UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2018

Changes in BMI in a Cohort of Irish Children: Some Decompositions and Counterfactuals

David Madden, University College, Dublin

WP18/02

January 2018

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Changes in BMI in a Cohort of Irish Children: Some Decompositions and Counterfactuals

David Madden
(University College Dublin)

January 2018

Abstract: This paper examines the change in body mass index for a cohort of Irish children as they aged from 9 to 13 and decomposes the change into parts attributable to changes in observable characteristics and changes in returns to observable characteristics. The decomposition is carried out over the whole of the distribution, with a particular focus on the upper percentiles and a number of different decomposition techniques are applied and compared. The overall increase in BMI at higher percentiles is modest and is over-explained by the change in characteristics and is not sensitive to the adopted technique. The paper also carries out a number of partial equilibrium counterfactuals examining the impact of non-marginal changes in variables such as exercise and maternal education. The impact of these counterfactuals is limited.

Keywords: Overweight; obesity; decomposition; counterfactual.

JEL Codes: I14, I32

David Madden
School of Economics
University College Dublin,
Dublin 4,
Ireland.
david.madden@ucd.ie

Changes in BMI in a Cohort of Irish Children: Some Decompositions and Counterfactuals

1. Introduction

This paper has three aims: first of all, to provide a decomposition of the change in body mass index (BMI) in Ireland for a cohort of children, as they grow from aged 9 to aged 13. As is becoming standard in the literature, we carry out the decomposition across the whole of the distribution and not at the mean. Secondly, we provide a methodological contribution in that we examine the sensitivity of the results to the chosen method of decomposition. Finally, we implement a number of counterfactuals to provide some indication of how the distribution of BMI may be influenced by non-marginal changes in certain, key, covariates.

BMI is probably the most frequently used measure of obesity and overweight in both children and adults (although it is by no means a perfect measure, see Cawley and Burkhauser, 2008). A value of BMI in excess of a key threshold can indicate whether a child is overweight or obese (note that for children and adolescents these thresholds vary with respect to age and gender, as is discussed in more detail below), and typically these thresholds are located somewhere in excess of the 70th percentile. Hence analysis of the distribution of BMI which concentrates on measures of central tendency, such as the mean or median, may not fully capture developments in the incidence or severity of obesity or overweight.

When analysing the change in BMI over time, it can be helpful to adopt a decomposition approach. BMI is modelled as a function of a number of observable characteristics. The change in BMI between two periods can then be decomposed into an "explained" portion, arising from changes in these observed characteristics, and an "unexplained" portion which arises from changes over time in the effect of these characteristics on BMI and also from

changes in unobserved characteristics. The best known of such decompositions is the Blinder-Oaxaca (BO) decomposition which is evaluated at the mean of the distribution. However, as discussed above, from a policy perspective we may well be interested in developments elsewhere in the distribution and hence a decomposition across the whole of the distribution may be preferable (for a review of decomposition methods see Fortin et al, 2011).

There are a number of alternative methods of carrying out decompositions across the distribution and we apply three of these approaches in our analysis: the re-weighting method of Dinardo, Fortin and Lemieux (henceforth DFL, 1996), the distribution regression method of Chernozhukov et al (henceforth CVM, 2013) and the re-centered influence function (henceforth RIF) approach of Firpo et al (2007). All of these methods have their advantages and disadvantages (which we briefly review below and which are discussed in Fortin et al) and it is worthwhile to see how the results from the models differ in practice (thus this portion of our paper is in the spirit of Jones et al, 2015).

Finally, the CVM and RIF approaches also permit more detailed analysis of decompositions whereby the influence of individual characteristics can be isolated. In addition, the CVM approach can be used to examine other counterfactuals, in particular counterfactuals involving non-marginal changes. For example, assuming maternal education is identified as a factor affecting child BMI, then what would be the effect on the BMI distribution of imposing the highest level of education on all mothers? Later in the paper we analyse the effect of a number of such counterfactuals.

The remainder of the paper proceeds as follows: In section 2 we briefly review the decomposition methods we apply. In section 3 we discuss our data and present results for the

different decompositions and for a number of counterfactuals. Section 4 provides concluding comments.

2. Decompositions Across the Distribution

In this section we briefly review the decomposition methods. We commence with the simple BO decomposition and then proceed to decompositions of the whole distribution.

Suppose we have an outcome, y_t (e.g. BMI in period t) and y_t is a linear function of K variables (characteristics). We wish to obtain a decomposition of the difference in outcomes between the two periods. Thus we have

$$y_t = X_t\beta_t + v_{t,}. \quad E(v_t) = 0, t \in \{1,2\}$$

where X represents a vector of characteristics and β is a vector of returns to characteristics (or slope parameters of the relationship, including the intercept). Since $E(v_t) = 0, t \in \{1,2\}$, then following the well-known result of Blinder (1973) and Oaxaca (1973) the difference in average outcomes $\Delta_y^\mu = \bar{y}_2 - \bar{y}_1$ can be decomposed as follows:

$$\Delta_y^\mu = E(y_2) - E(y_1) = E(X_2)\beta_2 - E(X_2)\beta_1 + E(X_2)\beta_1 - E(X_1)\beta_1.$$

$$= (E(X_2)[\beta_2 - \beta_1]) + ([E(X_2) - E(X_1)]\beta_1).$$

where $E(X_2)\beta_1$ is the unconditional counterfactual mean outcome i.e. the outcome if the pattern of returns in period 1 were applied to the characteristics of period 2. The second term on the right hand side above, $([E(X_2) - E(X_1)]\beta_1)$, shows that part of the gap which arises

owing to differences in the characteristics over the two periods and is sometimes referred to as the "explained" portion of the gap. The first term on the right hand side, $(E(X_2)[\beta_2 - \beta_1])$, is that part of the gap which arises owing to differences in the returns to characteristics and differences in unobservables, and is sometimes referred to as the "unexplained" portion of the gap. It is also possible to further decompose both the explained and unexplained portions of the gap to obtain the contribution of each covariate. This is sometimes called the "detailed decomposition".[12]

Note that in the decomposition above, in the explained portion of the gap, the differences in characteristics are weighted by the returns from period 1. An alternative decomposition, essentially the mirror image of the decomposition above, is also possible where the difference in characteristics are this time weighted by the returns from period 2. The key issue here is essentially the choice of a reference vector of returns coefficients between the two periods. In general, when looking at a change over time the convention appears to be to isolate the change in outcomes arising from changes in characteristics. Thus the counterfactual which is examined is the outcome which would have arisen with period 2 characteristics and period 1 returns, and this is the counterfactual which is the basis of our analysis below.

In the case of BMI we may also be interested in gaps and decompositions at parts of the distribution other than the mean, given that key BMI thresholds for overweight and obesity are typically at percentiles well above 50. One possible approach might be to carry out quantile regressions at the desired quantiles and then apply the BO decomposition.

---

[1] Detailed decompositions of the unexplained portion can also be sensitive to the choice of omitted category for categorical variables. See Fortin et al (2011).

[2] It is also possible to have a three-way decomposition. This recognises the fact that typically both characteristics and returns will differ between the two periods simultaneously, and so a third interaction term takes account of this.

In that case, given our outcome, y, the conditional quantile function is assumed to be linear of the form

$$Q_\theta(y|X) = X_i'\beta_\theta \text{ for each } \theta \in (0,1)$$

where X_i represents the set of covariates for individual i and β_θ is the coefficient vector for the θth quantile. The quantile coefficients can be seen as capturing the return of each covariate across the distribution of y. Given the assumption of linearity, it is possible to estimate the conditional quantile of y by linear quantile regression for each $\theta \in (0,1)$. The conditional quantiles for periods 1 and 2 are then $Q_{1\theta}(y_1|X_1) = X_{1i}'\beta_{1\theta}$ and $Q_{2\theta}(y_2|X_2) = X_{2i}'\beta_{2\theta}$ respectively.

Can we reconstruct the counterfactual unconditional distribution of outcomes $Q_\theta^C = X_{1i}'\beta_{2\theta}$ using estimates from the conditional quantile regressions, and hence carry out a BO type decomposition of the gap evaluated at each quantile? This is straightforward when dealing with the mean, since the law of iterated expectations tells us that $E(y) = E_X(E(y|X))$. Thus the OLS estimate for covariate X_i provides the effect of the covariate on either the conditional or unconditional mean of y. However, the law of iterated expectations does not hold in the case of quantiles and so $Q_\theta(y) \neq E_X[Q_\theta(y|X)]$ where $Q_\theta(y)$ is the θth quantile of the unconditional distribution and $E_X[Q_\theta(y|X)]$ is the corresponding conditional quantile. Thus, in terms of a decomposition, the differences in unconditional quantiles will not be the same as the difference in conditional quantiles and hence it is not straightforward to recover (and decompose) the gap between unconditional quantiles.

A number of different approaches have been put forward to overcome this problem or else to provide alternative decompositions (focusing on the conditional distribution), and we discuss three of them here. First we have the DFL "re-weighting" method:

let $F_{y_1|X_1}(y|X)$ and $F_{y_2|X_2}(y|X)$ represent the conditional, cumulative density functions for periods 1 and 2 respectively. The counterfactual we wish to obtain is the conditional density function for period 1 but with the distribution of covariates for period 2. DFL showed that this could be achieved via the computing of a reweighting factor. Thus the counterfactual we obtain is

$$F_{y_1|X_1}(y|X) = \int F_{y_1|X_1}(y|X)\Psi(x)\, dF_{X_1}(X)$$

where the re-weighting factor $\Psi(X) = \frac{dF_{X_2}(X)}{dF_{X_1}(X)}$. This weighting factor can be calculated relatively simply since it is possible to show that

$$\Psi(X) = \frac{\Pr\,(X|t=2)}{\Pr\,(X|t=1)} = \frac{\Pr(t=2|X)\,/\Pr\,(t=2)}{\Pr(t=1|X)\,/\Pr\,(t=1)}$$

where, for example, $\Pr\,(t=2|X)$ is the probability that from a pooled sample and given the observed set of covariates, an observation comes from period 2 (as opposed to period 1). $\Pr\,(t=2)$ is simply the unconditional probability that any observation from the pooled sample comes from period 2. As outlined below, since we are dealing with a balanced panel here, $\Pr(t=1) = \Pr\,(t=2)$. The conditional probabilities can be estimated from a simple logit model and then, given the calculation of the weighting factor, the counterfactual distribution can be obtained using period 1 observations, but reweighted by the weighting factor. Using terminology from the BO approach, this gives the explained part of the decomposition and given this, it is then straightforward to obtain the unexplained part. A

detailed decomposition can obtained by carrying out the process one covariate at a time. However, results are path dependent.

An alternative approach to estimating the conditional distribution is provided by CVM (2013). They term their approach "distributional regression" and again it is based upon estimating the conditional distribution of the outcome in question (BMI), $F_{y_1|X_1}(y|X)$, and then integrating this conditional distribution over the distribution of *X* in period 2 to obtain the counterfactual.

They suggest a distribution regression approach to estimate the conditional distribution, with a separate regression model estimated for *each* value of *y* in period 1. The precise regression model depends upon a link function, Λ, whereby $F(y|X) = \Lambda(X\beta(y))$. Possible specifications for Λ include the logit, probit, or linear probability models. Estimating the model for period 1 provides a set of parameter estimates $\widehat{\beta_1(y)}$, from which predicted probabilities $\Lambda(X_i\widehat{\beta_1(y)})$ are obtained. Averaging over these predicted probabilities over the period 2 sample gives the counterfactual:

$$F_{y_1|X_2}(y) = \frac{1}{N_2}\sum_{i\in 2}\Lambda(X_i\widehat{\beta_1}(y))$$

Thus the parameters from the estimated model for period 1 are applied to the period 2 sample, thus providing the counterfactual whereby the conditional distribution for period 1 applies to the sample with period 2 characteristics. As with the DFL approach, this provides the explained part of the decomposition and the function can be inverted to obtain conditional quantiles.

CVM also suggest an alternative whereby if conditional quantiles are obtained from quantile regression and then inverted to obtain the conditional distribution function, we obtain:

$$F_{y_1|X_2}(y) = \int_0^1 1\left\{X\widehat{\beta_1}(\acute{u}) \leq y\right\} du$$

where $\hat{\beta}(u)$ is the quantile regression estimator. CVM (2013) discuss the factors which should influence the choice between these two approaches. They suggest a pragmatic, case-by-case approach but recommend that the quantile approach should only be used if the dependent variable is smooth and continuous (thus it would not be suitable for estimating a counterfactual wage distribution where there is a minimum wage with a lot of mass concentrated at that point).

The final approach we consider is the RIF approach of Firpo et al (2009). They suggest an OLS-based regression method which estimates the impact of changes in an explanatory variable on the unconditional quantile of the outcome variable, via the regression of a transformation of the outcome variable on the set of explanatory variables. The transformation in question is based on the *influence function* (IF), which provides the influence of an individual observation on the distributional statistic of interest (such as the variance, or a particular quantile). In the case of the mean, for example, the influence function is the demeaned value of the outcome variable i.e. y-μ. What is known as the re-centered influence function (RIF) is obtained if the original distributional statistic of interest is added back to the IF. Thus in the case of the mean, the RIF=y-μ+μ=y.

More generally (and dropping type subscripts for convenience), if $F(y)$ is the cumulative distribution of the outcome variable and if $T(.)$ is the distributional statistic in question, e.g., a quantile, then the influence function is the directional derivative of $T(F)$ at F (Essama-Nssah and Lambert, 2011). By adding the IF to the original distributional statistic, we obtain the RIF. By construction, the RIF obeys the law of iterated expectations and thus

$E[RIF(y;T(.),F(y)] = T(.)$ and it is this which is regressed against the covariates in the X vector.

For the case where the distributional statistic is a specific quantile, Q_θ, the IF is defined as

$$IF(y;Q_\theta) = \frac{(\theta - I[(y \le Q_\theta])}{f_y(Q_\theta)}$$

where θ is the quantile in question, *I(.)* is an indicator function taking on the value of 1 if the expression in parentheses is satisfied, $Q_\theta(y)$ is the θth quantile of the unconditional distribution of the outcome variable and $f_y Q_\theta(y)$ is the density of the marginal distribution of y evaluated at Q_θ (see Essama-Nssah and Lambert, 2011). The RIF is then

$$RIF(y;Q_\theta) = Q_\theta + \frac{(\theta - I[y \le Q_\theta])}{f_y(Q_\theta)}$$

Having calculated the value of RIF for all observations in this way, the RIF regression model is then defined as $E[RIF(y;Q_\theta)|X] = X'\beta$, and can be estimated by OLS. The estimated coefficients of the vector β then give the effect of each covariate on the unconditional θ^{th} quantile of y. The regression can be estimated for different values of θ and for different types, and counterfactuals can be constructed as with the standard BO decomposition, including a detailed decomposition.

We thus have four approaches to constructing a decomposition of the change in the distribution of BMI over time, the DFL re-weighting approach, the CVM approaches using either quantile regression or distributional regression and finally the RIF approach of Firpo et al. In the next section we discuss our data and also carry out the decompositions in question.

3. Data and Results

Our data comes from the first two waves of the Growing Up in Ireland child cohort. This tracks the development of a cohort of children born in Ireland in the period November 1997-October 1998 (see Williams et al, 2009 and Quail et al, 2014). The sampling frame of the data was the national primary school system, with 910 randomly selected schools participating in the study. Weight was measured to the nearest 0.5 kg using a medically approved flat mechanical scales and children were advised to wear light clothing. Height was measured to the nearest mm using a height measuring stick.

In all, the original sample in wave 1 consisted of 8568 children. When we drop observations where either (a) height or weight were not measured appropriately (b) where observations on other relevant characteristics were missing or (c) where observations dropped out between waves 1 and 2, we are left with a balanced panel of 6936.

The most common measure of overweight/obesity used for adults is derived from BMI (weight in kilos divided by height in metres squared). The World Health Organisation suggests a threshold BMI of 25 for "overweight", a threshold of 30 for "obesity" and a threshold of 40 for "severely obese".

While the BMI thresholds for adults have general acceptance and do not differ by age or gender, the same is not true for children, where BMI can change systematically with age and gender. For example, at birth median BMI is around 13, this increases to 17 at age 1, decreases to 15.5 at age 6 and increases to 21 at age 20 (Cole et al, 2000). Cole et al (2000) provide a set of cutoff points for BMI for childhood based upon international data and which they suggest should be used for international comparisons. They obtain these by drawing centile curves which pass through the adult cut-off points at age 18 and which then can be traced back to provide "equivalent" cut-off points for different ages and genders. The cutoffs are obtained by averaging data from large nationally representative surveys from Brazil,

Great Britain, Hong Kong, the Netherlands, Singapore and the US, with in total nearly 200,000 observations aged from birth to 25.

The cutoffs are provided at half-yearly intervals. Thus for the first wave of our data, the vast majority of children are aged 9. Assuming that age is distributed uniformly within the cohort of nine year olds, it seems appropriate to take the cut-off for age 9.5. Similarly for the second wave of our data (who are mostly 13 year olds) we use the cut-off for age 13.5. For the very small number of children aged 8 and 10 we use the 8.5 and 10.5 cutoffs respectively and similarly for the second wave we use the 12.5 and 14.5 cut-offs for those aged 12 and 14. The age and gender specific cutoffs are provided in table 1. These cutoffs have also been used in previous studies which have analysed child obesity using GUI e.g. Layte and McCrory (2011).

There is one final adjustment we make to the data which facilitates our analysis. As the obesity and overweight thresholds for BMI change (since the sample is now four years older) a simple comparison of BMI can be misleading. Consequently we compare *normalized* BMI figures, where BMI is divided by the appropriate overweight threshold.[3] Thus for example, if a child has a normalized BMI of 1., then this indicates that the child had a BMI which was 1.1 times the relevant threshold for their age and gender. This facilitates comparisons across time and gender where these thresholds differ.

As we propose to decompose the change in BMI into that part arising from characteristics, we need to define what we view as the relevant characteristics. We employ a relatively parsimonious set of characteristics: maternal age (quadratic), maternal education, number of siblings, if mother was smoker or drinker at time of pregnancy, maternal economic status,

[3] We use the overweight threshold as evidence suggests that health problems associated with high BMI tend to start at the overweight threshold. See Aume et al (2016) and Global BMI Mortality (2016).

health of study child, an index of study child's exercise and (log of) equivalized family income. We assume a linear relationship between normalized BMI and these variables. Table 2 provides summary statistics for these variables for periods 1 and 2.

Comparing waves 1 and 2, we see that the sample is obviously older, the mothers are better educated, rates of maternal smoking and drinking have both dropped, a higher fraction of mothers are working, child illness is virtually unchanged and income has fallen. Probably the biggest change is observed in the index of child exercise which has fallen from around 6.4 to 5 (we discuss the change in exercise in more detail below when exploring counterfactuals). We also note that normalized BMI is unchanged while there is a marginal rise in the fraction overweight but a marginal fall in the fraction who are obese.

We are primarily interested in that part of the sample who are either overweight or at least close to overweight. This is effectively applying what in the poverty literature is known as the principle of "focus", whereby we are only concerned with developments in that range of the BMI distribution where health problems from being overweight can arise. Given that about 25% of the sample is overweight, we choose to focus our analysis on those at the 70th percentile of the BMI distribution and above. This captures those who are overweight and also the next 5% below the threshold.[4]

Figure 1 shows the quantile functions for the counterfactuals based upon the decompositions.[5] The vertical axis shows the value of normalized BMI corresponding to the quantile (on the horizontal axis). Wave 1 and wave 2 refer to periods 1 and 2 (when children were aged around 9 and 13 respectively). It can be seen that between the 70th and 80th percentile the quantile functions are virtually identical (hence the very similar rates of

[4] In this paper we are concerned with health problems arising from *too high* a BMI. We fully acknowledge, but choose not to address, health problems arising from having *too low* a value of BMI.

[5] The tables underlying these figures are in the appendix.

overweight, which is at around the 75th percentile). After about the 80th percentile however, the wave 2 quantile function lies above that of wave 1 indicating that BMI values in the upper parts of the distribution are higher in wave 2 than in wave 1. This would typically not be detected by simple prevalence measures of overweight or obesity which tell us the fraction of the sample above key thresholds but tell us nothing about how far *above* the thresholds are those observations which are overweight/obese. Above the 80th percentile the gap between the two waves averages at around 0.012 units of normalized BMI. This corresponds to less than 0.25 actual units of BMI (though the precise number of units will depend upon the age and gender of the individual involved).

The remaining quantile functions in figure 1 show the different counterfactuals on the basis of the conditional distribution of period 1 applied to period 2 characteristics. Thus the gap between the period 1 quantile function and the relevant counterfactual shows the impact of the change in characteristics between periods 1 and 2, while the gap between the counterfactual and the period 2 quantile function shows the impact of the change in the conditional distribution function. It is noticeable that in all cases the counterfactual lies above wave 1 for all quantiles under consideration. Thus the impact of characteristics alone is to increase BMI at each quantile, with an effect in the region of 0.04-0.05 units of normalized BMI at higher quantiles. This corresponds to around one unit of actual BMI (again depending upon age and gender of individual involved).

In turn, the gap between the counterfactual and the period 2 quantile function is about -0.02 units of normalized BMI, indicating the change in the conditional distribution, or in the "return" to characteristics, acts to *lower* BMI at each quantile in the range of interest. This lowering however is not sufficient to offset the effect of characteristics, hence the overall effect is that BMI in the upper regions of the distribution has increased over the period.

A further notable feature of figure 1 concerns the sensitivity of results (or lack of) to the choice of decomposition method. Simple eye-balling reveals that the quantile functions for the four different methods (DFL, RIF and the two versions of CVM) are virtually superimposed upon each other, suggesting that for this particular decomposition exercise at least, results are not sensitive to choice of method.

Figure 2 explores the effect of characteristics in more detail, by presenting the 95% confidence intervals for the characteristics effects for the different decompositions. It is noticeable that these intervals overlap each other, reinforcing the point made above, that decomposition results are not sensitive to choice of method. Furthermore, we can see that as "zero" lies outside of the bands of the confidence intervals that the effect of characteristics is statistically significant, though it is debatable as to whether it could be regarded as significant in an economic or health-related sense, since the effects are quite small.

We now apply the CVM method to estimate the effect of different counterfactual experiments. We examine three experiments, involving changes in maternal education, income and exercise respectively. We first describe the counterfactual for maternal education. We partition maternal education into three categories: category 1 includes those mothers who do not complete the full secondary school cycle (in Ireland secondary school completion occurs when someone takes what is known as the Leaving Certificate exam, which usually happens at the age of eighteen). Category 2 is mothers who do complete the secondary school cycle, while category 3 is those mothers who complete third level education. In previous work where we have specifically investigated the gradient of overweight with respect to maternal education we also included a category for mothers who obtain a degree or certificate intermediate between the Leaving Certificate and third level education (Madden, 2017). However, outcomes for this group proved to be almost identical to those from category 2, so we choose to include that group in category 2 and proceed with

just three categories. As we see, by the time children have reached aged 13, about 20% of mothers are in category 1, nearly 60% in category 2 and just over 20% in category 3. There is a general trend upwards in educational achievement between the two periods, particularly reflecting younger mothers managing to complete secondary school education. The counterfactual we apply is to examine the impact of *all* mothers having third level education.

The first income counterfactual we analyse is a mean preserving compression. The measure of income available in GUI is equivalised after-tax disposable family income. We apply a mean preserving compression to this income such that $y_{i,cf} = \bar{y} + 0.5 * (y_i - \bar{y})$ where $y_{i,cf}$ is counterfactual income for individual i and $\bar{y}$ is average income. This compression reduces the Gini coefficient of equivalised income from about 0.27 to 0.13, thus considerably reducing income inequality.

We also apply a second income counterfactual. This time, we raise everyone's income to the highest observed level. Thus not only is the income distribution equalized, but average income is also increased (by about 240%).

These counterfactuals are motivated by the frequently observed socioeconomic gradient of obesity, whether measured with respect to income or maternal education. In all cases the distribution of resources (income or education) is equalized. In the case of the education counterfactual it is equalized "upwards" in the sense that all mothers are raised to the highest level of education, thus the average level of education is raised and the distribution is equalized. In the case of the income counterfactuals, average income is unchanged in counterfactual 1 while it increases in counterfactual 2.

The final counterfactual we apply is with respect to exercise. We construct an index of exercise in the following way: principal carers are asked how many times in the previous fourteen days the study child had undertaken (a) strenuous exercise for at least 20 minutes

and (b) light exercise for at least 20 minutes. Each incidence of strenuous exercise is given a score of "1" while an incidence of light exercise is given a score of "0.5". The sum of these then gives the total exercise score for the study child. Figure 3 shows the histogram for exercise by period. In period 1 there is considerable mass around a value of 7.5, indicating that a high fraction of individuals had at least seven instances of heavy exercise or equivalent in the previous fourteen days. By period 2, this had fallen to around 12% and around half of the sample had five or less instances of heavy exercise or equivalent. The counterfactual we apply is that every individual has an exercise index of 7.5 (the highest observed). Thus, similar to the education counterfactual, the average level is raised and the distribution is equalized.

Before discussing the results from the different counterfactuals, it must be pointed out that these are strictly partial equilibrium, first round effects and it is quite possible that additional second round effects might be observed. Thus for example, were the counterfactual of every mother having the highest level of education applied, then it is likely that income too would be raised. However, the counterfactuals examined here merely change one factor in isolation.

All counterfactuals are applied to period 2 and the effect on each quantile of BMI is shown in figure 4. Thus we are estimating what the effect of each counterfactual would be on the distribution of BMI in period 2. We can see straightaway that income counterfactual 1 has virtually no impact for most of the distribution and then leads to a marginal fall for quantiles above the 80th. Even then the impact is tiny, about 0.002 units of normalized BMI, which corresponds to about 0.04 units of actual BMI.

Income counterfactual 2 has a greater impact but even then it could hardly be described as dramatic. It first increases and then reduces normalized BMI across the distribution with the greatest reduction of about 0.025 units of normalized BMI. This corresponds to a reduction

of around 0.5 of a unit of regular BMI. Note however, that the confidence intervals are quite wide for this counterfactual.

The education counterfactual has a similar impact to income counterfactual 2. Here the maximum impact is slightly greater with a fall in normalized BMI of about 0.029, but this corresponds to only just over 0.5 units of regular BMI and confidence intervals are not as wide as for the Income 2 counterfactual.

Finally, the exercise counterfactual has the biggest impact. Once again, the effect increases with higher quantiles and is greatest at the 95th quantile where it implies a reduction of 0.05 units of normalized BMI, which corresponds to about just under one unit of regular BMI. The effect also appears to be statistically significant.

Overall, it seems fair to say that the effect of all three counterfactuals on the BMI distribution is relatively marginal.

4. Conclusions

In this paper we have applied a number of decomposition techniques to examine the change in BMI for a sample of Irish children/adolescents as they aged from nine to thirteen. By applying the decomposition across all of the distribution (as opposed to evaluating it just at the mean) we were able to focus in on the top 30% of the distribution, that part where evidence suggests that health problems can arise (the threshold for "overweight" corresponds to approximately the 75th percentile). Overall we find that there is relatively little change in the fraction of the population with BMI above the overweight threshold. However, conditional upon being overweight, BMI appears to rise slightly. The traditional decomposition of this change into that attributable to changes in observed characteristics and that attributable to changes in the returns to characteristics or changes in unobserved

characteristics suggests that changes in characteristics over-explain the rise in BMI above the 70th percentile. Changes in unobservables and the returns to observables tend to bring down BMI. But overall, the change in BMI and in the constituent parts of the decomposition are relatively small.

In the final parts of the results section we also apply some simple non-marginal counterfactuals in the areas of income, exercise and maternal education. Bearing in mind that these are partial equilibrium effects, it is still noticeable that the impacts upon the distribution of BMI are relatively modest. In turn, this raises issues as to the efficacy of policy initiatives in these areas. For example, it might be thought that if all 13 year olds were to be brought up to the maximum observed level of exercise, then this would have a significant impact upon BMI. The counterfactual here however suggests that the impact would be comparatively limited. The effects of other counterfactuals such as higher maternal education or changes in the level and distribution of income are even more limited. It should be stressed however, that these are results for a specific age cohort in a specific country and it is quite possible that such policies applied to a different group might produce different outcomes.

Table 1: Age and Gender Specific Cutoffs for Overweight and Obesity from Cole et al

	Male		Female	
Age	**Overweight**	**Obese**	**Overweight**	**Obese**
8.5	18.76	22.17	18.69	22.18
9.5	19.46	23.39	19.45	23.46
10.5	20.20	24.57	20.29	24.77
12.5	21.56	26.43	22.14	27.24
13.5	22.27	27.25	22.98	28.20
14.5	22.96	27.98	23.66	28.87

Table 2: Summary Statistics

Variable	**Wave 1**	**Wave 2**
Normalised BMI	0.919	0.919
Fraction Overweight	0.253	0.257
Maternal Age	39.405	41.157
Fraction Obese	0.059	0.057
Lower Secondary Education	0.295	0.199
Complete Secondary Education	0.535	0.588
Third Level Education	0.170	0.213
Smoker	0.249	0.237
Drinker	0.458	0.409
Working	0.533	0.580
Child Ill	0.107	0.110
Exercise	6.426	5.002
Log Income	9.731	9.556

Figure 1: Quantile Functions for periods 1, 2 and impact of changes in characteristics

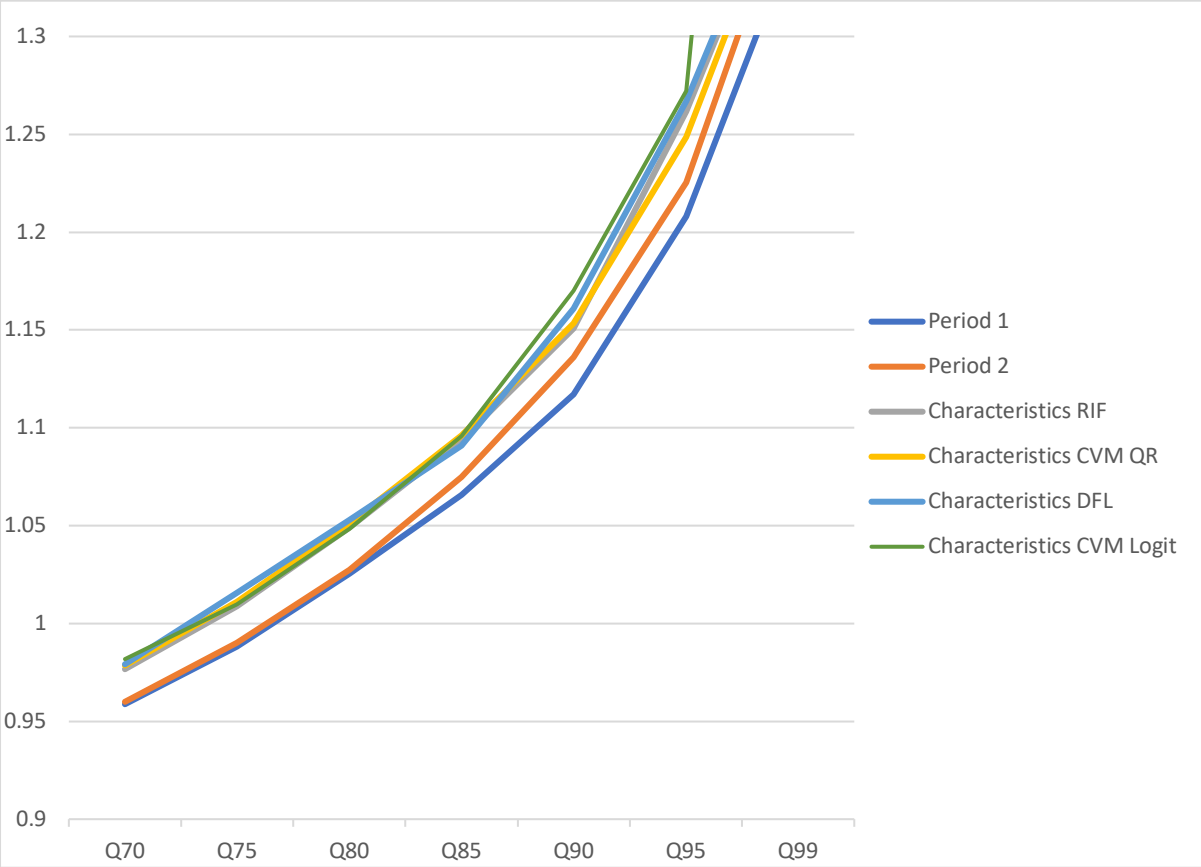


Figure 2: Characteristics Effects – Confidence Intervals

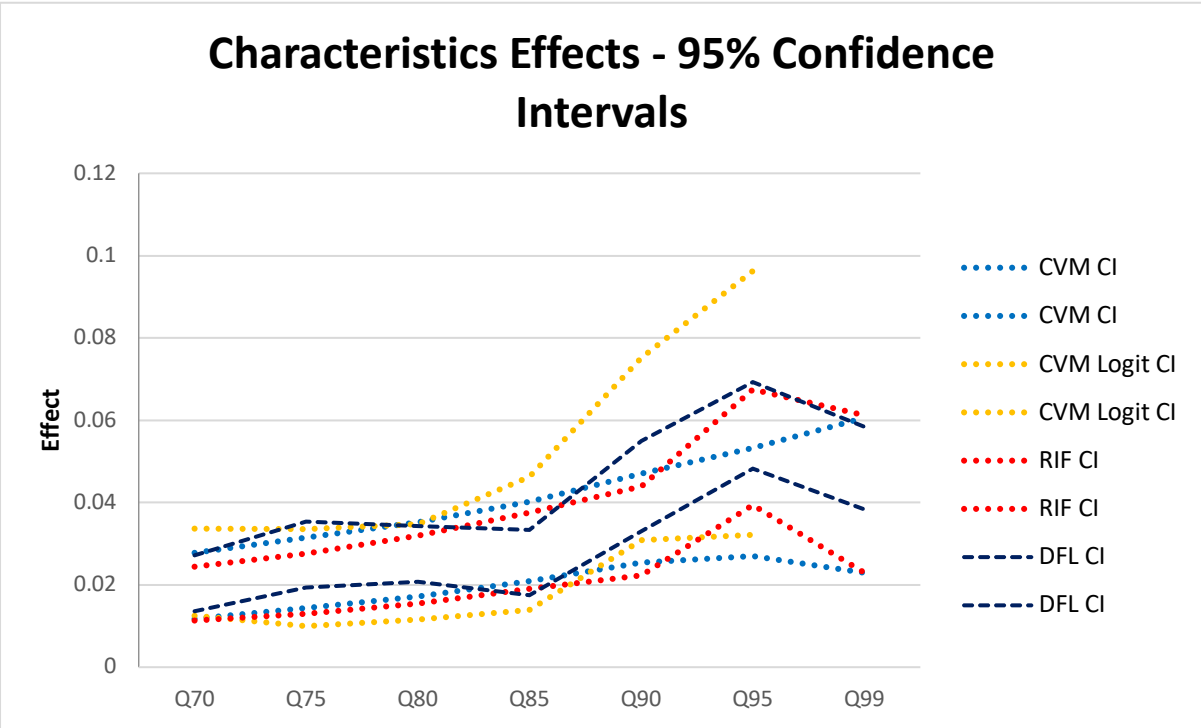
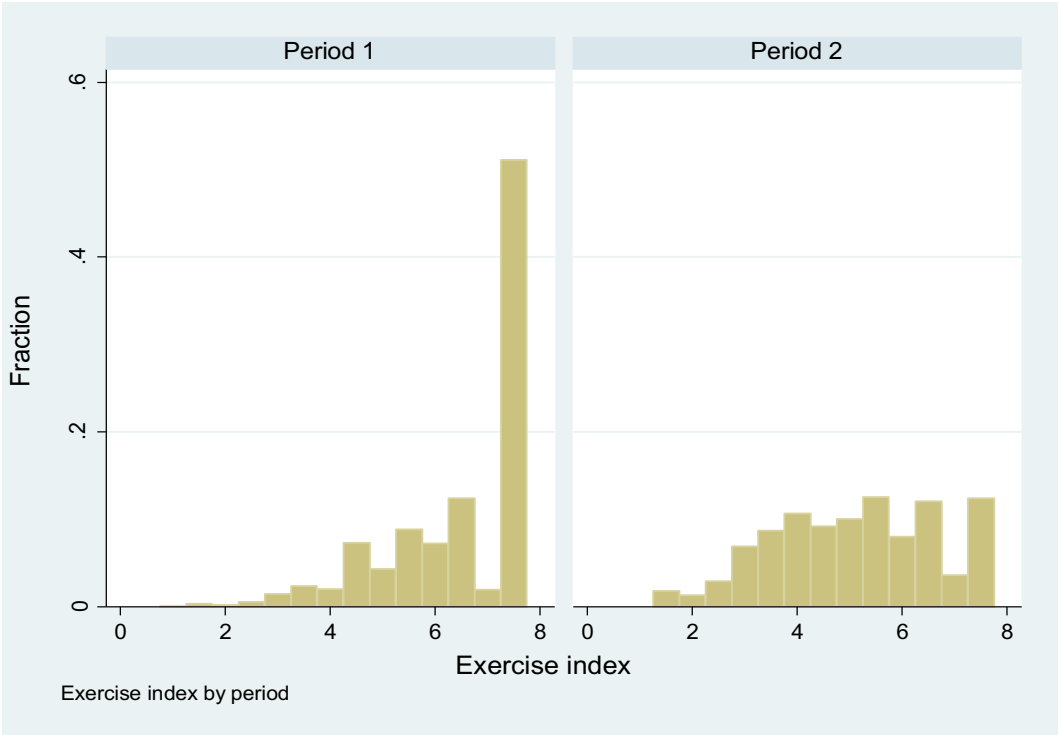
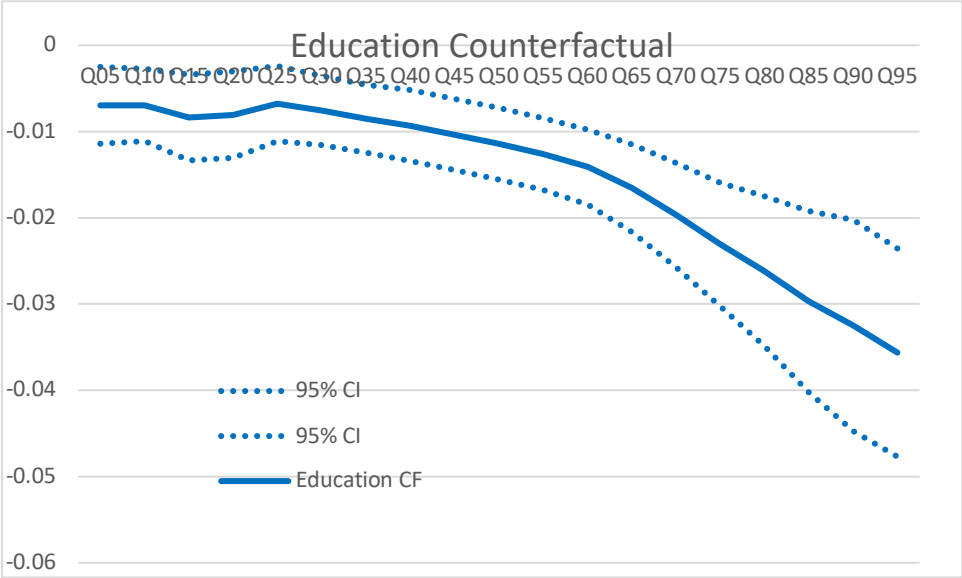
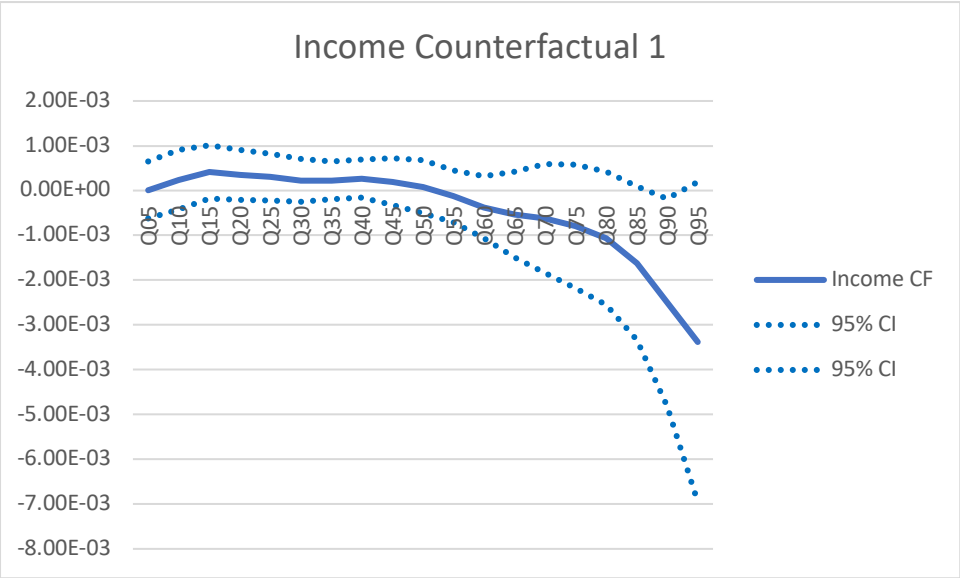


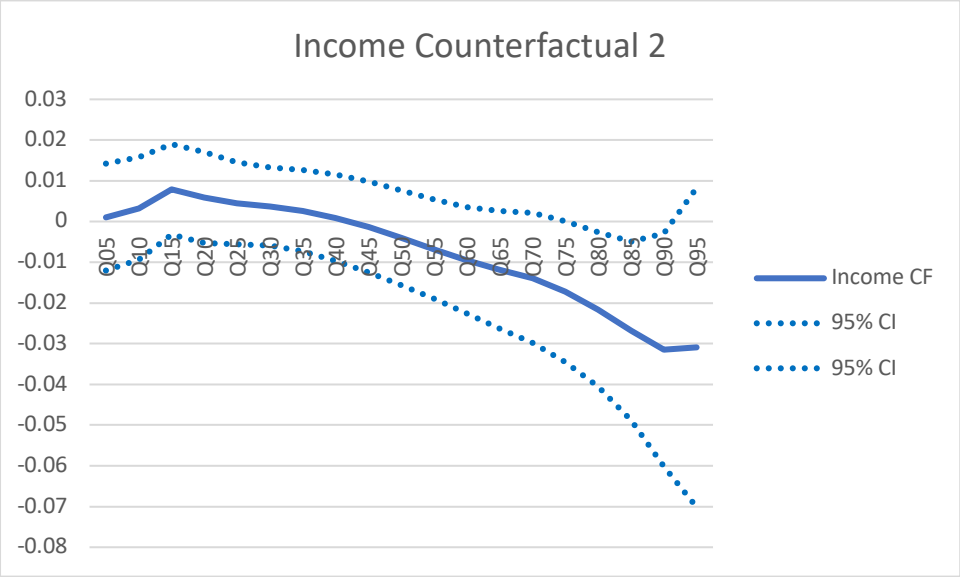
Figure 3: Histogram of exercise index by period

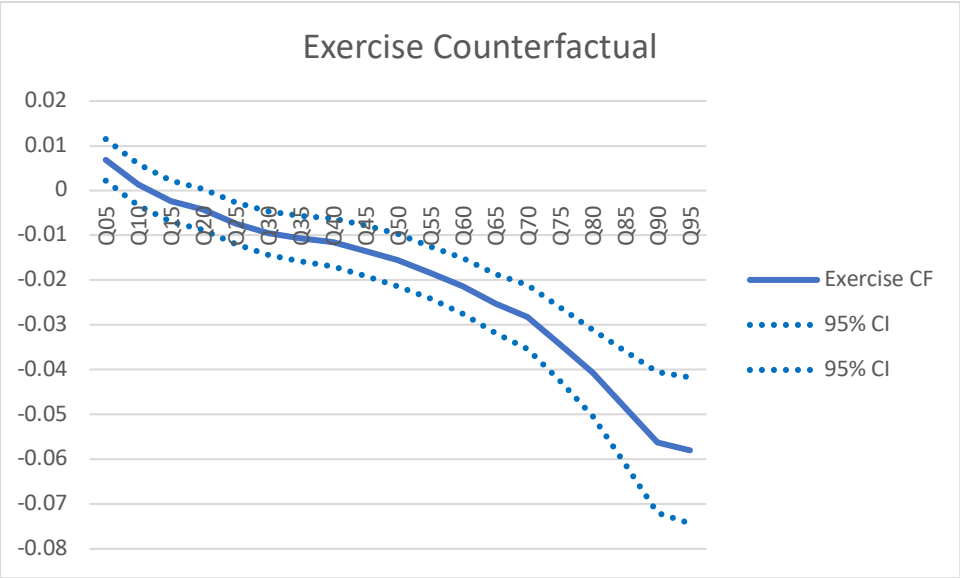


Figures 4a-4d: Effect of Counterfactuals on period 2 BMI by quantile




Income Counterfactual 1
2.00E-03
1.00E-03
0.00E+00
-1.00E-03
-2.00E-03
-3.00E-03
-4.00E-03
-5.00E-03
-6.00E-03
-7.00E-03
-8.00E-03
Q05
Q10
Q15
Q20
Q25
Q30
Q35
Q40
Q45
Q50
Q55
Q60
Q65
Q70
Q75
Q80
Q85
Q90
Q95
Income CF
95% CI
95% CI


Income Counterfactual 2
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
-0.07
-0.08
Q05
Q10
Q15
Q20
Q25
Q30
Q35
Q40
Q45
Q50
Q55
Q60
Q65
Q70
Q75
Q80
Q85
Q90
Q95
Income CF
95% CI
95% CI


Exercise Counterfactual
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
-0.07
-0.08
Q05
Q10
Q15
Q20
Q25
Q30
Q35
Q40
Q45
Q50
Q55
Q60
Q65
Q70
Q75
Q80
Q85
Q90
Q95
Exercise CF
95% CI
95% CI

Appendix Table 1

Q'tile	Wave 1 BMI	Wave 2 BMI	Total Effect	Characteristics Effect and 95% Confidence Interval			
				CVM QR	CVM Logit	RIF	DFL
70	0.9587	0.9599	0.0012	0.0197 0.0117-0.0277	0.0230 0.0124-0.0336	0.0178 0.0113-0.0244	0.0203 0.0135-0.0271
75	0.9881	0.9901	0.0019	0.0229 0.0144-0.0314	0.0217 0.0099-0.0335	0.0018 0.0129-0.0275	0.0274 0.0194-0.0354
80	1.0253	1.0275	0.0021	0.0262 0.0172-0.0352	0.0231 0.0115-0.0347	0.0236 0.0154-0.0319	0.0275 0.0207-0.0343
85	1.0656	1.0748	0.0091	0.0305 0.0209-0.0401	0.0300 0.0138-0.0463	0.0282 0.0189-0.0375	0.0254 0.0175-0.0333
90	1.1170	1.1360	0.0189	0.0362 0.0253-0.0471	0.0530 0.0308-0.0751	0.0331 0.0223-0.0438	0.0440 0.0330-0.0549
95	1.2080	1.2255	0.0174	0.0400 0.0269-0.0531	0.0642 0.0321-0.0962	0.0534 0.0393-0.0674	0.0587 0.0482-0.0692
99	1.3554	1.3876	0.0321	0.0417 0.0228-0.0605	0.5020	0.0420 0.0228-0.0612	0.0483 0.0383-0.0584

All measures here are in normalised BMI i.e. BMI divided by relevant age and gender thresholds

References:

Aune, Dagfinn, Abhijit Sen, Manya Prasad, Teresa Norat, Imre Janszky, Serena Tonstad, Pål Romundstad, and Lars J. Vatten (2016). BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *British Medical Journal*. 353: i2156.

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

Burkhauser, R. And J. Cawley (2008): Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics*, 27, pp. 519-529.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly (2013): "Inference on counterfactual distributions." *Econometrica,* Vol. 81, no. 6, pp 2205-2268.

Cole, T., M. Bellizzi, K. Flegal and W. Dietz (2000): Establishing a Standard Definition for Child Overweight and Obesity Worldwide: International Survey. *British Medical Journal*. 320, pp. 1-6.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica,* Vol. 64, pp 1001-1044.

Essama-Nssah, B., and P. Lambert(2011): Influence Functions for Distributional Statistics. Society for the Study of Economic Inequality Working Paper. ECINEQ WP 2011-236.

Firpo, S., N. Fortin and T. Lemieux (2009): Unconditional Quantile Regressions. *Econometrica.* Vol. 77, pp. 953-973.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (2011): "Decomposition methods in economics." In *Handbook of labor economics* (eds. Ashenfelter & Card) pp. 1-102.

Global BMI Mortality Collaboration (2016). Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *The Lancet* 388, no. 10046, pp. 776-786.

Jones, Andrew Michael, James Lomas, Peter Moore, and Nigel Rice (2016). "A quasi-Monte Carlo comparison of developments in parametric and semi-parametric regression methods for heavy-tailed and non-normal data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* pp. 951-974.

Layte, R., and C. McCrory (2011): *Obesity and Overweight Among Nine Year Olds*. Government Publications. Dublin.

Madden, David (2017). "Childhood obesity and maternal education in Ireland." *Economics & Human Biology* Vol. 27 pp. 114-125.

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

Quail, A., J. Williams, M. Thornton and A. Murray (2014): *A Summary Guide to Wave 2 of the Child Cohort (at 13 Years) of Growing Up in Ireland.* ESRI. Dublin.

Williams, J., S. Greene, E. Doyle et al (2009): *The Lives of 9 Year Olds: Growing Up in Ireland, National Longitudinal Study of Children (Report 1 of the Child Cohort).* Dublin; The Stationery Office.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP17/08 Ronald B Davies and Zuzanna Studnicka: 'The Heterogeneous Impact of Brexit: Early Indications from the FTSE' May 2017
WP17/09 J Peter Neary and Cormac Ó Gráda: 'Brendan M. Walsh (1940-2016): The Economist at Work' May 2017
WP17/10 Morgan Kelly and Cormac Ó Gráda: 'Speed under Sail, 1750–1830' June 2017
WP17/11 Morgan Kelly and Cormac Ó Gráda: 'Technological Dynamism in a Stagnant Sector: Safety at Sea during the Early Industrial Revolution' June 2017
WP17/12 Kate Hynes, Yum K Kwan and Anthony Foley: 'Local linkages: The interdependence of foreign and domestic firms' June 2017
WP17/13 Cormac Ó Gráda: 'Notes on the Demography of the Famine in Ulster' June 2017
WP17/14 Sarah Parlane and Yingyi Tsai: 'Optimal Management of Supply Disruptions when Contracting with Unreliable, Risk-averse, Suppliers' June 2017
WP17/15 Orla Doyle: 'The First 2,000 Days and Child Skills: Evidence from a Randomized Experiment of Home Visiting' July 2017
WP17/16 Christopher Dixon-O'Mara and Lisa Ryan: 'Energy efficiency in the food retail sector: Barriers, drivers, and acceptable policies' July 2017
WP17/17 Andrew E Clark, Orla Doyle, and Elena Stancanelli: 'The Impact of Terrorism on Well-being: Evidence from the Boston Marathon Bombing' September 2017
WP17/18 Kate Hynes, Jie Ma and Cheng Yuan: 'Transport Infrastructure Investments and Competition for FDI' September 2017
WP17/19 Kate Hynes, Eric Evans Osei Opoku and Isabel KM Yan: 'Reaching Up and Reaching Out: The Impact of Competition on Firms' Productivity and Export Decisions' September 2017
WP17/20 Tamanna Adhikari, Michael Breen and Robert Gillanders: 'Are New States More Corrupt? Expert Opinions vs. Firms' Experiences' October 2017
WP17/21 Michael Spagat, Neil Johnson and Stijn van Weezel: 'David Versus Goliath: Fundamental Patterns and Predictions in Modern Wars and Terrorist Campaigns' October 2017
WP17/22 David Madden: 'Mind the Gap: Revisiting the Concentration Index for Overweight' October 2017
WP17/23 Judith M Delaney and Paul Devereux: 'More Education, Less Volatility? The Effect of Education on Earnings Volatility over the Life Cycle' October 2017
WP17/24 Clemens C Struck: 'On the Interaction of Growth, Trade and International Macroeconomics' November 2017
WP17/25 Stijn van Weezel: 'The Effect of Civil War Violence on Aid Allocations in Uganda' November 2017
WP17/26 Lisa Ryan, Karen Turner and Nina Campbell: 'Energy Efficiency and Economy-wide Rebound: Realising a Net Gain to Society?' November 2017
WP17/27 Oana Peia: 'Banking Crises and Investments in Innovation' December 2017
WP17/28 Stijn van Weezel: 'Mostly Harmless? A Subnational Analysis of the Aid-Conflict Nexus' December 2017
WP17/29 Clemens C Struck: 'Labor Market Frictions, Investment and Capital Flows' December 2017
WP18/01 Catalina Martínez and Sarah Parlane: 'On the Firms' Decision to Hire Academic Scientists' January 2018

UCD Centre for Economic Research Email economics@ucd.ie