

## Chapter 21

### Latent Trait and Latent Class Models Applied to Survey Data

*David J. Bartholomew, Lilian M. de Menezes and Panagiota Tzamourani*

London School of Economics and Political Science

#### 1. Introduction

Latent trait models were developed for use in educational testing where they are widely used and are the subject of an enormous literature. Applications in social research have been relatively rare, but some recent work will be found in Knott, Albanese and Galbraith (1991), Bartholomew and Schuessler (1991) and Leung (1992). Latent class models, on the other hand, were introduced by Lazarsfeld (1950) in a sociological context and have continued to find fruitful applications in that field. Some of the principal developments will be found in Goodman (1979), Clogg (1993) and Langeheine and Rost (1988).

Both kinds of model have a common theoretical structure and their usefulness turns on whether the associations between categorical variables can be explained by their common dependence on underlying latent variables which may be continuous or categorical.

The present paper reports work on a project, supported by the U.K. Economic and Social Research Council (ESRC), to develop methods for the analysis of large social data sets where the use of models involving latent variables is called for. Extensive analyses are being carried out on a wide variety of data sets - mainly obtained from large social surveys. The aim is to see how far existing models and software are adequate for the purpose and hence to identify areas for methodological innovation. The dialogue between model and data works in two directions. On the one hand, we may wish to generalize existing models by incorporating such things as the effect of background variables and additional latent variables. On the other, the simpler models may identify deficiencies in the wording or administration of questions. Both aspects are covered in what follows. The main innovation on the methodological side is the use of models allowing for two latent variables and different ways of coping with missing values.

The two data sets on which we shall draw relate to attitudes of individuals. The first set, from Schuessler (1982), was used by him to construct attitude scales without the benefit of the models used in this paper. We have re-analyzed the data using new methodology. The other has been extracted from a large survey for use in this study and has not hitherto been analyzed by latent variable methods.

#### 2. The Data Sets

The first set on *social life feelings* (SLF) is described in Schuessler (1982). A replicate study carried out in Germany is reported in Krebs and Schuessler (1987), but here we use the American data from the first study. The data were derived from a survey of 1500 individuals. The questionnaire included 237 items (questions) on what the author calls social life feelings

and 31 social background items. The questions used in our analyses are given in section 4. Most items required a binary response (yes/no or agree/disagree) but some called for ordered categorical responses. In the data available to us, all had been reduced to binary responses which were coded 0,1. Just over 3% of responses were missing in the original data set. These missing values had been replaced by responses drawn with probabilities proportional to the marginal frequencies on the variable in question. If the responses were missing at random, this procedure would preserve the correct one-way marginal distributions, but would reduce the magnitude of any association between variables. Since only 3% of responses were missing any resulting biases are likely to be very small. Alternative methods of dealing with missing values have been used in the other data set.

The second set is derived from the British Social Attitudes (BSA) - 1990 Survey. This is the seventh survey of a series started in 1983 and carried out annually with the exception of 1988. It covers a wide range of attitudes including such things as party politics, the welfare state, healthcare, crime and such like. The 1990 survey had two components: a face-to-face interview lasting a little over an hour and a self-completion supplement to fill in after the interview. The standard questions cover major topic areas such as defence, the economy, labour market participation and the welfare state together with a wide range of background variables. The remainder of the questionnaire is devoted to specific issues which vary slightly from year to year. The sampling procedure uses a multi-stage design and aims at a representative sample of adults (18 and over) living in Great Britain. The sample sizes vary with the topic covered. More details may be found in Brook, Taylor and Prior (1990).

### 3. The Models

The models we have used are, for the most part, familiar and were given in Bartholomew (1987). The extension to missing values is given in Knott and Tzamourani (1996), and a more detailed description of the methods in the TWOMISS Manual (Albanese and Knott, 1990). Here we give a brief explanation of the additional features, which is sufficient to establish what is needed for the analyses which follow.

In the latent variable (trait) model used here, the probability of a positive response is supposed to depend on two continuous latent variables  $\boldsymbol{\theta} = (\theta_1, \theta_2)$ . This probability for the  $i^{\text{th}}$  item is denoted by  $\pi_i(\boldsymbol{\theta})$  and is known as the response function. If the associations among the responses,  $x$ 's, are wholly accounted for by their common dependence on  $\boldsymbol{\theta}$ , then, conditional on  $\boldsymbol{\theta}$ , they will be independent. Given  $\boldsymbol{\theta}$ , the item response,  $x_i$ , is therefore a Bernoulli random variable with probability function

$$p(x_i) = \{\pi_i(\boldsymbol{\theta})\}^{x_i} \{1 - \pi_i(\boldsymbol{\theta})\}^{1-x_i} \quad (x_i = 0 \text{ or } 1; i = 1, 2, \dots, k).$$

The remaining elements in the model are the specification of  $\pi_i(\boldsymbol{\theta})$  and the assumption of a prior distribution for  $\boldsymbol{\theta}$ . In the former case we use the logit function, for reasons given in Bartholomew (1987), writing

$$\text{logit } \pi_i(\boldsymbol{\theta}) = \sigma_i + \beta_{i1} \theta_1 + \beta_{i2} \theta_2 \quad (i = 1, 2, \dots, k) \quad (1)$$

(cf. equation (28) in chapter 1).

The parameter  $\sigma_i$  measures the „extremeness“ of the item; increasing/decreasing its value moves the response probability towards 1 or 0. In our analyses we sometimes find it more informative to use

$$\pi_i = \frac{1}{(1 + e^{-\sigma_i})} = p(x_i = 1 | \boldsymbol{\theta} = 0). \quad (2)$$

This is the probability that an individual at the median point of both latent scales gives a positive response. The coefficients  $\beta_{i1}$  and  $\beta_{i2}$  play the role of „factor loadings“.

An alternative parameterization which makes the link with factor analysis more transparent is

$$\beta_{i1}^* = \beta_{i1} / A, \quad \beta_{i2}^* = \beta_{i2} / A, \quad A = (1 + \beta_{i1}^2 + \beta_{i2}^2)^{1/2}. \quad (3)$$

The rationale for this parameterization is given in Bartholomew (1987). It depends on the fiction that the  $x$ 's have been produced by dichotomizing continuous underlying normally distributed variables. That establishes a correspondence with the linear factor model and enables us to interpret the  $\beta^*$ 's as correlations between the underlying variable and the latent variables.

In the case of a two-factor model there is an indeterminacy about the solution. The axes of the latent space may be rotated in exactly the same manner as in the normal linear factor model, and for the same reason. Factor analysis programs overcome this problem by including a constraint designed to yield a unique solution: other equivalent solutions are then obtained by rotation. The TWOMISS program does not impose any constraints and the solution reached will depend, in general, on the initial values used to begin the iteration. When there are only two latent variables, the estimated loadings can be plotted and an appropriate rotation selected by eye. We have used this approach in some of our examples, but we have also developed a routine for finding a canonical solution for later incorporation into our computer program.

The  $\theta$ 's are conventionally assumed to be independent standard normal variables. It is immediately clear from (1) that changes in the origin and scale of the  $\theta$ 's can be absorbed in the  $\beta$ 's, and so the choice of unit variance and zero mean involves no loss of generality. The form of the prior distribution is known to be relatively unimportant and so the choice of the normal form is essentially a matter of convenience.

Various methods are available to deal with missing responses. If they are relatively few in number, individuals with any missing response can be omitted from the analysis. However, the TWOMISS program which we use provides three alternative methods of dealing with the situation of which two are used here. The first treats the omission of a response as a random phenomenon and obtains estimates by maximizing the likelihood with the terms which would arise with the missing values left out. The second method allows information on the latent (attitude) variable to be inferred from the pattern of non-response. For each item an „indicator item“, which is equal to one if the corresponding item is known and zero otherwise, is generated. Two factors are considered: attitude and expression of opinion. The coefficients of each indicator item on the expression factor are constrained to be zero in the latent (attitude) variable, and thus we estimate the following:

$$\begin{aligned}\text{logit}\{\pi_{a_i}(\boldsymbol{\theta})\} &= \sigma_i + \beta_{i1} \theta_a, & i = 1, \dots, k, \\ \text{logit}\{\pi_{e_i}(\boldsymbol{\theta})\} &= \sigma_i + \beta_{i1} \theta_a + \beta_{i2} \theta_e, & i = k + 1, \dots, 2k,\end{aligned}$$

where  $\theta_a$  is the latent variable in question and  $\theta_e$  is the expression of opinion factor. The coefficients  $\beta_{i1}$  ( $i = k + 1, \dots, 2k$ ) give information on how attitude affects opinion.

We have used three methods of judging the goodness of fit of the models. The first is to compare the observed and expected frequencies for the  $2^k$  response patterns using the log likelihood ratio statistic ( $G^2$ ). As the number of manifest variables gets larger, the expected frequencies become smaller and a point is soon reached where response patterns have to be grouped in order to have large enough expected frequencies to justify the  $\chi^2$  approximation to the distribution of the test criterion. This quickly eliminates the degrees of freedom making no test possible.

Our second method compares the observed and expected two-way margins. That is, for each pair of items ( $i, j$ ) we compare the number of responses of the form (1,1) say with the expected. We do this for each observed (O) and expected (E) response frequency by computing  $(O-E)^2/E$ . These cannot be summed to give a valid chi-squared criterion, but they have proved useful as a means of detecting the source of poor fits. Even if the overall fit is poor, a model capable of successfully predicting the two-way margins captures an important part of the correlation structure of the data. Having looked at the two-way margins, we then inspect the three-way margins in the same fashion.

The third method is also informal and involves looking at the reduction achieved in the value of the log-likelihood by fitting a chosen model. If we denote by  $H_0$  the hypothesis that the variables are independent, and by  $H_1$  the latent variable model, and let  $G^2$  be the log-likelihood goodness of fit statistic, then

$$\{G^2(H_0) - G^2(H_1)\} / G^2(H_0)$$

may be regarded as a measure of the extent, to which the original departure from independence is accounted for by the proposed model.

We have also fitted latent class models to the data. The latent class model can be linked with the latent trait model as follows. We start with the latent trait specification of a one-dimensional latent variable with an arbitrary continuous prior distribution. Then we imagine the continuum divided into  $q$  mutually exclusive and exhaustive intervals. If the  $g$ th latent class is now defined to consist of those members of the population whose latent position is in the  $g$ th interval then  $\pi_g$  will be the area under the prior density in the  $g$ th interval (cf. equation (37) in chapter 1). The link is completed by supposing that the response function is constant within each interval and this constant level is then the conditional probability that a member of the  $g$ th class gives a positive response to the item.

By this means we can see that the latent class model is a latent trait model with an item response curve in the form of a step-function. Since any continuous curve can be approximated by a step function we might anticipate that when a latent trait model fits so will a latent class model.

#### 4. Social Life Feelings Data

By a variety of methods, Schuessler (1982) was able to select sub-sets of items, each of around a dozen items, which he believed could be used to construct scales for a number of social life feelings. Our objective here is to see how successful this can be assumed to have been when judged against the models of the last section. We present only a small part of the analyses carried out. They have been chosen to illustrate general points which have emerged.

We consider the items used to form two scales: SLF2 (items 1 to 8), *Doubt about trustworthiness of people* and SLF7 (items 7 to 12), *People cynicism*.

1. *It is hard to figure out who you can really trust these days.*
2. *There are few people in this world you can trust, when you get right down to it.*
3. *Most people can be trusted.*
4. *Strangers can generally be trusted.*
5. *Most people are fair in their dealings with others.*
6. *Most people don't really care what happens to the next fellow.*
7. *Too many people in our society are just out for themselves and don't really care for anyone else.*
8. *Many people are friendly only because they want something from you.*
9. *In a society where almost everyone is out for himself, people soon come to distrust each other.*
10. *Most people know what to do with their lives.*
11. *Many people in our society are lonely and unrelated to their fellow human beings.*
12. *Many people don't know what to do with their lives.*

Two items (numbers 7 and 8) were common to both scales. We chose this pair of scales because they are rather similar and we wished to see whether the selection made by Schuessler would emerge from our analysis. Specifically, we have aimed to answer the following questions.

- a) Can one successfully fit a one-variable latent variable model to the response data for the sets of items on each scale?
- b) If we consider all 12 items, does a two-variable model fit, and does it identify the same items as did Schuessler as belonging to each scale?
- c) Does a latent class model provide a better fit?

If the answers to (a) and (b) are in the negative, we shall wish to see whether the analysis shows how to obtain better scales.

The parameter estimates for the model with one latent variable are given in Tables 1 and 2. In both cases the estimates are very much what one would expect from a scale with a good range of values of  $\sigma_i$  and  $\beta_{i1}$ . However, the fit as measured by the likelihood ratio statistic based on comparing the observed and expected frequencies for the response patterns is very poor in both cases ( $G^2 = 285.36$  on 71 df. for SLF2 and  $G^2 = 102.1$  on 33 df. for SLF7). This might be because more than one latent variable is needed, but when we fit the model with two latent variables the situation is not greatly improved. We therefore examine the fit of the one variable model in more detail.

Item	$\sigma_i$	$\pi_i$	$\beta_{i1}$		$\beta_{i1}^*$
1	0.90 (.10)	.71	1.89	(.14)	.88
2	0.10 (.09)	.52	2.03	(.15)	.90
3	-1.12 (.11)	.25	2.39	(.18)	.92
4	0.75 (.07)	.68	1.21	(.10)	.77
5	-1.72 (.13)	.15	2.04	(.16)	.90
6	-0.13 (.09)	.47	1.91	(.14)	.89
7	0.92 (.11)	.71	2.26	(.17)	.91
8	-0.15 (.07)	.46	1.43	(.11)	.82

**Table 1.:** SLF2: Parameter estimates for the one-factor model, standard errors in brackets

Item	$\sigma_i$	$\pi_i$	$\beta_{i1}$		$\beta_{i1}^*$
7	0.83 (.10)	.70	1.90	(.19)	.88
8	-0.16 (.08)	.46	1.50	(.15)	.83
9	2.19 (.14)	.90	1.38	(.15)	.81
10	-0.03 (.06)	.49	0.77	(.09)	.61
11	1.73 (.11)	.85	1.27	(.14)	.79
12	1.81 (.12)	.86	1.46	(.15)	.82

**Table 2:** SLF7: Parameter estimates for the one-factor model, standard errors in brackets

We compare the observed and expected frequencies for the two-way margins, as previously described. In particular, we focus on responses coded (1,1) to each pair of items. In the case of SLF2, the largest discrepancies correspond to the pairs (2,1) and (5,3) with a further, less serious deviation in (6,7). In all these cases, the numbers giving positive answers to both questions are higher than would be expected from the common dependence on the latent variable alone. The same situation arises in the case of scale SLF7 and the pair (10,12). This suggests that there is some other reason why these pairs of items are correlated. When we examine the questions themselves, a possible reason is apparent. In SLF7, questions 10 and 12 are:

(10) *Most people know what to do with their lives.*

(12) *Many people don't know what to do with their lives.*

These are essentially the same question except for the qualifiers „many“ and „most“. It may well be that respondents remember the answer which they gave to the first, when they answer the second, and are influenced by a perceived need to be consistent when they answer the second.

A similar situation may be observed in scale SLF2. Questions 1 and 2 are:

(1) *It is hard to figure out who you can really trust these days.*

(2) *There are few people in this world you can trust, when you get right down to it.*

These are sufficiently similar to suggest that there is a carry-over effect from one to the other. Questions (3) and (5) are not so close, but both begin with the words „most people“. Question 6, which is also implicated in a fairly large deviation, also begins with these words. The hint that questions beginning with „most people“ may have a tendency to obtain assent regardless of the respondents latent position finds some support from inspection of the three-way margins, where it appears that there is an excess of respondents giving positive answers

to all three questions. However, this is not the most significant deviation and we shall return to the matter below.

These observations lead us to suspect that the latent variable model fails because of the inadequacy of the conditional independence assumption due to a separate source of correlation between certain pairs of items. If this is so, the fit ought to be improved by deleting one member of each pair from the set and re-fitting the model. This has been done by omitting the member of the pair whose parameter  $\pi_i$  differed most from 0.5. A summary of the results is given in Tables 3 and 4.

The parameter estimates are very similar to those obtained in Tables 1 and 2, but the fit is now much improved. For Table 3, we have  $G^2 = 79.9$  on 39 df. and, for Table 4,  $G^2 = 20.74$  on 17 df. The latter denotes a very good fit; the former is less so, though all of the two-way and three-way margins are well fitted by the model.

<i>Item</i>	$\sigma_i$	$\pi_i$	$\beta_{i1}$	$\beta_{i1}^*$
2	0.09 (.08)	.52	1.80 (.14)	.88
3	-1.04 (.11)	.26	2.15 (.18)	.91
4	0.73 (.07)	.67	1.12 (.10)	.75
6	-0.14 (.09)	.47	2.18 (.17)	.91
7	1.00 (.12)	.73	2.55 (.23)	.93
8	-0.15 (.07)	.46	1.45 (.11)	.82

**Table 3:** SLF2: Parameter estimates for the one-factor model with items 1 and 5 omitted, standard errors in brackets

<i>Item</i>	$\sigma_i$	$\pi_i$	$\beta_{i1}$	$\beta_{i1}^*$
7	0.93 (.13)	.72	2.29 (.30)	.92
8	-0.16 (.08)	.46	1.64 (.18)	.85
9	2.18 (.13)	.90	1.36 (.15)	.81
10	-0.30 (.06)	.49	0.61 (.08)	.52
11	1.64 (.10)	.84	1.10 (.13)	.74

**Table 4:** SLF7: Parameter estimates for the one-factor model with item 12 omitted, standard errors in brackets

Our next step is to see whether a two-factor model fitted to the combined set of items would separate them into the two groups of Schuessler's scale. Since it is possible that this might resolve the problem of fit revealed by the one-factor models, we did not, at this stage, delete any items except for the duplicates which appeared in both scales. The parameter estimates are given in Table 5, where the questions are ordered so that the overlapping items occur in the middle of the list.

As in factor analysis, the solution can be rotated without changing the fit, but the situation is clear from the unrotated solution. The questions in SLF7 which are *not* common to the two scales (i.e., 9,10,11,12) seem to go together, with the position of the second overlapping question(8) being less clear. This suggests that, with the possible exception of 8, the items of SLF2 form a scale, but that SLF7 should exclude the common items.

<i>Item</i>	$\sigma_i$	$\pi_i$	$\beta_{i1}$	$\beta_{i2}$	$\beta_{i1}^*$	$\beta_{i2}^*$
1	0.90 (.07)	0.71	1.01 (.11)	1.62 (.11)	0.47	0.75
2	0.10 (.07)	0.53	1.17 (.13)	1.72 (.12)	0.51	0.74
3	-1.13 (.08)	0.24	1.22 (.14)	2.12 (.14)	0.46	0.80
4	0.76 (.07)	0.68	0.76 (.10)	1.01 (.09)	0.47	0.63
5	-1.74 (.08)	0.15	1.05 (.13)	1.79 (.11)	0.45	0.78
6	-0.13 (.07)	0.47	0.66 (.11)	1.81 (.13)	0.30	0.83
7	0.99 (.08)	0.73	0.59 (.12)	2.41 (.15)	0.22	0.90
8	-0.15 (.07)	0.46	0.33 (.10)	1.48 (.11)	0.17	0.82
9	2.11 (.08)	0.89	-0.08 (.11)	1.26 (.09)	-0.05	0.78
10	-0.03 (.06)	0.49	-0.36 (.09)	0.92 (.09)	-0.25	0.65
11	1.68 (.08)	0.84	-0.47 (.10)	1.08 (.09)	-0.31	0.70
12	3.01 (.12)	0.95	-2.13 (.16)	2.50 (.16)	-0.62	0.73

**Table 5:** Combined SLF2 and SLF7: Parameter estimates for a two-factor model, standard errors in brackets

However, this should not be pressed too far because the overall fit of the model is poor ( $G^2 = 1162.96$  on 83 df.) and inspection of the observed and fitted margins of low order show some marked discrepancies. These include the questions which caused problems in the individual fits. The most striking difference is the large excess of observed over expected frequency for the group of questions 3,5 and 10 ( $(O - E)^2/E = 14.9$ ). These are three of the four questions beginning „most people“. This tends to confirm our conjecture that question wording issues are vital if attitude scales are to be successfully constructed from questionnaire items.

We have attempted further fits of the two-factor model omitting the items 1, 5 and 12 in various combinations. The percentage of  $G^2$  explained increased from 61% with all variables to 78% when 1, 5 and 12 were all omitted but, in every case, the overall fit was poor. The percentages compare with one-factor fits of 94% for SLF2, omitting items 1 and 5, and 95% for SLF7, excluding item 12.

It is interesting to compare the foregoing results with those for latent class models. These have been fitted to SLF2 and SLF7. If we fit a succession of models, increasing the number of classes by one at each step, we usually find that the results conform to a hierarchical structure with later classes in the sequence being formed, largely, by sub-division of one or more classes in the previous level.

A two-class solution will usually consist of one group of „high“ responders and one of „low“, and subsequently these divide to reveal more subtle groupings. In the present case, this occurs with both scales; items which we have noted as being more highly correlated than one would expect from a one-factor model tend to go together. An example is given in Table 6.

	<i>Two Classes</i>		<i>Three Classes</i>			<i>Four Classes</i>			
Item	1	2	1	2	3	1	2	3	4
1	.38	.90	.21	.76	.94	.22	.50	.91	.97
2	.22	.82	.08	.59	.90	.09	.37	.72	.93
3	.07	.64	.04	.25	.89	.03	.55	.12	.88
4	.45	.83	.33	.70	.90	.34	.79	.67	.90
5	.04	.48	.02	.16	.70	.01	.43	.05	.69
6	.18	.78	.09	.50	.89	.10	.60	.48	.88
7	.34	.92	.18	.73	.97	.21	.70	.75	.97
8	.23	.72	.18	.44	.87	.19	.36	.48	.87
$\pi_g$	.50	.50	.31	.42	.27	.33	.15	.26	.26
	$G^2 = 652.6$ (238 df.)		$G^2 = 395.5$ (229 df.)			$G^2 = 307.4$ (220 df.)			

**Table 6:** SLF2: Probabilities of individuals in class g giving positive responses to item i

If all the items are pooled, the four-class model is readily interpretable and the parameter estimates are given in Table 7.

	<i>Four-Class Model</i>			
Item	1	2	3	4
1	0.35	0.16	0.78	0.94
2	0.19	0.06	0.61	0.90
3	0.07	0.04	<b>0.26</b>	0.91
4	0.42	0.33	0.70	0.91
5	0.04	0.02	<b>0.17</b>	0.72
6	0.08	0.13	0.54	0.89
7	0.13	0.25	0.77	0.97
8	0.09	0.25	0.48	0.87
9	0.58	<b>0.75</b>	0.90	0.97
10	0.10	<b>0.59</b>	0.44	0.75
11	0.44	<b>0.81</b>	0.83	0.92
12	0.24	<b>0.97</b>	0.83	0.93
$\pi_g$	0.15	0.19	0.41	0.25
	$G^2 = 1880.97$ (4044 df.)			

**Table 7:** SLF2 and SLF7 (items 1 to 12): Probabilities of individuals in class g giving positive responses to item i.

The first group, comprising 15% of the population, has a relatively high probability of responding „no“ on all items. Similarly, the last group (25%) has a high probability of positive answer on all items. The two intermediate classes are, perhaps, the most revealing. One contains high probabilities on the last four questions only, these are the four which are unique to SLF7. The remaining class is a somewhat weaker version of the high probability class, except for having a low positive response probability on questions 3 and 5 of SLF2. Again, this links these two questions, but we find the latent class analysis less illuminating for this example than the latent trait model. This is what one would hope with items designed to measure a continuous scale.

## 5. British Social Attitudes

Unlike the social life feelings scales, the BSA questions were not designed to construct scales, but rather to simply record the pattern of responses and their changes over time. Since, however, there are groups of questions on particular topics, they do lend themselves to multivariate analysis. It is therefore possible to explore the dimensionality of the various attitude spaces, and latent variable models provide one way of doing this.

We have investigated five topics relating to the countryside, the environment, the National Health Service, internationalization and industrial relations. Here we focus on the National Health Service, but we shall briefly allude to other topics.

The following question was addressed to a sample of 2797 individuals:

*From your own experience, or from what you have heard, please say how satisfied or dissatisfied you are with the way in which each of these parts of the National Health Service runs nowadays*

- i) local doctors or GPs?*
- ii) National Health Service dentists?*
- iii) health visitors?*
- iv) district nurses?*
- v) being in hospital as an in patient?*
- vi) attending hospital as an out patient?*

Responses were within the scale 1 (very satisfied) to 5 (very dissatisfied). These were recoded to binary: 1 (satisfied) and 0 (dissatisfied). Non-response was recorded: 19% of respondents had at least one item missing.

It should be noted that there is a natural pairing in the questions with doctors and dentists, health visitors and district nurses, and hospital services being groups on which respondents might be expected to have similar views. Overall there was greatest satisfaction with doctors and dentists (80% and 69% respectively). For the other categories the percentages varied between 43% and 63%. Here, of course, we are interested in the way responses to the items are linked, and particularly, in how far their correlations can be explained by latent variables or classes.

One might expect varying degrees of satisfaction with the health services in general, but a one-factor model with non-response assumed to be random proved to be a poor fit. The exclusion of cases with at least one missing value gave rise to a better fit so other factors are clearly at work. One possibility which we investigated was that non-response was not random, for this we used the model which allows the probability of non-response to depend both on a „tendency to non-response“ and the main factor - here presumed to be general satisfaction with the health service. The resulting fit was much better with 85% of  $G^2$  explained, though goodness of fit as measured by the likelihood ratio statistic was still poor ( $G^2 = 730.3$  on 60 df.) There was a particularly large excess of observed over expected frequencies for the two-way margins of variables 1 and 2 (doctors and dentists) and 5 and 6 (hospital services). This supports our expectation that the natural linking of these pairs might account for part of the correlation between them. The parameter estimates are given in Table 8.

The values of  $\beta_{i1}$  for the first six items in the table suggest a common factor to which all variables contribute positively with items 3 and 4 carrying the most weight. This is consistent with the existence of a general satisfaction factor reflecting mainly those functions which involve home visits. Of particular interest are the loadings of the last 6 items which are concerned with whether a person responded or not. These coefficients are all negative and this implies that those who are most likely to respond are positioned low on the satisfaction scale. This shows why it would be inappropriate to treat non-response as a random phenomenon in this case. The large positive coefficients in the  $\beta_{i2}$  column confirm that the second factor does, indeed, reflect the variation in tendency to respond.

Item	$\sigma_i$	$\pi_i$	$\beta_{i1}$	$\beta_{i2}$	$\beta_{i1}^*$	$\beta_{i2}^*$
1	1.63 (.05)	.84	.82 (.06)	0	0.64	0
2	.99 (.04)	.73	.67 (.06)	0	0.55	0
3	.27 (.08)	.57	3.22 (*)	0	0.96	0
4	1.19 (.08)	.77	3.66 (*)	0	0.96	0
5	.80 (.05)	.69	1.05 (.07)	0	0.72	0
6	.14 (.04)	.54	.87 (.07)	0	0.67	0
7	14.67 (*)	1.00	-1.54 (.95)	4.86 (*)	-0.30	0.94
8	5.04 (*)	.99	-.83 (.15)	1.84 (.09)	-0.37	0.82
9	5.83 (*)	1.00	-1.47 (.12)	5.19 (*)	-0.27	0.95
10	6.56 (*)	1.00	-1.11 (.16)	5.91 (*)	-0.18	0.97
11	5.45 (*)	1.00	-.62 (.15)	2.29 (.09)	-0.24	0.89
12	6.18 (*)	1.00	-1.01 (.17)	2.46 (.11)	-0.36	0.87

**Table 8:** NHS: Parameter estimates and standard errors (an asterisk indicates a value too large to be estimated accurately) for the main factor plus non-response factor model.

Given the room for improvement in fit, it would be interesting to add a further factor, but our program is not yet able to do this. We can, however, fit a two-factor model excluding those who failed to respond on any item. This two-factor model provides a better fit in terms of the comparisons between observed and expected frequencies, with 93% of  $G^2$  explained and likelihood ratio statistic equal to 128.5 on 33 degrees of freedom. The reason that the fit is not even better appears to be the persistence of the tendency to give more positive responses to items 1 and 2 than the model predicts. Nevertheless, the interpretation is illuminating and the parameter estimates are given in Table 9.

Item	$\sigma_i$	$\pi_i$	$\beta_{i1}$	$\beta_{i2}$	$\beta_{i1}^*$	$\beta_{i2}^*$
1	1.69 (.06)	.84	0.74 (.08)	0.60 (.07)	.54	.43
2	0.95 (.05)	.72	0.41 (.07)	0.54 (.06)	.34	.45
3	-3.33 (.51)	.03	1.51 (.35)	13.94 (.94)	.11	.99
4	0.50 (.06)	.62	0.41 (.08)	2.33 (.11)	.16	.91
5	0.99 (.07)	.73	2.21 (.15)	0.93 (.09)	.85	.36
6	0.12 (.07)	.53	2.32 (.20)	0.75 (.08)	.88	.29

**Table 9:** NHS: Parameter estimates and standard errors for the two-factor model with missing values excluded.

The interpretation comes out most clearly from the standardized  $\beta$ 's columns. The first factor is heavily weighted towards satisfaction with hospital services and the second with health visitors and district nurses.

Once more the first and second items (local doctors and NHS dentists) were critical in terms of fit. The comparison of observed and expected responses to this pair of items indicate very large discrepancies ( $3.79 \leq (O - E)^2 / E \leq 26.99$ ). The exclusion of one item did not lead to a better fit and only by eliminating the first two items was a small improvement in fit achieved.

Because earlier results suggested that responses were grouped as follows: (doctors, dentists), (health visitors, district nurses) and (in-,out-patient), analyses were carried out using one item from each pair. Better fits were obtained in this way. In general, two-factor models confirmed two dimensions, „hospital“ and „visiting“ services, whereas the models with missing values showed that the respondents who scored low in satisfaction were more likely to respond.

Latent class models were fitted to the subset of those who responded to every item (2272 individuals). For small numbers of classes the results were poor, but a satisfactory fit was obtained for the models with 4 classes, despite  $G^2$  being equal to 190 on 36 degrees of freedom.

Class 4, comprising 39% of the sample, shows a high degree of satisfaction with all aspects of the health service. On the contrary, Class 1 represents those who have very low overall satisfaction; these are however a much smaller group (11%). Between the extremes there are two classes, whose satisfaction is more discriminating in terms of the natural grouping of items. Class 2 members (28%) are generally satisfied with NHS doctors and dentists, but mostly dissatisfied with other services. Class 3 (21%) represents individuals who are generally satisfied with services, except for „visiting“ services.

	<i>Four-Class Model</i>			
Item	1	2	3	4
1	0.00	1.00	0.83	0.90
2	0.31	0.71	0.68	0.83
3	0.13	0.19	0.12	1.00
4	0.26	0.28	0.32	1.00
5	0.17	0.32	0.97	0.81
6	0.06	0.22	0.87	0.70
$\pi_g$	0.11	0.28	0.21	0.30

**Table 10:** NHS: Probabilities of individuals in class  $g$  giving positive responses to item  $i$ .

This solution was obtained through different sets of initial conditions and should be identifiable according to the extended condition given by Madansky (1960) and described in Goodman (1974).

This analysis excludes individuals with any missing responses, so it can only be meaningfully compared with the two-factor solution obtained for the same group. The 4 classes show a scale of satisfaction and highlight the differing roles which specific pairs of items play. This much was revealed by the latent variable model, but the latent class model goes further. However, the two-factor model is more parsimonious and increasing the number of factors would, perhaps, capture other features in the data.

## 6. Conclusions

Our analysis has provided clear indications about the answers to the questions which we raised at the outset.

First, it has not proved practically feasible to distinguish between latent class and latent trait models. It is often possible to find a model of either class to fit a given set of data, and the interpretations are usually consistent with one another. In any particular case, one or other interpretation may seem more natural, but it is clear that any attempt to impose too fine a structure on the latent space is empirically unjustified.

We have successfully obtained good fits to a variety of data-sets. Nevertheless, even in cases of poor fit, an examination of the discrepancies between the observed and expected two-way margins has been revealing, since it has not only led to possible reasons for failure but has also hinted at question wording effects. Therefore, we stress the importance of more detailed diagnostic procedures in the process of understanding survey data.

As expected, fitting latent trait models has suggested some refinements of Schuessler's scales. In addition, the results from this study indicate that satisfactory scales can be successfully constructed from general survey questions.

Finally, we have demonstrated the importance of taking account of non-response. It is evident that the probability of responding is not only likely to vary from one individual to another, but also to depend on the individual's position in the latent space.

## Acknowledgements

This work was supported by the UK Economic and Social Research Council (ESRC) - grant number HS 519255002 and the BSA data were supplied by the ESRC Data-Archive. We are grateful to Karl Schuessler for making his data-set available to us. We thank Samantha Firth for her careful typing, Martin Knott and Jane Galbraith for their valuable comments and encouragement.

## References

- Albanese, M.T. and M. Knott (1990). "Twomiss: A Computer Program for Fitting a One- or Two- Factor Logit-Probit Latent Variable Model to Binary Data When Observations May be Missing." London School of Economics and Political Science.
- Bartholomew, D.J. (1987). *Latent Variable Models and Factor Analysis*. Vol. 40. Griffin's Statistical Monographs and Courses, ed. Alan Stuart. London: Charles Griffin & Company Ltd.
- Bartholomew, D.J. and K.F. Schuessler (1991). "Reliability of Attitude Scores Based on a Latent Trait Model." *Sociological Methodology* 21: 97-123.
- Brook, L., B. Taylor, and G. Prior (1990/1991). *British Social Attitudes, Survey*. SCPR, Technical Report
- Clogg, C.C. (1993). "Latent Class Models: Recent Developments and Prospects for the Future." In *Handbook of Statistical Modeling in the Social Sciences*, ed. G. Arminger, C.C. Clogg, and M.E. Sobel. New York: Plenum.
- Goodman, L.A. (1979). "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 74: 537-552.
- Goodman, L.A. (1974). "Exploratory latent structure analysis using both identifiable and unidentifiable models", *Biometrika* 61: 215-231

- Knott, M., M.T. Albanese, and J. Galbraith (1991). "Scoring Attitudes to Abortion." *The Statistician* 40: 217-223.
- Knott, M., P. Tzamourani. "Fitting a Latent Trait Model for Missing Observations to Racial Prejudice Data.", (in this volume).
- Krebs, D. and K.F. Schuessler (1987). *Soziale Empfindungen : ein interkultureller Skalenvergleich bei Deutschen und Amerikanern*. Monographien: Sozialwissenschaftliche Methoden, Frankfurt/Main, New York: Campus Verlag.
- Langeheine, R. and J. Rost (1988). *Latent Trait and Latent Class Models*. New York London: Plenum Press.
- Lazarsfeld, P.F. (1950). "The Logical and Mathematical Foundation of Latent Structure Analysis." In *Measurement and Prediction*, ed. S.A. et al. Stouffer. New York: John Wiley & Sons.
- Leung, S.O. (1992). "Estimation and Application of Latent Variable Models in Categorical Data Analysis." *British Journal of Mathematical and Statistical Psychology* 45: 311-328.
- Madansky, A. (1960). "Determinant Methods in Latent Class Analysis", *Psychometrika* 25: 183-198
- Schuessler, K.F. (1982). *Measuring Social Life Feelings.*, The Jossey-Bass Social and Behavioral Science Series, Jossey-Bass.