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

March 2019

Abstract

Given the increasing prevalence of adult obesity, furthering the 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. Although commonly used, 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 continued 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 have attracted increasing interest from both academics and policy-makers (Chou, Grossman, and Saffer 2002, 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 predominantly 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* (or finite mixture) framework (Deb, Gallo, Ayyagari, Fletcher, and Sindelar 2011, Greene, Harris, Hollingsworth, and Maitra 2014).

Latent class, or finite mixture modelling has been popular in economics, and especially in empirical models of health economics (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. In this way, the same explanatory variables can have differing effects across the model/classes (Bago d'Uva and Jones 2009). Latent class modeling has been used in explaining BMI before, and in this paper, we contribute to the this literature by proposing an extension to latent class modelling, which leads to a more detailed depiction of the determinants of BMI across a range of BMI types (or classes), and that we believe is better suited to such an application.

The broad latent class modelling contribution starts from the observation that although the classes are latent - by definition - researchers often label them *ex post* according to a tangible attribute such as an expected value (EV) - broadly defined - within each class. Indeed, uncovering evidence of the distinguishing features of the latent classes is a fundamental 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. 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.

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.

Although our focus lies in modelling BMI (a continous, cardinally ordered variabe), the approach is generally applicable to the analysis of any output variable that embodies an intrinsic notion of ordering (either cardinal or ordinal). The results suggest a more detailed picture of the determinants of BMI, with five classes being supported by our proposed approach (and ratified by the traditional one). Furthermore, we find considerable variation in the determinants of BMI across the five classes identified by our new approach. Although both approaches supported a 5-class model, all of the metrics clearly supported the newly proposed one. And moreover, quite dramatic differences in numerous *ex post* quantities of interest were found across the two, clearly suggesting that th choice of appropriate approach is very important. Fianlly, as noted, although our particular example is using continuous data, it is generically applicable to *any* output variable of interest that embodies any kind of cardinal, or ordinal, ordering.

2 Econometric framework

The model of interest is a latent class model (LCM), or finite mixture regression model, with predictors in the class proportions and the response densities. Note that the literature often uses the terms latent class and finite mixture interchangably. Here, we follow the more precise definition that the former corresponds to having covariates in the class equations, and is explicitly concerned with the drivers and the behaviouss within the various classes. On the other hand, we take finite mixture models, to (generally) be typified by just having constants in the class equation, and moreover to be more simply a technique concerned with better fitting the overall density by use of the sub-classes, which by themselves have no inherent interest.

This part of the empirical strategy is concerned with identifying the BMI-types, as discussed above. 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 *ordinal* instead of a *nominal* regression model for the class proportions/equations. This is undertaken in the following manner. 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_q. 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

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

Interest here is in the class of LCMs that are increasingly used when the researcher has some prior reasoning as to the determinants of class membership; that is, those with predictors (z_i) in the class proportions. Here 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). Indeed, the use of IC here appears entirely appropriate. Although AIC appears to be somewhat favoured in practice, the BIC can be shown to be consistent in the sense that $\Pr\left(\widehat{Q} = Q^*\right) \to 1$ as $N \to \infty$ (Cavanaugh and Neath 1999).

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. The ranking of the classes is paramount in much of the relevant previous literature (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 a key output of the modelling process, this ordering of the classes is not ensured during the estimation process. Here we suggest a simple way to do so, and thereby be explicitly consistent with the research question at hand. Thus, with regard to the modelling of observed BMI outcomes within this setting, 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 $f(y_i|x_i, \theta_q)$. In a linear regression model as in the case of BMI, $y_{i,q} = y^*_{i,q}$. Regardless of the model, EVs *on the assumption of underlying ordinality or cardinality of observed* $y_{i,q}$, are monotonically related to the index $x'_i\beta_q$. That is, this generic approach would be similarly applicable to *any* outcome variable of interest (that is, not just a linear regression model as considered here), assuming it embodies some underying 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 $EV_{i,q=1} \leq EV_{i,q=2} \leq \cdots \leq EV_{i,Q}$. As noted, such an *ex post* labelling of classes is very common in the LCM literature. Below we suggest an easy way in which this can be ensured in estimation.

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 $E\left(y_{q=1}\,|x\right) = x'_i\beta_{q=1}$ and $E\left(y_{q=2}\,|x\right) = E\left(y_{q=1}\,|x\right) + \exp\left(x'_i\beta_{q=2}\right)$. 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 $EV_{i,q=1} \leq EV_{i,q=2} \leq \cdots \leq EV_{i,Q}$. Continuing this progression we have

$$\begin{aligned} EV^*_{i,q=1} &= EV_{i,q=1}, \\ EV^*_{i,q} &= EV^*_{i,q-1} + \exp\left(x_i'\beta_q\right),\ q = 2, \ldots \end{aligned} \tag{5}$$

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:

$$\begin{aligned} &EV_1^* = x'\beta_1;\ \partial EV_1^*/\partial x_k = \beta_{1,k}, \\ &EV_q^* = EV_{q-1}^* + \exp\left(x'\beta_q\right);\ \partial EV_q^*/\partial x_k = \exp\left(x'\beta_q\right)\beta_{q,k} + \partial EV_{q-1}^*/\partial x_k\ ,\ q = 2, \ldots \end{aligned} \tag{6}$$

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 $\exp(x'\beta_2)$. The signs of these partial effects are not constrained by the $\exp(\cdot)$ transformation to be positive, but will be differentiated by the signs and magnitudes of their various components. In this case, 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. Indeed, it could be argued that as the classes in the MNL set-up have no inherent meaning (until after some *ex post* analysis), class-specific partial effects may not be particularly useful. However, in the suggested Ordered Probit (OP) approach, the classes have known characteristics and therefore class-specific partials here are, arguably, much more informative.

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. Note that 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 classses have been estimated, which should be evidened by the model metrics discussed in this paper. Moreover, even if EV's are very similar across classes, this doesn't 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 components of this. This is similarly true of the traditional approach however.

Although a simple, easy to implement and generic 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, π_q, in equation (1) may be a substantive part of the model construction. In the simplest case of a finite mixture model, the weights can be viewed simply as a component of the functional form,

$$f(y_i|x_i,\boldsymbol{\theta}) = \sum_{q=1}^{Q} \pi_q \times f_q(y_i|x_i,\boldsymbol{\theta}) = \sum_{q=1}^{Q} \pi_q \times f(y_i|x_i,\boldsymbol{\theta}, class = q) = \sum_{q=1}^{Q} \pi_q \times f(y_i|x_i,\theta_q), \quad (7)$$

in which θ_q is a subvector of $\boldsymbol{\theta}$. The functional form is common to the classes. They are differentiated by the indexed parameter vector. The standard 'mixture of normals' model (Pearson 1894) in which $f_q(y_i|\boldsymbol{\theta}) = N(\mu, \sigma^2)$ or $f_q(y_i|x_i,\boldsymbol{\theta}) = N\left(\alpha_q + x_i'\beta_q, \sigma_q^2\right)$ is a familiar example (Dayton and Macready 1988, Leroux 1992, Chen 2017). In this instance, π_q are nonnegative weights that sum to one. For purposes of estimation, the weights are treated as parameters that are calibrated subject to these two restrictions. A convenient, commonly used approach to impose the two restrictions is to map the weights to multinomial logit style probabilities, via $\pi_q = \exp\left(\gamma_q\right) \Big/ \sum_{q=1}^{Q} \exp\left(\gamma_q\right)$, with $\gamma_Q = 0$ (for the usual normalisation).

An alternative view of π_q treats them as probabilities attached to class assignment of a set of Q latent classes or segments of the population; $\pi_q = prob(individual\ i\ is\ a\ member\ of\ class\ q) = Prob_i(class = q)$. In this case, the subgroups, $q = 1, \ldots, Q$, represent meaningful, latent segments of the population under study. In a recent study for example

(Greene, Harris, Hollingsworth, and Maitra 2014). the latent class model of again BMI was suggested to be based on a latent trait, the presence of the (unobservable) FTO gene, for which observable characteristics (country of origin, for example), z_i, might contain relevant information. In this case, an approach that ties the class assignments to the observable data might be appealing. A common exercise in this setting that emerges in the presence of latent segments is to deduce (as well as possible), the class assignment, at least probabilistically, through the posterior probabilities (see Karvelis, Spilka, Georgoulas, Chudacek, Stylios, and Lhotska (2015), for example)

$$\begin{aligned} PostProb_i(class &= q\,|y_i, x_i\,) = \pi_{q|i} = \frac{\pi_q L_{iq}}{\sum_{q=1}^{Q} \pi_s L_{is}}, \qquad (8) \\ &\text{where} \\ L_{is} &= Likelihood\,(\theta_s\,|y_i, x_i\,)\,. \end{aligned}$$

One might also examine the correlation (or regression) of these posterior class probabilities ($PostProb_i$), with the exogenous information, z_i. Since z_i does not appear directly in π_q, the natural assumption would be that any explanatory power of a regression of $\pi_{q|i}$ on z_i is induced by correlation between z_i and (y_i, x_i). This two-step approach might be subject to a bias in that the omission of z_i from $Prob_i(class = q)$ might systematically skew the class probabilities and transmit that bias to secondary computations such as $\pi_{q|i}$.

It is difficult to predict how this mis-specification will impact the results computed based on the observed included factors, including parameter estimates and predictions. It does seem appropriate that if the class assignments are driven (in part) by the observables, z_i, then these should naturally be included in the specification of the model to begin with. To continue our application, one would want to parameterize the BMI class assignments with as many genetic proxies as possible. Although the unconditional form of the class probabilities is common in received applications, there are many that explicitly incorporate observed factors (z_i) directly in the prior probabilities, $Prob_i(class = q)$; see, Bago d'Uva and Jones (2009), Greene, Harris, Hollingsworth, and Maitra (2014) and Fabrizi, Montanari, and Ranalli (2016), for example.

The conditional (on z_i) usual multinomial logit (MNL) form,

$$\pi_{iq} = \frac{\exp\left(z_i'\gamma_q\right)}{\sum_{q=1}^{Q} \exp\left(z_i'\gamma_q\right)}, \qquad (9)$$

where one of the γ_q vectors is normalised to zero (usually either the first or last, although the choice is inconsequential), is a convenient choice that has been used in many studies. Indeed, The latent class model with heterogeneous class probabilities is now standard in the received applications, and has been built into popular software such as *Latent Gold* (Statistical Innovations, 2018, https://www.statisticalinnovations.com), *NLOGIT* (Econometric Software, Inc., 2016, http://www.nlogit.com) and some packages in R (Grun and Leisch 2007). Such an approach has the advantage of being relatively unrestrictive. It is also a particularly convenient form for the EM algorithm (see, Alfo, N. Salvati, and Ranalli (2017) and Friedl and Kauermann (2000), for example) For our purposes, the MNL form has two disadvantages. First, it is very heavily, perhaps excessively, parameterized. Second, it does not connect to an inherent ordering of the classes.

The MNL parameterization proliferates parameters – each additional class adds $k_z + 1$ parameters (including a constant term). The specification search for LC models is typically driven by information criteria such as BIC (as noted above), rather than directly by the log-likelihood. All IC criteria penalize large models such as the MNL. It follows that the MNL form is at a disadvantage to a more parsimonious one. We find that in the bulk of empirical exercises, the preferred number of classes is less than or equal to three. It may well be that 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 contaminated by a merging of heterogeneous classes.

When there are no covariates in the class probabilities, any parameterization with $Q - 1$ parameters that allows unrestricted probabilities will suffice (we noted the MNL form earlier). Some authors eschew the transformation approach and simply estimate Q weights subject to the constraints $0 < \pi_q < 1$ and $\sum_q \pi_q = 1$. However, we suggest that ordered probabilities of the form $\pi_{iq} = F\left(\mu_q - z_i'\gamma\right) - F\left(\mu_{q-1} - z_i'\gamma\right)$ with $q = 1, \ldots, Q$ and where $\mu_0 = -\infty$, $\mu_Q = \infty$, μ_1 is freely estimated (but there is no constant in z), γ is a free parameter vector and $F(.)$ is a cumulative distribution function (CDF), such as the normal or logistic, will suffice as well. In all cases where there are no covariates in the class equation(s), there are $Q - 1$ free structural parameters for the $Q - 1$ free probabilities. However, with observed covariates in the class probabilities, the different forms of the class probabilities will have substantive implications, as we see below.

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

$$\ln \frac{Prob\left(q \leq c \,|z_i\right)}{Prob\left(q > c \,|z_i\right)} = \mu_c + z_i'\gamma \tag{10}$$

that would be appropriate if an unobserved continuous variable is assumed to underlie the class assignment. This is useful for our purposes, 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 formulation for the prior probabilities is convenient for the purpose;

$$\pi_{iq} = \Phi\left(\mu_q - z_i'\gamma\right) - \Phi\left(\mu_{q-1} - z_i'\gamma\right), \tag{11}$$

where Φ is the standard normal *CDF*. The suggested approach has the advantage of a more parsimonious specification. The addition of another class to the 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

$$Prob_i(class = q\,|z_i) = Prob\left(\mu_{q-1} < z_i'\gamma + \varepsilon_i < \mu_q\right),\ \varepsilon_i \sim N\left(0, 1\right). \tag{12}$$

We should note that our contribution here is *not* in entering covariates into the class equations, as this approach has been in the literature for a long time (see, for example, Bartolucci, Farcomeni, and Pennoni (2012), and references therein), but *how* they enter into such.

The description thus far notes an implicit ordering that informs the class assignments. A distinction can be made between "latent class models," which are mixtures of possibly different models such as zero inflation models (Lambert 1992), switching regression models (Fair and Jaffee 1972), discrete choice models (Greene, Harris, Hollingsworth, and Maitra 2014, Greene, Harris, and Hollingsworth 2015) and a wide variety of others, some mentioned in McLachlan and Peel (2000). The "finite mixture" model typically blends weighted sums of like models, such as Chen (2017), Alfo, N. Salvati, and Ranalli (2017) and Fabrizi, Montanari, and Ranalli (2016). There might also be an implicit consideration of ordering within the class models themselves. Alfo, N. 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. In the model considered in this paper, the class specific outcomes, themselves, are ordered across the classes – for class q, we have

$$E\left(BMI\,|class = q, x_i\right) = E\left(BMI\,|class = q-1, x_i\right) + \exp\left(x_i'\beta_q\right), \tag{13}$$

recalling equations (3) to (5). This implies an external ordering of the class specific models. This is distinctly not a finite mixture model. The class specific models are constrained across the classes. There are a small number of similar applications among the received studies, though this feature appears to be uncommon. Heckman and Singer (1984) "random effects" form of a duration model imposes a common slope vector in a latent class accelerated failure time model while allowing only the constant term to differ across classes. In Collins, Greene, and Hensher (2013), the class specific multinomial choice models in a latent class model of attribute nonattendance differ only in the different positions of zeros in a restricted coefficient vector – for example, three classes are defined by $\gamma_1 = (0, \beta_2, \beta_3)$, $\gamma_2 = (\beta_1, 0.\beta_3)$ and $\gamma_3 = (\beta_1, \beta_2, 0)$. In our ordered outcomes, the means across classes all share some common parameters.

It might be suggested that the ordered choice form of the class probabilities is restrictive relative to the MNL form. Our experience suggests that the opposite is also plausible – that the MNL model with covariates over-fits the data. In the simulation experiments presented in the online Appendix, even with the data generated by a MNL process, applying the OP format does not adversely affect the results. Indeed, in all other cases, researchers typically do not use the MNL format when the data are naturally ordered. The format may be likewise out of place here. There would be 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.

2.3 Extension to a random effects panel specification

The application here involves two waves of the British Household Panel Survey (2004, 2006, see below). In general, the extension of the latent class specification to panel data involves treating the several waves jointly, holding constant over time the elements of the model that are specific to the individual. In the simplest cases, this becomes equivalent to treating the model parameters θ_q as a random parameter vector with discrete support (over the classes).

For the model considered in this paper, the 'pooled' starting point is

$$\begin{aligned} f(y_{it}|x_{it}) &= \sum_{q=1}^{Q} \pi_{iq} \times f(y_{it}|x_{it}, \theta_q), \\ & \quad \text{with} \\ \pi_{iq}(\mu, \gamma) &= \Phi\left(\mu_q - z_i'\gamma\right) - \Phi\left(\mu_{q-1} - z_i'\gamma\right), \end{aligned} \tag{14}$$

and the corresponding log-likelihood would be

$$\ln L = \sum_{i=1}^{N} \sum_{t=1}^{T} f(y_{it}|x_{it}) \tag{15}$$

with $i = 1, \ldots, N$, individuals observed over $t = 1, \ldots, T$, time periods. Note that the suggested approach of equation (14) *vis-à-vis* the traditional approach of equation (1), does not affect the contribution of the outcome variable (BMI) to the likelihood, but simply offers different functional forms for π_{iq} and $f(y_i|x_i, \theta_q)$.

For panel data, assuming conditional (on θ_q) independence, the joint density for the T_i joint observations for individual i is

$$f(y_{i1,\ldots,}y_{i,T_i}|x_{i1,\ldots,}x_{i,T_i}) = \sum_{q=1}^{Q} \pi_{iq} \times \prod_{t=1}^{T_i} f(y_{it}|x_{it}, \theta_q), \tag{16}$$

and the corresponding log-likelihood would now be

$$\ln L = \sum_{i=1}^{N} \ln f(y_{i1,\ldots,}y_{i,T_i}|x_{i1,\ldots,}x_{i,T_i}). \tag{17}$$

As noted above, the nature of the response variable of interest will, in most part, determine $f(.)$ in equations (14) to (17). Below we consider a standard linear regression model for f for our continuous variable of interest (BMI), although this approach is generally applicable for any f where there is any inherent ordering in the response variable y. Recall, as well, that this generic set-up differs from the usual approach in that monotonically increasing EVs are enforced as described in Section 2.1.

3 Data

We analyse data drawn from the British Household Panel Survey ($BHPS$). The $BHPS$ 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). The $BHPS$ is a rich source of information on labour market outcomes, socio-demographic and health variables. In only two waves 14 (2004) and 16 (2006), was information collected on weight and height, which we use to calculate individuals' BMI. Accordingly our data set comprises of 22,430 observations covering individuals aged 16 and over, who would at most be observed for two waves. The average BMI in the sample is 27.06, with a standard deviation of 5.45 (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 socio-economic characteristics relating to the individuals family background. Specifically we control for the respondent's father and mother's occupation when the respondent was aged 14 distinguishing between; professional and managerial; skilled non-manual; skilled and partly skilled manual; unskilled (with no occupation as the omitted category). Similarly, we include controls for parent's education: university level; further education; and school qualifications (with no qualifications as the omitted variable). Finally, we include time invariant controls for personality, specifically we include the *Big Five* personality traits, namely, agreeableness, conscientiousness, extraversion, neuroticism and openness to experience. In the psychology literature, it has been argued that the personality traits included in the *Big Five* taxonomy are stable over the life cycle; see, for example, Caspi, Roberts, and Shiner (2005) and Borghans, Duckworth, Heckman, and ter Weel (2008). There is however still some debate in the literature. Hence, we follow the standard practice in the existing literature to mitigate against the potential problem of life cycle effects influencing personality traits, and condition each personality trait (*i.e.*, one of the $Big\ Five = 1, ..., 5$) on a polynomial in age. The resulting residuals are standardised (zero mean and unit standard deviation) and used as indicators of personality traits net of life cycle influences (Nyhus and Pons 2005).

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 a quadratic in age, number of children, marital status, household income, employment status, highest level of educational attainment and region. Finally, we also include a set of eleven controls capturing a wide range of health problems, namely 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. However, due to high collinearity across these, as well as some some very sparse outcomes, we follow a lot of the relevant literature and consider a composite variable ($Comorbidities$); see for example Blustein, Hanson, and Shea (1998), Banks, Blundell, and Emmerson (2015), Marquesa, Cruzb, Regob, and da Silvab (2016), Lugo-Palacios and Gannon (2017) and Ha, Harris, Preen, Robinson, and Moorin (2018). To be specific, we set x in equation (2) to this set of control variables.

Descriptive statistics for the variables included in the empirical analysis (in the estimation sample) are presented in Table 1. The sample is evenly split by gender; just over half of the sample are married; and nearly 60% are in full-time employment. The average age of the sample is 48 and that for the number of children is just over a half. Having a vocational qualification is the most common highest educational attainment category, and the average number of comorbidities is just over one-and-a-quarter.

4 Results

4.1 Model comparison

We firstly compare a range of different models (all panel data verisons) using standard IC metrics in order to ascertain the preferred approach. Note that all estimations were obtained using author-written *Gauss* script utilising the *cmlMT* (constrained) maximum likelihood add-in module (identical results for the MNL variants are obtained from current standard software, such as *Nlogit 5; Stata v15*); a template *Gauss* code for estimation, as well as the proceedure file used for estimation, are freely available at: https://drive.google.com/drive/folders/1rtoYfs5c

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)
$Agreeableness$	-0.002	(1.00)
$Conscientiousness$	-0.003	(1.00)
$Extraversion$	-0.002	(1.00)
$Neuroticism$	0.004	(1.00)
$Openness\ to\ experience$	-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)
$Cormobidities$	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[a]	$121,086$	$120,936$	$121,105$	$120,985$	19
2-class (unrestricted)[b]	$115,537$	$115,064$	$115,597$	$115,219$	60
3-class (restricted)[c]	$113,182$	$112,552$	$113,262$	$112,758$	80
3-class (unrestricted)[d]	$113,308$	$112,512$	$113,409$	$112,773$	97
4-class (restricted)[e]	$111,717$	$110,929$	$111,817$	$111,187$	100
4-class (unrestricted)[f]	$112,022$	$110,903$	$112,164$	$111,269$	142
5-class (restricted)[g]	$\mathbf{111,143}$	$110,197$	$\mathbf{111,263}$	$\mathbf{110,507}$	120
5-class (unrestricted)[h]	$111,484$	$\mathbf{110,041}$	$111,667$	$110,513$	183
$Voung\ (BIC)\,;\ g\ vs\ h$	278.9	$p = 0$			
$Voung\ (AIC)\,;\ g\ vs\ h$	278.9	$p = 0$			
$Vuong\ (BIC, AIC)\,;\ g\ vs\ h$	278.9	$p = 0$			

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

pdhLJOZZEa56?usp=sharing. Also, some further estimation issues, including starting values and a discussion of maximum likelihood techniques *versus* the EM algorithm (which turns out to be invalid here), are discussed in the online appendix. After discussing the summary metrics, we then present detailed estimation results based on our preferred specification.

We start with a one class linear regression model and then successively increase the number of latent classes within both a standard framework (*unrestricted*) and our new proposed framework (*restricted*). We stopped searching for more potential classes at $Q = 5$, as these were already heavily parameterised models. Therefore in total, we consider 8 potential models.

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, and do not restrict ourselves to any *a priori* fixed number. Although AIC appears to be somewhat favoured in practice, it is well-known to favour larger models. Indeed, in the Monte Carlo experiments (see Online Appendix: Finite sample performance), the results suggested that the AIC, in this particular setting anyway, should really be avoided. On the other hand, as noted above, the BIC has appealing aymptotic properties. Although not as common, in this particular setting anyway (as evidenced by the Monte Carlo results), the $HQIC$ metric also appears extremely useful tool. In light of all of this, it is extremely reassuring to see that all of the BIC, $CAIC$ and $HQIC$ metrics unanimously favour the 5-class restricted (OP)

model, whilst only the (unreliable) AIC one favours the (much more heavily parameterised; by some 60 additional parameters) 5-class unrestricted (MNL) one.

To confirm the superiority of the restricted approach, we also consider three variants of the $Vuong$ (Vuong 1989) test for non-nested models. Based on the two metrics most commonly used in the related literature (AIC, BIC) : $Voung$ (BIC) considers a choice between the two top-performing models according to BIC; $Voung$ (AIC) the top two according to AIC; and $Vuong$ (BIC, AIC) , the top one from each. Although in the Online Appendix: Finite sample performance, Monte Carlo evidence showed a clear superiority of $Voung$ (BIC) , in the current instance this was immaterial as each one selected a choice between the the 5-class restricted model (labelled g in Table 2) and the 5-class unrestricted one (h) . The $Vuong$ clearly rejects the latter. This finding along with those of the IC metrics (in conjunction with the Monte Carlo results), makes a very compelling case for the 5-class restricted model.

In Table 3, we present some interesting summary statistics for each model: EVs by BMI class; average posterior class probabilities; and finally class-specific dispersion parameters. Note that class-specific EVs were evaluated at the sample means of all covariates and at the regression coefficients for that class (averaged individual EVs gave very similar results). "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 5 for the restricted 5-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 remarkably quite similar across the preferred restricted and unrestricted models. For example, the EV in class 1 (EV_1) is 20.6 compared to 20.7; with a probability of 14% (15) ; and with a dispersion parameter of 1.542 (1.538) . However, very interesting, and marked, differences occur in the top two classes of both.

In particular, we see that the EVs for class 4 restricted and unrestricted both lie in the very end of the WHO defined *overweight* range $(20 - 29.99)$; and at 28.8 and 29.5, respectively, are very close. Startlingly different though, is the proportion of individuals estimated to be in this class, reflected by the posterior probabilities, and the spread of individuals' BMIs within-class: 33% *c.f.* 12% and 3.4 *c.f.* 1.4, respectively). The same is somewhat true of the biggest EV BMI class. EVs varying marginally here (33.9 compared to 31.5, for restricted and unrestricted, respectively), as do dispersion parameters (6 compared to 6.2). But more-so, we witness large differences in proportions: 15% compared to 31%. Thus even before conducting any further analysis is undertaken, it is clear that the choice of approach is far from inconsequential.

Focussing on the preferred 5-class restricted results, we can see for class 1, the EV of around 20.5 sits at the low end of the WHO defined range of *normal weight* $(18.5 - 24.99)$. Based on the posterior probability, the smallest amount of individuals fall in this class (at 14%); and based on the estimated dispersion parameter, this class sits in the middle with respect to the spread of the distribution within it (we revisit the spread of these distributions below). Given the position of the mean within this class, this suggests that if anything, individuals within this category are more likely to slip into the *underweight* one, as opposed to the *overweight* one. Turning to the next class, with a mean of 23.5, this also sits within the *normal weight* range, but at the higher end of the scale. Judging by the spread of this distribution however (the lowest of any class at 1.133), individuals within this class have a relative low of moving far from the mean. Based on the posterior probability, about 20% of individuals would be in this group. Class 3, with a mean of 26, falls into the very lowest part of the *overweight* range, meaning that although the dispersion is small here (at 1.14), many of these individuals would still be in the healthy weight range. Around 17% of the population are estimated to be in this class.

Probably of more concern, and interest, however, are classes 4 and 5. With means of these distributions being at 29 and 34 respectively, these would fall into the (very high ends of) *overweight* and *moderately obese* $(30 - 34.99)$. Moreover, for class 4 the average posterior probability is "large" (at 0.33), suggesting that a worryingly large proportion of the population lie in this class. The dispersion of this distribution is relatively large (at 3.4), especially compared with classes 1-3. This does tend to imply 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 some in this range do also have the chance of becoming *moderately obese*. However, given the placement of the mean with respect to this range, it is unfortunately more likely that, if anything, individuals within this class will fall into the *moderately obese range* than the *healthy weight* one. Finally, there is a worryingly large proportion in the *moderately obese* class (15%) ; the spread within this distribution is very large, again suggesting that for these individuals that lifestyle factors could well be used to move themselves into much healthier weight ranges;

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

	$Q=5; OP$			$Q=5; MNL$		
	Expected Value	*Post. prob.*	*Dispersion* (σ_q)	*Expected Value*	*Post. prob.*	*Dispersion* (σ_q)
Class1	20.58 (0.05)**	0.14	1.542 (0.03)**	20.73 (0.05)**	0.15	1.538 (0.03)**
Class 2	23.50 (0.03)**	0.21	1.133 (0.03)**	23.69 (0.03)**	0.21	1.138 (0.03)**
Class 3	26.07 (0.04)**	0.17	1.140 (0.03)**	26.32 (0.04)**	0.21	1.248 (0.03)**
Class 4	28.79 (0.08)**	0.33	3.399 (0.05)**	29.46 (0.05)**	0.12	1.397 (0.04)**
Class 5	33.86 (0.29)**	0.15	6.204 (0.12)**	31.52 (0.10)**	0.31	5.985 (0.06)**
Overall	26.85 (0.05)**	–	–	27.76 (0.04)**	–	–

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

although obversely the converse is also true.

4.2 Parameter estimates

It is apparent that the class membership equation is reasonably well-specified (see Table 4), with the birth cohort controls, personality traits and, to a lesser extent, childhood conditions generally driving the statistical significance. Unlike the MNL approach, where such parameters are difficult to interpret, given our set-up, this is much more straightforward here. As positive (negative) OP coefficients imply higher probabilities of being the highest (lowest) classes (with the intervening ones being less clear: Greene and Hensher (2010)). Thus we can see that birth cohorts 1970 and 1980/90, are much more likely to be in the top (*moderately obese*) than any others (and more so for the latter). Individuals with higher levels of $Conscientiousness$, $Neuroticism$ and *Openness to experience* (a slightly weaker effect), are more likely to be in the *normal weight* class; whereas higher levels of $Extraversion$ are associated with higher probabilities of being in the *moderately obese* class. The indicators for both $Mother\ some\ education$ and $Mother\ further\ education$ (a stronger effect) are both associated with higher probabilities of being in the *normal weight* class; whereas that for $Father\ manual/unskilled$ are more likely to be in the *moderately obese* one. **[SARAH, I WONDER IF THIS SECTION NEEDS A BIT MORE STORY-TELLING BEING ADDED? ALOS IF THE TRAITS WERE ENTERED AS REISDUALS (?) ACN WE INTERPRET THESE AT ALL?!]**

In Table 5, we present the partial effects associated with our preferred 5-class model. Note that we label these classes according to the above analyses based on the EVs within

Table 4: Class membership equation; preferred specification

Variable	Estimated coefficient	Standard error
$Female$	-0.323	$(0.02)^{**}$
$Birth\ cohort$ 1940	-0.016	(0.04)
$Birth\ cohort$ 1950	-0.067	(0.04)
$Birth\ cohort$ 1960	0.031	(0.04)
$Birth\ cohort$ 1970	0.167	$(0.05)^{**}$
$Birth\ cohort$ $1980-1990$	0.184	$(0.06)^{**}$
$Agreeableness$	-0.006	(0.01)
$Conscientiousness$	-0.078	$(0.01)^{**}$
$Extraversion$	0.073	$(0.01)^{**}$
$Neuroticism$	-0.047	$(0.01)^{**}$
$Openness\ to\ experience$	-0.024	$(0.12)^{*}$
$Father\ some\ education$	-0.047	(0.04)
$Father\ further\ education$	-0.029	(0.03)
$Mother\ some\ education$	-0.057	$(0.03)^{*}$
$Mother\ further\ education$	-0.099	$(0.04)^{**}$
$Father\ professional/managerial$	0.010	(0.05)
$Father\ skilled\ non-manual$	-0.002	(0.05)
$Father\ manual/unskilled$	0.085	$(0.03)^{**}$
$Mother\ professional/managerial$	0.059	(0.04)
$Mother\ skilled\ non-manual$	0.008	(0.04)
$Mother\ manual/unskilled$	0.125	$(0.03)^{**}$
μ_1	-1.215	$(0.05)^{**}$
μ_2	-0.509	$(0.04)^{**}$
μ_3	-0.055	(0.04)
μ_4	0.959	$(0.05)^{**}$

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

each one. As would be expected, the partial effects differ dramatically across the five 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 five classes and increasing in magnitude from class 1 to class 5. Those of the squared term again differ dramatically across classes, and generally increase in magnitude with class. There appears to be a distinct "hill-shaped" effect of age within all of the classes: that is, within each class individuals weight 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 5 (*moderately obese*), one's BMI only increases by some 0.004 per year. On the other hand, this number is much larger for class 4 at 0.136, implying that, all other things equal, over the life-cycle from say 20 to 60, and individuals BMI is likely to rise by some (40×0.136) 5.44 units for individuals in this category. This is potentially quite concerning such that from a policy perspective this group should potentially be focussed on for potentially changing lifestyle factors/choices.

Whilst the effect of the number of dependent children appears to have no statistical effect, the effect of being married appears to quite significantly (both in economic and statistical terms) raise BMI in all but class 5. These positive partial effects appear to decline somewhat as the within-class EV increases, dropping from 0.462 (class 1) to 0.379 (class 4). Somewhat surprisingly, income only has a strongly significant effect in class 1, with a 10% rise in income leading to a 0.0153 increase in BMI. Being employed has a large and significant positive effect in class 1 and a smaller one in class 2, with no effects in the remaining ones. On the other hand, not being in the labour force, has quite large and negative effects (-0.307 and -0.277), but only in the *higher normal* and *lower over* classes.

There appears to be a lot of heterogeneity across educational attainment and classes. For example, having a degree as the highest level of educational attainment has a large, and statistically significant negative effect, but only for the *higher over* and *moderately obese* classes, with the effect being much stronger in the latter class. Having a vocation degree, only has an effect (positive, but much smaller compared to the $Degree$ effects) in the *normal* classes. $A - level$ has a positive (negative) effect for the *lower normal* (*higher over*) class; whereas $GCSE$ only has an effect (positive) in the *lower over* class. Indeed, such 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).

As noted, above, we control for health conditions by entering the composite $Cormobidities$

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 highly significant) negative effect in the lower normal class, which is presumably a result of worse health in general, being associated with lower weight levels for individuals already in a healthy weight range. This effect is mirrored in the *higher normal* class (to small and positive), and then successively increasing as the classes do, peaking at a large value of 0.761 for the *moderately obese* class. At these more unhealthy weight-range classes, the effect is presumably positive, and more pronounced as the within class EVs increase as rising ill-health means 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 more readily interpreted as correlations rather than causations. However, we do return to this point below in our robustness checks.

Overall, it is clearly evident that there is indeed much heterogeneity across the population with respect to the effects of covariates, thoroughly justifying a LCM approach in general, and specifically our suggested new approach to such.

Thus while although the above results clearly illustrate how such an approach can highlight interesting differential partial effects across classes, the LCM 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 simply be interested only in overall partial effects (and not those split by class). Moreover, if the overall partials from both the 5-class restricted and unrestricted models were very similar, it could be argued that our suggested approach has very little benefit and/or effect in practice. So, to explore this issue, Tables 6 compares the overall (prior probability weighted) partial effects across the two models.

Note now, as compared to the within class partial effects, now variables in the class equation(s) do 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 (possibly suggesting that the unrestricted model may be yielding unreliable results). For example, take the variables in the class equation(s) first: females, for example, have a significantly negative overall negative effect in both, but of quite distinctly different magnitudes $(-1.3, -0.3)$ .None of the birth cohort variables have an effect in the MNL approach, whereas two of

Table 5: Class-specific partial effects

Variable	Class 1 *(lower normal)*	Class 2 *(higher normal)*	Class 3 *(lower over)*	Class 4 *(higher over)*	Class 5 *(moderately obese)*
$Age/10$	1.590	2.323	2.828	3.348	5.371
	(0.13)**	(0.11)**	(0.12)**	(0.25)**	(0.56)**
$Age^2/1000$	−1.312	−1.929	−2.421	−2.067	−5.553
	(0.13)**	(0.10)**	(0.12)**	(0.25)**	(0.60)**
Age	0.033	0.047	0.050	0.136	0.004
$Number\ of\ children$	−0.020	−0.011	0.014	0.014	0.108
	(0.04)	(0.03)	(0.03)	(0.06)	(0.13)
$Married$	0.462	0.458	0.385	0.379	0.131
	(0.08)**	(0.06)**	(0.07)**	(0.13)**	(0.31)
$(Log\ of)\ household\ income$	0.153	−0.006	0.088	0.187	−0.048
	(0.05)**	(0.04)	(0.05)*	(0.10)*	(0.23)
$Employed$	0.655	0.187	−0.117	0.162	−0.677
	(0.11)**	(0.08)**	(0.10)	(0.20)	(0.46)
$Not\ in\ the\ labour\ force$	0.186	−0.307	−0.277	−0.243	−0.144
	(0.13)	(0.09)**	(0.11)**	(0.22)	(0.47)
$Degree$	−0.043	−0.094	−0.160	−0.781	−1.359
	(0.11)	(0.09)	(0.10)	(0.22)**	(0.50)**
$Vocational\ degree$	0.183	0.164	0.001	−0.026	−0.307
	(0.09)*	(0.07)**	(0.08)	(0.15)	(0.37)
$A-level$	0.282	0.125	0.028	−0.593	−0.071
	(0.12)**	(0.10)	(0.11)	(0.23)**	(0.42)
$GCSE$	0.058	0.122	0.236	0.053	−0.638
	(0.12)	(0.08)	(0.09)**	(0.18)	(0.43)
$Cormobidities$	−0.054	0.072	0.143	0.313	0.761
	(0.02)**	(0.02)**	(0.02)**	(0.04)**	(0.09)**
$Midlands$	−0.070	−0.050	0.071	0.186	0.806
	(0.13)	(0.10)	(0.11)	(0.22)	(0.47)*
$North$	−0.007	−0.208	−0.324	0.118	0.014
	(0.12)**	(0.09)**	(0.10)**	(0.18)	(0.44)
$Wales$	0.224	0.308	0.539	0.520	0.976
	(0.10)**	(0.07)**	(0.09)**	(0.19)**	(0.40)**
$Scotland$	0.144	0.157	0.067	0.271	0.248
	(0.10)	(0.08)**	(0.09)	(0.18)	(0.45)
$Northern\ Ireland$	0.295	0.605	0.833	0.678	0.775
	(0.11)**	(0.08)**	(0.09)**	(0.20)**	(0.44)*

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

Table 6: Overall partial effects: OP vs MNL

	$Q=5;OP$		$Q=5;MNL$	
$Female$	-1.266	$(0.10)^{**}$	-0.339	$(0.08)^{**}$
$Birth\ cohort$ 1940	-0.062	(0.15)	-0.165	(0.13)
$Birth\ cohort$ 1950	-0.255	(0.16)	-0.109	(0.13)
$Birth\ cohort$ 1960	0.119	(0.16)	0.097	(0.13)
$Birth\ cohort$ 1970	0.636	$(0.18)^{**}$	-0.090	(0.15)
$Birth\ cohort$ $1980-1990$	0.702	$(0.22)^{**}$	-0.179	(0.18)
$Agreeableness$	-0.022	(0.22)	-0.256	(0.53)
$Conscientiousness$	-0.295	$(0.17)^{*}$	0.001	(0.33)
$Extraversion$	0.279	(0.25)	0.075	(0.75)
$Neuroticism$	-0.180	(0.17)	-0.170	(0.64)
$Openness\ to\ experience$	-0.090	(0.20)	0.110	(0.54)
$Father\ some\ education$	-0.180	(0.14)	0.112	(0.12)
$Father\ further\ education$	-0.109	(0.11)	0.262	$(0.09)^{**}$
$Mother\ some\ education$	-0.215	$(0.12)^{*}$	0.088	(0.11)
$Mother\ further\ education$	-0.379	$(0.14)^{**}$	-0.042	(0.12)
$Father\ professional/managerial$	0.038	(0.14)	0.134	(0.12)
$Father\ skilled\ non-manual$	-0.007	(0.19)	-0.235	(0.18)
$Father\ manual/unskilled$	0.323	$(0.12)^{**}$	-0.028	(0.10)
$Mother\ professional/managerial$	0.335	(0.16)	0.015	(0.14)
$Mother\ skilled\ non-manual$	0.031	(0.15)	0.322	$(0.13)^{**}$
$Mother\ manual/unskilled$	0.477	$(0.11)^{**}$	-0.385	$(0.10)^{**}$
$Age/10$	3.101	$(0.14)^{**}$	3.547	$(0.12)^{**}$
$Age^2/1,000$	-2.808	$(0.14)**$	-3.285	$(0.03)^{**}$
Age	0.041		0.039	
$Number\ of\ children$	0.018	(0.03)	0.072	$(0.03)^{**}$
$Married$	0.372	$(0.07)^{**}$	0.278	$(0.07)^{**}$
$(Log\ of)\ house\ holdincome$	0.091	$(0.05)^{*}$	0.039	(0.05)
$Employed$	0.062	(0.10)	-0.078	(0.11)
$Not\ in\ the\ labour\ force$	-0.191	$(0.11)^{*}$	-0.145	(0.12)
$Degree$	-0.513	$(0.11)^{**}$	-0.559	$(0.11)^{**}$
$Vocational\ degree$	0.006	(0.08)	-0.030	(0.08)
$A-level$	-0.142	(0.12)	-0.067	(0.11)
$GCSE$	0.000	(0.10)	-0.193	(0.10)
$Cormobodities$	0.249	$(0.02)^{**}$	0.336	$(0.02)^{**}$
$Midlands$	0.172	(0.12)	0.124	(0.12)
$North$	-0.060	(0.10)	-0.293	(0.10)
$Wales$	0.506	$(0.10)^{**}$	0.443	$(0.10)^{**}$
$Scotland$	0.191	$(0.10)^{*}$	0.119	(0.10)
$Northern\ Ireland$	0.653	$(0.10)^{**}$	0.735	$(0.10)^{**}$

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

them do in the OP one. Both approaches generally agree on the non-importance of the personality traits with respect to observed BMI levels (as oposed to class memebership). 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. Contrasting this, the OP approach finds significance for nearly half of these. It is hard to speculate on what is causing this in general (it may well be the case that this varies from case-to-case). However, it may be that in estimating multiple parameters per covariate compared to one in the OP approach, adversely affects 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 variables in 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 significant variables: ten in the OP approach compared to just eight in the MNL one. Again, it is hard to generalise here, but this could well be a function of the more parsimonious approach of the former. Thus with respect to 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). There is a 0.1 difference in the effect of being married, and around the same in the effect of comorbidities. And there are divergences of similar magnitudes for both of the $Wales$ and $Northern\ Ireland$ indicators.

Thus overall, these results taken in conjunction with the summary ones presented above, clearly suggest that there are broad consistencies across approaches: both suggest a 5-class model is optimal and there is general agreement of the direction (and approximate magnitude) of overall effects. However, summary statistics regarding the EVs within each class, the relative size and dispersion of these classes, combined with the magnitudes and significance of covariates on overall BMI levels, suggests that the choice across the two approaches is clearly important.

To further explore BMI behaviours within the estimated classes, and also to ascertain the overall appropriateness of our approach, we take a closer look at some estimated densities.

Firstly, in Figure 1, we present a kernel density for the raw BMI data.

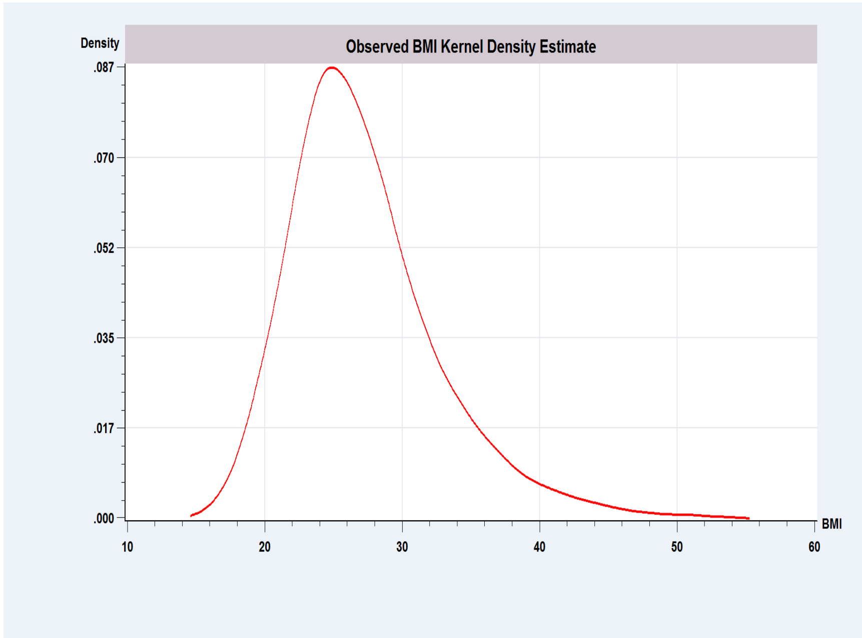

Figure 1: Observed Density of BMI Values

Then in Figure 2 we plot the implied estimated densities by class for the new 5-class OP approach. The (enforced) ordering in these densities is evident, as their measures of central tendency (and generally dispersion) clearly increase over classes. Based on posterior probabilities, individuals have a relatively low chance of being in the lowest BMI, *lower normal*, range class (14%). However, individuals in this group are clearly likely to have low expected BMI levels (around 20), and with quite a tight distribution around this mean. Individuals in the *higher normal* group, again have a very low probability of being particularly far from the EV (of around 24) here, having the least amount of dispersion across all of distributions. There is quite a high chance (1 in 5) of an individual being in this class. The *lower over* class again exhibits a relatively tight distribution around its EV of 26. Given the tighness of the distributions in these first 3 classes, it is very unlikely that individuals within these will stray far from their EVs.

The same cannot be said though, of the *higher over* class. Although the mean is 29, it is clear that individuals within this group could easily be anywhere between the low 20's to the low 30's; where the former would be nudging *Normal weight* and the latter well into the *Moderately obese* range. However, this effect is even more pronounced in class 5, the *moderately obese* group. Firstly, it has a markedly higher EV, at nearly 34; but secondly, it is clearly characterised by a very high level of dispersion. Individuals in this inherent group could easily find themselves anywhere between mid 20's (*Normal/Overweight*), but also as high as the low to mid 40's. The upper end of this range could well place individuals in the *Very severely obese* and *Morbidly obese*, WHO categoties.

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

Finally, in Figure 3 we present the actual density for comparison, along with: our preferred 5-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 in the OP approach compared to the MNL one (by over some 60 fewer parameters). Again, we would 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; potentially 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 highlights the possibility that an inappropriate modelling strategy may lead to incorrect inference and

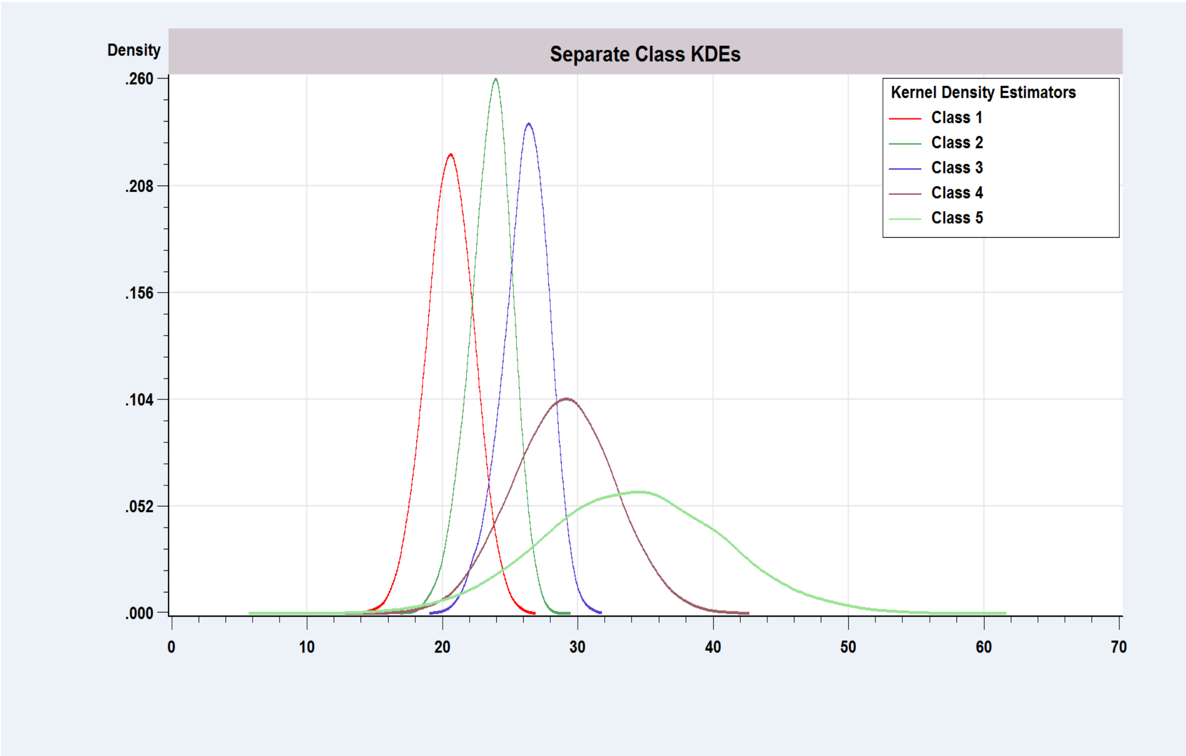

Figure 2: Individual Class Densities

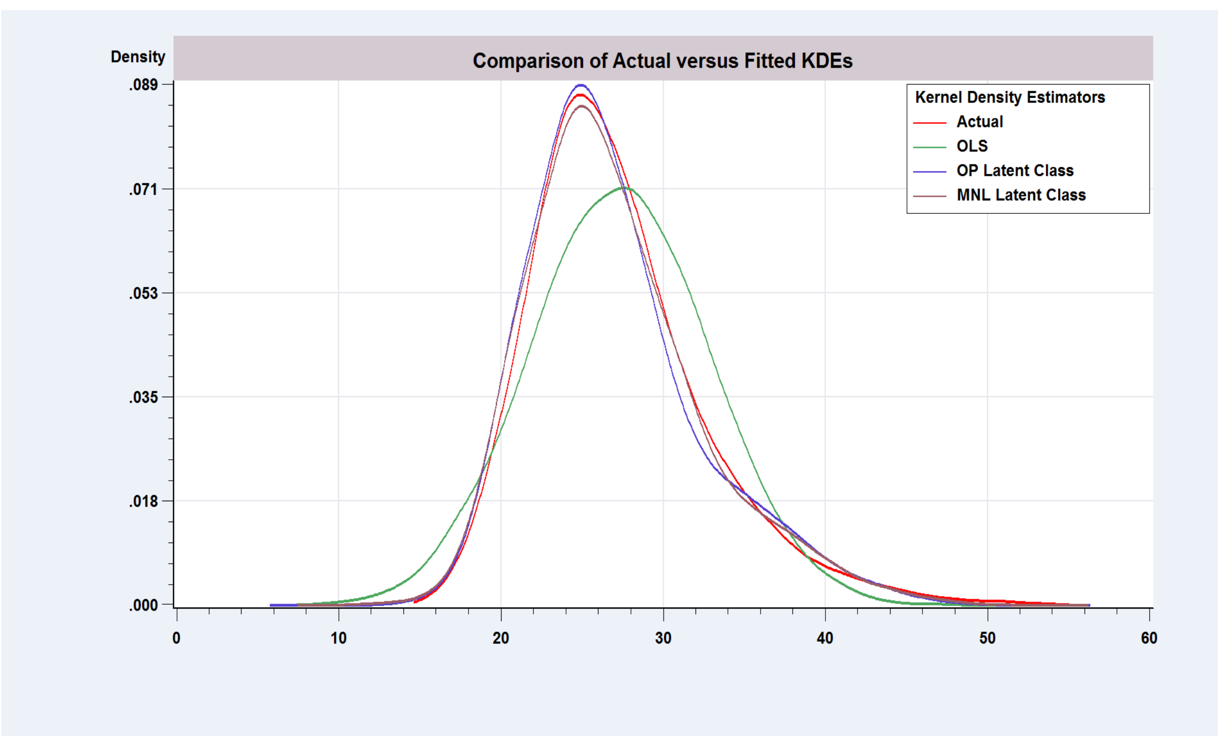

Figure 3: Comparison of Actual versus Predicted Densities

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

	BIC	AIC	$CAIC$	$HQIC$
5-class (panel)	$111,143^{**}$	$110,197^{**}$	$111,263^{**}$	$110,507^{**}$
2-class (pooled)^a	$117,788$	$117,314$	$117,848$	$117,469$
3-class (pooled, restricted)^b	$117,518$	$116,887$	$117,598$	$117,094$
3-class (pooled, unrestricted)^c	$117,544$	$116,748$	$117,645$	$117,009$
4-class (pooled, restricted)^d	$117,435^{*}$	$116,646$	$117,535^{*}$	$116,904$
4-class (pooled, unrestricted)^e	$117,642$	$116,522$	$117,784$	$116,889$
5-class (pooled, restricted)^f	$117,500$	$116,554$	$117,620$	$116,864^{*}$
5-class (pooled, unrestricted)^g	$117,842$	$116,399^{*}$	$118,025$	$116,871$
$Voung\ (BIC);\ d\ vs\ c$	1.7	$indeterminate$		
$Voung\ (AIC);\ f\ vs\ g$	524.4	$p = 0$		
$Vuong\ (BIC, AIC);\ d\ vs\ g$	524.3	$p = 0$		

Note: overall preferred model for each metric denoted by **; preferred model for the pooled model only by *.

policy prescriptions relating to measures to tackle high BMI levels and obesity. And finally there was overwhelming support for the new approach over the traditional one.

4.3 Robustness checks

The panel data approach employed here, being based on multiple observations per individual, should intuitively be better be able to identify the inherent classes. 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). However, in Table 7 we present the model selection metrics from this exercise, along with those for our preferred model.

Across all of the pooled variants, only the (unreliable) AIC favours a MNL variant (5-class). Both the BIC and $CAIC$ again select the 4-class OP variant, whereas the $HQIC$, the 5-class OP one. The $Vuong$ tests provide inconclusive evidence comparing model d vs c; prefer the 5-class rectricted over the 5-class unrestricted; and also the 4-class restricted over the 5-class unrestricted one. Thus, in summary the choice over the pooled variants clearly favours the OP variants, although the choice between the 4- and 5-class models is somewhat unclear.

Importantly, comparing the pooled results with the preferred panel one though, given

Table 8: Model selection metrics; comparison with constants-only approach

	BIC	AIC	$CAIC$	$HQIC$
5-class (restricted)	$111,143^{**}$	$110,197^{**}$	$111,263^{**}$	$110,507^{**}$
2-class (constants)	$115,336$	$114,813$	$115,406$	$114,989$
3-class (constants)	$113,118$	$112,332$	$113,225$	$112,596$
4-class (constants)	$111,887^{*}$	$110,836$	$112,030^{*}$	$111,189^{*}$
5-class (constants)	$112,105$	$110,790^{*}$	$112,284$	$111,232$

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

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.

Another 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 (that is, we estimate here what many researchers would take to be a standard LCM, or more corectly, a finite mixture model, estimable using *Stata* or *NLogit*, for example). To this end we re-estimate our model removing all covariates from the class equations, and include these variables in the mean equation (apart from the birth cohorts as we already include a quadratic in age). Again we treat the model as a panel data one. For reasons of space, we do not report the full set of results from this exercise (available on request). However, in Table 8 we present the model selection metrics from this exercise, along with the ones for our preferred model.

Thus we can see that for all IC measures, our approach consistently out-performs all possible contenders for the constants-only version. There is disagreement across the IC metrics as to the preferred number of classes for the panel constants-only approach, with BIC, $CAIC$ and $HQIC$ all favouring 4-classes and AIC five. In light of these, we take the 4-class constants-only version as the preferred specification here.

So the constants-only approach appears to favour a smaller number of BMI classes, and in terms of the metrics considered, appears to perform worse than our preferred approach. However, again, if the researcher is primarily interested in overall partial effects, then if the two approaches yield very similar results in this respect, one would presumably favour the less complicated approach. To address this, in Table 9 we compare (prior probability weighted) overall marginal effects from the 4-class constants-only approach with those previously presented from our preferred approach. In the final column we also present percentage differences in these.

It is clear that the approach undertaken (constsnts only *versus* our suggested OP approach) is substantive for these summary partial effects. For example, we find both large absolute and relative changes in partials across the approaches, and moreover even changes in signs and significance. For example, the effects of all of the following variables actually change sign: $Agreeableness$; $Father\ skilled\ non{-}manual$; $Number\ of\ children$; $Vocational\ degree$; $GCSE$; and $North$. Some examples of very large relative changes are: $Vocational\ degree$ (300%); $Agreeableness$ (188%); $Number\ of\ children$ (186%); $North$ (120%); and $Father\ skilled\ non-manual$ (103%). Moreover, we also witness very large absolute differences, such as the variables: $Age/10$ (0.76); $Age^2/1,000$ (0.632); $Female$ (0.445); $North$ (0.366); $Midlands$ (0.295); $Mother\ manual/unskilled$ (0.246); and so on. Indeed, in absolute value, nearly half of the variables exhibited differences in excess of 0.1. However, we also note that large differences are not evident across-the-board: for example, the difference in the effect of $Neuroticism$ is very small in both absolute and relative terms, as is $Conscientiousness$ and $Married$.

As noted before, we would surmise that the variables exhibiting the largest differences are probably those most severely affected by ignoring the omitted covariates in the class equation: that is, presumably the most highly correlated with the omitted drivers of the class equation; and those where the change is negligible, would be less affected (and presumably less strongly related to the omitted class covariates).

The next robustness check we consider, is that in our (BMI) output equation we include the composite health indicator, $Cormobodities$, with the rationale that BMI is affected by this general proxy for "health". However, clearly the strong possibility of reverse causation exists here, with the 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 othogonal 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 here, so instead we simply remove this variable and re-estimate the model. Reassuringly the broad results are effectively unchanged: indeed now, all of the metrics (and $Vuong$ tests) favour

Table 9: Overall partial effects; comparison with constants-only approach

	5-class restricted	4-class constants only	Difference (%)
Female	−1.266	−0.821	54.2
Agreeableness	−0.022	0.025	−188.0
Conscientiousness	−0.295	−0.284	3.9
Extraversion	0.279	0.204	36.8
Neuroticism	−0.180	−0.187	−3.7
Openness to experience	−0.090	−0.123	−26.8
Father some education	−0.180	−0.057	215.8
Father further education	−0.109	−0.059	84.7
Mother some education	−0.215	−0.153	40.5
Mother further education	−0.379	−0.168	125.6
Father professional/managerial	0.038	0.188	−79.8
Father skilled non − manual	−0.007	0.221	−103.2
Father manual/unskilled	0.323	0.260	24.2
Mother professional/managerial	0.225	0.471	−52.2
Mother skilled non − manual	0.031	0.247	−87.4
Mother manual/unskilled	0.477	0.723	−34.0
$Age/10$	3.101	2.341	32.5
$Age^2/1,000$	−2.808	−2.176	29.0
Age	0.041	0.025	64.0
Number of children	0.018	−0.021	−185.7
Married	0.372	0.350	6.3
(Log of) house holdincome	0.091	0.043	111.6
Employed	0.062	0.207	−70.0
Not in the labour force	−0.191	−0.067	185.1
Degree	−0.513	−0.586	−12.5
Vocational degree	0.006	−0.003	−300.0
A − level	−0.142	−0.114	24.6
GCSE	0.000	−0.060	−
Cormobodities	0.249	0.352	−29.3
Midlands	0.172	0.467	−63.2
North	−0.060	0.306	−119.6
Wales	0.506	0.449	12.7
Scotland	0.191	0.128	49.2
Northern Ireland	0.653	0.520	25.6

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

the OP 5-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): 21 (21); 24 (24); 26 (26); 29 (29); and 35 (34).

Similar reverse causation arguments could however, also be levelled at the personality traits. In general, these are generally assumed to be fixed for most of an individual's life but especially for younger individuals. it could be that BMI levels potentially affect personality traits. So, as a further robustness check, we also remove these variables as well from the model. Once more, results are remarkably robust: all ICs (with the exception of the AIC) still favoured the 5-class OP approach (as did all of the $Vuong$ tests), and EVs were remarkably similar (at 21, 24, 26, 29 and 35).

5 Analysis by gender

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. For females, all IC metrics and $Vuong$ tests favoured the 5-class OP variant; for males the ICs again, with the exception of AIC, all select the same 5-class OP model, as do all of the $Vuong$ tests. Thus, in summary, across all the pooled and gender split models, there is overwhelming support for the 5-class OP model over all other OP models and importantly all MNL ones.

Although all samples agree on the number of classes and approach, in Table 10 we examine whether there is much heterogeneity across genders with respect to the behaviour within classes, along with the equivalent ones from the pooled sample (full model estimation results for the two gender samples are available on request). In general, there is a remarkable consistency across all samples. They all identify the lowest BMI range class as *normal weight*, with EVs of around $20 - 22$, with probabilities in the $0.12 - 0.14$ range, and with similar dispersion ranges, with $\widehat{\sigma}_1 \approx 1.5$ (with the male class here sample being slightly less heterogeneous). Similarly, all three samples identify the next BMI range with an $EV \approx 23-24$; slightly more females appear to be in this class however (with a posterior probability of just over 0.25 compared to 0.22 for males and 0.21 overall), and this female class also appears more dispersed $(\widehat{\sigma}_{m,2} = 1.07, \widehat{\sigma}_{f,2} = 1.24)$.

All samples identify the next class as being in the (low) *overweight* range (although for females this is right on the border); and essentially the same proprtion of individuals (18%)

Table 10: Expected values, averaged posterior probabilities and dispersion parameters by gender

	EV	*Post. prob.*	*Dispersion* (σ_q)
		Combined	
Class 1; *lower normal*	20.58 (0.05)**	0.14	1.542 (0.03)**
Class 2; *higher normal*	23.50 (0.03)**	0.21	1.133 (0.03)**
Class 3; *lower over*	26.07 (0.04)**	0.17	1.140 (0.03)**
Class 4; *higher over*	28.79 (0.08)**	0.33	3.399 (0.05)**
Class 5; *moderately obese*	33.86 (0.29)**	0.15	6.204 (0.12)**
Overall	26.85 (0.05)**	–	–
		Male	
Class 1; *lower normal*	21.59 (0.06)**	0.14	1.316 (0.04)**
Class 2; *higher normal*	24.44 (0.04)**	0.22	1.069 (0.03)**
Class 3; *lower over*	27.26 (0.04)**	0.18	1.166 (0.04)**
Class 4; *higher over*	29.55 (0.15)**	0.38	4.526 (0.11)**
Class 5; *moderately obese*	34.83 (0.83)**	0.09	7.096 (0.35)**
Overall	27.42 (0.07)**	–	–
		Female	
Class 1; *lower normal*	19.72 (0.08)**	0.12	1.480 (0.05)**
Class 2; *higher normal*	22.76 (0.05)**	0.25	1.241 (0.03)**
Class 3; *lower over*	25.54 (0.05)**	0.18	1.161 (0.04)**
Class 4; *higher over*	28.43 (0.09)**	0.33	3.024 (0.05)**
Class 5; *moderately obese*	34.95 (0.29)**	0.12	5.268 (0.12)**
Overall	26.20 (0.06)**	–	–

are estimated to be in this category, and dispersion parameters (at around $1.1 - 1.2$) are also incredibly similar. There is again, much alignment with *EV*s in the *higher over* class (29, 30, 28). However, the proportions do vary a bit here, with 38% of males predicted to be in this class compared to 33% of females, and the distribution of males within this class is slightly more dispersed $(\widehat{\sigma}_{m,4} = 4.53, \widehat{\sigma}_{f,4} = 3.02)$. Finally, *EV*s in the top class, *moderately obese*, are again extremely close, at $34 - 35$, although once more the distribution of males within this class is slightly larger $(\widehat{\sigma}_{m,5} = 7.10, \widehat{\sigma}_{f,5} = 5.27)$, although fractionally more (3 percentage points), females are estimated to be in this class.

Thus overall, we can see that splitting the sample by gender has no real substantive affect on our results. Indeed, a likelihood ratio test did not reject the null of equality of coefficients across genders ($p = 0.75$).

6 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 econometric approach leads to the estimation of five BMI classes (as did the traditional approach). 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 five classes in terms of sign, size and statistical significance. All metrics employed, clearly favoured the newly suggested approach (with the exception of AIC; which was shown in Monte Carlo experiments, to have very poor performance). Indeed, the experimental evidence (provided in the online appendix), suggested that, in general, the BIC, $HQIC$ and the $Vuong$ test statistics, are all very useful in correctly selecting the appropriate model, whereas the AIC should be avoided.

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 suggests that choosing incorrectly could easily lead to incorrect associations in terms of the magnitude and even 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. It is apparent that in order 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. 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.

Blustein, J., K. Hanson, and S. Shea (1998): "Preventable hospitalizations and socioeconomic status," *Health Affairs*, 17(2), 177–189.

Borghans, L., A. Duckworth, J. Heckman, and B. ter Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43(3), 972–1059.

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.

Caspi, A., W. Roberts, and R. Shiner (2005): "Personality Development: Stability and Change," *Annual Review of Psychology*, 56, 453–84.

Cavanaugh, J., and A. Neath (1999): "Generalizing the derivation of the Schwarz information criterion," *Communications in Statistics - Theory and Methods*, 28(1), 49–66.

Chen, J. (2017): "On Finite Mixture Models," *Statistical Theory and Related Fields*, 1(1), 15–27.

Chou, S., M. Grossman, and H. Saffer (2002): "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System," *Journal of Health Economics*, 23(3), 565–587.

——— (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.

Collins, A., W. Greene, and D. Hensher (2013): "Accounting for Attribute Non-Attendance and Common-Metric Aggregation in a Probabilistic Decision Process Mixed Multinomial Logit Model: A Warning on Potential Confounding," *Transportation*, 40(5), 1003–1020.

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

Dayton, C., and G. B. Macready (1988): "Concomitant variable latent-class models," *Journal of the American Statistical Association*, 83(401), 173–178.

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.

Fair, R., and D. Jaffee (1972): "Methods of Estimation for Markets in Disequilibrium," *Econometrica*, 40, 497–514.

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, and B. Hollingsworth (2015): "Inflated responses in measures of self-assessed health," *American Journal of Health Economics*, 1(4), 461–493.

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.

Grun, B., and F. Leisch (2007): "FlexMix: An R Package for Finite Mixture Modelling," Discussion paper, Faculty of Commerce, University of Wollongong, Faculty of Commerce, University of Wollongong.

Ha, N., M. Harris, D. Preen, S. Robinson, and R. Moorin (2018): "Identifying patterns of general practitioner service utilisation and their relationship with potentially preventable hospitalisations in people with diabetes: The utility of a cluster analysis approach," *Diabetes Research and Clinical Practice*, 138, 201–210.

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.

Karvelis, P., J. Spilka, G. Georgoulas, V. Chudacek, C. Stylios, and L. Lhotska (2015): "Combining Latent Class Analysis Labeling with Multiclass Approach for Fetal Heart Rate Categorization," *Physiological Measurement*, 36, 1001–1024.

Lambert, D. (1992): "Zero Inflated Poisson Regression with an Application to Defects in Manufacturing," *Technometrics*, 34, 1–14.

Leroux, B. (1992): "Consistent Estimation of a Mixing Distribution," *Annals of Statistics*, 20(3), 1350–1360.

Lugo-Palacios, D., and B. Gannon (2017): "Health care utilisation amongst older adults with sensory and cognitive impairments in Europe," *Health Economics Review*, 7(44), 1–15.

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.

McLachlan, G., and D. Peel (2000): *Finite Mixture Models*. Wiley, Canada.

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.

Pearson, K. (1894): "Contributions to the mathematical theory of evolution," *Philosophical Transactions of the Royal Society of London A*, 185, 71–110.

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.

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.

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.

7 Online Appendix: Finite sample performance

To examine the validity of our modelling approach, we undertake a Monte Carlo (MC) analysis. We generated under two scenarios: one where the MNL is the true data generating process (dgp) and the other where the OP is. For both scenarios we based the dgp on the actual data and specification used above, and coefficients from the appropriate model. In both cases we generate according to a 3-class model and in estimation search for up to a 4-class one in each repetition. The number of MC repetitions was set at 250 in each case, although results were essentially unchanged from 100 onwards. A range of outputs was collected, but only for the (correct) 3-class model (OP and MNL variants).

Thus Table 11 presents the proportion of times the IC metrics choose the respective 3-class model across all models estimated (Overall IC); the proportion of times they selected the model when only considering all OP or MNL variants, respectively (Within IC); and the average proportion of correct class predictions based on the maximum - posterior - probability rule ($Correct$). In addition to these, as in the the empirical example(s) in the main body of the paper, we also report the proportion of rejection probabilities of three versions of the $Vuong$ (Vuong 1989) statistic. Again the three variants correspond to the $Vuong$ test between the optimal two models: chosen by BIC only; by AIC only; and by both BIC and AIC.

As with any MC experiment these results cannot necessarily be generalised to all other situations; however, they do clearly demonstrate the validity of the approach in the current context and moreover, give greater confidence to the empirical findings.

Considering firstly the OP dgp, Columns $(1-2)$, we can see the preferable properties of BIC (as compared to AIC), in that it correctly selects the OP 3-class model in every instance. That is, when the OP is the true model, BIC always correctly selects the 3-class OP approach across all models considered (all 1-, 2-, 3- and 4-class OP and MNL models). On the other hand, AIC, tending to favour more highly parameterised models, only correctly selects the OP 3-class model in some 40% of instances; whereas for $CAIC$ and $HQIC$, these rejection probabilities are much more favourable (at 100 and 90%, respectively). Mirroring these results, we can see that across all models and variants, none of the IC metrics would (erroneously) select the 3-class MNL model.

Looking at the selection probabilities *within* all OP models only, we see that the performance of the AIC metric improves somewhat (to 62%), suggesting that it was incorrectly selecting the MNL variants in choosing across both OP and MNL models. Looking at IC selection performance only within all MNL models, there appears to be quite a range of them "correctly" predicting the 3-class model, from a low of 10% ($CAIC$) to a high of 99% $(HQIC)$.

There is some evidence of the $Vuong$ test being useful in this selection. Indeed, based on the BIC variant of this, we see that the correct model is selected 62% of the time. That is, if one compares the best two performing models using the BIC metric, the $Vuong$ test will correctly select the OP 3-class model in 62% of instances. The AIC version never selects the correct model, presumably as it very rarely selects the correct OP 3-class model. Finally, the BIC/AIC version, also has some power in selecting the correct model: selecting the best model (BIC) *versus* the best according to AIC, will result in the correct model being chosen in 40% of instances.

Turning to the $Vuong$ test for the 3-class MNL mode, we can see that the variants based on the best AIC models and the best BIC and AIC models would essentially never pick the 3-class MNL model. On the other hand, using the BIC variant, the test would incorrectly pick the 3-class MNL model about a quarter of the time (although this still may be of use, if the alternative model was a 4-class one, for instance).

In summary of these OP *dgp* experiments: all of the BIC, $CAIC$ and $HQIC$ have very good performance in correctly selecting the 3-class OP model, as does the $Voung$ (BIC) test; overall the AIC metric has poor performance; if one is interested in just selecting the MNL model with the correct number of classes, the $HQIC$ *metric* is extremely useful.

When the true data is generated by a 3-class MNL model, Columns $(3-4)$, the performance of the IC metrics in correctly selecting this is less convincing (Column 4). Now the standout performer is the $HQIC$ metric, with correct selection probabilities of just under 100%. The AIC continues to perform badly (at only 16%), and even the BIC appears marginally worse than a 50/50 bet. Mirroring these results are the equivalent ones for the OP model (Column 3); here we see that reliance on the BIC metric would erroneously cause one to select the 3-class OP model just over 50% of the time (with the $CAIC$ faring even worse at nearly 80%). AIC would never wrongly select the OP 3-class model (but neither would it ever substantially select the correct 3-class MNL one). Again the $HQIC$ metric appears a good choice here, in that whilst correctly selecting the 3-class MNL model nearly ubiquitously, it also never incorrectly selects the 3-class OP model.

If one were primarily concerned with simply picking the correct number of classes within either all OP or MNL models, then all of the BIC, $CAIC$ and $HQIC$ metrics appear very useful here (with the poorest performance being afforded by the $HQIC$ within the OP approaches; but still quite reasonable at some 80%). However, once more the AIC appears quite unreliable here.

In short, with respect to the MNL dgp, only the $HQIC$ metric appears to be able to reliably select this across all OP and MNL models (although all of the BIC, $CAIC$ and $HQIC$ do well in selecting within only MNL models). However, if one considers the $Vuong$ statistics, we see that the BIC variant of this, has a 100% success rate. That is, if one chooses the optimal two models with respect to BIC values, the $Vuong$ test will correctly select the 3-class MNL model each time.

In summary of these model selection metrics then, in general one can relatively safely disregard the AIC. For the OP dgp, the BIC, $CAIC$ and $HQIC$ all work very well (especially the two former) in selecting the correct model, as does the $Vuong$ (BIC) test (although slightly less-so). On the other hand, for the MNL dgp the $HQIC$ appears to dominate, and the $Vuong$ (BIC) now has a perfect success rate. Obviously in practice one will not know how the data was generated. However, the above results clearly suggest that if one focusses on the BIC and $HQIC$ metrics, combined with the $Vuong$ (BIC) statistic, one is very likely to correctly choose the right model, regardless.

Turning next to the proportion of observations correctly allocated to each class (Table 11) - generally a very important aspect of LCM applications - there is effectively little difference between the two approaches (as noted, classes were predicted on the basis of the maximum, posterior, probability rule). So, when the OP approach is true, the OP model correctly predicts, on average, 69% of observations, compared to 68% for the MNL one (based upon the assumption that the correct class model has been selected). A similarly matched performance across the two approaches is found for the MNL dgp, at 72 and 73%, respectively.

Thus all of the above results give us great confidence in relying on these metrics and tests, with the exception of AIC, both in general, and in particular for our modelling BMI.

Finally, prediction of within-class EV's is also considered in Table 12: comparing actual EV's with estimated ones based on sample means of covariates, $EV(\overline{x})$; and averaged across observations by predicted class $EV(x_i)$. For the latter, individuals were allocated to classes

Table 11: Monte Carlo results: summary statistics

	OP dgp		*MNL dgp*	
	OP	*MNL*	*OP*	*MNL*
Overall *IC*	(1)	(2)	(3)	(4)
BIC	1.00	0.00	0.55	0.45
AIC	0.40	0.00	0.00	0.16
CAIC	1.00	0.00	0.78	0.22
HQIC	0.90	0.00	0.00	0.99
Within *IC*				
BIC	1.00	0.39	1.00	1.00
AIC	0.62	0.17	0.01	0.16
CAIC	1.00	0.10	1.00	1.00
HQIC	0.90	0.99	0.81	0.99
Correct	0.69	0.68	0.72	0.73
Vuong (*BIC*)	0.62	0.26	0.00	1.00
Vuong (*AIC*)	0.00	0.06	0.00	0.16
Vuong (*BIC*, *AIC*)	0.40	0.00	0.00	0.16

based on the maximum posterior probability rule (in instances where there were no individuals predicted to be in any particular class, simple sample averages across all observations were used). Once more, somewhat reassuringly, no matter the approach used, both at $\overline{x}$ and also averaged over individuals, class-specific EV's are very accurately estimated (again on the assumption that the correctly classed model has been selected). For example, when generating under OP, the true $EV_1 = 23.1$, which is estimated exactly by the OP approach at $\overline{x}$, and only 0.1 out when averaged over individuals; the MNL ones are similarly also very favourable here, at 23.1 for both. These results essentially hold throughout, implying that within class EV's will be accurately, and similarly, estimated regardless of the true *dgp*, on the assumption that the correct class of model has been selected.

It is intuitive to relate these findings back to the analysis of BMI. Fortunately in the empirical example(s), there was invariably much consensus across all IC variants, with the exception of the AIC, where all of the former tended to select the suggested OP variants. Indeed, given the results of these current findings, one would do well to steer away from the AIC suggestions. So for our primary example (as well as generally for the robustness checks), all but the (unreliable) AIC metrics, as well as all of the *Vuong* tests considered, lent support to the 5-class OP model, again giving us great confidence in the results in our suggested approach and the consequent findings.

The fact that these MC experiments show that both approaches often have quite similar

Table 12: Monte Carlo results: summary statistics

	OP			MNL		
			OP dgp			
EV's	Actual	EV $(\overline{x})$	EV (x_i)	Actual	EV $(\overline{x})$	EV (x_i)
EV class 1	23.1	23.1	23.2	23.1	23.1	23.1
EV class 2	26.9	26.8	26.9	26.9	27.0	27.1
EV class 3	33.2	33.0	33.0	33.2	33.5	33.2
			MNL dgp			
EV class 1	22.6	22.5	22.4	22.6	22.6	22.5
EV class 2	26.7	26.6	26.7	26.7	26.7	26.7
EV class 3	33.2	32.4	32.1	33.2	33.3	32.6

performance across dgp's is not to be taken as an indication that the form of the approach taken will be inconsequential in practice. Indeed, the results from modelling BMI make this quite clear, with many quantities of interest being quite distinct across approaches. In reality, it is likely that neither of these approaches represent an exact description of the true dgp, but the choice is more so of which one more closely mimics this reality in a parsimonious manner, as compared with the "clinical laboratory" conditions of the MC experiments. Given the results presented in this paper, it is our conjecture that this will, more often than not, be provided y the newly suggested OP approach.

8 Online Appendix: Estimation considerations

8.1 Estimation algorithms; maximum likelihood *versus* EM

Model parameters were estimated by maximum likelihood. The algorithm used is predominantly the $BFGS$ gradient method. Likelihoods for latent class models are sometimes maximized by the EM algorithm. However, this method cannot be used for this model because the class specific functions are not separable: due to the imposed ordering across classes, β_1 appears in the conditional mean function of *all* Q class specific functions. The so-called M step of the EM algorithm involves computation for each class separately, which would not impose the cross-class equality constraints required here. On the other hand, maximum likelihood estimation is generally routine and conveniently allows the construction of the full model. In estimation we also used algorythmic derivatives, whereas analytical ones are likely to improve convergence performance. We also used numerical procedures where appropriate, to evaluate relevant quantities of interest, and corresponding standard errors were obtained using the Delta method. Robust standard errors were calculated using the usual outer product of the gradient (OPG) estimator for the parameters of the model.

8.2 Start values

Starting values for the OP procedure were obtained in the following manner.

1. The MNL 2-class model was firstly estimated, using OLS values for the regression and variance terms (perturbed for one of the classes), β_q and $V(\varepsilon_q) = \sigma_q^2$, $q = 1, 2$. In estimation to ensure well-defined variances/standard deviations, these entered the likelihood functions as $\sigma_q = \exp(\omega_q)$, where ω_q is freely estimated. Starting values for the single parameter vector γ required for a 2-class model, was obtained by a random draw from $N(0,1)/10$. Note that here, and elsewhere where appropriate, the user-written *Gauss* code was benchmarked against the available commercial software (*c.f., Limdep/Nlogit* and *Stata*).

2. Based on $\hat{\gamma}$ from 1., a 2-class restricted variant was estimated where start values for $\beta_{q=2}\,(restricted)$ were given by $\widehat{\beta}_{q=2}\,(unrestrricted)/100$. We note here that we do not consider this as a valid OP variant, as due to the 2-class nature of the model, one class by definition must embody a higher (lower) EV than the other one; and moreover the probabilistic expressions for both will be identical. However, we use it simply as a tool for providing sensible start values for the 3-class variant.

3. For the 3-class OP variant, we require start values for γ, μ_1, μ_2, β_q and ω_q, $q = 1, 2, 3$. We set the $\boldsymbol{\mu}$ values to simply split the standard normal dustribution into equal parts: $\mu_1 = \Phi^{-1}(1/3)$ and $\mu_2 = \Phi^{-1}(1/3)$. We set the start value for $\beta_1 = \widehat{\beta}_1$ from 2.; all other start values were set equal to zero. Note that in estimation we used the in-built *Gauss cmlMT* inequality constraint function to ensure the requisite ordering in the μ_q parameters throughout. If such a function is unavailable, one could equivalently use $\mu_q = \mu_{q-1} + \exp(a_q)$, where a_q would be freely estimated.

4. For the 4-class model, a similar progression was followed for start values: $\mu_1 = \Phi^{-1}(1/4)$; $\mu_2 = \Phi^{-1}(1/2)$; $\mu_3 = \Phi^{-1}(3/4)$; $\beta_1 = \widehat{\beta}_1$ (from 3.); all other start values were set to zero.

5. Start values for the 5-class model continued this progression.

We should note that, so long as sensible start values were given, the maximum likelihood estimates ended-up at the same values, but speed of convergence was sometimes affected. However, the procedures described above may not necessarily be optimal for all applications. In practice it might be advisable to try a range of different start values, and to enter previously solved final estimates as new start values to ensure that the likelihood has achieved a global, and not local, maximum.

Note also that *Gauss* code is freely available at:

https://drive.google.com/drive/folders/1rtoYfs5qfwcI4NcFpq0-pdhLJOZZEa56?usp=sharing.