STATISTICS IN TRANSITION-new series, Spring 2013
Vol. 14, No. 1, pp. 171–182

MODEL OF LATENT PROFILE FACTOR ANALYSIS FOR ORDERED CATEGORICAL DATA

Piotr Tarka[1]

ABSTRACT

In the literature factor analysis is admittedly a well-known and effective multivariate method in the reduction of extensive and broad data, e.g., in the analysis of too many variables. It is also known for the process of unidimensional or multidimensional scale/s construction. Typically, in many studies (especially those pertaining to market research area) a common factor analysis solution is used (based on continuous data). However, there are rarely ever undertaken studies pertaining to latent variable models where other type of data is used based on discrete variables. One of these models might be called Latent Profile Factor Analysis - LPFA. In this article author's main objective is to propose and discuss its (LPFA) main assumptions. In order to prove the model's functionality in practice of market research, a brief example of LPFA model for ordered categorical data (based on one-factorial solution) in reference to hedonic consumption data is given at the end of the paper.

Key words: latent profile factor analysis model, ordered categorical data.

1. Introduction

Most of professional researchers in the socio-economic field, when analyzing market and people-customers' traits, often conduct projects in statistical research based on qualitative data. Most of them are thus forced to describe customers by simply asking questions (including prepared earlier set of items) about their hidden and unknown structure concerning for example personal attitudes, feelings or values. In consequence, in order to examine internal structure of customers, they need to implement an appropriate model for the purposes of data reduction, facing the problems of broad data, e.g. including too many variables. Researchers struggle also with the selection of appropriate method in order to increase precision level in the analysis according to the type of collected data. Solutions, as

[1] Poznan University of Economics, Department of Marketing Research, al. Niepodległości 10, Poland, e-mail: piotr.tarka@ue.poznan.pl.

usual, come with latent variable models based on multivariate complexity. And because in social sciences and in many surveys (undertaken within market research) collected data is mainly of categorical nature, and categorical variables are definitely more used than continuous variables, hence they need a more sophisticated latent variable model to examine this type of data as compared to classical solution based on common factor analysis (Vasdekis et al., 2008). By term “ordered categorical” we will refer to type of data being measured on ordinal variables. For instance in market survey, respondents are often asked to characterize their opinions or attitudes (e.g. about products, etc.) on measurement scales where answers are ranging from “strongly disagree” to “strongly agree”. This is a common example of such data. This type of data is also known as the *ordinal logit, ordered polytomous logit, constrained cumulative logit, proportional odds* (Borooah, 2002; Cohen et al., 2003; DeMaris, 2004; Hoffmann, 2004; Long and Freese, 2006). The most natural way to view structure in ordinal data is to postulate the existence of an underlying latent (unobserved) variable associated with each respondent’s response – observed variable. Unfortunately as it often happens in research practice, the analysis of such data is performed without regard to their ordinal nature (Agresti 2007). For this reason in this article the author investigates the most important characteristics and specificity of Latent Profile Factor Analysis (LPFA). LPFA model is designed for data, that is originating strictly from ordered categorical responses. At the end of paper a practical example of this model is given.

2. Generalized Linear Latent Variable Model

Generalized Linear Latent Variable Model (GLLVM) could be approximately a framework or some kind of a background for construction of Latent Profile Factor Analysis Model (LPFA) for ordered categorical responses. As far as the GLLVM model is concerned, it includes (Moustaki and Knott 2000; Moustaki 2003):

- the random component in which each of the *p* random response variables, $\left(x_1,...,x_p\right)$ has a distribution from the exponential family such as Bernoulli, Poisson, Multinomial, Normal, Gamma,
- the systematic component in which latent variables vector and covariates vector $\mathrm{z}'=\left(z_1,...,z_q\right)$, $\mathrm{x}'=\left(x_1,...,x_r\right)$ produce a linear predictor η_i corresponding to each category of x:

$$\eta_i=\alpha_{i0}+\sum_{j=1}^{q}\alpha_{ij}z_j+\sum_{l=1}^{r}\beta_{il}x_l,\qquad i=1,...,p. \tag{1}$$

- the links between the systematic component and the conditional means of the random component distributions:

$$\eta_i=\upsilon_i\left(\mu_i\left(\mathrm{z},\mathrm{x}\right)\right) \tag{2}$$

where:

$$\mu_i\left(\mathrm{z},\mathrm{x}\right)=E\left(x_i\middle|\mathrm{z},\mathrm{x}\right) \quad (3)$$

and υ_i is called the link function which can be any monotonic differentiable function and may be different for different manifest variables x_i $i=1,\ldots,p$.

We shall also assume that $\left(x_1,x_2\ldots,x_p\right)$ denotes a vector of p manifest variables where each variable has a distribution in the exponential family taking the form:

$$g_i\left(x_i;\theta_i,\phi_i\right)=\exp\left\{\frac{x_i\theta_i-b_i\left(\theta_i\right)}{\phi_i}+d_i\left(x_i,\phi_i\right)\right\},\quad i=1,\ldots,p, \quad (4)$$

where $b_i\left(\theta_i\right)$ and $d_i\left(x_i,\phi_i\right)$ are specific functions taking a different form depending on the distribution of the response variable x_i.

Because of the existence of different types of collected responses (depending on type of used measurement scale) there will be different distribution forms, which we rearrange respectively to their specific transformation functions (Table 1).

Table 1. Distributions and transformation functions from Generalized Linear Model approach

Scale type x_i	**Distribution** $f\left(x_i\middle|\theta\right)$	**Transformation** $g\left[E\left(x_i\middle|\theta\right)\right]$
Dichotomous	Binomial	Logit
Nominal	Multinomial	Logit
Ordinal	Multinomial	Restricted logit
Count	Poisson	Log
Continuous	Normal	Identity

Source: Vermunt and Magidson 2005.

And from perspective of GLLVM approach we may further assume a four-fold classification including sub-models of latent variables. They are: *Factor Analysis (FA), Latent Trait Factor Analysis (LTFC), Latent Profile Factor Analysis (LPFA), and Latent Class Analysis (LCA),* as shown in Table 2. The fundamental distinction in this classification is the one between continuous and discrete latent variables, so that a researcher has to decide whether to treat the underlying latent variable(s) as continuous or discrete. In case of LPFA model, the latent variable is assumed to be discrete and to come from a multinomial distribution.

Table 2. Classification of Latent Variable Models

Manifest Variables	Latent variables	
	Continuous	Categorical
Continuous	**Factor Analysis (FA)**	**Latent Profile Factor Analysis (LPFA)**
Categorical	**Latent Trait Factor Analysis (LTFA)**	**Latent Class Analysis (LCFA)**

Source: own construction based on Bartholomew and Knott 1999.

3. Latent Profile Factor Analysis (LPFA) against a background of other useful models

Classical Factor Analysis (FA) is a popular used tool in market research where in a given set of manifest variables one wants to find a set of latent variables $\xi_1,...,\xi_k$, fewer in number than the manifest variables, which contain essentially the same information. Although FA is meant in general for continuous observed indicators, it is often used by researchers with ordinal models which are based on other types of discrete variables. This mistake yields in the end results that might be incorrect. Not only parameter estimates may be biased, but also goodness-of-fit indices cannot be trusted (Moustaki and Jöreskog, 2006).

Latent Profile Factor Analysis (LPFA) differs from standard Factor Analysis mainly in the sense that the observed variables are either ordered categorical variables (e.g.: "very much", "a little", "not very much") or measured on attitudinal statements (such as: "strongly disagree", "disagree", "agree", "strongly agree"). These answers collected from survey fall into only one category. Such categorization makes the data of ordinal nature. However, as already mentioned, assumptions of ordinality of data in practice of market research is often ignored and numbers such as 1, 2, 3, 4, 5 representing ordered categories are treated as numbers having metric properties 1-2-3-4-5 which yields incorrect results. In consequence, ordered categorical data (which has for example number of five or seven categories) is by mistake of many analysts treated as if there were some kinds of interval level variables in it. Indeed, proceeding in that way with standard factor analysis allows them to compute correlations on the basis of so-called *pseudo-continuous variables*. Moreover, this uncritical approach to application of factor analysis associated with ordered categorical data is likely to give biased estimates of the factor loadings. Hence, the better solution in finding relationships between ordered categorical data comes with minor modifications of Item Response Theory Models where one assumes that the responses to the ordinal items are independent conditional on the latent variables (conditional independence) (Bartholomew 2002). For ordered-response categories (which appear in LPFA) IRT models[1] are definitely more informative and reliable than simply scored items by FA.

However, IRT solution is not yet enough. In order to construct a good model of LPFA we need to focus additionally on Latent Trait Factor Analysis for binary data, where we usually analyze the probability of a randomly selected individual giving a positive response to an item as a function of the latent variables. In case of ordinal data, where more than two categories exist, we simply need to specify probabilities for each category. As a result the observed ordinal variables are denoted by $x_1,\dots,x_p$. Let us suppose that there are m_i categories for variable i labelled $(1,\dots,m_i)$. For binary items $m_i = 2$ (for each i) the category labels are usually denoted as 0 and 1 but they could equally well have been marked as 1 and 2. In LPFA we need to redefine a response probability for each category. Let now $\pi_{i(s)}(\mathrm{F})$ be the probability so that given $\mathbf{F}$ a response falls in category s for i-th variable. The position with two categories can be then compared with the general case as follows:

Categories	0	1
Response probability	$1-\pi_i(\mathrm{f})$	$\pi_i(\mathrm{f})$

Categories	1	2	…	s	…	m_i
Response probability	$\pi_{i(1)}(\mathrm{f})$	$\pi_{i(2)}(\mathrm{f})$	…	$\pi_{i(s)}(\mathrm{f})$	…	$\pi_{i(m_i)}(\mathrm{f})$

In both cases, the response probabilities sum to one. The question is now on how to use logit model (expressing the logit of probability of response in category as a linear function of *f*) for more than just two categories. Suppose we divided categories into two groups with categories $(1,2,\dots,s)$ - into one group and $(s+1,s+2,\dots,m_i)$ - into other group and were merely to report into which of these two groups the response fall. We would thereby need to reduce the polytomous variables to a binary variable. Therefore, it seems reasonable to infer that any model we choose for polytomous case should be consistent with the one

[1] IRT model(s) may be characterized by a few options such as (Embretson and Reise, 2000): *Graded Response Model* (Samejima, 1969), *Modified Graded Response Model* (Muraki, 1992), (which is used with questionnaires that have a common rating scale format (e.g., all item responses scored on a five-point scale). These two models are considered as “indirect” models because a two-step process is needed to determine the conditional probability of the response in particular category. The other remaining models are considered as “direct” IRT models because only a single equation is needed to describe the relationship between respondent response level and the probability of responding in particular category. Specifically there are two polytomous models that are extensions of the Rasch model, e.g. *Partial Credit Model* (Masters, 1982) and *Rating Scale Model* (Andrich, 1978 a-b).

which we use also for the binary case. As a result, in order to make ordered categorical model (in LPFA) more effective, we need to apply binary logit model. To do so we must split the binary model where the probabilities of a response fall into the first and second group, which may be written as follows (Bartholomew, 2002):

$$\gamma_{i(s)}(\mathrm{f}) = \Pr(y_i \leq s) = \pi_{i(1)}(\mathrm{f}) + \pi_{i(2)}(\mathrm{f}) + ... + \pi_{i(s)}(\mathrm{f}), \tag{5}$$

and

$$1 - \gamma_{i(s)}(\mathrm{f}) = \Pr(y_i > s) = \pi_{i(s+1)}(\mathrm{f}) + \pi_{i(s+2)}(\mathrm{f}) +,...,+ \pi_{i(m_i)}(\mathrm{f}), \tag{6}$$

where: s denotes the category into which the i-th variable falls.

$\gamma_{i(s)}(\mathrm{f})$ - the probabilities are referred to as cumulative response probabilities .

Next we need to define the model, supposing that binary logit model holds for all possible divisions of the m_i categories into two groups. We can do this by specifying the model in terms of logit $\gamma_{i(s)}(\mathrm{f})$ or logit $\left(1 - \gamma_{i(s)}(\mathrm{f})\right)$.

The model is thus expressed as follows:

$$\log\left[\frac{\gamma_{i(s)}(\mathrm{f})}{1 - \gamma_{i(s)}(\mathrm{f})}\right] = \alpha_{i(s)} - \sum_{j=1}^{k} \alpha_{ij} f_j, \tag{7}$$

where: $(s = 1,...,m_i - 1;\ i = 1,..., p)$.

For a positive factor loading α_{ij} the higher the value of an individual on the latent variable f_j, the higher the probability of that individual responding in the higher categories of item *i*. The intercept parameter $\alpha_{i(s)}$ is one for each category. The ordering of the categories implies that the intercept parameters are also ordered:

$$\alpha_{i(1)} \leq \alpha_{i(2)} \leq ... \leq \alpha_{i(m_i)}. \tag{8}$$

In consequence the factor loadings remain the same across categories of the same variable. Otherwise, the discriminating power of the item does not depend on where the split into two groups was made. The π's are obtained from the γ's by:

$$\pi_{i(s)}(\mathrm{f}) = \gamma_{i(s)}(\mathrm{f}) - \gamma_{i(s-1)}(\mathrm{f}) \qquad (s = 2,...,m_i), \tag{9}$$

where $\gamma_{i(1)}(\mathrm{f}) = \pi_{i(1)}(\mathrm{f})$ and $\gamma_{i(m_i)}(\mathrm{f}) = 1$. We refer to $\gamma_{i(s)}(\mathrm{f})$ as cumulative response function and to $\pi_{i(s)}(\mathrm{f})$ as category response function.

4. Goodness-of-fit in Latent Profile Factor Analysis – (LPFA)

The LPFA model should be fitted in the same way as the binary latent trait model using the method of maximum likelihood. Goodness-of-fit can likewise be judged using the same criteria based on the likelihood ratio G^2 and the Pearson chi-squared χ^2 statistics calculated from the whole response patterns as follows:

$$G^2 = 2\sum_{i=1} O_i \log \frac{O_i}{E_i} \tag{10}$$

$$\chi^2 = \sum_{i=1} \frac{\left(O_i - E_i\right)^2}{E_i} \tag{11}$$

where O_i and E_i are the observed and expected frequency of response pattern *i*. When the sample size *n* is large and *p* small, the statistics under the hypothesis that the model fits follow a chi-square distribution with degrees of freedom the number of response patterns minus the number of independent parameters minus one. As the number of items increases, the chi-square approximation to the distribution of either goodness-of-fit statistic ceases to be valid. Parameter estimates are still valid but it is difficult to assess the model.

The goodness-of-fit in Latent Profile Factor Analysis can be also assessed by looking at the two or three-way margins. The pairwise distribution of any two variables is then displayed as a two or three-way contingency table, and chi-squared residuals are constructed by comparing the observed and expected frequencies. The differences are computed using G^2 and χ^2 statistic. If there are small differences, it means the associations between all pairs of responses are well predicted by the model.

5. Example on construction *LPFA one-factorial model* in reference to hedonic consumption data

For demonstration purposes the data set that was extracted from the earlier study conducted by the author (Tarka, 2010) was prepared. The data included the responses given by 232 individuals to four below listed items concerning attitude to hedonic consumption-oriented lifestyle. For each item (statement), respondents were asked the following response alternatives based on four-point scale: [1] = strongly disagree, [2] = disagree to some extent, [3] = agree to some extent, [4] = strongly agree. And the chosen questions were given:

1: *I'm money-oriented person and looking for wealth in my life* [money]
2: *I'm striving to be free in my private life with no family frontiers* [freedom]
3: *I'm having a good time and enjoying only things I like and prefer* [party]
4: *I'm looking for adventurous and risky life* [full of life]

The output of the one-factor analysis for above ordered categorical data is given below. For calculations the author used LAMI software which contains an interface that allows users of the GENLAT and LATCLASS programs to run their analyses conveniently than using the original DOS programs directly. The program fits a latent trait model for ordinal observed variables with up to two latent variables. The program computes parameter estimates, standard errors, chi-squared residuals, and scoring methods.

In order to start program input file parameters were specified as follows:

One-Factorial Model = 1
Number of Observed Variables = 4
Number of Ordinal Variables = 4
Number of Cases Sampled = 232
Proportion of Response Patterns with at Least One Missing Observation = 0,00
Number of Quadrature Points Used = 8
Maximum Number of Iterations Permitted = 2000
Convergence Tolerance For The Relative Likelihood Value = 0.00000000
………….
NFAC: Number of factors (1)
INIT: 0 if the initial parameter values are set in the program or 1 if the initial parameter estimates are to be read from file
ITER: Number of iterations (maximum is 2000)
PREC: Precision for maximization (e.g. 0.0000001, convergence tolerance of the EM algorithm)
SCOR: 1 if scoring results to be printed, 0 otherwise.

Source: Own construction based on LAMI software.

Finally, we obtained the following estimated scores (according to printed version in LAMI software. These results are shown in Tables 3-8. From Table 3 we can observe that percentage of individuals *agreeing to some extent* or *agreeing strongly* (categories 3 and 4) is larger as compared to other two response categories denoted by 1 and 2. Having inspected the results (in case of binary data), we looked at pairwise associations between four mentioned above items which suggest there exists some real common underlying factor including all four items. They can be considered as indicators for measuring respondents' attitude to lifestyle based on hedonic consumption.

Table 3. Items – category frequencies

Item 1		Item 2	
1	0,0043103	1	0,0732759
2	0,0689655	2	0,2413793
3	**0,6810345**	**3**	**0,5689655**
4	**0,2456897**	**4**	**0,1163793**

Table 3. Items – category frequencies (cont.)

Item 3		Item 4	
1	0,0431034	1	0,0387931
2	0,1810345	2	0,2413793
3	**0,5474138**	**3**	**0,5172414**
4	**0,2284483**	**4**	**0,2025862**

Source: Own calculations based on LAMI software.

Also the parameters and standard errors based on maximum likelihood estimates for particular categories associated with respective items (Tab. 4) indicate that the strongest relationships appear mainly in the third category containing positive values and smaller standard errors (S.E.). This result simply means that third category of the respective item (I) will compose to a greater part our considered one factorial-model.

Table 4. Maximum likelihood estimates of item parameters and standard errors (S.E.)

Item	Category	(I - Item, J - Factor)	S.E
1	1	-5,888	6,498
1	2	-2,920	2,523
1	**3**	**1,328**	**0,328**
1	4	-3,420	3,613
2	1	-3,035	1,059
2	2	-1,003	0,743
2	**3**	**2,481**	**0,371**
2	4	-2,820	1,583
3	1	-4,587	1,876
3	2	-2,019	1,304
3	**3**	**1,964**	**0,369**
3	4	-4,423	3,523
4	1	-3,616	2,470
4	2	-1,112	0,803
4	**3**	**1,624**	**0,314**
4	4	-1,920	0,423

Source: Own calculations based on LAMI software.

Now, if we decide to fit this type of one-factor model based on hedonic consumption data, we need to obtain the estimates given in Table 5. The Alpha's are simultaneously representing factor loadings. They are defined in the literature as discriminating parameters. If the values of factor loadings are large and they all are positive but the standard errors are small, then there is an underlying factor which is common to all items. And this is purely visible in our case. The high values of standardized loadings (Table 6) also suggest that the single factor model provides a good explanation for all four ordinal (ordered categorical) items, especially for item number 3. However, before putting too much weight on this conclusion, we need to look at how well the model fits.

Given the sparsity of the data (at total frequency of 232 spread over multiple response categories), it is not feasible to carry out global tests. Instead, we need to look at the fits to the margins. Therefore, for each pair of items, (see Table 7) we need to calculate the sum of the chi-squared residuals over each pair of item categories. Sixteen chi-squared residuals for each pair of items were generated, since each variable had four response categories.

Table 5. Alpha as factor loadings and standard errors (S.E.) for items

Items	Alpha(1,I)	S.E
1	0,982	0,241
2	1,167	0,313
3	**2,009**	0,398
4	1,000	0,212

Source: Own calculations based on LAMI software.

Table 6. Standardized Loadings for items

Items	St. Alpha(1,I)
1	0,7008
2	0,7592
3	**0,8952**
4	0,7072

Source: Own calculations based on LAMI software.

Table 7 shows how the entry 20,47 (due to calculations based on two-way margins of selected items "Money" and "Full of Life" of Table 8) is computed. The sum of the entries of Table 8 is 20,47. In similar way we computed the sums of chi-squared residuals for other two-way tables including another pairs of items. In order to confirm if the model is correct, we need to check the chi-squared residuals. Values greater than about 4 would indicate a poor fit. For instance, as observed from Table 8, values larger than 4 do not appear. For the best part of cells they are considerably below 4. In other words these associations make up a good configuration for our items in the model.

Table 7. Sum of chi-squared residuals for all pairs of items derived from the two-way margins for one-factorial model

Items	Money	Freedom	Party
F.o.life	**20,47**	10,58	23,32
Money		17,21	12,45
Freedom			9,56

Source: Own calculations based on LAMI software.

Table 8. Chi-squared residuals for the two-way margins of selected pair of items "Money" and "Full of Life"

Category		**Money** 1	2	3	4
F.o.Life	1	0,87	1,65	0,34	0,90
	2	2,40	3,40	2,10	1,09
	3	2,34	1,13	1,09	0,56
	4	1,02	1,30	0,21	0,94

Source: Own calculations based on LAMI software.

Since the sum of these residuals over all the cells in a two-way marginal table is analogous to Pearson's chi-squared statistic for goodness-of-fit, but because the model has been fitted to the full multi-way table, the standard chi-squared test does not apply. We may still use this sum, as a diagnostic, e.g. D. Larger value of D would then suggest that the associations in two-way table are not well explained. As a rule of thumb that D is too large we need to take into account value that is greater than upper 1% point of a chi-square distribution with $\left[\left(m_i \times m_j\right) - 1\right]$ degrees of freedom. And as observed from the results (Table 7), each pair of all analyzed six entries has values of D less than 28,58 (the upper 1% point of chi-square with 15 degrees of freedom). Therefore, the fit to each two-way marginal table (pair of item) appears satisfactory. Overall, the one-factor model appears to give an adequate description of data. Therefore, we can use this factor as a summary measure of attitude to hedonic consumption issues.

6. Conclusions

Latent Profile Factor Analysis (LPFA), being a part of four latent variable models, is a powerful and useful tool for researchers. However, this model has been languishing too long on the borders of statistics and most importantly in research practice. It is slowly and surely taking its right place in the main stream, stimulated in part by the recognition of its greater value and sound foundations which have been given to it within a statistical framework. Assuredly, this new solution clarifies, simplifies and reduces broad data as far as the ordered categorical responses are concerned into more simple form than the previous model based on classical factor analysis. LPFA model would not be for sure possible without earlier progress of Item Response Theory which supported to a greater extent the development of Latent Profile Factor Analysis.

REFERENCES

AGRESTI, A., (2007). An introduction to categorical data analysis, 2nd ed. New Jersey: John Wiley and Sons.

ANDRICH, D., (1978a). Application of psychometric model to ordered categories which are scored with successive integers, "Applied Psychological Measurement", 2, pp. 581–594.

ANDRICH, D., (1978b). A rating formulation for ordered response categories, "Psychometrika", 43, pp. 561–573.

BARTHOLOMEW, D. J., KNOTT, M., (1999). Latent variable models and factor analysis. London: Arnold.

BARTHOLOMEW, D. J., (2002). Old and new approaches to latent variable modelling, [in:] MARCOULIDES, G.A., MOUSTAKI, I. (Eds.). Latent variable and latent structure models, New Jersey: Lawrence Erlbaum Associates, pp. 1–15.

COHEN, J., COHEN, P., WEST, S. G., AIKEN, L. S., (2003). Applied multiple regression/correlation analysis for the behavioral sciences, (3rd ed.). Mahwah, New York: Lawrence Erlbaum.

DEMARIS, A., (2004). Modeling continuous and limited response variables. Hoboken, New York: John Wiley and Sons.

EMBRETSON, S. E., REISE, S., (2000). Item response theory for psychologists, New Jersey: Lawrence Erlbaum Associates.

JÖRESKOG, K. G., MOUSTAKI, I., (2006). Factor analysis of ordinal variables with full information maximum likelihood, unpublished report.

LONG, J. S., FREESE, J., (2006). Regression models for categorical dependent variables using Stata, (2nd ed.). College Station, TX: Stata Press.

MAGIDSON, J., VERMUNT, J. K., (2005). Factor analysis with categorical indicators – a comparison between traditional and latent class approaches, [In:], VAN DER ARK, L. A., CROON, M. A., SIJTSMA, K., (Eds.), New developments in categorical data analysis for the social and behavioral sciences, New Jersey: Lawrence Erlbaum Associates, pp. 41–62.

MASTERS, G. N., (1982). A Rasch model for partial credit scoring, "Psychometrika", 47, pp. 149–174.

MOUSTAKI, I., KNOTT, M., (2000). Generalized latent trait models, "Psychometrika", 65, pp. 391–411.

MOUSTAKI, I., (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables, "British Journal of Mathematical and Statistical Psychology", 56, pp. 337–357.

SAMEJIMA, F., (1969). Estimation of latent ability using a response pattern of graded scores, "Psychometrika Monograph Supplement", 17, pp. 1–97.

TARKA, P., (2010). Latent variable models - issues on measurement and finding exact constructs – "Przegląd Statystyczny", 4, pp. 142–167.