Chapter 24

Nonparametric Latent Factor Analysis of Occupational Inventory Data

A. Jan Kutylowski

1. Introduction

Occupational prestige is a crucial sociological concept, and considerable effort has been devoted to its conceptualization and measurement (Wagener, 1992). Occupational inventory data that are typically used for construction of an index of occupational prestige, consist of ratings of a number of occupations thought representative with respect to possible variations in prestige (or desirability) by respondents randomly selected from a large, often national, population. Such data are formally analogous to rating data in educational testing analyzed with item response theory models, where one typically postulates a single underlying latent variable (factor) which „explains" the joint frequency distribution of ratings in the sense of conditional independence. This analogy was not pursued by sociologists, and measurement issues in the investigation of occupational inventory data have largely remained frozen at the level achieved in the early 1960s.

In section 2 we critically review traditional scoring of occupational ratings. Section 3 outlines a strategy for the analysis of ordinal multiresponse data as exemplified by occupational inventory data. It presents an application of that strategy that involves measurement of a specific aspect of occupational desirability. In particular, we demonstrate that a model developed by Ramsay (1991, 1995), implementing the strategy on the basis of kernel smoothing, closely corresponds to the notions used in traditional analysis of occupational inventories. It also gives valuable insights with respect to both the estimation of the latent factor underlying occupational ratings, and differential evaluations of the ratings. section 4 reports on application of that model to Polish occupational inventory data, followed by a brief discussion in section 5.

2. Ordinal multiresponse data and the measurement of occupational prestige

2.1 Occupational inventory data

Occupational inventory data typically involve ratings that are (i) *homogeneous* (or identical) for each response (i.e., occupation or item), and that (ii) aim at eliciting supposedly equidistant gradings of response categories. The latter feature distinguishes *Likert multiresponse data* from other instances of ordinal multiresponse data. Although (i) is often violated by sociological inventories, we shall mainly focus, for brevity of notation but without loss in generality, on homogeneous ratings.

Occupational inventory data comprise a cross-classification of k ordinal response variables (i.e., evaluations of an occupation), where item categories (*reactions*) are indexed in increasing order by c(i), $1 \le c(i) \le C(i)$. For homogeneous items C(i) = C for all i. In section

4 we shall analyse data from an inventory employed in a number of studies conducted at the *Instytut Filozofii i Socjologii* (Institute of Philosophy and Sociology), Polish Academy of Sciences: this inventory will be referred to as *IFiS Occupational Inventory* (IFiS–OI). The data are an example of homogeneous Likert multiresponse data consisting of evaluations of 29 occupations and occupational positions collected as part of a large-scale survey investigating macrosocietal structure in Poland in 1987 (Slomczynski et al., 1989). In that survey the targeted sample consisted of 2000 employed adults with age range of 20–65 for men, and 20–60 for women (men used to retire at the age of 65 , and women at 60). The number of respondents reached in the sample was 1894, 47.5% of them women. Since 9 respondents did not report any ratings, the effective sample size is N = 1885. Five evaluations concerning general esteem of each occupational title by the respondent were available: very low (c = 1), low (c = 2), medium (c = 3), high (c = 4), and very high (c = 5). Detailed list of titles can be found in the Appendix.

After discarding from the analysis respondents who gave no evaluations, $n_{c(i)}$ evaluations by N respondents are available concerning a given occupational category. Let $\mathbf{x}_v$ denote the vector of frequencies of (suitably indexed) *reaction profiles* (or patterns), i.e., combinations of the response categories of all k occupations, observed for respondent v, $1 \le v \le N$. The respective stochastic random variable $\mathbf{X}_v$ is Multinomial with index 1 and probability function $\Pr(\mathbf{x}_v)$. Since the total number of reaction profiles is typically much larger than N, the data are highly sparse. Stochastic variables $\mathbf{X}_{vi}$, representing categorical response variables that comprise an occupational inventory, can be viewed as components of the vector $\mathbf{X}_v$, $\mathbf{X}_v = (\mathbf{X}_{v1},\ldots,\mathbf{X}_{vi},\ldots,\mathbf{X}_{vk})$. Omitting index v we note that, for a given i, of distribution of response $\mathbf{X}_i$ is parametrized by vector $\boldsymbol{\pi}_i$ containing reaction probabilitites $\pi_{c(i)}, \Sigma_{c(i)=1}^{C} \pi_{c(i)} = 1$.

2.2 Some substantive aspects of occupational inventory data

Reputational method of prestige scoring (Reiss, 1961) has become something of an industry standard in empirical studies of prestige. After elimination of all „don't know“ answers, one calculates the proportions $p_{c(i)} = n_{c(i)}/N$ of those responding in a specific category of item i as

$$p_{c(i)} = \sum_{v=1}^{N} w_v y_{vc(i)}, \quad w_v = 1/N, \tag{1}$$

where $y_{vc(i)}$ is 1, when response category is c(i) for respondent v, and 0 otherwise. For homogeneous items those proportions are subsequently weighted (Reiss, 1961, p. 81) by *simple prior weights* $w^*_{c(i)}$, with the purpose of obtaining what can be described as *simple raw item* (or occupation) *scores* (**sris**)

$$s^*_i = \sum_{c(i)=1}^{C} w^*_{c(i)} p_{c(i)}, \quad w^*_{c(i)} = c(i)/C. \tag{2}$$

For the five category ratings in the IFiS–OI data, the simple prior weights are: 1/5 for „very low“, 2/5 for „low“, 3/5 for „average“, 4/5 for „high“, and 5/5 for „very high". In practice one uses simple raw percentage item scores.

Acceptance of the substantive and methodological tenets of the simple reputational scoring has not been complete. In particular, arguments were put forward in favour of a more cautious interpretation of the variable underlying **sris** as *general desirability of occupations* (Goldthorpe and Hope, 1972). If occupational ratings do not really measure „prestige“ but rather some overall „desirability“ of occupations, it is legitimate to focus on the mechanisms underlying evaluation of occupations. For such evaluation to take place, one must learn either (i) the specific attributes of each occupation, (ii) the criteria for evaluating them, or (iii) societal evaluations for each occupation itself. In either case, as noted by Reiss (1961), the knowledge respondent has is likely to vary according to characteristics related to the societal position of the person in a society. It is therefore of interest to measure not only collective evaluation of single occupations, but also respondents' evaluations of all occupations in an inventory. We propose to interpret the latter evaluations as representing *crystallization of occupational desirability* (COD). Assuming that an occupational inventory comprises a sufficiently wide range of occupations, estimated COD scores represent resondents' overall level of perception of occupations in terms of their desirability. Since the ability to discriminate societal distributions is generally more pronounced for high status respondents than for low status respondents, COD can be expected relevant with respect to making comparisons of how respondents comprehend stratification phenomena, given the established fundamental role of occupational differences in societal stratification. Measurement of COD is thus clearly of sociological interest but, as will be shown in the following, estimated COD scores can only provide a (partial) rank order of respondents.

Since **sris** may be viewed as a function of COD, estimation of COD should also be helpful in evaluating dissensus in prestige ratings. Although many sociologists basically deny importance of dissensus in occupational ratings, Sawinski and Domanski (1991) conclude that the 1987 IFiS–OI data show nonnegligible dissensus. Analysis by those authors, however, largely relies on correlation coefficients applicable to linear relations, whereas relations involving probabilities will normally be nonlinear. Since comparison of evaluations of a given occupation with respect to some stratifying variables requires that the COD variable is kept „constant“, an analysis of dissensus in occupational ratings presupposes a model for COD.

2.3 Formal interpretation of the latent factor

It follows from the above discussion that substantive research hypotheses concerning occupational inventory data should not ignore explanation of the structure of association among the responses with the help of a *latent factor* variable. An explanation of associations among the response variables is achieved if there *is conditional independence* of stochastic responses $\mathbf{X}_i$ given a metrical latent factor Θ. This implies that the joint distribution of **X** and Θ decomposes into the *response structure* distribution represented by $p(\mathbf{x}|\theta)$, and the *factor structure* distribution represented by the prior $Q(\theta)$:

$$P(\mathbf{x},\theta) = p(\mathbf{x}|\theta)Q(\theta) = \prod_{i=1}^{k} p_i(\mathbf{x}_i|\theta)Q(\theta). \qquad (3)$$

According to (3), a model is required for each single response distribution $p_i(\mathbf{x}_i|\theta)$ only. Thus we need to model vectors of probabilities $\boldsymbol{\pi}_i$, where each reaction probability $\pi_{c(i)}$ is a suitable function of, θ, $\pi_{c(i)}(\theta)$ say. We can now define the *expected item* (or occupation) *score* (**eis**) as a function of the latent factor:

$$s_i(\theta_v) = \sum_{c(i)=1}^{C} w_{c(i)}\, \pi_{c(i)}(\theta_v), \tag{4}$$

where $w_{c(i)}$ are suitable prior weights that take into account ordinality of the response variable. In particular, following arguments in Likert (1932), the prior weights can be defined as $x_{c(i)} = c(i)-1$. For a five-category item, Likert prior weights are 0, 1, 2, 3 and 4 .

The *raw item score* can be obtained as sample counterpart of (4) by ignoring the dependency of reaction probabilities on θ_v, and taking them to be equal to respective sample proportions. *Expected total* (scale, or respondent) *score* (**ets**) is defined as

$$t(\theta_v) = \sum_{i=1}^{k} s_i(\theta_v). \tag{5}$$

The variance of **eis** is approximately equal to

$$V[s_i(\theta_v)] = \sum_{c(i)=1}^{C} w_{c(i)}^2 V[\pi_{c(i)}(\theta_v)], \tag{6}$$

and similarly for the variance of **ets**. The raw total score $t_v^{\dagger}$ can be obtained by summing Likert prior weights for each item:

$$t_v^{\dagger} = \sum_{i=1}^{k} x_{vc(i)}\,. \tag{7}$$

The quantitites defined above may be seen as a generalization of the simple reputational approach, since simple prior weight $w_{c(i)}^{*}$ is a linear function of Likert weight $x_{c(i)}$. However, both the estimation of latent factor scores θ_v and of reaction probabilities $\pi_{c(i)}(\theta_v)$ is necessary, before one can make use of *estimated expected item scores* (**eeis**) and *estimated expected total scores* (**eets**).

In parametric modelling *reaction* probabilities (or *curves*) $\pi_{c(i)}(\theta_v)$ are thought to be functions of Θ that are assumed known up to the values of certain parameters, $\pi_{c(i)}(\theta_v) = h_{c(i)}(\boldsymbol{\beta}_i;\boldsymbol{\theta}_v)$. Here $\boldsymbol{\beta}_i$ stands for a vector of item parameters describing the dependency of reactions on the factor Θ. In other words, a reaction curve is defined in terms of both the response function $h_{c(i)}$ and the parameters $\boldsymbol{\beta}_i$. Factor scores for all respondents, considered as either fixed but unknown quantitites, or as values of a random variable, are contained in $\boldsymbol{\theta} = \{\theta_v\}$. Given extreme sparseness of the data, satisfactory estimation of factor scores $\boldsymbol{\theta}$, in the presence of $\boldsymbol{\beta} = \{\beta_i\}$, is quite a difficult task. Consequently, only simple forms of response functions, which impose stringent conditions on the data, have been considered in the literature. Analysis of data that do not meet those conditions results in unwarranted inferences, whereas sparseness of the data makes it difficult to develop and apply diagnostic methods that would lead to the choice of a suitable parametric model that adequately matches properties of the data. These difficulties alone justify paying attention to nonparametric modelling strategies, where instead of imposing certain functions for reaction curves, we estimate those curves from the data.

3. A strategy and a model for latent factor analysis

3.1 A nonparametric methodological strategy

Interpretation of the latent factor as a random variable has serious consequences for model specification that are often overlooked. This section outlines a strategy that explicitly recognizes those consequences. This strategy does not appear to have been previously formulated for the purpose of guiding substantive applications: although Bartholomew (1983, 1988) and Ramsay (1991, 1995) discuss the underlying principles and some implementations.

First, model specification needs to deal with the indeterminacy of the factor structure. Since Θ is intrinsically unobservable, any one-to-one order preserving transformation of whatever Θ values are used must be seen as equally valid (Bartholomew, 1983). Consequently, on the basis of the model, factor Θ can only be observed in terms of ranks. Hence a „scale“' of COD cannot yield more than (partial) rank order information about the respondents, and applications of the random factor approach should be based on quantities that are invariant with respect to monotone transformations of Θ. This implies in turn that instead of choosing specific functions $h_{c(i)}$ for the purpose of describing reaction curves, as in parametric modelling, we should instead *estimate* the values of reaction curves. Since it is a curve that is of interest, and parameters only serve to define a curve in a particular way, we may want to directly estimate reaction probabilities as functions of Θ. Efficient estimation of reaction curves $\pi_{c(i)}(\theta)$ is, therefore, a central modelling objective, in the sense that it is the starting point for interpretation of the data.

Second, there is interdependency between estimating reaction curves and estimating factor scores, since we cannot a priori expect that a simple model (e.g., a Rasch model) will be sufficient to describe response structure in satisfactory detail. The interdependency can be dealt with by alternating between estimation of the response structure and estimation of the factor structure. As factor scores are determined up to a monotone transformation, one can initially use a *reference factor* that is related in a suitable manner to the latent factor.

Given initial factor scores that represent the reference factor, the modelling strategy has the following steps: (**a**) Assuming a given distribution of factor scores, and using initial factor scores, obtain initial estimates of the factor structure. (**b**) Next obtain initial estimates of the response structure, preferably by a suitably chosen smoothing algorithm, given results obtained in the first step. (**c**) Using results obtained in (**b**), estimate latent factor scores. (**d**) Using latent factor scores resulting from (**c**), increment the alternation counter and go to (**b**), repeating the alternations in (**d**) until current response structure and factor structure do not appreciably change in comparison with the structures obtained in previous alternations. This will be referred to as ALternating REsponse and Factor structure modelling strategy (ALREF).

Any particular implementation of the strategy will be based on choices in (a), (b) and (c). The common assumption concerning (a) is that $\Theta \sim N(0,1)$, see Bartholomew (1988) for general arguments and some evidence supporting this choice. Considering (**c**), and given that response structure has been estimated in some suitable manner, one can estimate factor structure using the Bayes theorem, which gives the posterior density $q(\theta|\mathbf{x})$ as

$$q(\theta|\mathbf{x}) = Q(\theta)\, p(\mathbf{x}|\theta) / \Pr(\mathbf{x}), \tag{8}$$

where $p(\mathbf{x}|\theta)$ decomposes as in (3). Maximization of the posterior density with respect to elements of θ gives modal Bayes estimates (see e.g., Mislevy, 1986). These estimates may be interpreted as maximum likelihood (ML) estimates assuming that „true“ response structure is known.

3.2 Implementation of the strategy based on kernel smoothing

Nonparametric modelling of ordinal multiresponse data is still in its infancy. Ramsay (1995) reports on user-friendly implemenatation, in the form of the TESTGRAF95 program, of the ALREF strategy for ordinal multiresponse data based on kernel smoothing of the reaction curves, N(0,1) prior distribution of the latent factor, and modal Bayes estimation of Θ scores. See also Ramsay (1991) for a similar exposition involving non-ordinal multiresponse data. In the following we shall give a brief outline of the ALREF method based on kernel smoothing.

At alternation 0 the *initial* factor values, taken as Likert raw total scores $\{t_v^\dagger\}$, are used to preliminarily rank the respondents with respect to the latent factor. Those initial ranks are then replaced by the corresponding scores of the reference factor $\{\hat{\theta}_v^\dagger\}$ taken as quantiles of the assumed factor distribution, here N(0,1). Then, for fixed i and each c, one estimates $\{\pi_{c(i)}(\hat{\theta}_v)\}$ by kernel smoothing the relationship between the Bernoulli variates $\{y_{vc(i)}\}$ and the (current) reference factor scores $\{\hat{\theta}_v\}$. The kernel smoothing formula employed in TESTGRAF95 keeps the probabilities within the [0,1] limits by using suitably defined kernel function ker( ) and weights ω_v. It can be written as

$$\tilde{\pi}_{c(i)}\left(\hat{\theta}_{[q]v}\right) = \sum_{v=1}^{N} \omega_{[q]v}\, y_{vc(i)}, \quad \omega_{[q]v} = \frac{\ker\left[\left(\hat{\theta}_v - \hat{\theta}_{[q]}\right)/h\right]}{\sum_{v=1}^{N} \ker\left[\left(\hat{\theta}_v - \hat{\theta}_{[q]}\right)/h\right]}. \tag{9}$$

Index [q] refers here to suitably defined grouped (or binned) values of $\{\hat{\theta}_v\}$, that are introduced for the purpose of speeding up the computations, yet without significant loss in accuracy. Binning is advisable considering that in some applications N can be very large. As the result of binning, each respondent v is uniquely allocated the corresponding value of $\hat{\theta}_{[q]}$. In TESTGRAF95 the maximal value of q, $1 \le q \le Q$, is fixed at $Q = 101$. For brevity of notation, index q is omitted in the following. The bandwidth parameter h is adjustable with default value $h = 1.1\, N^{-1/5}$.

Formula (9) is similar to (1) and generally close in spirit to local averaging (Altman, 1992). Using the smoothed probabilities $\{\tilde{\pi}_{c(i)}(\hat{\theta}_v)\}$ one calculates latent factor scores as modal Bayes estimates. After the alternation counter is incremented, respondents are ranked again using the newly obtained factor scores, and the smoothing procedure (9) is repeated. In most cases, repetition of the procedure stabilizes the estimates of both reaction curves and factor scores after a small number of alternations. The smoothing procedure is not iterative, and the estimated reaction curves, the result of applying (9) at the last alternation, are not ML estimates of the response structure. However, factor scores can be viewed as ML estimates conditioned on the assumption that smoothed reaction curves are „true“ reaction curves.

Given the estimates of latent scores $\{\hat{\theta}_v\}$, and replacing $\pi_{c(i)}(\theta_v)$ by $\tilde{\pi}_{c(i)}(\hat{\theta}_v)$, we obtain estimated expected item scores (**eeis**) from (4), and estimated expected total scores (**eets**) from (5). At this stage, the crucial difference between dealing with ordinal responses and nominal responses lies in the use of Likert prior weights $x_{c(i)}$ in the former case. In the latter case, the weights are binary and distinguish only between „correct“ reaction and the incorrect ones.

Since the smoothed reaction probabilities are linear functions of the data, it is straightforward to obtain approximate standard errors of $\{\tilde{\pi}_{c(i)}(\hat{\theta}_v)\}$. These can be further used in obtaining approximate pointwise confidence intervals for estimated expected item scores based on (6).

Smoothed reaction curves $\{\tilde{\pi}_{c(i)}(\hat{\theta}_v)\}$ can now be plotted against estimated expected total scores (**eets**) for the purpose of displaying response structure for each i (Figure 1). In these displays **eets**, which is usually a monotone function of the standard Normal quantiles used in the smoothing procedure (9), is treated as a display factor representing the latent variable. Other natural choices for the display factor are standard Normal quantiles (**snq**), and Likert raw total scores (7). Similarly, we can plot estimated expected item scores (**eeis**) (4) against a display factor, for the purpose of assessing the relation between aggregate scores for a given item and the latent variable. Information concerning the precision with which expected scale scores are estimated (Figure 1) is provided by the use of approximate pointwise confidence limits for each point $\hat{\theta}_v$ of the display factor.

4. Analysis of IFiS Occupational Inventory Data

4.1 Interpretation of response structure

This section discusses some selected aspects of interpretation of the results from an application of TESTGRAF95-based ALREF strategy to the IFiS–OI data. A discussion comprising full substantive interpretation will be presented elsewhere.

Main results of the analysis are represented by 29·5 = 145 plots of *reaction curves*, and 29 related plots of *item score curves.* The latter show estimated expected item scores (**eeis**) versus estimated expected total scores (**eets**). Reaction curves are referred to in the text using $c_{\{i\}}$ as an index for response categories (index i can be omitted in graphs), so that $1_{\{2\}}$ means „very low esteem“ for the miner, see also Appendix. Figure 1 shows plots of $\tilde{\pi}_{c(i)}(\hat{\theta}_v)$ against the display factor **etss**, together with corresponding plots of **eeis** against **eets** for three selected occupations: (*technician*) *electrician* {23}, *miner* {2}, and (*defence*) *lawyer* {26}. In Figure 1, all plots (based on 10 alternations) use Q = 101 points. In comparison to the TESTGRAF95 default value of the bandwidth parameter h = 0.24, the size of h has been raised to h = 0.40 for the purpose of avoiding certain oscillations in reaction curves that are overpronounced for lower values of h. Those oscillations do not seem to affect the overall shape of the curves. Vertical dashes indicate respectively 5%, 25%, 50%, 75% and 95% percentiles of the distribution of COD as represented by the display factor **eets**. Hence, at least 50% of the respondents viewed the miner as occupation with „high“ or „very high“

desirability, and over 90% of the respondents gave the miner an **eeis** of at least 3 , roughly corresponding to „high esteem".

In the item score plots, the cross-hatching indicates approximate pointwise 95% confidence limits on the position of the **eeis**, which informs how well such a position is defined. These positions are generally poorly defined for very low and very high COD scores, as few respondents choose reactions „very low" and „very high". Examination of plots for all occupations suggests that respondents show a strong tendency to avoid allocating low desirability evaluations to most of the occupations, so that there is fairly little differentiation of desirability. Titles {2,11,17,18,19,29} appear outstanding with respect to the propensity with which respondents evaluate them as deserving „high" or „very high" esteem.

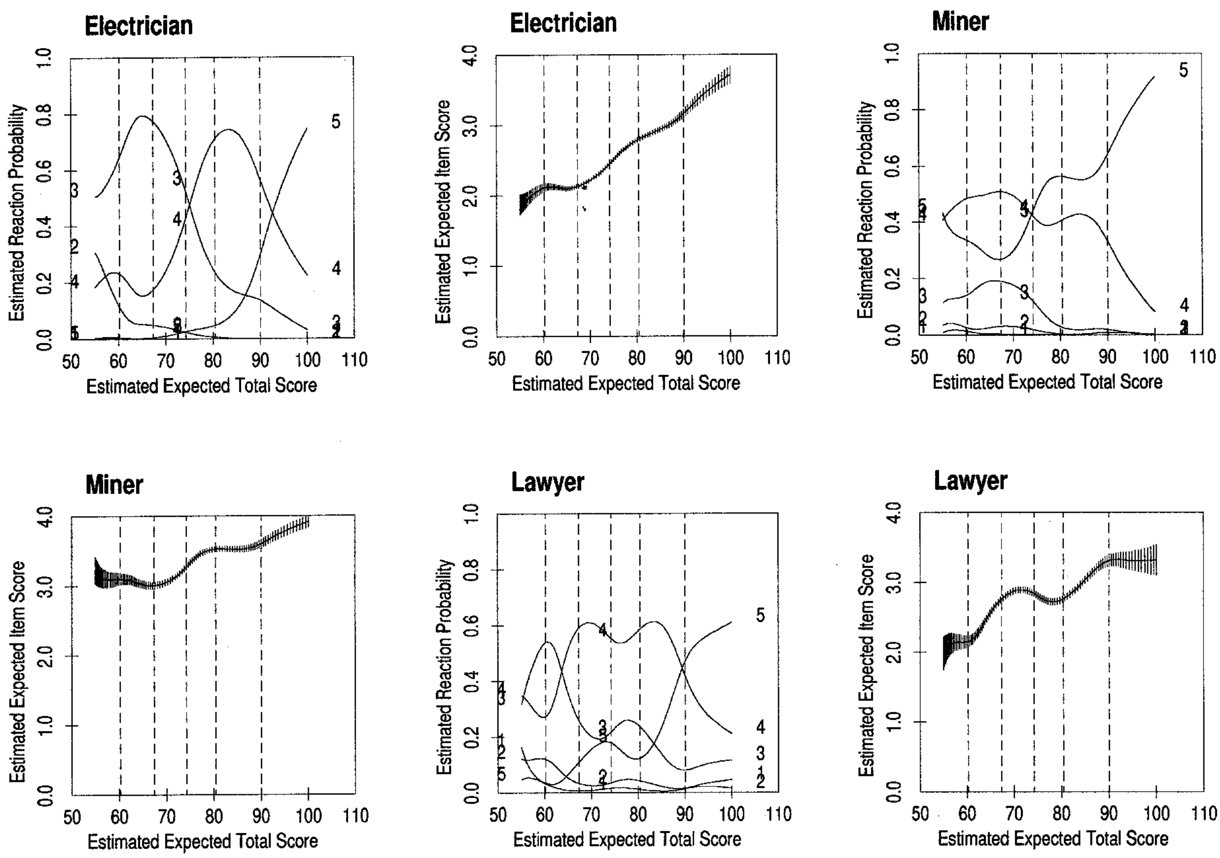

Figure 1: Response structures for selected occupational titles in the IFiS–I data.

In line with the general tendency, few respondents view (*technician*) *electrician* {23} as meriting „very low" or „low" esteem. Response structure for this occupation is remarkably regular, and could well have been captured by a parametric model, because the reaction curves are (almost everywhere) monotone and symmetric. This occupation discriminates rather well between respondents with low and high COD, thus contributing considerably to the estimation of COD scores. However, with respect to the extent of monotonicity and symmetry of the curves, only items {1,14,27} show roughly similar response structures. Thus, parametric-like response structures constitute more of an exception rather then the rule in the IFIS–OI data. Reaction curves for other occupations show deviations from the typical assumptions of parametric models. For instance, reaction curves for (*defence*) *lawyer* {26} are locally nonmonotone and generally non-symmetric: one particular feature of the $4_{\{23\}}$ curve is the absence of a peak corresponding to the median of the **eets**. Application of a typical parametric model that assumes symmetric reaction curves would, in this case, lead to imposition of highly regular curves instead of the „actual“ ones shown in Figure 1, with corresponding bias of the estimates of parameters purported to describe the curves.

As the result of nonmonotonicities in the reaction curves for (*defence*) *lawyer*, **eeis** is not a nonmonotone function of the **eets** for the first quantile of **eets**. Nonmonotonicity of relation between **eeis** and the display factor is much more pronounced for {26} than it is for {23} or {2}. Oscillations of reaction curves similar to those in the plots for {26} are highly pronounced in the case of *government minister* {11}, *physician* {17}, and *university* (*full*) *professor* {19}. One possible interpretation underlying those observations may derive from inherent lack of crystallization of societal status for those titles. In an autocratic polity, incumbents of {11} have much power and influence but little „popular“ legitimacy, whereas incumbents of {19} and {17} would presumably score better if they did not suffer from low base salaries paid them by the communist state. At any rate, the ALREF-based kernel smoothing model clearly diagnoses irregularities with respect to occupational desirability, pointing out the need for more detailed investigation.

Even though nonmonotonicities in the reaction curves may be indicative of possible multidimensionality of the factor structure, or reflect dependency of reaction curves on some exogeneous variables, the sociologist may be willing to accept the response structures for the items that do not show extreme nonmonotonicity of the **eeis** with respect to **eets**. From this point of view occupations {11,12,25} appear to be foremost candidates for removal and/or further analysis.

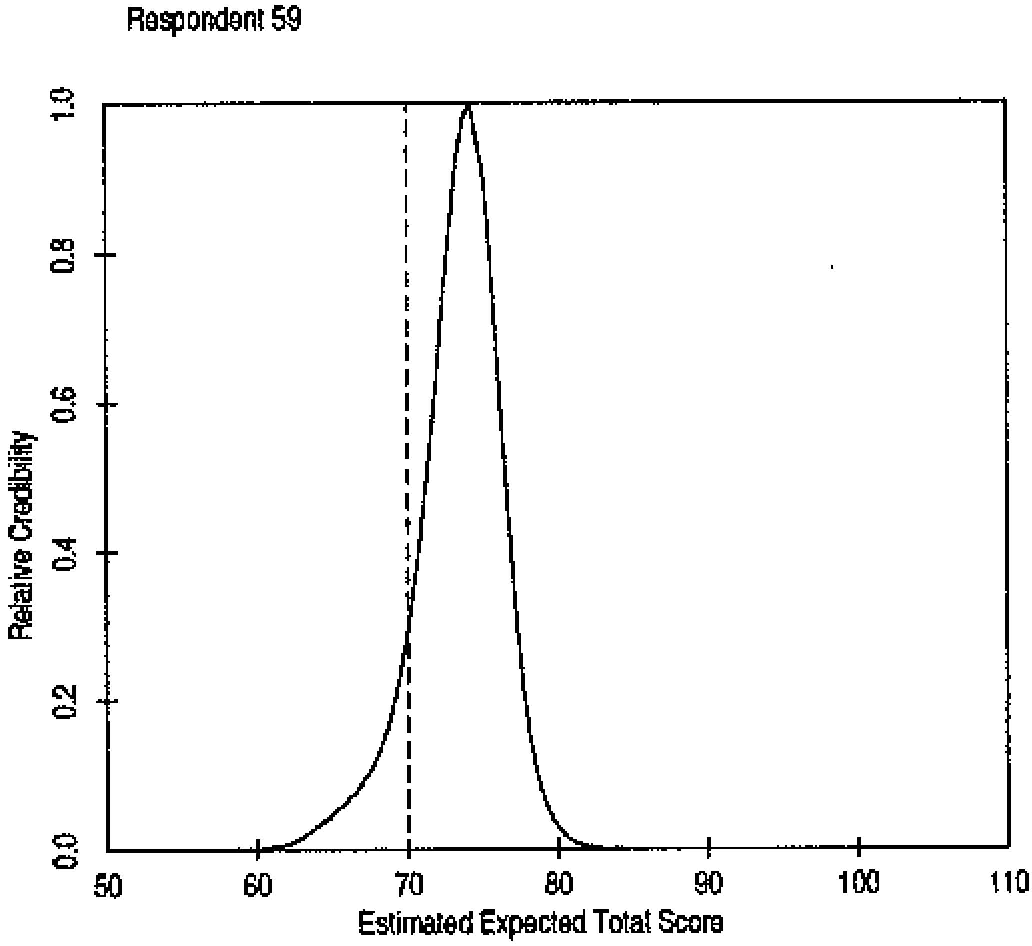

Figure 2: Plot of relative credibitliy curve against display factor **eets**, showing an example of bias that would be introducted by raw total score compared to model-based estimation of COD. The solid curve shows suitably scaled relative likelihood of respondent's true (display) factor score being at various values. The dashed line shows the Likert raw total score $t^{\dagger}_{59} = 70$.

The presence of patterns of variation in the reaction curves that do not correspond to rigid conditions imposed by parametric models on the data may not only lead to biased estimates of item parameters, but also to bias in the estimation of factor scores $\{\theta_v\}$, and thus possibly to wrong rankings of respondents in terms of COD. This can be seen from Figure 2, which shows *relative credibility curve* (**rcc**) with respect to (display) factor score $t(\hat{\theta}_{59}) = 74.3$ corresponding to the Likert raw total score for respondent v = 59. This particular respondent is characterized by reaction profile that can be written, in terms of $x_{c(i)}$, as $\langle 3,4,1,1,2,3,3,0,0,2,3,2,2,2,2,3,4,4,4,2,2,3,3,2,2,3,2,3,3 \rangle$. The relative credibility curve is the relative likelihood curve (see e.g., Lindsey, 1973) that has been scaled for the ease of interpretation, so that 1 is the maximum of the curve for each respondent. Availability of **rcc** is one of the consequences of the interpretation of latent factor as a random variable. The differences between observed total scores and **eets** arise from the use of information about all reaction profiles in the estimation of factor scores, and as the result of alternations. Those differences also reflect the fact that the impact of the Likert raw total scores (the initial values in the kernel smoothing procedure) on the final estimated factor scores is rather low. This impact is further diminished by the purely ordinal interpretation of the factor scores in terms of ranks.

4.2 Sensitivity with respect to exogeneous variables

Graphical representation of response structures facilitates diagnosis and interpretation of dissensus in occupational ratings. A specific aspect of the dissensus derives from possible lack of invariance of occupational evaluations with respect to some exogeneous variables. Much of the evidence supporting the thesis of high degree of consensus in occupational inventory data is based on summaries (such as simple raw item scores), which ignore the dependency on COD. Direct comparison of reaction curves for selected categorical stratifying variables is particularly relevant in this context.

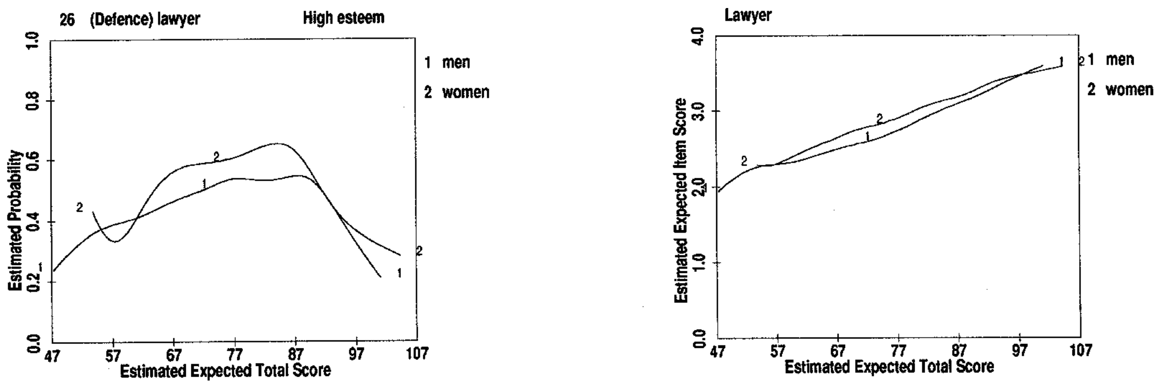

Figure 3: An instance of slight dissensus in the evaluation of occupational desirability between men and women.

A general claim has been made that „[a]n important source of differential prestige judgements seems to be gender“ (Wagener, 1992, p. 254), although this hypothesis is not shared by adherents of reputational methodology. In the following, we shall examine possible systematic differences between men and women in their evaluation of occupational titles in the IFIS–OI. Once a latent factor model for occupational inventory data has been established, a direct method allowing testing for lack of invariance is provided by the psychometric technique of *differential item functioning* (DIF), where the COD scores are „kept“ constant.

Comparison of 290 reaction curves for men and women, such as those shown in Figure 3, shows remarkable similarities with some interesting exceptions. In particular, positions of authority such as {5,11} seem to be afforded higher esteem by women than by men in certain circumstances. For women with **eets** > 80 this applies in particular to $4_{\{5\}}$, i.e., high esteem for the *foreman*, and for women with **eets** in the interval (60,90) to $4_{\{26\}}$, i.e., high esteem for (*defence*) *lawyer*. The latter difference is chiefly responsible for slight differences in **eeis** for the two genders (Figure 3). In comparison with the present model, the highly aggregate character of the „classical“ **sris** tends to obscure qualitative differences in evalution of occuptions to a large extent. An interesting example of the latter is the fact that the probability of perception that tailor has a „high“ esteem is ca. 10% higher for women than for men when **eets** > 80 , although the shapes of the reaction curves are nearly identical for both genders.

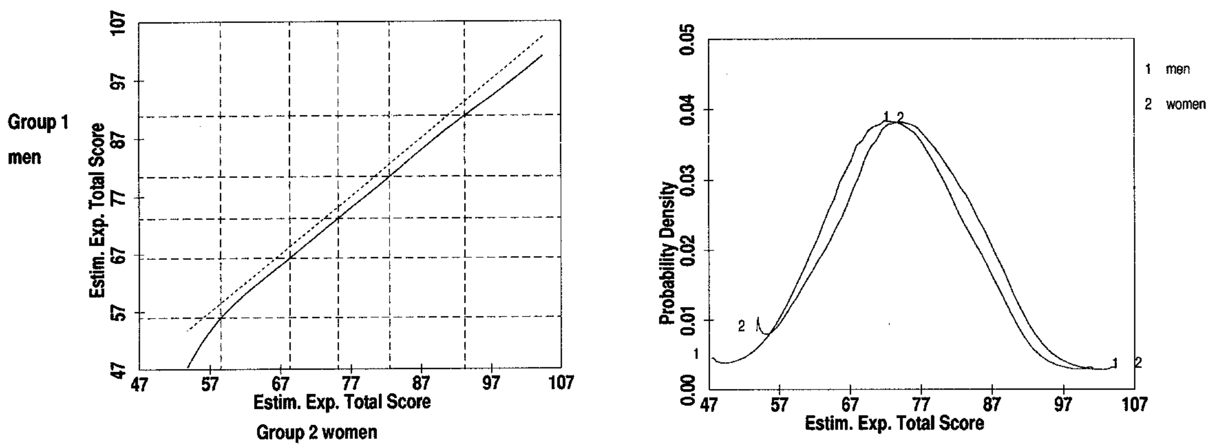

Figure 4: Comparison of the crystallization of ocupational desirability for women and men in the 1987 IFiS–OI data. Left: women compared with men using **eets**. Right: women compared with men in terms of latent factor densities.

We conclude that, contrary to the initial hypothesis, there are no major differences between women and men in the evaluations of the IFiS–OI inventory. However, with the exceptions mentioned above, women are generally more cautious in selecting higher ratings than men (Figure 4).

5. Discussion

Since Likert prior weights $x_{c(i)}$ appear in ordinal Rasch models (chapter 1), the presence of those weights in Ramsay's model may suggest that interpretation of ordinality in both models is somewhat similar. However, an important difference lies in the fact that prior weights $x_{c(i)}$ enter the „systematic component“ of ordinal Rasch models, whereas in Ramsay's model the smoothing of reaction curves does not depend on those weights. An interesting line of inquiry concerns further clarification of similarities and differences among those two types of models, including the suitability of parameters from Rasch models as numerical summaries of important aspects of the reaction curves, when the latter are sufficiently „well-behaved“. Such inquiries should also include comparative analyses of sociological inventories, with the view towards possible differences in the substantive interpretation of results obtained on the basis of Rasch and nonparametric models.

One consequence of the fact that in Ramsay's model the choice of prior weights does not enter in the smoothing stage is the possibility of choosing weights other than the Likert weights $x_{c(i)}$. For instance, in the case of occupational inventories, substantive considerations suggest choosing simple prior weights used in (2). This choice would not have affected conclusions presented in this paper. Non-equidistant weights can be specified as well. A further consequence of the availability of the choice of prior weights is the capability of TESTGRAF95 to model nonhomogeneous Likert multiresponse data that arise in sociology and elsewhere. This can be done by waiving restriction C(i) = C in (4), (5), (6) and (7), and by specifying different prior weights for different items.

One possible disadvantage of the ALREF strategy, and hence the kernel smoothing model, lies in the bias of factor scores arising from their dependency, in the smoothing stage, on the raw total scores (which e.g., can be considerably skewed in some cases). However, the possible bias is substantially reduced by the subsequent use of ranks in the smoothing procedure, and the alternations between RE and F steps in the model fitting. Further reduction in bias is obtained when the resulting factor scores are categorized for the purpose of providing an ordinal variable that would subsequently be used as an „observed" variable in structural models such as graphical models. In those models relations between COD and other variables (e.g., education or societal class) are of definite sociological interest (Kutylowski 1992, 1994).

It has been stated that „[t]here is much to be said for adopting an approach which explicitly recognizes that the latent variables can only be scaled ordinally" (Bartholomew, 1983, p.238). By recognizing lack of identifiability of the latent factor scores, the ALREF strategy in general, and Ramsay's kernel-smoothing model in particular, implement such an approach in a consequent manner. The ALREF strategy appears also promising for further development of flexible modelling tools that remove the necessity to commit oneself to a given, restrictive parametric class of latent factor models.

Acknowledgements

I wish to express my thanks to J. O. Ramsay for clarification of issues arising in the use of the TESTGRAF95 package. Thanks are also due to K. M. Slomczynski and H. Domanski for making the data available.

APPENDIX: Titles comprising IFiS Occupational Inventory

The occupational titles are: {1} turner; {2} miner; {3} construction worker; {4} (unskilled) agricultural (farm) worker; {5} factory foreman; {6} factory engineer (with higher university-type education); {7} agricultural engineer (with higher university-type education); {8} office (female) secretary; {9} official („white collar" clerk); {10} office manager/supervisor; {11} government minister; {12} „private" farmer (medium size farm); {13} locksmith (small proprietor); {14} tailor (small proprietor); {15} merchant (shop owner); {16} (Catholic) priest; {17} physician; {18} teacher; {19} university (full) professor; {20} executive manager of the state agricultural unit (large farm); {21} military officer (captain); {22} journalist; {23} electrician (technician); {24} shop assistant /saleswoman; {25} (female) cleaner; {26} (defense) lawyer; {27} truck or van driver; {28} nurse; and {29} director (executive manager).

References

Altman, N.S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician, 46*, 175–85.

Bartholomew, D. (1983). Latent variable models for ordered categorical data. *Journal of Econometrics, 22*, 229–43.

Bartholomew. D. (1988). The sensitivity of latent trait analysis to choice of prior distribution. *British Journal of Math. and Statist. Psychology, 41*, 101–7.

Goldthorpe, J.H., Hope, K. (1972). Occupational grading and occupational prestige. In Hope, K. (ed.): *The Analysis of Social Mobility. Methods and Approaches*. Oxford: Carendon Press.

Kutylowski, A.J. (1992). Sequential models for ordinal response data. In: van der Heijden, P.G.M., Jansen, W., Francis, B., and Seber, G.U.H. (eds.): *Statistical Modelling*. Amsterdam: North Holland.

Kutylowski, A.J. (1994) A comparison of strategies for the modelling of ordinal multiresponse data. Paper presented at the *9th International Workshop on Statistical Modelling*, University of Exeter, 11–15 July.

Likert, R. (1932). A technique for the measurement of attitudes. *Arch. Psychol. No. 140*.

Lindsey, J.K. (1973). *Inferences from Sociological Survey Data: a Unified Approach*. Amsterdam: Elsevier.

Mislevy, R. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics, 11*, 3–31.

Ramsay, J.O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika, 56*, 611–30.

Ramsay, J.O. (1995). *Testgraf. A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data.* (January 4.) Department of Psychology, McGill University, Montreal.

Reiss, J.A. (1961). *Occupations and Social Status*. Glencoe, IL: Free Press.

Sawinski, Z., Domanski, H. (1991). Dissensus in assessments of occupational prestige: the case of Poland. *European Sociological Review, 7*, 253–265.

Slomczynski, K.M., Bialecki, I., Domanski, H., Janicka, K., Mach, B.W., Sawinski, Z., Sikorska, J., and Zaborowski, W. (1989). *Struktura Spoleczna: Schemat Teoretyczny i Warsztat Badawczy*. [Social Strucure: Theoretical Scheme and Research Toolbox.] Warsaw: Institute of Philosophy and Sociology, Polish Academy of Sciences.

Wagener, B. (1992). Concepts and measurement of prestige. *Annual Review of Sociology, 18*, 253–80.