Mixture models for ordinal data

Breen, R.; Luijkx, R.

Published in:
Sociological Methods and Research

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Breen, R., & Luijkx, R. (2010). Mixture models for ordinal data. *Sociological Methods and Research*, *39*(1), 3-24.

Mixture Models for Ordinal Data

Richard Breen and Ruud Luijkx

Sociological Methods & Research 2010 39: 3 originally published online 14 May 2010
DOI: 10.1177/0049124110366240

Sociological Methods & Research
39(1) 3–24

DOI: 10.1177/0049124110366240
http://smr.sagepub.com
SAGE

Mixture Models for Ordinal Data

Richard Breen[1] and Ruud Luijkx[2]

Abstract

Cumulative probability models are widely used for the analysis of ordinal data. In this article the authors propose cumulative probability mixture models that allow the assumptions of the cumulative probability model to hold within subsamples of the data. The subsamples are defined in terms of latent class membership. In the case of the ordered logit mixture model, on which the authors focus here, the assumption of a logistic distribution for an underlying latent dependent variable holds within each latent class, but because the sample then comprises a weighted sum of these distributions, the assumption of an underlying logistic distribution may not hold for the sample as a whole. The authors show that the latent classes can be allowed to vary in terms of both their location and scale and illustrate the approach using three examples.

Keywords

ordered probability models, mixture models, latent class, odds ratios

Cumulative probability models are widely used for the analysis of data where the dependent variable is ordinal. The ordered logit model, on which we focus here, rests on the assumption that the observed dependent variable, Y, is a discretized observation of an underlying continuous logistically distributed variable, Y^*, which has a common scale parameter but whose location parameter differs across units of observation in the population

[1]Yale University, New Haven, CT, USA
[2]Tilburg University, The Netherlands

Corresponding Author:
Richard Breen, Department of Sociology and Center for Research in Inequalities and the Life Course, Yale University, P.O. Box 208265,New Haven,CT06520-8265, USA
Email: richard.breen@yale.edu

according to their values on measured covariates. Other assumptions about how Y^* is distributed give rise to other forms of cumulative probability models having different so-called link functions, such as the cumulative normal and complementary log-log. In this article we present an empirical modeling approach based on the less restrictive assumption that the population of interest comprises a number of subpopulations, and each of these has its own baseline logistic distribution of the underlying latent dependent variable, differing in either or both their location or scale. Thus, we propose mixture models in which the distribution of the dependent latent variable in the population as a whole is a mixture of these separate distributions. Estimation of these models involves determining the number of subpopulations, the share of the population in each of them, and the parameters of their distributions. In practice the subpopulations are captured as latent classes, and so the model we present here is based on the idea of adding latent classes to conventional cumulative probability models.

We use three data sets to explain and to illustrate our proposed approach. Although we refer to the ordered logit model throughout, our results apply equally to any other choice of link function that might be used with cumulative probability models.

Cumulative Probability Models and the Ordered Logit

Let Y be a dependent ordinal variable with categories indexed $j = 1, \ldots J$, and X a vector of explanatory variables. We write the probability that the value of Y for the ith observation, $y_{\mathrm{i},}$ is less than or equal to j, given X, as $\gamma_j(x_i)$. There exists a family of statistical models that sets $g(\gamma_j(x_i)) = t_j - \beta' x_i$ (McCullagh 1980), where g is a link function (e.g., the logit) that maps the (0,1) interval into $(-\infty, \infty)$ and t_j are a set of thresholds. Differences in $g(\gamma_j(x))$ between observations (dropping the i subscript for convenience) depend only on their values of x and the β parameters and are independent of the location of the thresholds. A cumulative probability model with a logit link sets $g(\gamma_j(x)) = \ln \frac{(\gamma_j(x))}{(1-\gamma_j(x))}$, and the log odds ratio of exceeding, rather than failing to exceed, any particular level of Y, as between respondents i and i, is

$$g(\gamma(x)) - g(\gamma(x')) = \ln \frac{\gamma(x)}{1-\gamma(x)} - \ln \frac{\gamma(x')}{1-\gamma(x')} = \beta(x_i - x_{i'}). \qquad (1)$$

The log odds ratio is independent of the thresholds and does not vary over them (reflected in the absence of a j subscript on β). In the case where X is a single variable taking a finite set of values, the model's assumptions imply that Y follows a logistic distribution in each category of X and that these distributions vary only in their location.

In his 1980 paper, McCullagh (p. 119) extended this model by allowing the scale of the logistic distribution, as well as its location, to differ according to values of X. This model is

$$g(\gamma_j(x)) = (t_j - \beta' x_i)/\tau(x_i). \tag{2}$$

where τ is the scale parameter. In McCullagh's example, there was one categorical X variable, and so each category of X had its own scale and location parameters. The scale parameters are not separately identified but their relative values are: So typically one would fix $\tau(x = 1) = 1$. Even so, McCullagh's model is not identified. The statistical model that generalizes this idea—namely, a model allowing the scale parameter to be a function of covariates—is sometimes called the ordered logit with heteroskedasticity and is available as an option in the LIMDEP program (Greene 2002). This model can be identified by making the scale parameter a function of variables that do not also affect the location.

Ordered Logit Mixture Models

An extension of this idea, which we develop here, is to allow one or both of the scale or location parameters to vary according to the value of a latent variable. Because our latent variable is categorical, this gives rise to simple mixture models in which different subgroups of the population have a latent dependent variable, Y^*, whose baseline logistic distribution is characterized by subgroup specific location and/or scale parameters. In its most general form our model can be written:

$$g(\gamma_j(x)|k) = \frac{t_j - \beta' x_i - \alpha_k}{\tau_k}. \tag{3}$$

Here k denotes membership of the kth latent class. We normalize our estimates by setting $t_1 = -\infty$ (and so t_j, $j > 1$, is the threshold separating the $j-1$th and jth categories) $\alpha_1 = 0$ and $\tau_1 = 1$. We estimate the parameters, α_k and τ_k $(k = 2, \ldots, K)$ and the distribution of the population across each of the K latent classes. Substantively the model allows different subpopulations to have distributions of the latent dependent variable that vary in both their average value and in their dispersion around this value. We provide examples of this later in the article.

A simpler model allows only the location of the underlying logistic distribution to vary over latent classes:

$$g(\gamma_j(x)|k) = t_j - \beta' x_i - \alpha_k. \tag{4}$$

Another possibility is to allow the dispersion in the responses to vary across latent classes by permitting heterogeneity in the thresholds but not in the β parameters:

$$g(\gamma_j(x)|k) = \frac{t_j}{\tau_k} - \beta' x_i - \alpha_k. \tag{5}$$

The advantage of (5) over (3) is that it allows for latent heterogeneity but keeps the effects of the *X*s the same regardless of latent class membership (whereas in (3) they vary over latent classes because their effect is equal to β/τ_j). We refer to (5) as the heterogeneous threshold model. In this case, if $\tau_k < 1$, the interthreshold spacing increases: Positive thresholds become larger and negative ones become smaller (further from zero). Conversely, if $\tau_k > 1$, all the thresholds are moved closer together. Compressed thresholds increase the likelihood that responses will fall into the more extreme categories, while more dispersed thresholds lead to responses clustered around the middle category. This makes it clear that in this model, in contrast to model (3), we are changing the scale of the observed responses and not of the underlying logistic variable.

But this is not the whole story because this model also allows the mean to differ over latent classes (captured by α_k) and, taken together with the rescaling of the thresholds, this allows us to capture a wide range of distributions of responses. For example, if the mean is larger in latent class *k* and $\tau_k < 1$, this will lead the distribution of responses to be clustered on a category above the middle one. Similarly, if $\tau_k > 1$, instead of responses being clumped at both extremes we may instead find that they pile up in the top categories and are almost absent from the bottom ones. We shall see examples of this later and the properties of the model will be further discussed and illustrated.[1]

Population Quantities

Population models can be derived from the mixture model by summing over the latent classes, though because the cumulative logit model is nonlinear, this is less straightforward than in some other mixture models. However, we can derive the following expression for the probability that $y_{i,}$ is less than or equal to j, given X, in the population as a whole:

$$\gamma_j(x) = \sum_k \pi_k \frac{\exp(g(\gamma_j(x)|k)}{1 + \exp(g(\gamma_j(x)|k)}. \quad (6)$$

Specific formulations of the mixture model (as in equations (3), (4), and (5)) could then be inserted in place of $g(\gamma_j(x)|k)$, and from (6) we could, if we wished, compute the corresponding log odds for the population, $g(\gamma_j(x_i))$.

Interpretation

It is well known (Holm, Jaeger, and Pedersen 2009; Lindsay 1983a, 1983b) that any form of unobserved heterogeneity can be arbitrarily well approximated by a finite number of latent classes, and thus our approach, which involves estimating a mixing distribution via the addition of latent classes, can also be seen as a way of modeling unobserved heterogeneity. One consequence of unmeasured heterogeneity is that estimates of the effects of the explanatory variables may be biased, and in nonlinear models, unlike in linear models, this can hold even when such heterogeneity is uncorrelated with the measured variables. The model can therefore be motivated in two ways. On the one hand, we can see it as a way of dealing with the failure of the conventional ordered logit model to give a good account of the data because of unmeasured heterogeneity arising from omitted variables. In this case we might want to consider the latent classes as representing real categories of observations in the data, especially when membership of these categories is difficult to measure directly: for example, in attitude questions distinguishing those who feel strongly about the issue in question from those who do not. On the other hand, we could see the latent classes as a means of providing a more flexible functional form and in this case we should probably not want to interpret the latent classes in any substantive way.

Comparisons

Equations (3) through (5) apply to a single table, but they are easily generalized to use in the context of comparisons across samples that come from different countries or different birth cohorts. In this case we would usually allow the threshold parameters to differ across samples to allow for differences in the marginal distribution of Y, and our interest would lie chiefly in whether the β parameters also differ between them. Letting S denote the different samples, the counterpart to model (4), with βs constant across samples, can be written

$$g(\gamma_j(x,s)|k) = t_{js} - \beta' x_i - \alpha_k, \tag{7}$$

and allowing the βs to vary over samples we have:

$$g(\gamma_j(x,s)|k) = t_{js} - \beta'_s x_i - \alpha_k. \tag{8}$$

In addition, the latent classes in all these models can be independent of S and X (as in all the examples we present) or they can be correlated with them. The advantage of independence of the latent classes is that odds ratios formed by comparing observations with different values of X in the same latent class will be identical in all latent classes. So, for example, if, as in one of our examples, X measures social class origins, class origin inequalities in the odds ratio of being in one category of Y rather than another will be the same regardless of the latent class membership of the respondents being compared.

Identification and Estimation

Given panel data, the use of latent classes independent of X is a discrete counterpart to a random effects model: A continuous random effect is replaced by a discrete approximation.[2] But here we are concerned only with the use of latent classes in cross-sectional analyses. In an article dealing with the use of latent classes in binomial logit models, Holm et al. (2009) show that such a model is well identified using panel data but identification is fragile when using cross-sectional data. We can reasonably assume that the models we present are somewhat more robust than that of Holm et al. because of the greater information available about the latent variable in cumulative probability models compared with binomial logit models; nevertheless, the issue of identification raises some difficulties. The models we consider are not nonparametrically identified: Rather, when they are identified they are so only because of the distributional assumptions of the original ordered logit model (and the same would hold for the ordered probit, complementary log-log, etc.). But these assumptions do not guarantee identification, and in the absence of any general proofs, it is important to check local identification by inspection of the eigenvalues of the information matrix. In addition, because the likelihood function may have local maxima it is also important to run the model many times with different starting values. We estimate the models using the LEM program (Vermunt 1997) and the appendix shows the LEM syntax for several of the models fitted here (the online appendix is available at http://smr.sagepub.com/supplemental).

We illustrate the approach with three examples of increasing complexity. In the first we allow for two latent logistic distributions, differing in only their location. Our second example, applied to data in the form of a single table, investigates models that allow the scale and location to differ across latent classes and that allow for differences in location and for heterogeneity in thresholds. Our last example involves a comparison across countries and shows how the use of this approach can add new insights in comparative analyses.

Educational Inequality in Great Britain

Our first example uses data from one of the British birth cohort studies, the National Child Development Study (NCDS). This comprises data referring to all children born in Great Britain in a particular week in March 1958. The initial sample size was just over 17,000 and information has been collected on sample members at birth, ages 7, 11, 16, 23, 33, and 40. Here we focus on the relationship, among men, between parental social class, ability, and highest level of educational attainment. The measure of highest educational qualification used here has been coded into a set of categories known as the NVQ (National Vocational Qualification) levels.[3] The educational categories, from lowest to highest, are:

1. No qualifications;
2. Certificate of Secondary Education (CSE) or equivalent;
3. Ordinary Level General Certificate of Education (O-level), or equivalent;
4. Advanced Level General Certificate of Education (A-level), or equivalent;
5. Higher technical qualifications, subdegree qualifications;
6. University degree or higher.

Social class origins are defined using the original seven class version of the Goldthorpe class schema (Erikson and Goldthorpe 1992:chap. 2):

I. Upper service class—higher grade professionals, administrative, and managerial workers;
II. Lower service class—lower grade professionals, administrative, and managerial workers;

Table 1. Educational Attainment by Class Origins, British Men Born 1958 (National Child Development Study Data)

	Educational attainment					
Class origins	None	CSE	O-level	A-level	Subdegree	Degree
I	21	7	61	68	29	51
II	57	21	136	126	42	48
III	41	14	106	72	21	17
IV	47	11	48	46	13	5
V	41	10	68	46	12	8
VI	281	54	258	197	62	34
VII	239	42	164	83	25	9

III. Routine nonmanual workers;
IV. Petty-bourgeoisie—the self-employed and small employees;
V. Technicians and supervisors of manual workers;
VI. Skilled manual workers;
VII. Nonskilled manual workers.

The classes are derived from information about the occupational position of the head of the household when the respondent was age 11. This was originally coded to the British census's "Socio-Economic Group" classification from which a reliable approximation to the original Goldthorpe seven-class schema (Heath and McDonald 1987) can be obtained. The cross-tabulation of highest educational attainment and father's social class is shown in Table 1.

We use a continuous measure of ability in the analysis: This is based on the results of a general ability test administered at age 11. The results form an 80-point scale that has previously been used as a proxy for IQ (Breen and Goldthorpe 2001:84; Douglas 1967:33-6).

As Table 2 shows, an ordered logit model, with educational attainment as the dependent variable and the social classes entered as dummy explanatory variables, returns a log-likelihood of –25,904. Considered as a model applied to the data as shown in Table 1 it returns a deviance of 40.82 on 24 *df* ($p =$ .017) and thus fails to fit the data by some way. The coefficients for the social classes have the expected magnitudes: There is a clear gradient, with the log odds of failing to progress beyond any given level of education increasing as we move toward the less advantaged classes. The addition of ability to the model has a substantial impact, reducing the log-likelihood by over 300 points for the loss of one degree of freedom. The negative coefficient reflects

Table 2. Parameter Estimates of Models for British Educational Attainment Data

	Model					
	Ordered logit		Ordered logit		Ordered logit with two latent classes	
Variables	Parameter	*SE*	Parameter	*SE*	Parameter	*SE*
Class origins						
II	0.5973	0.15	0.4651	0.15	0.539	0.20
III	0.9218	0.16	0.7889	0.16	0.9742	0.22
IV	1.3000	0.18	1.022	0.19	1.3644	0.26
V	1.1929	0.18	0.9546	0.18	1.2225	0.25
VI	1.4701	0.14	1.0612	0.14	1.3797	0.19
VII	2.0249	0.15	1.5648	0.15	2.049	0.21
Ability			–0.061	0.003	–0.0857	0.005
Latent class location parameters						
Class 1	—	—	—	—	0	fixed
Class 2	—	—	—	—	–3.4412	0.25
Latent class probabilities						
Class 1	—	—	—	—	0.672	—
Class 2	—	—	—	—	0.328	—
Log-likelihood	–25,904		–25,595		–25,570	

the lesser odds for those with greater ability of failing to exceed a given educational threshold. As we might have anticipated, including ability reduces the class effects: In particular we now see that much of the disadvantage of students from classes VI and VII is mediated through their lower measured ability. A two-class mixture model, in which the classes have different location but common scale parameter, further improves the fit of the model. For the loss of two further degrees of freedom the log-likelihood declines by 25 points.[4] The inclusion of the latent classes causes the estimates of the class origin effects to increase somewhat so that they are quite similar to those reported in the model without ability, though with larger standard errors. This model can be written as $g(\gamma_j(x,z)|k) = t_j - \beta_1 z_i - \sum_{l=2}^{7} \beta_l x_{il} - \alpha_k$ where z_i indicates ability and x_l class membership. The LEM syntax for this model is provided in the appendix.

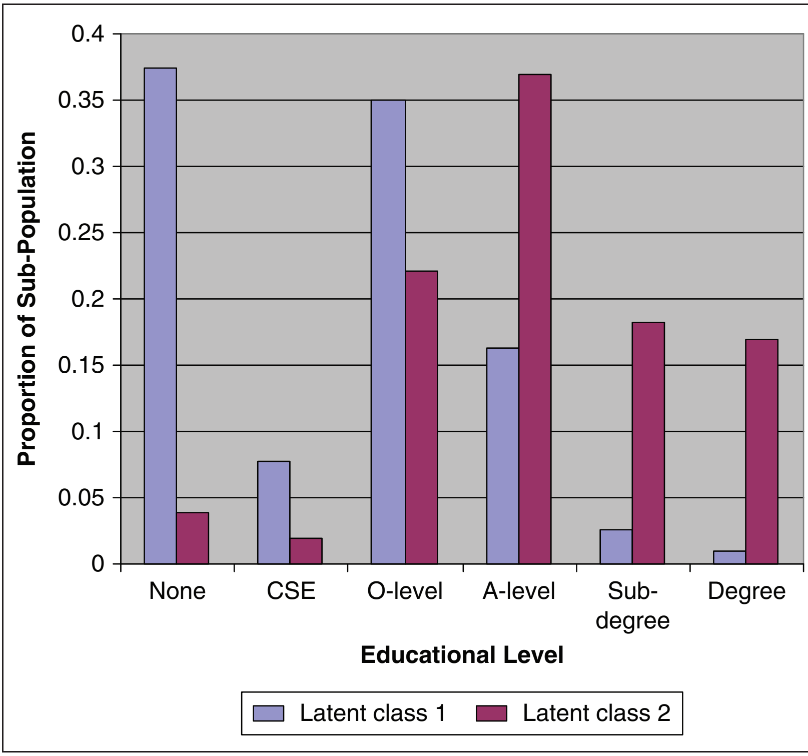

Figure 1. Distribution of educational attainment within latent classes

The first latent class accounts for two-thirds of the sample. Given the parameterization of the ordered logit model, the large negative location parameter of the second latent class means that its distribution is further to the right than that of the first latent class, so educational attainment is higher in the second class. This can be seen in Figure 1, which plots the distribution of educational attainment within each of the two latent classes. This makes clear that the population comprises a group of low achievers (members of latent class 1) and a group of high achievers (latent class 2).

Attitudes to Premarital Sex

Our second example concerns attitudes toward heterosexual premarital sex. The data were collected in Great Britain in 1983 (when this was a more controversial topic than nowadays) in the first wave of the British Social

Table 3. Attitudes Toward Premarital Sex by Gender and Age Group, British Social Attitudes 1983

	Age group						
	19 to 25	25 to 35	36 to 45	46 to 55	56 to 65	Older than 65	Total
Men							
Always wrong	1	0	3	4	7	6	21
Mostly wrong	0	0	3	7	4	6	20
Sometimes wrong	2	9	8	2	10	4	35
Rarely wrong	3	4	5	5	2	2	21
Not wrong	26	34	26	19	13	10	128
Total	32	47	45	37	36	28	225
Women							
Always wrong	1	2	4	11	8	25	51
Mostly wrong	0	2	3	10	16	14	45
Sometimes wrong	8	7	11	13	7	10	56
Rarely wrong	6	5	3	3	2	1	20
Not wrong	20	34	30	16	8	6	114
Total	35	50	51	53	41	56	286

Attitudes Survey Panel Study, 1983-1986.[5] The dependent variable is the answer to the item ‘‘If a man and a woman have sexual relations before marriage what would your general opinion be?’’ Six possible responses are provided, but we omit ‘‘it depends/varies’’ and this leaves us with five valid answers forming an ordinal scale: always wrong, mostly wrong, sometimes wrong, rarely wrong, and not wrong. We examine the variation in response according to gender and age group under the assumption that men will express more support for this item than women and that there will be a gradient of declining support with age. We distinguish six age groups: 19 to 25, 26 to 35, 36 to 45, 46 to 55, 56 to 65, and older than 65. In the models age and gender are dummy variables with the youngest age group and men being the omitted categories. The data are shown in Table 3 (cases with missing values on the variables and on the same dependent variable in the 1984 wave have been omitted).

The goodness of fit of various models applied to these data is shown in Table 4. The ordered logit model, with additive effects of gender and age, fits the data reasonably well. Adding interactions between gender and age makes little difference. The mixture model with differences between latent classes in slope and location (and additive effects of gender and age) also fits reasonably well, but is not a clear improvement on the simple ordered

Table 4. Goodness of Fit of Models Applied to British Social Attitudes Data

	Deviance	*df*	*p*
1. Ordered logit additive	52.00	38	.065
2. Ordered Logit With Gender × Age Interaction	45.21	33	.076
3. Ordered logit mixture model with scale and location differences	47.57	35	.076
4. Ordered logit mixture model with location differences	52.00	36	.041
5. Ordered logit mixture model with location differences and heterogeneous thresholds	37.40	35	.360

logit. Allowing only for differences in location (model 4) renders no improvement over the ordered logit. The best fitting model proves to be the one that allows the latent classes to differ in their location and to be heterogeneous in their thresholds: that is $g(\gamma_j(x)|k) = \frac{t_j}{\tau_k} - \beta' x_i - \alpha_k$, where x_i is a vector of values of gender and age and β is the corresponding vector of coefficients. The model has deviance of 37.4 on 35 *df*. The LEM syntax for this model is provided in the appendix.

Parameter estimates are reported in Table 5. The estimated effects of age and gender are as expected, with younger people (younger than 35) and men being less likely to consider premarital sex wrong and the probability of considering it wrong increasing with age. The two latent classes account for 28 percent and 72 percent of the sample, respectively. The estimate of τ_2 is 7.19, so the thresholds for the second latent class are scaled by a factor of 0.139 (which is what we report in Table 5). The rescaled thresholds for the second latent class are shown at the bottom right of Table 5 alongside the threshold values for the first latent class on the left. They display less variation than those for latent class 1 and this implies that the responses in latent class 2 are more dispersed and more likely to fall into the extreme categories. But class 2 also has a higher mean (recall that negative coefficients increase the log odds of exceeding a given threshold) and so this skews the responses such that a greater share falls into the upper extreme category (not wrong) compared with the lower extreme (always wrong). We can see this in Figure 2, which shows the estimated distribution of responses in each latent class, and indeed, in latent class 2 the responses are clustered at the two extremes but with many more giving the higher (not wrong) response. This leads to the

Table 5. Parameter Estimates From Model With Heterogeneous Thresholds, British Social Attitudes Data

Age group	Parameter	*SE*	*z*		Parameters for latent class = 2		
26 to 35	0.398	0.56	0.71		Parameter	*SE*	*z*
36 to 45	1.355	0.57	2.38	Location	−3.257	0.59	−5.51
46 to 55	2.561	0.61	4.18	Threshold scaling[a]	0.139	0.07	1.87
56 to 65	3.075	0.58	5.26				
Older than 65	3.663	0.59	6.24		Latent Class Probabilities		
Gender				Class 1	Class 2		
Women	0.823	0.22	3.75	0.278	0.722		
Estimated thresholds for class = 1				Estimated thresholds for class = 2			
2	−7.617	3.92	−1.94	−1.055			
3	−3.696	0.74	−4.97	−0.512			
4	−0.693	0.71	−0.98	−0.096			
5	1.156	1.85	0.63	0.160			

a. Threshold scaling = $1/\tau_2$.

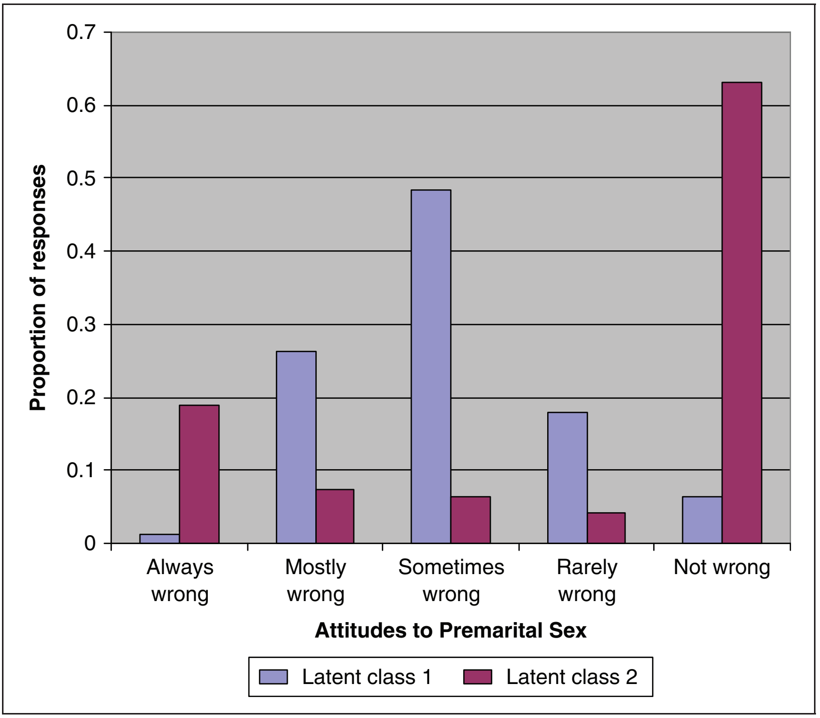

Figure 2. Distribution of responses by latent class, British Social Attitudes Survey Panel Study, 1983-1986 data

interpretation of these classes as distinguishing between the majority of respondents who have strong views about the rightness or otherwise of premarital sex (latent class 2) and the minority who do not. And, of those with strong views, most believe that premarital sex is not wrong.

To test the validity of this model, we estimated it using data from the next (1984) wave of the panel.[6] Since we have already selected only those cases with valid responses in both 1983 and 1984 the sample is the same. The mixture model with differing location and heterogeneous thresholds applied to the 1984 data yields an estimate of the threshold scaling parameter $(1/\tau_2)$ of .139 (exactly the same as in 1983) and of the location, α_2, of –2.49 (compared with –3.26). The second latent class comprises 76 percent of cases in 1984 compared with 72 percent in 1983. The similarity in the τ and α parameters means that the distribution of responses in the second latent class is very similar in the two years: Indeed, all the differences that we find fall within the margins of measurement error.[7] This gives us good grounds for thinking that the underlying distributions identified by the latent class analyses are valid for this sample.

Religion and Politics

Our final example addresses people's views of the relationship between belief in god and suitability for public office. We use data from the 1999/2000 wave of the European Values Study (www.europeanvaluesstudy.eu) for France, Germany, and Poland and from the 2000 wave of the World Values Survey (www.worldvaluessurvey.org) for the United States, and we analyze responses to the item: ''Politicians who don't believe in god are unfit for public office.'' Possible replies are (1) *agree strongly*, (2) *agree*, (3) *neither agree or disagree*, (4) *disagree*, and (5) *strongly disagree*: In simpler terms, higher responses reflect the more secular view that it is not necessary to believe in god in order to be fit for public office.

We assume that individuals' views on this issue are likely to differ according to their own characteristics and that different countries will tend to have different distributions of support for this item. The latter guided our choice of countries, and so we chose two countries that might be described as secular—France and Germany—and two countries in which we expect a close relationship between religion and politics—Poland and the United States. In terms of individual characteristics, we distinguish gender (men and women), education (three categories: primary and uncompleted lower secondary, secondary, and tertiary education), and religiosity, measured by how frequently the respondent attends religious services, coded (1) *once a month or more*, (2) *once a year or on special holidays*, and (3) *never or almost never*. These three variables are included in the models as dummies with the omitted categories being, respectively, men, primary education, and once a month or more. Dropping cases with missing values yields a sample of 5,664.

A model allowing the thresholds to differ between countries but assuming common effects of gender, education, and religiosity has a deviance of 377.23 on 267 *df*. Allowing the effects of gender, education, and religiosity to vary cross-nationally reduces the deviance to 306.63 on 252 *df*, and each of the variables shows significant differences in its effects. The ordered logit mixture model with two latent classes differing in scale and with heterogeneous thresholds has a deviance of 284.91 on 249 *df* ($p = .059$) while the simpler model allowing for only differences in location has deviance of 288.93 with 250 *df* ($p = .046$). There is little to choose between these models (difference $= 4.02$ with 1 *df*) but we interpret the results from the former model because they serve to illustrate the substantive insights that it can yield. This model is written as

$$g(\gamma_j(x)|k) = \frac{t_{js}}{\tau_k} - \beta_s' x_i - \alpha_k.$$

In this case, s denotes country, β_s is a country-specific vector of coefficients, and x_i is the vector of values of country, gender, education, and religiosity. Table 6 shows the parameter estimates from this model.

As in the previous examples, negative parameters mean a lower probability of failing to exceed a given threshold and imply higher expected responses on the dependent variable: that is, a view that belief in god is less necessary for those who wish to hold public office. With this in mind, we see that there is a great deal of qualitative commonality among the countries in the effects of the explanatory variables. Women think that belief in god is more important and there is a clear gradient associated with religiosity. There is also an educational gradient, with the more highly educated taking a more secular view. The magnitude of these differences, however, shows some cross-national variation. The gender gap is particularly large in the United States, while the religiosity effect is smaller here than elsewhere. In France both the gender and education differentials are quite small, while in Poland educational differences are very marked, especially between those with only primary education and those with more. Overall differences between the countries in the way in which responses are distributed are captured in the threshold parameters and here the major distinction lies between France and the United States. In France, the third and fourth thresholds, separating the responses *neither agree nor disagree* from *disagree* and *disagree* from *strongly disagree*, lie much further to the left of the underlying logistic distribution than in any other country, meaning that a much larger share of the French sample lies above these thresholds, or, more simply, disagrees or strongly disagrees with the statement about belief in god and public office. In the United States the opposite is true: In fact, here all the thresholds are further to the right than in any other country, so that a much larger share of the U.S. sample falls into the *agree strongly* and *agree* categories and a much smaller share in the *disagree* categories. It seems, then, that these data support our initial hypothesis that France is a secular country in which people tend to feel that politics and belief in god should be kept apart, whereas Americans tend to express the opposite view.

Table 6. Parameter Estimates of Models Applied to Values Data (Standard Errors in Parentheses)

	Ordered logit model, different thresholds and coefficients				Ordered logit mixture model, different thresholds and coefficients and two latent classes			
	France	Germany	Poland	United States	France	Germany	Poland	United States
Women	0.1073	0.1936	0.2546	0.4077	0.1414	0.2193	0.3305	0.5424
	(0.107)	(0.087)	(0.118)	(0.107)	(0.131)	(0.100)	(0.142)	(0.132)
Education								
Secondary	–0.340	–0.2291	–0.8172	–0.2959	–0.4161	–0.2577	–0.9767	–0.3255
	(0.140)	(0.091)	(0.135)	(0.151)	(0.166)	(0.103)	(0.164)	(0.188)
Tertiary	–0.3889	–0.6349	–1.0905	–0.5513	–0.4874	–0.6958	–1.2833	–0.7056
	(0.135)	(0.151)	(0.177)	(0.140)	(0.162)	(0.171)	(0.211)	(0.175)
Religiosity								
Once per year	–0.3914	–1.4892	–0.7352	–0.5342	–0.4775	–1.7848	–0.8283	–0.6407
	(0.168)	(0.125)	(0.179)	(0.141)	(0.228)	(0.169)	(0.204)	(0.173)
Never	–1.4605	–1.9497	–1.5986	–1.1507	–1.7974	–2.3079	–1.7807	–1.3737
	(0.151)	(0.112)	(0.224)	(0.137)	(0.217)	(0.177)	(0.264)	(0.168)
Thresholds								
1	–2.193	–2.597	–2.298	–1.114	–1.592	–2.055	–1.838	–0.548
2	–1.284	–0.787	–1.302	–0.009	–0.630	–0.081	–0.753	0.807
3	–0.203	0.549	–0.235	1.123	0.591	1.448	0.543	1.998
4	0.508	2.069	1.957	3.003	1.403	2.781	2.608	3.315
						Latent class location parameters		
					Class 1		Class 2	
					0	fixed	–4.5697	(1.574)
						Latent class threshold scale parameters[a]		
					Class 1		Class 2	
					1	fixed	2.2384	(0.927)
						Latent class probabilities		
					Class 1	0.6684		
					Class 2	0.3316		

a. Threshold scaling $= 1/\tau_2$.

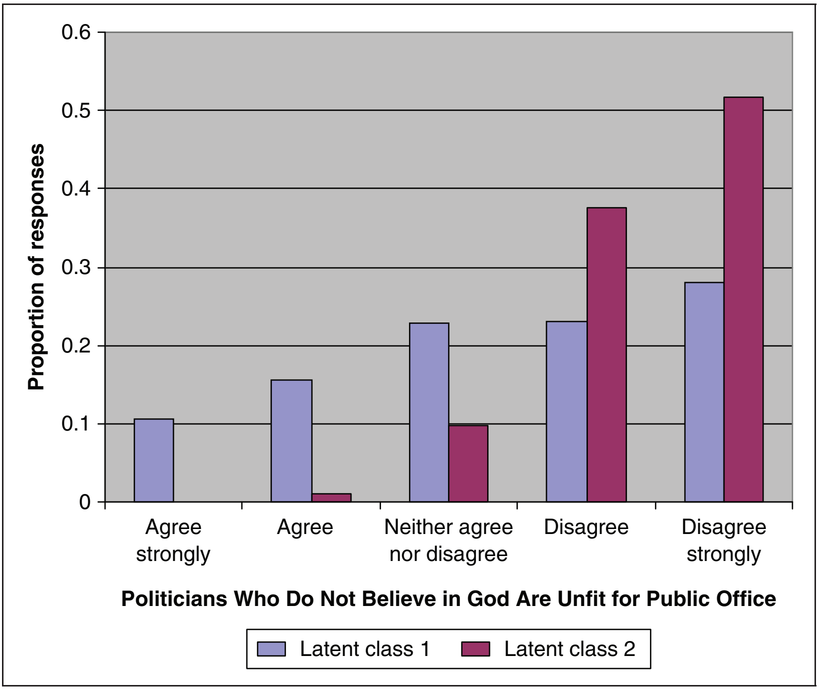

Figure 3. Distribution of responses by latent class, values data

All these results are evident in both sets of parameter estimates shown in Table 6, though as before, the effects are somewhat stronger in the mixture model. However, the mixture model can add some further insight. The distributions of responses to the original question differ quite markedly according to latent class membership. The differences in scale and the heterogeneity of the thresholds between the two latent classes are such that the smaller latent class, comprising roughly a third of the total, has a much less compressed distribution of the response thresholds ($\hat{\tau}_2 = 0.447$) and a higher mean of the underlying latent logistic variable ($\hat{\alpha}_2 = -4.570$). The effect of this on the responses can be seen in Figure 3: Almost all those from the smaller latent class (latent class 2) fall in the categories *disagree* and *disagree strongly* with almost no agree or agree strongly responses, whereas for the larger latent class, they are more evenly dispersed. This suggests that in all countries we find a sizeable minority of respondents who think that belief in god is not relevant in assessing suitability for public office. Furthermore, among this

Table 7. Estimated Probabilities of Exceeding Given Thresholds in Each Latent Class, Values Data

	Latent class 1				Latent class 2			
Threshold	2	3	4	5	2	3	4	5
France	0.17	0.35	0.64	0.80	0.87	0.98	1.00	1.00
Germany	0.11	0.48	0.81	0.94	0.70	0.99	1.00	1.00
Poland	0.14	0.32	0.63	0.93	0.80	0.98	1.00	1.00
United States	0.37	0.69	0.88	0.96	0.99	1.00	1.00	1.00

group, differences in education, gender, or religiosity have little impact on their response. We can see this if we note that all the coefficients in Table 6 are negative except those for women: Thus, the expected value of the underlying dependent latent variable, Y^*, is –4.57 for men with primary education who attend religious services most frequently (i.e., the omitted categories of the dummy variables) and who are in the second latent class. For men with more education or who attend less frequently, this expected value is more strongly negative and so –4.57 is the largest value among men in latent class 2. By the same argument, among women the largest values are –4.02, –4.24, –4.35, and –4.43 in the United States, Poland, Germany, and France, respectively.[8] At the same time, the smaller scale parameter for latent class 2 means that the thresholds are more widely dispersed than for latent class 1. Taking the United States as an example, if we rescale the threshold values by 2.238 ($=$ 1/0.4467; this is the value reported in Table 6), we find that they become –1.23, 1.81, 4.47, and 7.42. Now we can compute the probability of exceeding the *j*th threshold for our hypothetical man in latent class 2 with primary education who attends religious services most frequently, as $\exp(\gamma(\mathrm{t}_{j2}))/(1+\exp(\gamma(t_{j2}))$, where $\gamma(t_{j2}) \equiv t_{j2} - (-4.57)$. Here the 2 subscript refers to the second latent class. The estimated probabilities, for all four thresholds and all four countries, are shown in Table 7, together with the equivalent values for the first latent class. The probability of exceeding any threshold is much greater for the second latent class, implying that responses among its members will be clustered in the disagree categories (as shown in Figure 3) to a much greater extent than will members of class 1. The only group for whom these probabilities will be lower are women with the same level of education and religiosity, but even for them, the probabilities are still close to one if they are members of latent class 2. For any other combinations of education and religiosity, the probabilities will be greater. Thus, latent class 2 comprises respondents who disagree with the statement that ''politicians who don't believe in god are unfit for public office'' irrespective of their education or religiosity or gender. We might describe them as having an ideological commitment that overrides the usual differences associated with education, religiosity, and gender.

Conclusions

We have presented three examples of increasing complexity to illustrate the ordered logit mixture model. As these examples have shown, the model allows for a variety of specifications, of which in this article we have focused on three: allowing the mixing distributions or latent classes to have different locations and/or different scales and allowing the threshold parameters to be scaled differently in the different latent classes. We found that by combining the latter, which we call the heterogeneous threshold model, with differences in location we could capture a wide variety of possible distributions of responses over the categories of the ordinal dependent variable.

There are several ways in which the models presented here could be extended. Most simply, we could allow the latent classes to be correlated with some of the observed covariates instead of, as here, making them independent. An obvious application would seem to occur in our final example, where we might hypothesize that the distribution of the latent classes would differ across countries.[9] But here caution will be necessary because although making latent class membership depend on observed covariates that also affect the dependent variable may be an attractive idea, identification is likely to be fragile unless the latent class distribution also depends on covariates that do not affect the dependent variable: In other words, we need instrumental variables for identification in this case.

Acknowledgments

We are grateful to participants for helpful comments. We would also like to thank five *SMR* reviewers for their criticisms and suggestions.

Authors' Note

An earlier version of this article was presented to ASA Methodology Section Conference, Yale University, March 2007.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: Funding for this research was provided under the EU's 6th Framework Programme through the EQUALSOC Network of Excellence.

Notes

1. As we have introduced it here, this model permits the scaling of thresholds over latent variables but it has obvious applications that would involve scaling over observed covariates. To model temporal convergence or polarization in attitudes, say, we could set τ to be a function of period or cohort. This would be a parsimonious alternative to allowing the thresholds to vary freely over period or cohort.
2. There is a large literature on random effects models for data with repeated measurements, such as panel or hierarchical data. Hedeker and Gibbons (1994) presented the first general random effects model for ordinal regression models.
3. Information on both National Child Development Study (NCDS) and on the later British Cohort Study 1970 surveys can be found at http://www.cls.ioe.ac.uk/. The educational data used in this article come from the age 23 sweep of NCDS.
4. Applying the same latent class model to the data shown in Table 1 returns a deviance of 25.38 on 22 *df* ($p = .28$), while a three-class model has deviance $= 24.67$ on 20 *df*.
5. The data were kindly made available to us by the UK's Economic and Social Data Service based at the University of Essex available at http://www.esds.ac.uk/findingData/snDescription.asp?sn = 2197.
6. Details of this model are available from the authors on request.
7. A test of the hypothesis that the latent class structures of the 1983 and 1984 data are the same cannot be rejected: The test has a deviance of 1.3 on 3 *df*.
8. Differences between countries are captured in the country-specific threshold parameters while gender differences (which are country-specific) are captured by the parameters for women. Differences between men in different countries are thus absorbed into the thresholds, while country differences between women depend on both the thresholds and the parameters for the woman dummy variable.
9. In fact, there is some evidence that this is so, with the distribution differing in the United States from that in France, Germany, and Poland, though this difference does not reach statistical significance (using the .05 criterion).

References

Breen, Richard and John H. Goldthorpe. 2001. "Class, Mobility and Merit: The Experience of Two British Birth Cohorts." *European Sociological Review* 17:81-101.

Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford, UK: Clarendon.
Douglas, J. W. B. 1967. *The Home and the School*. 2nd ed. London: Panther Books.
Greene, William H. 2002. *LIMDEP 8.0*. New York: Econometric Software Inc.
Heath, Anthony F. and Kenneth McDonald. 1987 ''Social Change and the Future of the Left.'' *Political Quarterly* 53:364-77.
Hedeker, Donald and Robert D. Gibbons 1994. ''A Random-Effects Ordinal Regression Model for Multilevel Analysis.'' *Biometrics* 50:933-44.
Holm, Anders, Mads Meier Jaeger, and Morten Pedersen. 2009. ''Unobserved Heterogeneity in the Binary Logit Model With Cross-Sectional data and Short Panels: A Finite Mixture Approach.'' Unpublished manuscript.
Lindsay, Bruce G. 1983a. ''The Geometry of Mixture Likelihoods: A General Theory.'' *Annals of Statistics* 11:86-94.
Lindsay, Bruce G. 1983b. ''The Geometry of Mixture Likelihoods, Part II: The Exponential Family.'' *Annals of Statistics* 11:783-92.
McCullagh, Peter. 1980. ''Regression Models for Ordinal Data.'' *Journal of the Royal Statistical Society, Series B* 422:109-42.
Vermunt, Jeroen K. 1997. *LEM: Log-linear and Event History Analysis With Missing Data*. Tilburg, the Netherlands: Tilburg University.

Bios

Richard Breen is Professor of Sociology and Co-Director of the Center for Research on Inequalities and the Life Course at Yale University. He is also a Senior Research Fellow of Nuffield College, Oxford. His research interests are social stratification and inequality, the application of formal models in the social sciences, and quantitative methods. In 2009 he published ''Non-Persistent Inequality in Educational Attainment: Evidence From Eight European Countries'' (with Ruud Luijkx, Walter Müller, and Reinhard Pollak) in the *American Journal of Sociology.*

Ruud Luijkx is an Associate Professor of Sociology at Tilburg University (www.tilburguniversity.nl/webwijs/show/?uid=r.luijkx). His research focuses on the comparative research of educational outcomes, social mobility, and life courses. He is currently working on the European Values Study 2008.