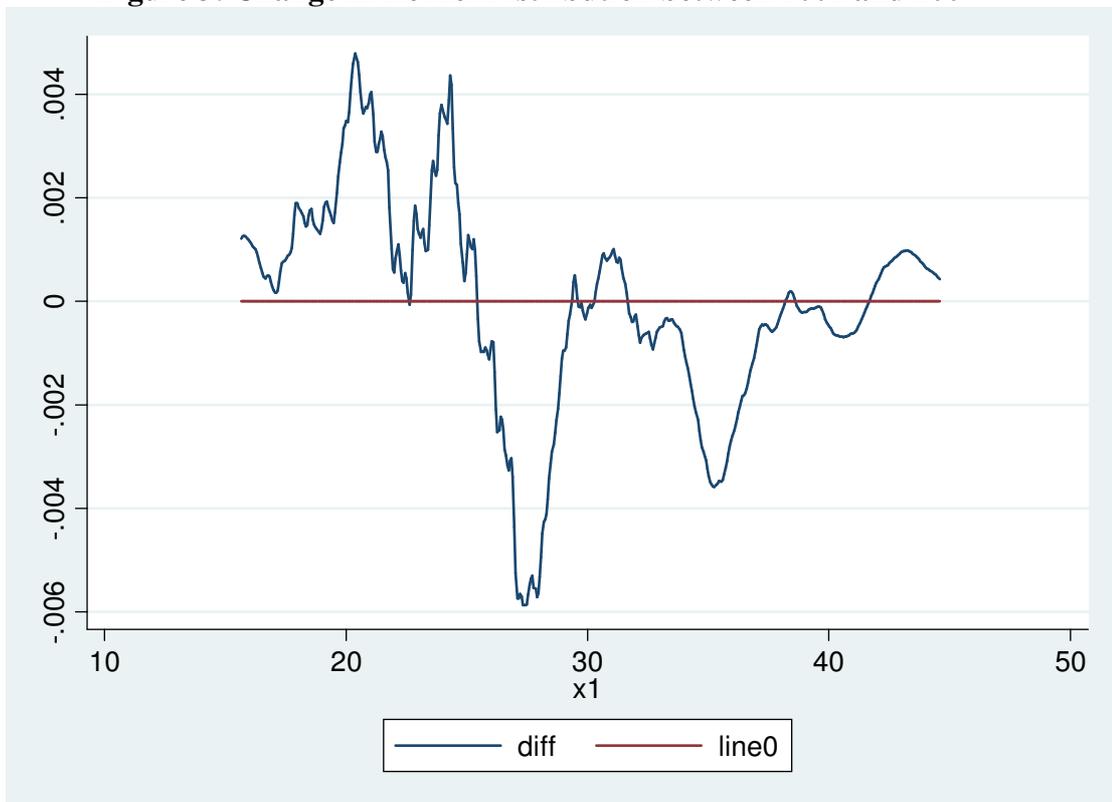
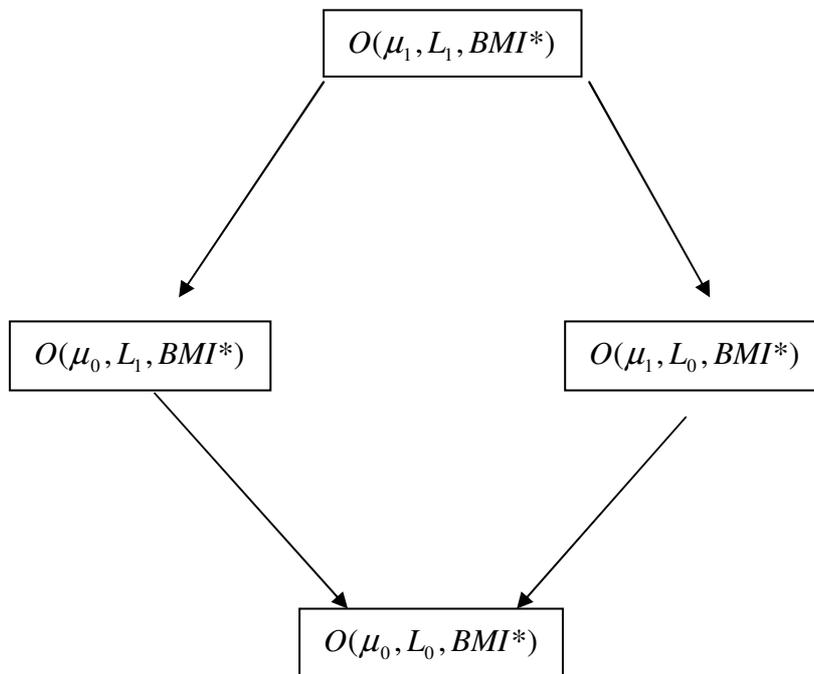**Figure 4: Difference in Cumulative Distribution Functions for BMI, 2002 and 2007**
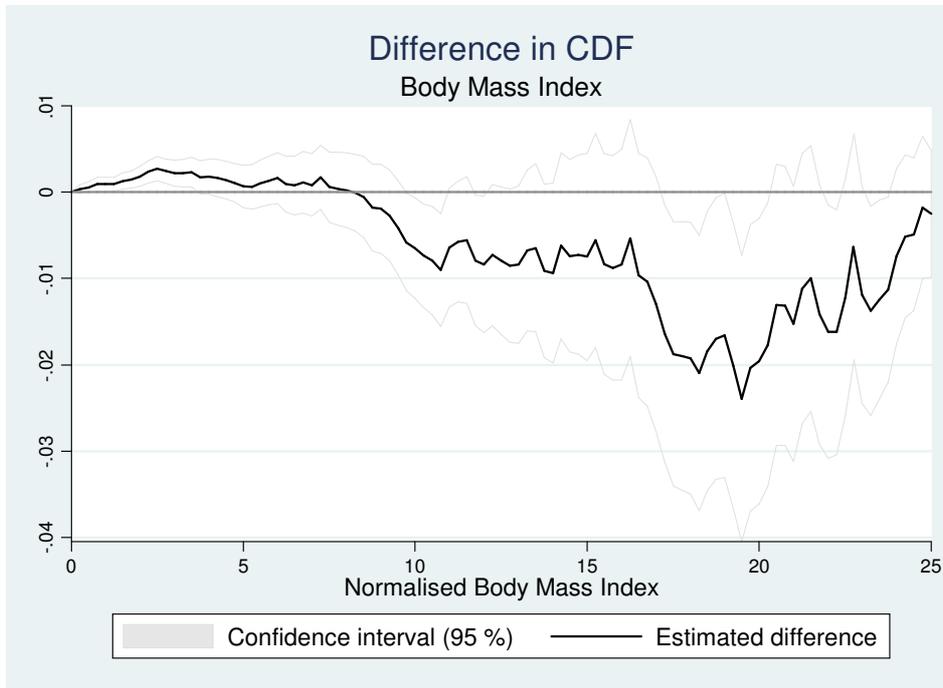
**Figure 5: Difference in Area under Cumulative Distribution Functions for BMI, 2002 and 2007**
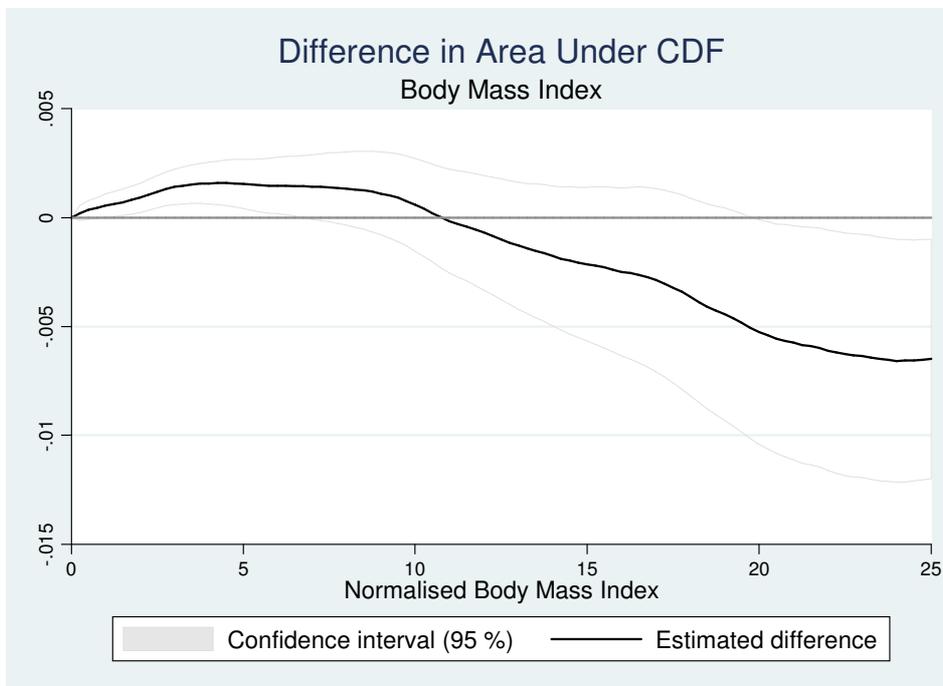
Figure 6: Changes in Mean and Distribution of BMI

Pdf

Obesity threshold

Pdf

Obesity threshold

BMI

Mean increases but "shape" of distribution is unchanged

Mean unchanged but "shape" of distribution changes
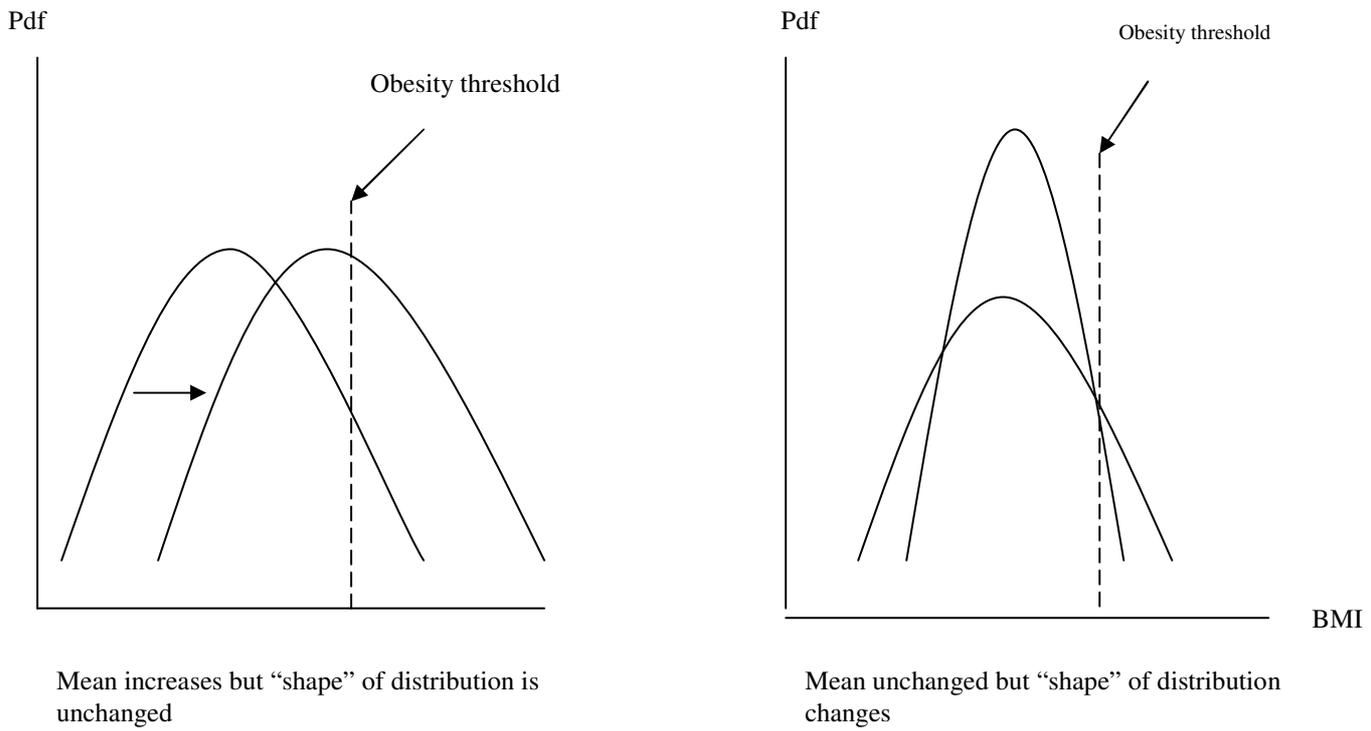
Figure 7: Two-way Shapley Decomposition

$$O(\mu_1, L_1, BMI^*)$$

$$O(\mu_0, L_1, BMI^*)$$

$$O(\mu_1, L_0, BMI^*)$$

$$O(\mu_0, L_0, BMI^*)$$

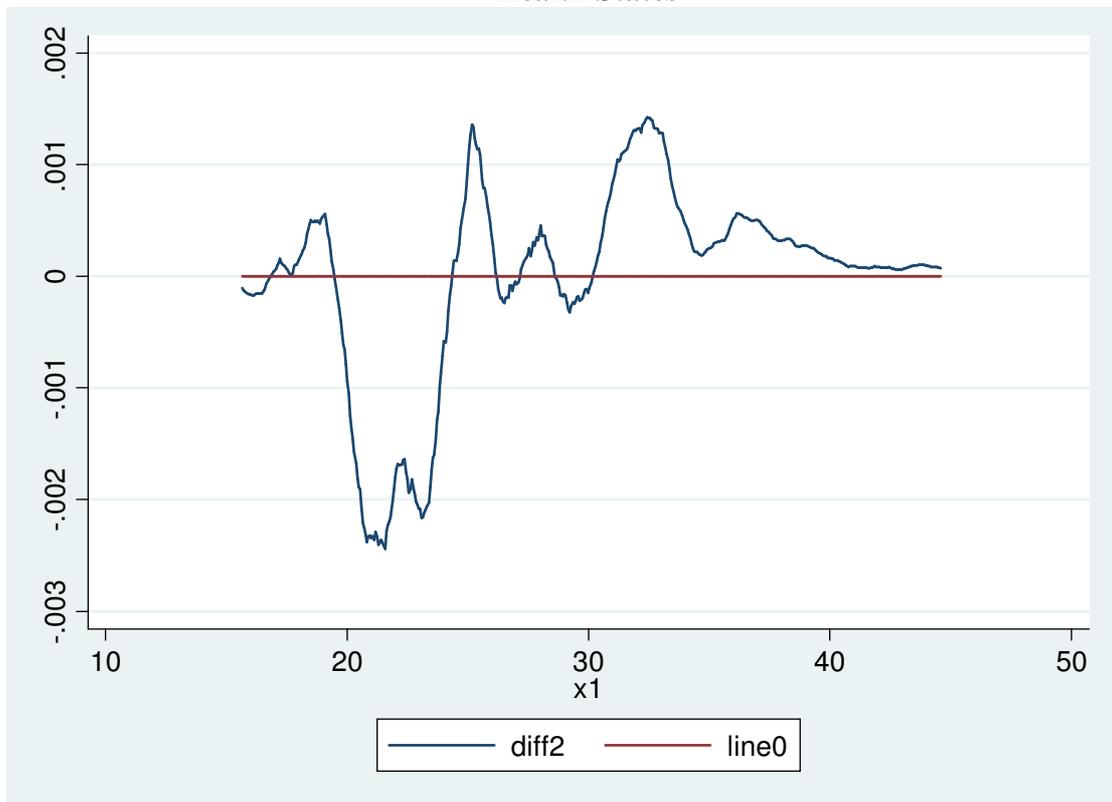**Figure 8a: Counter Factual Distribution – Socio-demographic Attributes**
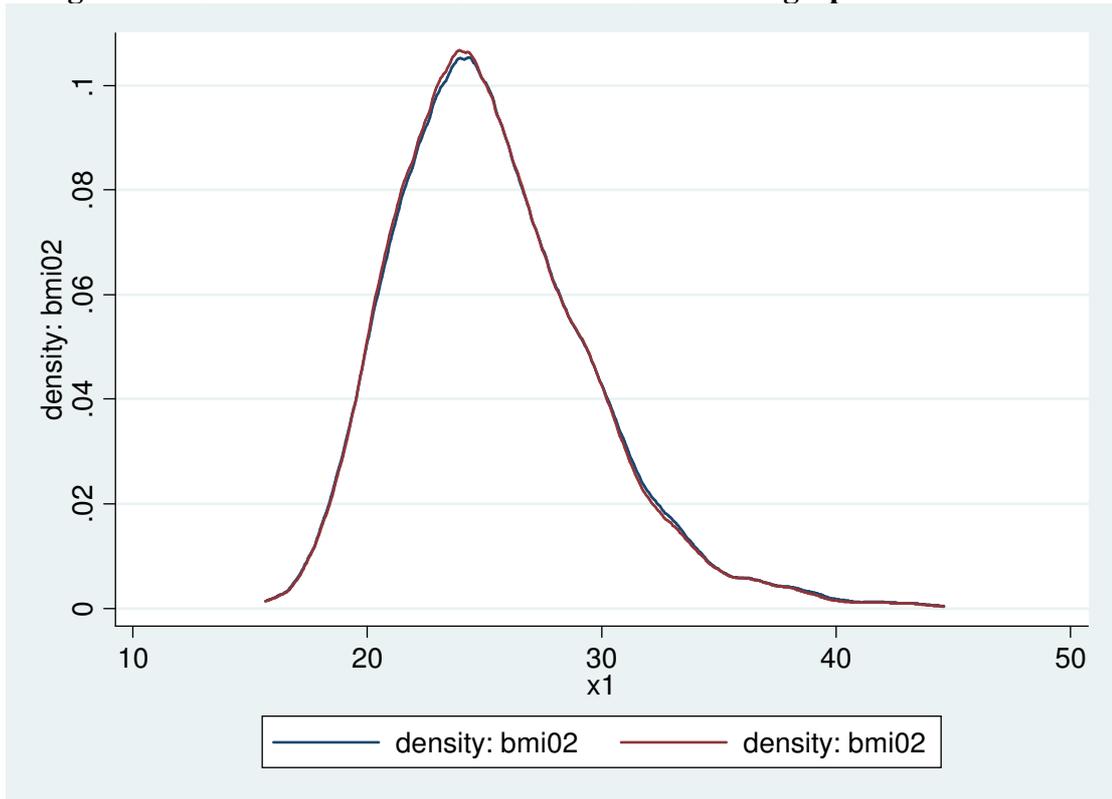
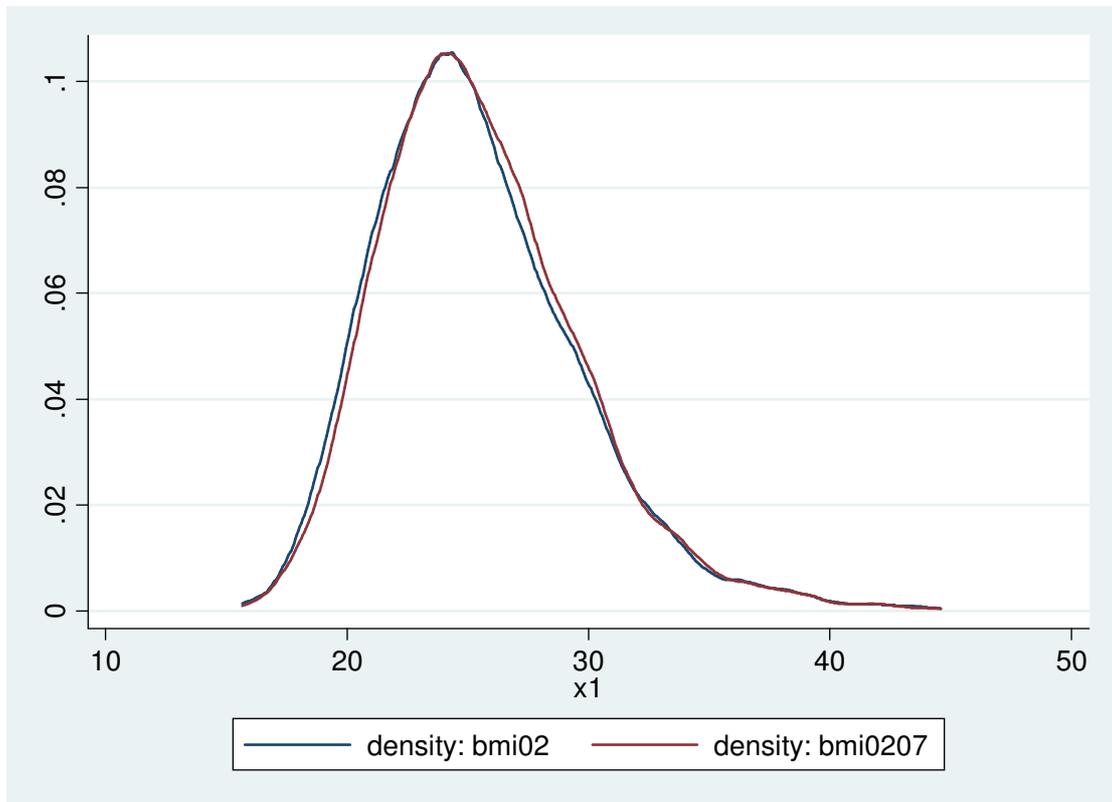**Figure 8b: Counter Factual Distribution Changes – Socio-demographic Attributes**
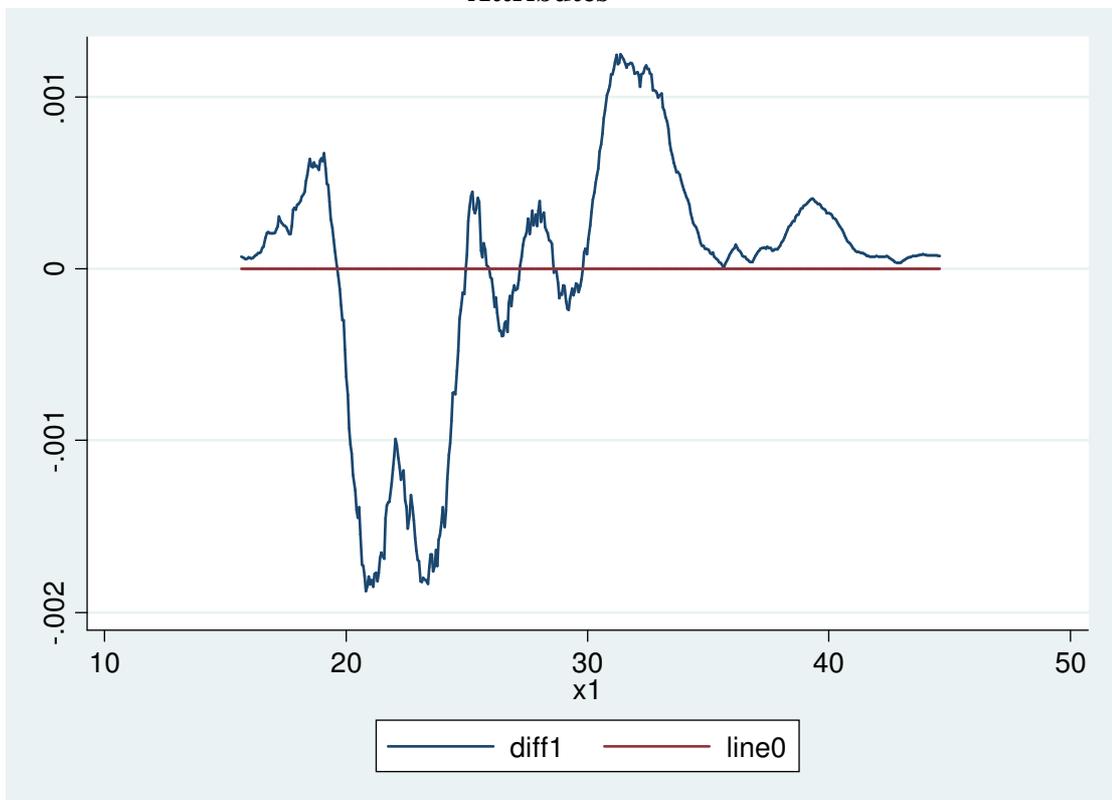
Figure 9a: Counter Factual Distribution – Socio Demographics and Health States
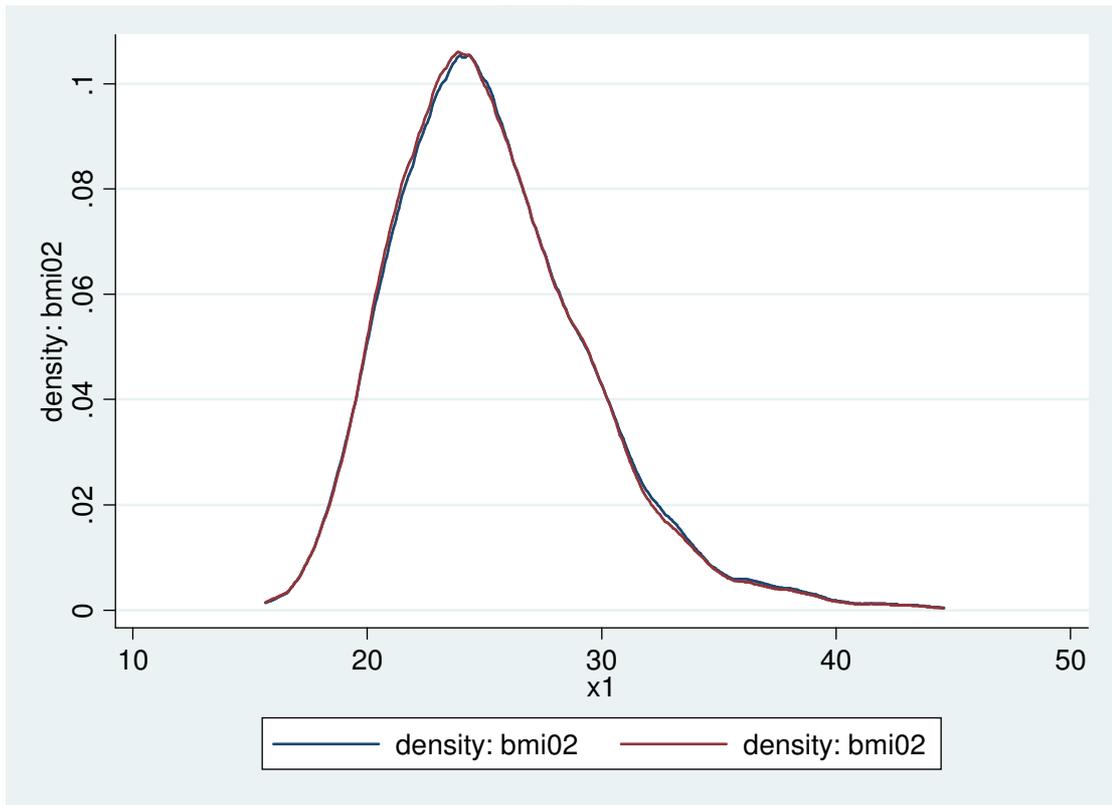
Figure 9b: Counter Factual Distribution Changes – Socio Demographics and Health States

Figure 10a: Counter Factual Distribution – Socio Demographics, Health States and BMI returns

Figure 10b: Counter Factual Distribution Changes – Socio Demographics, Health States and BMI returns