Article:
Brown, S., Greene, W.H. and Harris, M. (2020) A novel approach to latent class modelling: identifying the various types of body mass index individuals. Journal of the Royal Statistical Society: Series A, 183 (3). pp. 983-1004. ISSN: 0964-1998

https://doi.org/10.1111/rssa.12552

A novel approach to latent class modelling: Identifying the various types of Body Mass Index individuals

October 2019

Abstract

Given the increasing prevalence of adult obesity, furthering understanding of the determinants of measures such as Body Mass Index (BMI) remains high on the policy agenda. We contribute to existing literature on modelling BMI by proposing an extension to latent class modelling, which serves to unveil a more detailed picture of the determinants of BMI. Interest here lies in latent class analysis with: a regression model and predictor variables explaining class membership; a regression model and predictor variables explaining the outcome variable *within* BMI *classes*; and instances where the BMI classes are naturally ordered and labelled by expected values within class. A simple and generic way of parameterising both the class probabilities and the statistical representation of behaviours within each class is proposed, that simultaneously preserves the ranking according to class-specific expected values and yields a parsimonious representation of the class probabilities. Based on a wide range of metrics, the newly proposed approach is found to dominate the prevailing one; and moreover, results are often quite different across the two.

JEL Classification: C3, I12

Keywords: Body Mass Index (BMI), expected values, latent class models, obesity, ordered probability models.

1 Introduction and background

The World Obesity Federation (www.worldobesity.org) states that "the epidemic of obesity is now recognized as one of the most important public health problems facing the world today". This is not surprising given that the World Health Organisation (WHO) in 2011 reported that since 1980 adult obesity rates have doubled worldwide. Indeed, adult obesity is more prevalent than under-nutrition. Around 670 million adults are obese, and 98 million severely so (World Health Organisation 2014). Obesity is a condition of excessive body weight in the form of fat, which is causally linked to a large number of debilitating and life-threatening disorders. The adverse physical and monetary costs of obesity are well-documented. It is generally argued by health experts that given the height of an individual, their weight should lie within a certain range. Accordingly, the most commonly used measure to assess whether an individual is obese is the Body Mass Index (BMI): the ratio of the individual's weight to the square of height. A widely recognised shortcoming of BMI is that it is not an ideal measure of weight-related health status: for example, it fails to distinguish between fat and muscle mass, and is affected by the distribution of fat. Nevertheless, its popularity is attributable to the fact that relative to more accurate anthropometric measurements (skin-fold tests, waist measurements) it is relatively cheap and easy to collect, and hence obtain from large-scale nationally representative samples (Wooden, Watson, and Freidin 2008).

Given the serious health related issues associated with obesity, it is not surprising that modelling BMI and obesity rates has attracted increasing interest from both academics and policy-makers (Cutler, Glaeser, and Shapiro 2003, Chou, Grossman, and Saffer 2004, Philipson and Posner 2008, Mills 2009, Madden 2012, Brown and Roberts 2013, Greene, Harris, Hollingsworth, and Maitra 2014, Hong, Yue, and Ghosh 2015). It is clearly important to select an appropriate modelling approach in the context of such a highly policy relevant application. There is evidence that individuals are essentially (primarily genetically) predisposed to be in particular weight-related health statuses (that is, BMI bands) as an obesity predisposing genotype has been found to be present in 10% of individuals (Herbert, Gerry, and McQueen 2006). That is, it is (medically) very likely that individuals are genetically predisposed to being in different BMI classes. Observed BMI outcomes will be then a combination of the underlying BMI-type range, but tempered by lifestyle choices. Moreover, these different BMI-type classes will undoubtedly react differently (with regard to their observed BMI levels) to a similar set of lifestyle characteristics. So, with regard to an

appropriate empirical strategy, which will simultaneously account for, and identify, these different BMI types, and allow for them to react differently to a similar set of characteristics, several authors have suggested a *latent class* framework (Deb, Gallo, Ayyagari, Fletcher, and Sindelar 2011, Greene, Harris, Hollingsworth, and Maitra 2014).

Latent class modelling has been especially popular in health research (Deb and Trivedi 2002, Bago D'Uva 2005b, Bago D'Uva 2005a, Reboussin, Ip, and Wolfson 2008, Bago d'Uva and Jones 2009, Deb and Holmes 2000, Deb, Gallo, Ayyagari, Fletcher, and Sindelar 2011, Chung, Anthony, and Schafer 2011). It involves probabilistically splitting the population into a finite number of homogeneous classes, or types. Typically, within each of these the same statistical model applies, but with differing parameters allowing the same explanatory variables to have differing effects across the model/classes (Bago d'Uva and Jones 2009).

The latent class modelling contribution starts from the observation that although the classes are latent - by definition - researchers often label them *ex post* according to an observed attribute such as an expected value (EV) within each class. Uncovering evidence of the distinguishing features of the latent classes is an important part of the modelling process. Moreover, a natural inconsistency arises as the (unrestricted) probabilities driving these class allocations will typically not respond to this eventual ordered labeling of them. As an explicit contribution to the literature, we propose a simple way of parameterising both the class probabilities and the statistical representation of behaviours within each class, that simultaneously preserves their ranking according to class-specific EVs and which yields a parsimonious representation of the class probabilities, which is also consistent with the inherent ordering in such. We do this by explicitly enforcing an ordering in the EVs across classes combined with an ordered probabilistic specification for the class assignments. This specification is both consistent with the ordering in the EVs across classes and offers a natural and informative representation of the class assignment probabilities. The results suggest a more detailed picture of the determinants of BMI, with six classes being supported by our proposed approach, as compared to five classes being supported by the standard approach. All of the metrics clearly support the new approach, and moreover, significant differences in *ex post* quantities of interest are found, suggesting that the choice of appropriate approach is, indeed, important.

In summary, interest here lies in latent class analysis, with: a regression model and predictor variables explaining BMI class membership; a regression model and predictor

variables explaining the outcome variable *within BMI classes*; and instances where the BMI classes are naturally ordered and labelled by expected values within class. Our aims are: to uncover both the true number and the underlying characteristics of the (predominantly) genetically determined BMI types (and moreover, how these relate to those determined by the WHO); and to determine the differing drivers of observed BMI outcomes within each of these classes. We are interested in ensuring: a parsimonious form for the class probabilities that is consistent with the inherent ordering in the classes; and to ensure that EVs are indeed ordered within each class.

The explicit contribution of the paper is an extension to the methodology of latent class models (LCMs). Received developments of LCMs include treatments of ordering in a latent tendency that relates to the probabilities of latent class membership. The structure developed here extends the concept of ordering to the cross class comparison of the distribution of the observed outcome. The first of these appears occasionally in the received literature; whilst the second is new. The two combined lead to a methodological contribution that ties the empirical model to the theory of the data generating process of the observed data. In our application, the end result appears to provide a better, and more parsimonious, fit to the data, although it is important to acknowledge that this will not always be the case. What we do develop here though, is a modeling framework in which the researcher can learn more about causal relationships, partial effects, and meaningful simulation of observed outcomes.

2 Statistical and modelling framework

The model of interest here is a LCM with predictors in the class proportions and the response densities; the use of covariates in the former has been well-studied and widely used (Vermunt 2010, Bartolucci, Farcomeni, and Pennoni 2012). However, our contribution here lies in *how* they enter. The suggested approach produces a solution in which the classes are ordered (with respect to EVs) for all possible predictor values. When the classes are ordered, it is logical to use an *ordered* regression model for these.

The overall density for individual i ($i = 1, \ldots, N$), $f(y_i|x_i, \boldsymbol{\theta})$, is assumed to be an additive mixture density of Q distinct sub-densities weighted by their appropriate mixing probabilities, π_{iq}. The outcome variable of interest is y_i, affected by the $(k_x \times 1)$ vector of covariates in the model, x_i, which have different effects in each q class, and $\boldsymbol{\theta}$ denotes all of the parameters

of the model. The corresponding mixed density is

$$f(y_i|x_i) = \sum_{q=1}^{Q} \pi_{iq} \times f(y_i|x_i, \theta_q). \tag{1}$$

Here interest is where π_{iq} is a function of predictors (z_i). A very common approach is to employ a multinomial logit (MNL) form to quantify the effects of z_i on the probabilities of class membership; and implicitly, probabilistically, to allocate individuals to the various classes (Greene 2012). An element of the specification search is determining the appropriate number of classes, Q^*. A common approach is to use information criteria (IC) metrics; such as BIC/SC (Schwarz 1978), AIC (Akaike 1987), corrected AIC, $CAIC$ (Bozdogan 1987), and Hannon-Quinn, $HQIC$, (Hannan and Quinn 1979).

2.1 Monotonically increasing expected values

In most empirical applications of LCMs there is an *ex post* labelling of the classes based upon estimated EVs within each of the $q = 1, \ldots, Q$ classes (Deb and Holmes 2000, Deb and Trivedi 2002, Bago D'Uva 2005b, Bago D'Uva 2005a, Bago d'Uva and Jones 2009, Deb, Gallo, Ayyagari, Fletcher, and Sindelar 2011). Although it is an important output of the modelling process, this ordering of the classes is not ensured during the estimation process. Here we suggest an easy to implement way to do so, and thereby be consistent with the research question at hand. Thus, with regard to the modelling of observed BMI outcomes, we simply wish to ensure that as classes "increase" with respect to EVs, the EVs do actually rise.

The properties of the output variable to be modelled will dictate the specific functional form for the specification of the density $f_q(y_i|x_i, \theta_q)$; given the continuous nature of BMI, for us this is a simple linear regression model. However, it is useful here to consider the determination of observed y_i *within* each class. Consider a latent index function of the form

$$y^*_{i,q} = x'_i\beta_q + \varepsilon_{i,q}, \tag{2}$$

where β_q are the response parameters and $\varepsilon_{i,q}$ a disturbance term. The $y^*_{i,q}$ of equation (2) will be related to observations within group $y_{i,q}$ via a mapping dictated by $f(y_i|x_i, \theta_q)$. Regardless of the model, EVs *on the assumption of underlying ordinality or cardinality of observed* $y_{i,q}$, are monotonically related to the index $x'_i\beta_q$. This generic approach would be similarly applicable to *any* outcome variable of interest, assuming it embodies some underlying form

of ordering, generally defined. Thus ensuring that $x_i'\beta_{q=1} \leq x_i'\beta_{q=2} \leq \cdots \leq x_i'\beta_Q$ will ensure that $EV_{i,q=1} \leq EV_{i,q=2} \leq \cdots \leq EV_{i,Q}$.

We define $EV_{i,q}^*$ as a function of the index $x_i'\beta_q$ (such that $EV_{i,q}^*$ will be positively, and monotonically related to the true EV, $EV_{i,q}$). Consider modelling the $EV_{i,q}^*$ in the first, or smallest EV, class ($q = 1$) as simply

$$EV_{i,q=1}^* = EV_{i,q=1}. \tag{3}$$

In a linear regression setting, this would amount to setting $EV_{i,q=1} = x_i'\beta_{q=1}$. Without the necessity of being model-specific we now want to express the "mean" function in $q = 2$ which, by construction we wish to be greater than that for $q = 1$,

$$EV_{i,q=2}^* = EV_{i,q=1}^* + \exp\left(x_i'\beta_{q=2}\right). \tag{4}$$

Therefore, in a simple regression setting, we would have $E\left(y_{q=1}\left|x\right.\right) = x_i'\beta_{q=1}$ and $E\left(y_{q=2}\left|x\right.\right) = E\left(y_{q=1}\left|x\right.\right) + \exp\left(x_i'\beta_{q=2}\right)$. As long as the relationship between EV and EV^* is monotonic, enforcing $EV_{i,q=1}^* \leq EV_{i,q=2}^* \leq \cdots \leq EV_{i,Q}^*$ will enforce $EV_{i,q=1} \leq EV_{i,q=2} \leq \cdots \leq EV_{i,Q}$. This progression is simply continued for subsequent classes. This approach ensures that the EV's (generally defined) are ordered across classes, whilst the specification of $EV_{i,q=1}$ is likely to be model-specific. For example, in a linear regression $EV_{i,q=1} = x_i'\beta_q$; whilst $EV_{i,q=1} = \exp\left(x_i'\beta_{q=1}\right)$ in a Poisson count model; and so on.

Assuming that the within class models are linear regressions, then within class 1 partial effects are given by the respective coefficients in that class (or the appropriate partial effect in nonlinear models). Coefficients $\beta_{q,k}$, $q > 2$, can be directly interpreted as differential effects with respect to $EV_{i,q-1}^*$. Take for example, the partial effect of x_k: in the linear regression case:

$$\begin{array}{l} EV_1^* = x'\beta_1;\ \partial EV_1^*/\partial x_k = \beta_{1,k}, \\ EV_q^* = EV_{q-1}^* + \exp\left(x'\beta_q\right);\ \partial EV_q^*/\partial x_k = \exp\left(x'\beta_q\right)\beta_{q,k} + \partial EV_{q-1}^*/\partial x_k\ ,\ q = 2, \ldots \end{array} \tag{5}$$

Thus the partial effect for x_k in $q = 2$ includes a differential effect to that of $q = 1$. If $\beta_{2,k}$ (*i.e.*, the coefficient of x_k in the second class) is negative, so will be the differential effect, and the magnitude given by the value of this coefficient and the weighting term $\exp\left(x'\beta_2\right)$. The signs of these partial effects are not constrained by the $\exp\left(\cdot\right)$ transformation to be positive, but will be differentiated by the signs and magnitudes of their various components. The signs of the differential effects from $q = q^*$ to $q = q^* + 1$ will be uniquely determined by

the sign of the coefficient in that class, $\beta_{q^*,k}$. The coefficients are not, as in most nonlinear models, direct estimates of partial effects; with the exception here of $q = 1$. A negative coefficient in a particular class does not necessarily imply a negative partial effect within that class.

Overall partial effects can be obtained by constructing a weighted average of EV's across classes, and differentiating this with respect to the covariate of interest. In our analysis of BMI, we use prior probabilities for weights along with numerical derivatives, and apply the delta method to obtain standard errors. It may be that in any particular application, neighbouring class EV's might converge and/or similarly boundary parameters. This could well be evidence that too many classes have been estimated, which should be evidenced by the model metrics discussed in this paper. Moreover, even if EV's are very similar across classes, this does not necessarily imply that partial effects will also be, as EV's are a function of the composite index $x'\beta_q$ as opposed to any single component of this. This is similarly true of the traditional approach.

We note here that a similar ranking could also be obtained by enforcing other restrictions. For example, response parameters within the class regressions could be forced to be equal across classes, and ordering imposed by simple ordering of the constant terms. However, in general we would recommend against such an approach, as it appears rather arbitrary and overly restrictive and would appear to have adverse consequences on overall model fit.

2.2 Class probabilities

The specification of the mixing weights can be a substantive part of the model construction. In a recent latent class study of BMI, Greene, Harris, Hollingsworth, and Maitra (2014) suggested a model that embodied a latent trait, the presence of the unobservable FTO gene, for which observable characteristics, z_i (such as country of origin), might contain relevant information. The conditional (on z_i) usual multinomial logit (MNL) form for the prior probabilities,

$$\pi_{iq} = \frac{\exp\left(z_i'\gamma_q\right)}{\sum_{q=1}^{Q}\exp\left(z_i'\gamma_q\right)}, \tag{6}$$

where one of the γ_q vectors is normalised to zero, is a convenient choice that has been used in many studies. Indeed, this is now standard in the received applications, and has been built into many popular software packages. The approach has the advantage of being relatively unrestrictive. It is also a particularly convenient form for the EM algorithm (see, Alfo,

Salvati, and Ranalli (2017) and Friedl and Kauermann (2000), for example). We extend equation (6) in two directions. First, we seek a more parsimonious functional form. The functional form in equation (6) adds a full $k_z + 1$ parameter vector (including a constant term) for each class. Model selection criteria that penalize specifications with a large number of parameters will tend to discriminate against the MNL model, possibly unduly so. Second, we connect the class probabilities to an inherent ordering of the classes. There is no obvious way to do so with equation (6), but it is relatively straightforward with the specification proposed below.

The specification search for $LCMs$ is typically driven by information criteria such as BIC. As stated above, IC metrics are structured so as to penalize large models. The MNL form is at a disadvantage to a more parsimonious one in that each new class adds potentially many parameters to the model. We find that in many empirical exercises, the preferred number of classes is less than or equal to three. It may well be that more classes could be identified if the analysis were based on a more compact form for the class probabilities. On this basis, class-specific results might be distorted by a merging of heterogeneous classes.

Although a variety of approaches appear in the received studies, the MNL form is by far the most common. However, Fabrizi, Montanari, and Ranalli (2016) do mention an ordered logit alternative of the form

$$\ln \frac{Prob\left(q \leq c \,|z_i\right)}{Prob\left(q > c \,|z_i\right)} = \mu_c + z_i'\gamma \tag{7}$$

that would be appropriate if an unobserved continuous variable is assumed to underlie the class assignment. This is a useful starting point for our extensions, as we have assumed not only that the class assignments are ordered in this fashion, but also that the ordering extends to the main outcomes in the classes through the means, $EV_q^* > EV_{q-1}^*$. An ordered probit (OP) formulation for the prior probabilities,

$$\pi_{iq} = \Phi\left(\mu_q - z_i'\gamma\right) - \Phi\left(\mu_{q-1} - z_i'\gamma\right), \tag{8}$$

where Φ is the standard normal CDF, appears particularly appropriate and has the advantage of a more parsimonious specification. The addition of another class to this formulation adds only a single additional cut point, μ, again, consistent with a partitioning of the range of an underlying continuous variable. The assumed form implies

$$Prob_i(class = q \,|z_i) = Prob\left(\mu_{q-1} < z_i'\gamma + \varepsilon_i < \mu_q\right), \varepsilon_i \sim N\left(0, 1\right). \tag{9}$$

It is conceivable that the ordered choice form of the class probabilities is restrictive relative to the MNL form. However, if the type of ordering suggested here is an intrinsic part of the data generating process, then a model that does not take advantage of that feature, such as the MNL with covariates, could overfit the data. In essence, the only structure that the MNL imposes on the discrete outcomes is that only one of them can occur – any inherent ordering is not modeled. In the simulation experiments presented in the Online Appendix, Section 1, even with the data generated by an MNL process, applying the OP format does not adversely affect the results. Indeed, researchers typically do not use the MNL format when the data are naturally ordered. The format may be likewise out of place here. There are other ways to restrict the MNL model, perhaps along the lines of Heckman and Singer (1984) with some device to impose an ordering on the constant terms. However, the OP approach has an intuitive appeal and is straightforward to implement.

It is interesting to compare how our suggested approach described above, relates to existing studies. Ordering in the class probabilities has appeared in numerous applications in the literature, such as Croon (2002), Fabrizi, Montanari, and Ranalli (2016), Vermunt (2010) and Karabatsos and Sheu (2004). In these studies, the model for the underlying class probabilities is built upon a latent variable that asserts an ordering to the classes. The class specific segment of the model is heterogeneous, but not otherwise ordered. Croon (2002), for example, examines a class specific multinomial distribution, with no implicit comparison across classes. More recently, Fabrizi, Montanari, and Ranalli (2016) use the segmentation to deduce the sizes of the latent classes. A class that is higher on the underlying scale is not necessarily larger. In a similar vein to our proposal, Alfo, Salvati, and Ranalli (2017) consider a mixture of quantile regression models. The specific quantile examined (for example, the median), however, is fixed in advance, and is common across the classes. The classes here are not ordered. The quantile regressions would seem to embody an ordering of sorts - within a class, the 90^{th} percentile of $y|x$ is necessarily greater than the 80^{th} quantile of $y|x$. However, it does not follow, for example, that if class 3 is ranked higher than class 2, that the 90^{th} percentile in class 3 is necessarily greater (or smaller) than the 90^{th} percentile of class 2. It is this latter comparison that interests us here.

2.3 Extension to a random effects panel specification

The application here involves two waves of the British Household Panel Survey, $BHPS$ (2004, 2006, see below). In general, the extension of the latent class specification to panel data involves treating waves jointly, holding constant over time the elements of the model that are specific to the individual. This becomes equivalent to treating the model parameters $\boldsymbol{\theta}$ non-parametrically as a random vector with discrete support – the discrete outcomes define the classes. Accordingly there is a single set of class probabilities, π_{iq}, $q = 1, \ldots, Q$, for each individual that is time invariant. For panel data, assuming conditional (on q) independence, the joint density for the T_i observations for individual i is

$$f(y_{i1,\ldots,}y_{i,T_i}|x_{i1,\ldots,}x_{i,T_i}) = \sum_{q=1}^{Q} \pi_{iq} \times \prod_{t=1}^{T_i} f(y_{it}|x_{it}, \theta_q), \tag{10}$$

with corresponding log-likelihood

$$\ln L = \sum_{i=1}^{N} \ln f(y_{i1,\ldots,}y_{i,T_i}|x_{i1,\ldots,}x_{i,T_i}). \tag{11}$$

Note, model identification of the new procedure is discussed in the Online Appendix, Section 2.1.

3 Data

We analyse data drawn from the $BHPS$, which is a longitudinal survey of private households in Great Britain, and was designed as an annual survey of each adult member of a nationally representative sample of households. The $BHPS$ sample design was based on a clustered stratified sample of addresses across Great Britain with individuals living at these addresses being identified as potential panel members. The first wave in 1991 achieved a sample of some 5,500 households, covering approximately 10,300 adults from 250 areas of Great Britain (Taylor 2010). In only two waves 14 (2004) and 16 (2006), was information collected on weight and height, which we use to calculate BMI. Our data comprises of 19,628 observations covering individuals aged 16 and over. Average BMI in the sample is 27.218, with a standard deviation of 5.43 (Table 1), which lies in the lower end of the overweight BMI category suggested by the WHO. The WHO classification assigns adults to either underweight, normal range, overweight or obese categories (WHO 2000); underweight

is $BMI < 18.5$; normal is $18.5 \leq BMI < 24.99$; overweight $25 \leq BMI < 29.99$; and obese $BMI \geq 30$.

We treat class membership as time-invariant and search for indicators for different genetic types to explain membership of these BMI classes. Such an approach would therefore be consistent with there being an obesity predisposing genotype present in individuals (Herbert, Gerry, and McQueen 2006). Following the related literature, we include all available time invariant characteristics, such as birth cohort and gender. We also control for socioeconomic characteristics relating to family background: the respondent's parents' occupation (at respondent age 14). Similarly, we include controls for parents' education. Finally, we include time invariant controls for personality: the *Big Five* personality traits. We follow the standard practice to mitigate against the potential problem of life-cycle effects influencing these and condition each personality trait on a polynomial in age (Nyhus and Pons 2005). The resulting residuals are standardised and used as indicators of personality traits.

In the outcome equation, we again follow the received literature (Cutler, Glaeser, and Shapiro 2003, Chou, Grossman, and Saffer 2004, Brown and Roberts 2013, Greene, Harris, Hollingsworth, and Maitra 2014) and control for age, number of children, marital status, household income, employment status, highest level of educational attainment and region. Finally, we control for a wide range of health problems: problems with: arms, legs, hands, *etc.*; sight; hearing; skin conditions/allergy; chest/breathing; heart/blood pressure; stomach or digestion; diabetes; anxiety, depression, *etc.*; migraine; and cancer. We follow the relevant literature and consider a composite variable (*Comorbidities*) denoting the number of reported health problems, see for example, Banks, Blundell, and Emmerson (2015) and Marquesa, Cruzb, Regob, and da Silvab (2016). Descriptive statistics are presented in Table 1.

4 Results

4.1 Model comparison

We firstly compare a range of different latent class models using standard IC metrics in order to ascertain the preferred approach. Note that as currently the suggested approach is not available in commercial software, all estimations were obtained using author-written *Gauss* script utilising the *cmlMT* (constrained) maximum likelihood add-in module (template *Gauss* code for estimation, as well as the procedure file used for estimation, are freely

Table 1: Descriptive statistics, $N = 19,628$

Variable	Mean	Standard Deviation
BMI	27.218	(5.43)
$Female$	0.503	(0.50)
$Birth\ cohort$ 1940	0.165	(0.37)
$Birth\ cohort$ 1950	0.179	(0.38)
$Birth\ cohort$ 1960	0.212	(0.41)
$Birth\ cohort$ 1970	0.165	(0.37)
$Birth\ cohort$ $1980-1990$	0.094	(0.29)
$Big\ 5: Agreeableness$	-0.002	(1.00)
$Big\ 5: Conscientiousness$	-0.003	(1.00)
$Big\ 5: Extraversion$	-0.002	(1.00)
$Big\ 5: Neuroticism$	0.004	(1.00)
$Big\ 5: Openness\ to\ experience$	-0.001	(1.00)
$Father\ some\ education$	0.152	(0.36)
$Father\ further\ education$	0.289	(0.45)
$Mother\ some\ education$	0.222	(0.42)
$Mother\ further\ education$	0.177	(0.38)
$Father\ professional/managerial$	0.224	(0.42)
$Father\ skilled\ non-manual$	0.069	(0.25)
$Father\ manual/unskilled$	0.490	(0.50)
$Mother\ professional/managerial$	0.092	(0.29)
$Mother\ skilled\ non-manual$	0.117	(0.32)
$Mother\ manual/unskilled$	0.203	(0.40)
$Age10$	4.804	(1.72)
$Number\ of\ children$	0.587	(0.96)
$Married$	0.587	(0.49)
$(Log\ of)\ household\ income$	10.213	(0.73)
$Employed$	0.608	(0.49)
$Not\ in\ the\ labour\ force\ (NILF)$	0.144	(0.35)
$Degree$	0.150	(0.36)
$Vocationaldegree$	0.303	(0.46)
$A-level$	0.117	(0.32)
$GCSE$	0.159	(0.37)
$Comorbidities$	1.267	(1.44)
$Midlands$	0.100	(0.30)
$North$	0.151	(0.36)
$Wales$	0.166	(0.37)
$Scotland$	0.175	(0.38)
$Northern\ Ireland$	0.167	(0.37)

Table 2: Model selection metrics

	BIC	AIC	$CAIC$	$HQIC$	$Parameters$
Linear Regression	121,086	120,936	121,105	120,985	19
2-class (unrestricted)	115,537	115,064	115,597	115,219	60
3-class (restricted)	113,182	112,552	113,262	112,758	80
3-class (unrestricted)	113,308	112,512	113,409	112,773	101
4-class (restricted)	111,717	110,929	111,817	111,187	100
4-class (unrestricted)	112,015	110,895	112,157	111,262	142
5-class (restricted)	111,143	110,197	111,263	110,507	120
5-class (unrestricted); model b	111,484	110,041	111,667	110,513	183
6-class (restricted); model a	**110,675**	**109,571**	**110,815**	**109,932**	140
6-class (unrestricted)	111,627	109,861	111,851	110,439	224
7-class (restricted)	111,910	110,649	112,070	111,062	160
7-class (unrestricted); model c	111,676	109,586	111,941	110,271	265
$Vuong\ (BIC)$; $a\ vs\ b$	8.48				
$Vuong\ (AIC)$; $a\ vs\ c$	14.4				
$Vuong\ (BIC, AIC)$; $a\ vs\ a$	–	–			

Note: preferred model for each metric in **bold**.

available - see the Online Appendix Sections 2 and 3 for details including an example likelihood function). Further estimation details, including starting values and a discussion of maximum likelihood techniques *versus* the *EM* algorithm (which turns out to be invalid here), are also discussed in the Online Appendix, Section 2.

We consider 12 models in total, with up to $Q = 7$ for both approaches (including a simple linear regression model). In Table 2 we present in **bold** for each IC metric, the favoured model (the $Parameters$ column details the total number of parameters estimated in each specification). As is usual in such exercises, we simply let the IC metrics dictate the optimal number of classes.

It is reassuring to see that *all* of the IC metrics unanimously favour the 6-class restricted (OP) model. However, in terms of identifying the appropriate unrestricted model for comparison purposes, there is some disagreement amongst the IC metrics with respect to selecting amongst the unrestricted models: BIC and $CAIC$ favour the 5-class, whilst AIC and $HQIC$ the 7-class. However, there is much evidence to suggest that AIC is inconsistent and tends to select models that are over-fitted; see for example, Koehler and Murphree (1988). In particular, in the mixture context, AIC tends to overestimate the correct number of components/classes (Soromenho 1994). On this basis we select the 5-class unrestricted

model for comparison purposes.

We also consider three variants of the $Vuong$ test for non-nested models. To take into account potentially large differences in model sizes, we use the BIC bias corrected version of the $Vuong$ statistic (Vuong 1989). Based on the two metrics most commonly used in the related literature (AIC, BIC): $Vuong$ (BIC) considers the two top-performing models according to BIC; $Vuong$ (AIC) the top two according to AIC; and $Vuong$ (BIC, AIC), the top one from each. Both the AIC and the BIC select the restricted 6-class model and, in both instances, this model is preferred to the competing unrestricted 5- and 7-class models. These findings make a very compelling case for the 6-class restricted model.

In Table 3, we present some summary statistics for the preferred restricted 6-class model and the 5-class unrestricted model: EVs by BMI class (evaluated at sample means); average posterior class probabilities; and finally class-specific dispersion parameters. "Overall" EVs were calculated as the (prior probability) weighted average of the class-specific ones. Table 3 presents the increasing pattern in the EVs from classes 1 to 6 for the restricted 6-class model and those from classes 1 to 5 (reported in increasing order) for the unrestricted 5-class one.

For classes 1 to 3, all of the EVs, posterior probabilities and dispersion parameters are similar across the preferred restricted model and the unrestricted model. For example, the EV in class 1 (EV_1) is 20.14 compared to 20.73; with a probability of 10% (15%); and with a dispersion parameter of 1.464 (1.538). Similarly, we see that the EVs for class 4 restricted and unrestricted both lie in the end of the WHO defined *overweight* range $(20 - 29.99)$; and at 27.73 and 29.46, respectively, are close. Furthermore, the proportion of individuals estimated to be in this class, reflected by the posterior probabilities, is the same for each model, at 12%, and, similarly, the spread of individuals' BMIs within-class, 1.2 *c.f.* 1.4, respectively, are also close.

There are some differences for the largest EV BMI classes. For example, for class 5, the EVs are relatively close (29.28 compared to 31.52, for restricted and unrestricted, respectively), and the posteriors are identical, although there are some are some differences in the dispersion parameters (4.19 compared to 5.99). The difference in the dispersion parameters could reflect the additional class for the restricted model, class 6, which is characterised by a relatively high dispersion parameter, at 6.43. From the perspective of these summary measures, it is clear that the choice of approach can make a significant difference.

Focusing on the results from the preferred 6-class restricted model, for class 1 the EV of 20.14 sits at the low end of the WHO defined range of *normal weight* $(18.5 - 24.99)$. Based on the posterior probability, this class is characterised by one of the smallest numbers of individuals (at 10%). Given the position of the mean within this class (and its dispersion), this suggests that individuals within this category are more likely to slip into the *underweight* one, as opposed to the *overweight* $(25 - 29.99)$ one. Turning to the next class, with a mean of 22.75, this also sits within the *normal weight* range, but at the higher end. Judging by the spread of this distribution however (the lowest of any class), individuals within this class have a relatively low probability of moving far from the mean. Based on the posterior probability, 18% of individuals are estimated to be in this group. Class 3 ($EV = 25.19$) falls into the very lowest part of the *overweight* range, meaning that although the dispersion is small here (at 1.08), many of these individuals would still be on the borderline of the normal/overweight range. Around 19% of the population are estimated to be in this class.

Of more concern however, are classes 4, 5 and 6. With means of 27.73, 29.28 and 35.14 respectively, these fall into the (mid and very high ends of) *overweight* and *obese* (> 30). Moreover, for class 5 the average posterior probability is "large" (at 0.31), suggesting that a worryingly large proportion of the population lie in this class. The dispersion of this distribution is relatively large (at 4.185), especially compared with classes 1-4. This implies that individuals genetically predisposed to be in this *overweight* class, can use lifestyle options to place themselves in healthier weight-related ranges (although, by symmetry, this also implies that there is significant risk of slipping into *obesity* as well). However, given the placement of the mean with respect to this range, it is unfortunately more likely that individuals within this class will fall into the *obese range* than the *healthy weight* one. Finally, there is a worryingly large proportion in the *obese* class (9%) ; the spread within this distribution is very large, again suggesting that for these individuals lifestyle factors, for example, could be used to move themselves into much healthier weight ranges.

4.2 Parameter estimates

The class membership equation is reasonably well-specified (Table 4), with gender, the birth cohort controls, personality traits and, to a lesser extent, childhood conditions generally driving the statistical significance. Positive (negative) OP coefficients imply higher probabilities of being in the highest (lowest) classes (with the intervening ones being less clear:

Table 3: Expected values, averaged posterior probabilities and dispersion parameters

	$Q=6; OP$			$Q=5; MNL$		
	Expected	*Post.*		*Expected*	*Post.*	
	Value	*prob.*	*Dispersion* (σ_q)	*Value*	*prob.*	*Dispersion* (σ_q)
Class 1	20.14 (0.06)**	0.10	1.464 (0.03)	20.73 (0.05)**	0.15	1.538 (0.03)**
Class 2	22.75 (0.04)**	0.18	1.074 (0.03)	23.69 (0.03)**	0.21	1.138 (0.03)**
Class 3	25.19 (0.04)**	0.19	1.077 (0.03)	26.32 (0.04)**	0.21	1.248 (0.03)**
Class 4	27.73 (0.04)**	0.12	1.165 (0.04)	29.46 (0.05)**	0.12	1.397 (0.04)**
Class 5	29.28 (0.11)**	0.31	4.185 (0.08)	31.52 (0.10)**	0.31	5.985 (0.06)**
Class 6	35.14 (0.45)**	0.09	6.429 (0.18)			
Overall	26.76 (0.05)**	–	–	27.76 (0.04)**	–	–

Notes: ** and * denote significant at 5, and 10% size. *Post. prob.* is posterior probability.

Greene and Hensher (2010)). Birth cohorts 1960, 1970 and 1980/90 are associated with being in the higher BMI classes; and *Conscientiousness*, *Neuroticism*, *Extraversion* and *Openness to experience* are also strong predictors of class membership. The indicator for *Mother further education* is associated with being in lower BMI classes; whereas that for *Father manual/unskilled* and *Mother manual/unskilled* are the opposite.

In Table 5, we present the class-specific partial effects. To aid interpretation, we label these classes according to the above analyses based on the EVs within each one (Table 3), and where these lie with respect to the WHO defined ranges: *low normal (class 1), high normal (class 2), low over (class 3), mid over (class 4)*, *high over (class 5)* and *obese (class 6)*. As would be expected, the partial effects differ dramatically across the 6 classes in terms of both size and statistical significance. In the case of age, the partial effects of the linear term are positive and statistically significant in all 6 classes and increasing in magnitude from class 1 to class 6. Those of the squared term again differ across classes, and increase in (absolute) magnitude with class. Within each class then, individuals' BMI initially rises with age, peaks, and then starts to decline. The single effect of age (Age), shows that for every year one ages in class 6, BMI only increases by some 0.005 per year. On the other hand, this number is much larger for class 4 at 0.055.

Whilst the number of dependent children appears to have no statistically significant effect across the six classes, the effect of being married appears to quite significantly (both in economic and statistical terms) raise BMI in all but class 6. Income has a strong significant positive effect in classes 1, 2 and 3. Being employed has a large and significant positive effect

Table 4: Class membership equation; preferred specification

Variable	Estimated coefficient	Standard error
$Female$	-0.324	$(0.02)^{**}$
$Birth\ cohort$ 1940	0.040	(0.04)
$Birth\ cohort$ 1950	0.014	(0.04)
$Birth\ cohort$ 1960	0.120	$(0.04)^{**}$
$Birth\ cohort$ 1970	0.239	$(0.04)^{**}$
$Birth\ cohort$ $1980-1990$	0.225	$(0.06)^{**}$
$Agreeableness$	-0.007	(0.01)
$Conscientiousness$	-0.076	$(0.01)^{**}$
$Extraversion$	0.071	$(0.01)^{**}$
$Neuroticism$	-0.050	$(0.01)^{**}$
$Openness\ to\ experience$	-0.027	$(0.01)^{**}$
$Father\ some\ education$	-0.048	(0.04)
$Father\ further\ education$	-0.031	(0.03)
$Mother\ some\ education$	-0.050	(0.03)
$Mother\ further\ education$	-0.092	$(0.04)^{**}$
$Father\ professional/managerial$	0.014	(0.04)
$Father\ skilled\ non-manual$	0.001	(0.05)
$Father\ manual/unskilled$	0.095	$(0.03)^{**}$
$Mother\ professional/managerial$	0.049	(0.04)
$Mother\ skilled\ non-manual$	0.000	(0.04)
$Mother\ manual/unskilled$	0.122	$(0.03)^{**}$
μ_1	-1.346	$(0.05)^{**}$
μ_2	-0.631	$(0.04)^{**}$
μ_3	-0.103	$(0.04)^{**}$
μ_4	0.206	$(0.04)^{**}$
μ_5	1.324	$(0.07)^{**}$

Notes: ** and * denote significant at 5, and 10% size, respectively.

in class 1; a significant negative effect in classes 3, 4 and (weakly) 6. On the other hand, not being in the labour force, has quite large and negative effects (-0.343 and -0.374), but only in classes 2 and 3, *i.e.*, the *high normal* and *low over* classes.

There appears to be considerable heterogeneity in the effects of educational attainment across the classes. For example, having a degree as the highest level of educational attainment has a large, and statistically significant negative effect for class 5 and positive significant effects for classes 2, 3 and 4. Having a vocational degree has an effect (positive, but smaller compared to the *Degree* effects) only in class 4. $A-level$ has a negative effect for class 5; whereas $GCSE$ has an effect (positive) in classes 2, 3 and 4. A "causal protective effect" of education on BMI has previously been found in the literature (Webbink, Martin, and Visscher 2010, Brunello, Fabbri, and Fort 2013).

We control for health conditions by entering the composite *Comorbidities* variable. Indeed, this variable is a very strong driver of BMI levels across all classes. As the number of comorbidities rises, it has a small (but significant) negative effect in class 1, being associated with lower BMI levels for individuals in the *low normal* category. The effect of the *Comorbidities* variable is positive and significant across classes 2 to 6 and increases in magnitude across the classes, from 0.06 (class 2) to 0.70 (class 6). At the higher BMI classes, the effect is more pronounced: as the within class EVs increase, the effects of worsening ill-health suggest that these individuals find it harder to maintain a healthy weight range, via reduced exercise levels and the like. With this health proxy, we note the clear potential for reverse causation and that our findings are interpreted as correlations rather than causation (we return to this below). Finally, the regional effects are often statistically significant, especially in classes 1-4, with considerable heterogeneity in terms of magnitude apparent across both classes and regions.

Although the above results illustrate how such a LCM approach can highlight differential partial effects across classes, the approach could also simply be used as a tool to allow for more unobserved heterogeneity in the modelling exercise. If so, one would assume that the researcher would primarily be interested only in overall partial effects. Moreover, if the overall partials from the 6-class restricted and 5-class unrestricted models were similar, it could be argued that our suggested approach has very little benefit and/or effect in practice. Hence to explore this issue, Table 6 compares the overall (prior probability weighted) partial effects across the two models. We also include simple OLS results here as well.

Table 5: Class-specific partial effects

Variable	Class 1 *(low normal)*	Class 2 *(high normal)*	Class 3 *(low over)*	Class 4 *(mid over)*	Class 5 *(high over)*	Class 6 *(obese)*
$Age/10$	1.123	2.046	2.529	2.722	3.567**	5.560**
	(0.14)**	(0.11)**	(0.11)**	(0.13)**	(0.30)	(0.78)
$Age^2/1000$	−0.876	−1.631	−2.090	−2.259	−3.293	−5.732**
	(0.13)**	(0.11)**	(0.10)**	(0.13)**	(0.33)**	(0.83)
Age	0.028	0.048	0.052	0.055	0.040	0.005
$Number\ of\ children$	−0.004	−0.054	−0.054	−0.066	−0.085	0.204
	(0.05)	(0.04)	(0.03)	(0.04)	(0.07)	(0.18)
$Married$	0.481	0.520	0.386	0.414	0.468	−0.008
	(0.09)**	(0.06)**	(0.06)**	(0.08)**	(0.14)**	(0.44)
$(Log\ of)\ household\ income$	0.148	0.101	0.101	−0.024	0.202	−0.068
	(0.06)**	(0.04)**	(0.04)**	(0.05)	(0.11)*	(0.32)
$Employed$	0.562	−0.031	−0.292	−0.351	0.212	−1.215
	(0.13)**	(0.08)	(0.08)**	(0.12)**	(0.24)	(0.66)*
$Not\ in\ the\ labour\ force$	−0.003	−0.343	−0.374	−0.100	−0.177	−0.237
	(0.14)	(0.09)**	(0.10)**	(0.13)	(0.27)	(0.69)
$Degree$	0.177	0.227	0.276	0.292	−0.898	−1.067
	(0.12)	(0.09)**	(0.10)**	(0.13)**	(0.26)**	(0.72)
$Vocational\ degree$	0.180	−0.015	−0.021	0.224	0.023	−0.162
	(0.11)*	(0.07)	(0.07)	(0.10)**	(0.14)	(0.54)
$A-level$	0.196	−0.094	−0.021	0.083	−0.589	0.260
	(0.15)	(0.10)	(0.10)	(0.14)	(0.23)**	(0.59)
$GCSE$	0.041	0.233	0.384	0.497	0.185	−0.651
	(0.13)	(0.08)**	(0.09)**	(0.11)**	(0.18)	(0.64)
$Comorbidities$	−0.062	0.056	0.127	0.185	0.376	0.703
	(0.03)**	(0.02)**	(0.02)**	(0.02)**	(0.04)**	(0.12)**
$Midlands$	0.254	0.539	0.718	0.936	0.216	0.795
	(0.14)*	(0.10)**	(0.09)**	(0.11)**	(0.31)	(0.70)
$North$	0.385	0.368	0.332	0.436	0.281	0.163
	(0.13)**	(0.09)**	(0.09)**	(0.11)**	(0.20)	(0.62)
$Wales$	0.433	0.696	0.749	0.368	0.670	1.039
	(0.12)**	(0.08)**	(0.08)**	(0.11)**	(0.17)**	(0.57)*
$Scotland$	0.225	0.265	0.354	−0.038	0.319	−0.020
	(0.12)*	(0.09)**	(0.08)**	(0.11)	(0.17)*	(0.66)
$Northern\ Ireland$	0.553	0.977	1.163	0.768**	0.743	0.713
	(0.12)**	(0.08)**	(0.08)**	(0.12)	(0.19)**	(0.63)

Notes: ** and * denote significant at 5, and 10% size, respectively.

Table 6: Overall partial effects: OP *vs* MNL *vs* Constants-only

	$Q=6;OP$		$Q=5;MNL$		OLS		$CONSTANTS$	
Female	−1.230	(0.09)**	−0.339	(0.08)**	−0.782	(0.08)**	−0.852	(0.07)**
Birth cohort 1940	0.153	(0.15)	−0.165	(0.13)	–	–	–	–
Birth cohort 1950	0.055	(0.15)	−0.109	(0.13)	–	–	–	–
Birth cohort 1960	0.456	(0.15)**	0.097	(0.13)	–	–	–	–
Birth cohort 1970	0.909	(0.17)**	−0.090	(0.15)	–	–	–	–
Birth cohort 1980 − 1990	0.853	(0.21)**	−0.179	(0.18)	–	–	–	–
Agreeableness	−0.027	(0.27)	−0.256	(0.53)	−0.038	(0.04)	0.024	(0.13)
Conscientiousness	−0.288	(0.32)	0.001	(0.33)	−0.336	(0.04)**	−0.248	(0.16)
Extraversion	0.271	(0.23)	0.075	(0.75)	0.267	(0.04)**	0.222	(0.23)
Neuroticism	−0.190	(0.16)	−0.170	(0.64)	−0.279	(0.04)**	−0.258	(0.31)
Openness to experience	−0.101	(0.23)	0.110	(0.54)	−0.061	(0.04)	−0.103	(0.35)
Father some education	−0.183	(0.14)	0.112	(0.12)	−0.033	(0.12)	0.161	(0.12)
Father further ed.	−0.118	(0.11)	0.262	(0.09)**	−0.025	(0.10)	0.165	(0.09)*
Mother some ed.	−0.189	(0.12)	0.088	(0.11)	−0.210	(0.11)*	−0.043	(0.10)
Mother further ed.	−0.348	(0.14)**	−0.042	(0.12)	−0.285	(0.12)**	−0.340	(0.12)**
Father prof./manage.	0.051	(0.14)	0.134	(0.12)	0.042	(0.13)	0.162	(0.12)
Father skill. non − man.	0.004	(0.19)	−0.235	(0.18)	0.123	(0.17)	−0.078	(0.17)
Father man./unskill.	0.362	(0.12)**	−0.028	(0.10)	0.431	(0.11)**	0.233	(0.10)**
Mother prof./manage.	0.187	(0.16)	0.015	(0.14)	0.198	(0.14)	0.148	(0.14)
Mother skill. non − man.	0.001	(0.14)	0.322	(0.13)**	−0.067	(0.12)	0.048	(0.12)
Mother man./unskilled	0.464	(0.11)**	−0.385	(0.10)**	5.219	(0.01)**	0.457	(0.09)**
Age/10	2.928	(0.14)**	3.547	(0.12)**	2.558	(0.15)**	2.656	(0.13)**
$Age^2/1,000$	−2.615	(0.14)**	−3.285	(0.03)**	−2.525	(0.15)**	−2.439	(0.13)**
Age	0.042		0.039		0.013		0.031	
Number of children	−0.038	(0.03)	0.072	(0.03)**	−0.030	(0.04)	−0.006	(0.04)
Married	0.414	(0.07)**	0.278	(0.07)**	0.391	(0.09)**	0.334	(0.08)**
(Log of) household inc.	0.107	(0.05)**	0.039	(0.05)	0.082	(0.06)	−0.012	(0.06)
Employed	−0.092	(0.11)	−0.078	(0.11)	0.059	(0.14)	−0.056	(0.12)
Not in the labour force	−0.225	(0.12)*	−0.145	(0.12)	−0.213	(0.16)	−0.194	(0.14)
Degree	−0.232	(0.12)**	−0.559	(0.11)**	−0.913	(0.14)**	−0.391	(0.15)**
Vocational degree	0.031	(0.08)	−0.030	(0.08)	−0.225	(0.11)**	−0.024	(0.11)
A − level	−0.157	(0.11)	−0.067	(0.11)	−0.347	(0.14)**	−0.099	(0.14)
GCSE	0.184	(0.09)**	−0.193	(0.10)	−0.121	(0.12)	0.122	(0.11)
Comorbidities	0.233	(0.02)**	0.336	(0.02)**	0.580	(0.03)**	0.297	(0.02)**
Midlands	0.515	(0.13)**	0.124	(0.12)	0.466	(0.14)**	0.357	(0.14)**
North	0.325	(0.10)**	−0.293	(0.10)	0.249	(0.12)**	0.094	(0.12)
Wales	0.663	(0.09)**	0.443	(0.10)**	0.527	(0.12)**	0.513	(0.11)**
Scotland	0.234	(0.09)**	0.119	(0.10)	0.195	(0.12)**	−0.048	(0.11)
Northern Ireland	0.850	(0.09)**	0.735	(0.10)**	0.880	(0.12)**	0.678	(0.11)**

Notes: ** and * denote significant at 5, and 10% size, respectively.

As compared to the within class partial effects, variables in the class equation(s) now also have effects on overall BMI values. Although the general pattern of results is broadly consistent across the two models, there are some substantive differences in terms of size and statistical significance for a number of explanatory variables (suggesting that the unrestricted model may be yielding unreliable results). For example, take the class equation(s) first: females, for example, have a significant negative overall effect in both, but of quite distinctly different magnitudes $(-1.23, -0.3)$. None of the birth cohort variables have an effect in the MNL approach, whereas three of them do in the OP one. Both approaches agree on the non-importance of the personality traits with respect to observed BMI levels (as opposed to class membership). There is a wide divergence in the significance of the parental variables; indeed, for the $Mother\ manual/unskilled$ control, which is significant in both, its effect actually switches in sign across approaches. Interestingly, of the 20+ variables in the class equation part of the model, the MNL approach suggests that only four have a significant (at 10% or above) effect. The OP approach finds significance for more of these. It is hard to speculate on what is causing this. It may be that estimating multiple parameters per covariate compared to one in the OP approach, adversely affects statistical significance as the MNL unnecessarily "over-fits" these class probabilities and/or that effects across classes possibly cancel each other out; either way, the OP approach will not suffer from such potential drawbacks.

Next considering the output equation, we can again see that the overall partial effects are "better explained" by the OP equation with respect to the number of statistically significant variables: thirteen in the OP approach compared to just eight in the MNL one. Thus with respect to statistical significance, there are several differences across the approaches, but there are also differences in the estimated magnitudes of significant variables (although direction of effect appears relatively consistent). For example, the implied nonlinear age profile appears quite different in shape across both (although the overall effect of age is quite similar). The effect of being married differs across the models (0.414 *versus* 0.278), whereas the effect of comorbidities is similar. Finally, there are divergences in magnitudes and statistical significance across the region indicators. With respect to comparisons with the OLS results, we see that although the general findings agree with directions of effects, it is evident that there are clear differences with regards to both magnitudes and significance levels.

The results so far suggest that the choice of approach is important. To further explore this we take a closer look at some estimated densities. In Figure 1 we plot the implied estimated densities by class for the new 6-class OP approach. The (enforced) ordering in these densities is evident, as their measures of central tendency (and generally dispersion) clearly increase over classes. From these, it is clear that the spread of classes 1-4 is very similar, and quite tightly centred around their respective means. The implication of these findings is that individuals within these classes are very unlikely to move from their respective expected values corresponding to WHO ranges of: (low) *normal*; (high) *normal*; (low) *overweight*; and (mid) *overweight*. However, the increased dispersion of class 5 ($EV = 29$), and even more-so, 6 ($EV = 35$) is also clearly evident, corresponding to WHO ranges of (high) *overweight* and *obese*.

An implication of these findings, is that although the two highest BMI range classes have high, and unhealthy, EVs, it does appear that behavioural choices, for example, could help these individuals into more healthy BMI ranges. On the contrary, individuals in the other, more healthy ranges, classes 1-3, appear to be very likely to be closely bound to their class-specific EVs (as are those in class 4). Given their spread, we can see that large parts of the distributions of classes 5 and 6 overlap with each other, as well as with those of both classes 3 and 4. This effect is probably more pronounced for the *obese* (class 6) group, who do have quite significant chances of moving themselves into more healthy weight ranges. Interestingly, as Figure 1 makes clear, an individual with an observed BMI of say 25, could conceivably be in any of these middle (2-5 class) groups. On the other hand, an observed BMI of say 35, is clearly only really likely to belong to either classes 5 or 6. Thus, from a policy perspective, it is extremely important to be able to identify which group any particular individual belongs to, which highlights the importance of the current research.

Finally, in Figure 2 we present the actual density of the raw BMI data for comparison, along with: that from our preferred 6-class OP approach (prior probability weighted of the above individual densities); that from the preferred MNL specification (5-class); and that from a simple linear regression. Clearly, a simple linear regression approach is not a sensible contender here. However, it is evident that the suggested approach does an excellent job in predicting the empirical density. Indeed, it is difficult to distinguish the actual from the predicted densities here. The same could also be said of the 5-class MNL approach though. However, such a similar extremely high "level of fit" is achieved much more parsimoniously

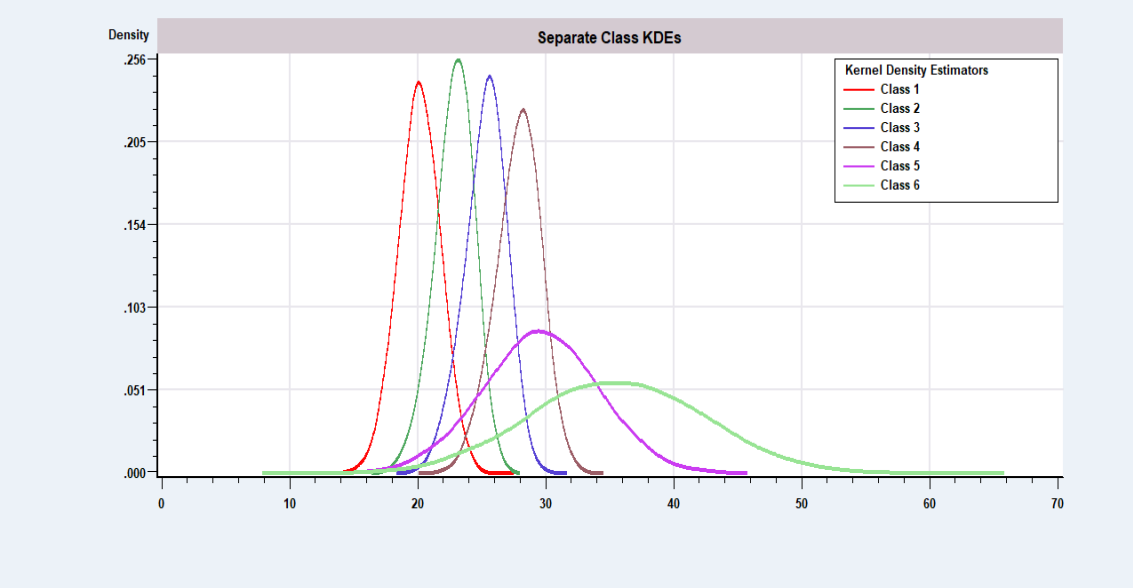

Figure 1: Individual Class Densities

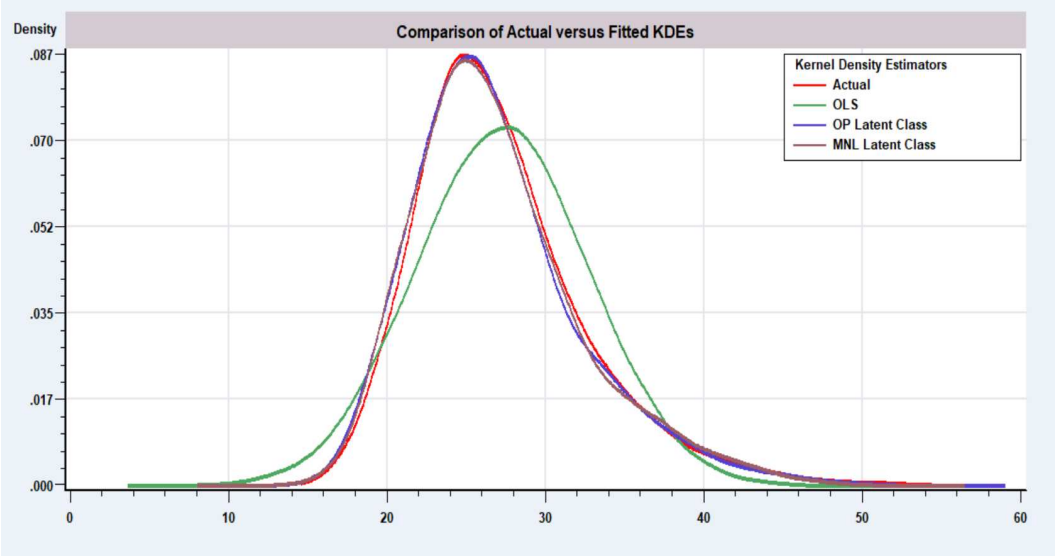

Figure 2: Actual versus Predicted Densities

in the OP approach compared to the MNL one (140 *versus* 183 parameters). Again, we suggest that this is a further validation of the suggested approach.

In summary, it is clear that the choice of approach matters: they imply quite different overall partial effects; a different number of classes; and different behaviours within each class. The new approach appears to provide just as good a fit as the much more heavily parameterised existing one. Differences in the overall partial effects highlight the possibility that an inappropriate modelling strategy may lead to incorrect inference and policy prescriptions relating to measures to tackle high BMI levels and obesity. And finally there is overwhelming support from the model selection metrics for the new approach over the traditional one.

4.3 Robustness checks

An obvious robustness check against which to compare our model results, is to consider a constants-only variant. So here, following much of the LCM literature, the class-assignment prior probabilities are simply modelled as constants, and there are no restrictions placed on the specifications of the mean function. We re-estimate our model removing all covariates from the class equations, and include these in the outcome equation (apart from the birth cohorts as we already include a quadratic in age). Again we treat the model as a panel data one. In Table 7 we present the model selection metrics from this exercise, along with the ones for our preferred model.

Once more we find strong evidence of a $Q = 6$ model being optimal, with all of the IC metrics similarly favouring the constants-only 6-class model. Thus, there appears to be strong evidence here for a 6-class model. Moreover, it is also clear that across-the-board our preferred OP approach is preferred to the constants-only approach. However, again, if the researcher is primarily interested in overall partial effects and, if the two approaches yield very similar results in this respect, one would presumably favour the less complicated approach. In Table 6 we compare (prior probability weighted) overall marginal effects from the preferred constants-only approach, along with those from the corresponding OP, MNL and OLS ones (previously discussed) under the $CONSTANTS$ heading.

It is clear that the approach undertaken can be substantial for these summary partial effects, with often large absolute and relative changes in magnitudes, and even changes in signs and significance levels. For example, the constants-only approach suggests a much

Table 7: Model selection metrics; comparison with constants-only approach

	BIC	AIC	$CAIC$	$HQIC$
6-class (panel)	$110,675^{**}$	$109,571$	$110,815^{**}$	$109,932^{**}$
2-class (constants)	$115,373$	$114,813$	$115,444$	$114,997$
3-class (constants)	$113,175$	$112,332$	$113,282$	$112,608$
4-class (constants)	$111,964$	$110,836$	$112,107$	$111,205$
5-class (constants)	$111,392$	$109,981$	$111,571$	$110,443$
6-class (constants)	$111,165^{*}$	$109,470^{*}$	$111,380^{*}$	$110,025^{*}$
7-class (constants)	$111,578$	$109,599$	$111,829$	$110,247$

Note: preferred model for each metric denoted by **; preferred model for the constants-only versions by *.

smaller gender effect compared to the preferred OP one (well under and over unity, respectively). The parental characteristics generally agree with respect to significance levels, but can be quite similar ($Mother\ further\ ed.$) or divergent ($Father\ man./unskill.$). Interestingly, ($Father\ further\ ed.$), whilst negative and insignificant in the OP model, is positive and (weakly) significant in the constants-only one. The nonlinear effect of age is much more pronounced in the OP model, as is the combined (linear) effect. The magnitude of the significant effect of *Degree* in both, is almost double in the OP approach; whereas that for $Married$ is quite similar (0.414 compared to 0.334). Finally, the regional effects appear to be much more prominent in the OP model, and indeed, the strong positive *Scotland* effect here, is not only insignificant in the constants-only, but also negative.

As noted before, we would surmise that the variables exhibiting the largest differences are probably those most severely affected by ignoring the omitted covariates (and possible mis-specification) in (of) the class equation. Also, as with the other comparisons considered, it is clear that the method chosen can quite often (but not across-the-board) have large consequences.

The panel data approach employed here, being based on multiple observations per individual, should intuitively be better able to identify the inherent classes than a pooled approach. However, if the model has been mis-specified in some manner, or individuals potentially move across classes over time, then the panel approach adopted could also be potentially mis-specified. Therefore an obvious robustness check is to compare our panel data model results against a pooled, or cross-sectional, variant. For reasons of space, we do not report the full set of results from this exercise (available on request). Instead we simply discuss the findings relating to the IC metrics. We find that amongst the pooled variants the IC metrics all favour the 7-class restricted model. Similarly, all the $Vuong$ statistics pro-

vide further evidence supporting the 7-class restricted model amongst the pooled variants. However, all of the pooled models are inferior to the 6-class panel model. Hence, comparing the pooled results with the preferred panel one, given the much improved IC metrics and likelihood values, one would clearly prefer the panel variant(s) to the cross-sectional ones. Fully utilising the repeated nature of observations of individuals within class therefore aids in better identification of/allocation to, the correct respective classes, and consequently results in a better specified/performing model.

The next robustness check we consider, is that in our (BMI) output equation we include the composite health indicator, $Comorbidities$, with the rationale that BMI is affected by this general proxy for "health". However, clearly the strong possibility of reverse causation exists here, with health not only causing the BMI level (in part), but also BMI levels (in part) contributing to the various health levels. If we had appropriate identifying variables for this composite health proxy, that could be considered orthogonal to BMI, we might be able to apply techniques for allowing for this endogeneity (Rivers and Vuong 1988, Terza, Basu, and Rathouz 2008). As always, such variables are hard to find and justify, so instead we simply remove this variable and re-estimate the model. Reassuringly the broad results are effectively unchanged: indeed the metrics generally favour the OP 6-class model, as above. Moreover, estimated EVs and other quantities of interest, are also all very similar. For example, EVs in this model were (compared to above): 20 (20) ; 23 (23) ; 25 (25) ; 28 (28) ; 29 (29) and 36 (35) .

Similar reverse causation arguments could however, also be levelled at the personality traits. In general, these are assumed to be fixed for most of an individual's life. It could be that BMI levels potentially affect personality traits. So, as a further robustness check, we also remove these variables from the model. Once more, the results are remarkably robust: the ICs still favoured the 6-class OP approach (as did all of the $Vuong$ tests), and EVs were remarkably similar (at 20, 23, 25, 28, 29 and 36).

Finally, clearly there is the potential for significant differences by gender, both in the number of BMI classes and the behaviour within these. Thus we restricted sub-samples to both males and females, and we find that overall, splitting the sample by gender has no real substantive effect on our results (available on request). For example, the $Q = 6$ OP model is strongly preferred for both genders; EVs are very similar across all of the split gender and the pooled samples, and indeed, all would fall into the same WHO BMI ranges.

5 Conclusions

To evaluate the health of the nation, policy-makers place a great deal of emphasis on BMI levels and the distribution of such. In this paper, we have furthered understanding of the determinants of BMI, a key indicator of health risk, by proposing an extension to the latent class methodology. Our extension allows for the ranking of expected values across classes in estimation as well as developing a functional form for the class probabilities that is more parsimonious than the familiar multinomial logit model. Our newly proposed approach leads to the estimation of six BMI classes. This compares very quite favourably with the four broad categories (*Underweight*, *Normal*, *Overweight* and *Obese*) as identified by the WHO. Moreover, the estimated partial effects differed dramatically across the classes in terms of sign, size and statistical significance. All metrics employed, clearly favoured the newly suggested approach. Indeed, the experimental evidence (provided in the Online Appendix), suggested that, in particular the BIC, $HQIC$ and $Vuong$ metrics/statistics, are all very useful in correctly selecting the appropriate model.

Furthermore, we find substantive differences in terms of size and statistical significance in the overall partial effects for many of the explanatory variables across the two approaches. These differences highlight the importance of selecting an appropriate approach for modelling BMI. Differing results across the two suggest that choosing incorrectly could easily lead to incorrect associations in terms of the magnitude and even the sign of the effect, which in turn may lead to inappropriate policy prescriptions. Overall, our findings serve to highlight the importance of selecting an appropriate modelling approach in the context of a policy-relevant area such as BMI. To design appropriate strategies for tackling high BMI levels and obesity, policy-makers need to fully understand their determinants and our proposed modelling approach, which is widely applicable across a wide range of research topics across the social sciences, is an important step in this direction.

References

Akaike, H. (1987): "Information Measures and Model Selection," *International Statistical Institute*, 44, 277–291.

Alfo, M., N. Salvati, and M. Ranalli (2017): "Finite Mixtures of Quantile and M-quantile Regression Models," *Statistical Computing*, 27, 547–570.

Bago D'Uva, T. (2005a): "Latent Class Models for Utilisation of Health Care," *Health Economics*, 15(4), 329–343.

——— (2005b): "Latent Class Models for Utilisation of Primary Care: Evidence from a British Panel," *Health Economics*, 14(9), 873–892.

Bago d'Uva, T., and A. Jones (2009): "Health care utilisation in Europe: New evidence from the ECHP," *Journal of Health Economics*, 28, 265–279.

Banks, J., R. Blundell, and C. Emmerson (2015): "Disability Benefit Receipt and Reform: Reconciling Trends in the United Kingdom," *Journal of Economic Perspectives*, 29(2), 173–190.

Bartolucci, F., A. Farcomeni, and F. Pennoni (2012): *Latent Markov Models for Longitudinal Data*. CRC Press.

Bozdogan, H. (1987): "Model Selection and Akaike's Information Criteria (AIC): The General Theory and its Analytical Extensions," *Psychometrika*, 52, 345–370.

Brown, H., and J. Roberts (2013): "Born to be wide? Exploring correlations in mother and adolescent body mass index," *Economics Letters*, pp. 413–415.

Brunello, G., D. Fabbri, and M. Fort (2013): "The Causal Effect of Education on Body Mass: Evidence from Europe," *Journal of Labor Economics*, 31(1), 195–223.

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

Chung, H., J. C. Anthony, and J. L. Schafer (2011): "Latent class profile analysis: an application to stage sequential processes in early onset drinking behaviours," *Journal of the Royal Statistical Society: Series A*, 174, 689–712.

Croon, M. (2002): "Ordering the classes," in *Applied Latent Class Analysis*, ed. by J. Hagenaars, and A. McCutcheon, chap. 5, pp. 137–162. Cambridge University Press.

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

Deb, P., W. T. Gallo, P. Ayyagari, J. M. Fletcher, and J. L. Sindelar (2011): "The Effect of Job Loss on Overweight and Drinking," *Journal of Health Economics*, 30, 317–327.

Deb, P., and A. Holmes (2000): "Estimates of use and costs of behavioural health care: a comparison of standard and finite mixture models," *Health Economics*, 9(6), 475–489.

Deb, P., and P. Trivedi (2002): "The Structure of Demand for Health Care: Latent Class versus Two-Part Models," *Journal of Health Economics*, 21(4), 601–625.

Fabrizi, E., E. Montanari, and M. Ranalli (2016): "A Hierarchical Latent Class Model for Predicting Disability Small Area Counts from Survey Data," *Journal of the Royal Statistical Society A*, 79(1), 103–131.

Friedl, H., and G. Kauermann (2000): "Standard Errors for EM Estimates in Generalized Linear Models with Random Effects," *Biometrics*, 56(3), 761–767.

Greene, W. (2012): *Econometric Analysis 7e*. Prentice Hall, New Jersey, USA.

Greene, W., M. Harris, B. Hollingsworth, and P. Maitra (2014): "A Latent Class Model for Obesity," *Economics Letters*, 123, 1–5.

Greene, W., and D. Hensher (2010): *Modeling Ordered Choices*. Cambridge University Press.

Hannan, E., and B. Quinn (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, B*, 41, 190–195.

Heckman, J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

Herbert, A., N. Gerry, and N. e. A. McQueen (2006): "A Common Genetic Variant Is Associated with Adult and Childhood Obesity," *Science*, 312, 279–283.

Hong, H., Y. Yue, and P. Ghosh (2015): "Bayesian estimation of long-term health consequences for obese and normal-weight elderly people," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3), 725–739.

Karabatsos, G., and C.-F. Sheu (2004): "Order-Constrained Bayes Inference for Dichotomous Models of Unidimensional Nonparametric IRT," *Applied Psychological Measurement*, 28(2), 110–125.

Koehler, A. B., and E. S. Murphree (1988): "A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order," *Applied Statistics*, 37, 187–195.

Madden, D. (2012): "A profile of obesity in Ireland, 2002-2007," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(4), 893–914.

Marquesa, W., V. Cruzb, J. Regob, and N. da Silvab (2016): "The impact of comorbidities on the physical function in patients with rheumatoid arthritis," *Revista Brasileira de Reumatologia*, 56(1), 14–21.

Mills, T. (2009): "Forecasting obesity trends in England," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1), 107–117.

Nyhus, E., and E. Pons (2005): "The effects of personality on earnings," *Journal of Economic Psychology*, 26, 363–384.

Philipson, T., and R. Posner (2008): "Is the obesity epidemic a public health problem? A review of Zoltan J. Acs and Alan Lyles's obesity, business and public policy," *Journal of Economic Literature*, 46(4), 974–982.

Reboussin, B. A., E. Ip, and M. Wolfson (2008): "Locally dependent latent class models with covariates: an application to under-age drinking in the USA," *Journal of the Royal Statistical Society: Series A*, 171, 877–97.

Rivers, D., and Q. Vuong (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

Schwarz, G. (1978): "Estimating the Dimensions of a Model," *Annals of Statistics*, 6(2), 461–464.

Soromenho, G. (1994): "Comparing approaches for testing the number of components in a finite mixture model," *Computational Statistics*, 9, 65–78.

Taylor, M. (2010): "British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices," Discussion paper, University of Essex.

Terza, J., A. Basu, and P. Rathouz (2008): "Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling," *Journal of Health Economics*, 3(3), 531–543.

Vermunt, J. (2010): "Latent Class Modeling with Covariates: Two Improved Three-Step Approaches," *Political Analysis*, 18(4), 450–469.

Vuong, Q. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–334.

Webbink, D., N. Martin, and P. Visscher (2010): "Does Education Reduce the Probability of Being Overweight?," *Journal of Health Economics*, 29(1), 29–38.

WHO (2000): "Obesity: preventing and managing the global epidemic," Discussion Paper 894, World Health Organization.

Wooden, M., N. Watson, and S. Freidin (2008): "Assessing the quality of the height and weight data in the HILDA survey," HILDA project technical paper series 1/08, University of Melbourne.

World Health Organisation (2014): "World Health Statistics," Technical report, World Health Organisation.