Uniwersytet
Ekonomiczny
w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 304 · 2016

Informatyka i Ekonometria 7

Ewa Genge

Uniwersytet Ekonomiczny w Katowicach
Wydział Ekonomii
Katedra Analiz Gospodarczych i Finansowych
ewa.genge@ue.katowice.pl

DICHOTOMOUS IRT MODELS IN MONEY-SAVING SKILLS ANALYSIS

Summary: Latent variable models include latent class models, item response theory (IRT) models, latent profile models or common factor models. We focus on item response models (latent variable models where the latent variable is continuous, whereas observed variables are categorical). Those kind of latent trait models are popular in educational testing, however, the paper presents an application of item response models in economic analysis which is relatively rare.

The aim of this paper is to find the most suitable IRT model and asses the Poles responses according to their money-saving skills (ability to save money), and the difficulty of the items (evaluation of the reliability of the item scale).

Keywords: latent variable models, dichotomous data, item response theory.

Introduction

Latent variables analysis is a powerful and useful statistical tool being unappreciated for a long time. It is now taking its place in the main stream, within a traditional statistical framework. Latent class analysis, along with latent trait analysis, have their roots in the work of the sociologist Paul Lazarsfeld in the 1960's [Lazarsfeld, Henry, 1968]. Although a latent trait model differs from latent class model only in the fact that the latent dimension is viewed as continuous rather than categorical it is considered separately because it owes its development to the particular types of application, i.e. educational testing which is based on the idea that human abilities vary and that individuals can be located on a scale of the ability under investigation by the answer they give to a set of test items [Bartholomew, 2002]. The latent trait model provides the link between the responses and the underlying trait. A seminal contribution to the theory was provided by Birnbaum [1968], but today there is an enormous literature, both applied and theoretical including multitude of articles [van der Linden, Hambleton, 1997; Baker, Kim, 2004; Alagumalai et al., 2005; Bezruczko, 2005; Panayides et al., 2010; Bond, Fox, 2013].

Many researches in the various sciences use questionnaires to measure properties that are of interest of them. Examples of properties include personality traits such as introversion and anxiety (psychology), political efficacy and motivational aspects of voter behavior (political science), attitude toward religion or euthanasia (sociology), aspects of quality of life (medicine) and preferences towards particular brands of products (marketing). Often, questionnaires consist of a number of questions, each followed by binary or a rating scale and the respondent is asked to mark the category that he thinks applies most to his personality, opinion or preference. The rating scales are scored in such a way that the ordering of the scores reflects the hypothesized ordering of the answer categories on the measured properties (called latent traits). The total score is well known from classical test theory [Lord, Novic,1968] and Likert scaling [1932], and is the test performance summary most frequently used in practice.

In the item response theory (IRT), latent traits are usually measured by employing probabilistic models for responses to sets of items. One of the most prominent examples for such an approach is the Rasch model [Rasch, 1960] which captures the difficulty (or equivalently easiness) of binary items and the respondent's trait level on a single common scale.

A main characteristic of some IRT models concerns the separation of two kinds of parameters, one that describes qualities of the subject under investigation, and the other relates to qualities of the situation under which the response of a subject is observed. Item response theory (IRT) is widely used in assessment and evaluation research to explain how participants respond to the questions. It is assumed that people respond to a test item according to their ability and the difficulty of the item. The main applications of IRT can be found in educational testing in which analysts are interested in measuring examinees' ability using a test that consists of several items (i.e. questions). Therefore, several IRT models have been proposed in the literature that deal with different kinds of questionnaires. Among the main contributors we remind: Rasch [1960], Masters [1982], Wright and Masters [1982], Duncan and Stenbeck [1987], Agresti [1993], Kelderman and Rijkes [1994], Kelderman [1996], Adams et al. [1997], Sagan [2002], Frick et al [2012], Bartolucci et al. [2014].

We will apply dichotomous IRT models in measuring money-saving skills in Poland (abilities to save money) using and comparing results of different kinds of dichotomous IRT models with continuous latent variable.

1. Latent variable models

The basic idea of latent variable analysis is to find for a given set of response variables $\mathbf{X} = X_1,\ldots,X_m$ a set of latent variables $\mathbf{Z} = Z_1,\ldots,Z_q$ (with $q \le m$). The latent variable regression models have usually the following form:

$$E(X_j \mid \mathbf{Z}) = g(\gamma_{j0} + \gamma_{j1}Z_1 + \ldots + \gamma_{jq}Z_q),\ j = 1,\ldots,m \qquad (1)$$

where:
$g(\cdot)$ – a link function;
$\gamma_{j0},\ldots,\gamma_{jq}$ – the regression coefficients for the response variables;
X_j – is independent of $X_{l1}(j \neq l)$ given $\mathbf{Z}$.

The popular factor analysis model assumes that the responses are continuous variables following a normal distribution with $g(\cdot)$ being an identity link function. In this paper we focus on dichotomous IRT models, in which $E(X_j|\mathbf{Z})$ expresses the probability of endorsing one of the response categories. In the IRT framework usually one latent variable (Z) is assumed[1].

2. IRT models for dichotomous data

Latent traits can be measured through a set of items to which binary responses are given. Success of solving an item or agreeing with it is coded as „1", while „0" codes the opposite response. The IRT modeling for dichotomous data is the model for the probability of positive (correct) response in each item given the ability level Z. A general model for this probability for the i-th respondent in the j-th item is the following [see i.e. Rizopoulos, 2015]:

$$P(x_{ij} = 1 \mid z_i) = c_j + (1 - c_j)g\{\alpha_j(z_i - \vartheta_j)\}, \qquad (2)$$

where:
x_{ij} – the dichotomous response variable;
z_i – denotes the respondent's level on the latent continuous scale;
c_j – the guessing parameter;
α_j – the discrimination parameter;
ϑ_j – the difficulty parameter.

[1] We will consider the latent level of saving ability in the empirical part of the paper.

The guessing parameter expresses the probability that the respondents with very low ability responds positively to an item by chance. The discrimination parameter quantifies how well the item distinguishes between subjects with low/high standing in the continuous latent scale, and the difficulty parameter expresses the difficulty level of the item.

The one-parameter logistic model, known as Rasch model [Rasch, 1960], assumes that there is no guessing parameter, i.e. $c_j = 0$, and that discrimination parameter equals one, i.e. $\alpha_j = 1$. The two-parameter logistic model allows for different discrimination parameters per item and assumes that $c_j = 0$. The Birnbaum's three parameter model [1968] estimates all three parameters per item.

The most common choices for $g(\cdot)$are the probit and the logit link, which correspond to the cumulative distribution function (*cdf*) of the normal and logistic distribution, respectively[2].

3. Parameter estimation

An important modeling issue is the parameter estimation. Three major methods, namely conditional, full, and marginal maximum likelihood, have been developed under maximum likelihood estimation. Estimation of model parameters has received a lot of attention in the IRT literature [Linacre, 1998; Martin, Quinn, 2006; Mair, Hatzinger, 2007]. A detailed overview of these methods is presented in Baker and Kim [2004] and a brief discussion about the different methods can be found in Agresti [2002]. In the empirical part of this article, we use ltm packge of R applying marginal maximum likelihood estimation (MMLE). Parameter estimation under MMLE assumes that objects represent a random sample from a population and their ability is distributed, according to a distribution function $F(Z)$.

The model parameters are estimated by maximizing the observed data log-likelihood:

$$l\ (\boldsymbol{\varphi}) = \log p(\mathbf{X};\boldsymbol{\varphi}) = \log \int p(\mathbf{X} \mid Z;\boldsymbol{\varphi}) p(Z) dZ, \tag{3}$$

where: $p(\cdot)$denotes a probability density function, $\mathbf{X}$ denotes the vector of responses for the ith sample unit, Z is assumed to follow a standard normal distribution and $\boldsymbol{\varphi} = (\alpha_j, \vartheta_j, c_j)$.

[2] In empirical part of our work, we will apply a logit link function.

The IRT models with different parameterization are compared on the basis of log-likelihood ratio test as well as information criterion such as Bayesian Information Criterion (BIC) [Schwarz, 1978] or Akaike Infromation Criterion (AIC) [Akaike, 1974].

4. Empirical analysis

The empirical part of the article is based on Social Diagnosis questionnaires. The Social Diagnosis (Objective and Subjective Quality of Life in Poland) is a diagnosis of the conditions and quality of life of Poles as they report it [Czapiński, Panek, red., 2016].

We considered questionnaire items about the Polish saving behaviours. The data concern 12 dichotomous response variables measured at the last year of the survey, i.e. 2015. In total, there is complete information on $n = 7399$ households. The public data set, available at *www.diagnoza.com* [see also: Czapiński, Panek, red., 2016].

All computations and graphics in this paper have been done in ltm [Rizopoulos, 2015] package of R. The following variables (questions) considering the different purposes of household's savings were used in the analysis:

- X_1 (*HF8_01*) – current consumer needs (e.g. food, clothes),
- X_2 (*HF8_02*) – regular fees (e.g. home payments),
- X_3 (*HF8_03*) – purchase of consumer durables,
- X_4 (*HF8_04*) – purchase of house, apartment, payments to the housing cooperative,
- X_5 (*HF8_05*) – renovation of house, apartment,
- X_6 (*HF8_06*) – medical treatment,
- X_7 (*HF8_07*) – medical rehabilitation,
- X_8 (*HF8_08*) – leisure (recreation),
- X_9 (*HF8_09*) – unexpected events ("rainy day"),
- X_{10} (*HF8_10*) – the children's future,
- X_{11} (*HF8_11*) – security for the old age,
- X_{12} (*HF8_12*) – business development.

We initially fitted the original form of the Rasch model that assumes known discrimination parameter fixed at one. Then we investigated two possible extensions of the unconstrained Rasch model – two parameter logistic model (2PL model) which assumes a different discrimination parameter per item and the unconstrained Rasch model with guessing parameter, i.e. Birnbaum's three parameter model. The optimal parametrization of dichotomous IRT models was chosen on the basis of likelihood ratio test as well as AIC and BIC information criteria (see: Table 1).

Table 1. Log-likelihood, AIC and BIC results for different dichotomous IRT models

Model	LL	AIC	BIC
Rasch	-40452.39	80928.79	81011.7
2PL	-39937.90	79923.80	80089.62
Birnbaum's three parameter	**-39812.94**	**79675.87**	**79848.60**

Source: Own calculations in R.

On the basis of a likelihood ratio test as well as BIC and AIC criterion, the Birnbaum's three parameter model was chosen. The results of ltm package of **R** related to selected model are presented in Table 2.

Table 2. The results of ltm package of **R**

Item	c_j	ϑ_j	$P(x_j = 1\|z)_3$
X_1 (current needs)	0.399	2.218	0.354
X_2 (fees)	0.113	2.253	0.131
X_3 (foods)	0.019	1.251	0.121
X_4 (home)	0.025	3.128	0.029
X_5 (renovation)	0.049	1.261	0.147
X_6 (treatment)	0.123	1.373	0.198
X_7 (rehabilitation)	0.011	2.027	0.040
X_8 (leisure)	0.000	1.061	0.138
X_9 (rainy day)	0.436	0.533	0.597
X_{10} (children future)	0.069	1.636	0.122
X_{11} (aging)	0.131	1.143	0.237
X_{12} (business_development)	0.018	2.627	0.029

Source: Own calculations in R.

We can observe the highest probability that Poles with very low money-saving skills (low ability to save money) respond correctly[3] to the X_9 and X_2 items (*rainy day* and *current needs*) by chance (the highest guessing parameter).

The items X_4 (*home*) and X_{12} (*business development*) are considered to have the highest difficulty level of the item. The probability of a positive response to those items is close to 0,03. In turn, the probability of a positive response to the ninth item (*rainy day*) for the average individual is 0.597.

[3] „Yes" answer.

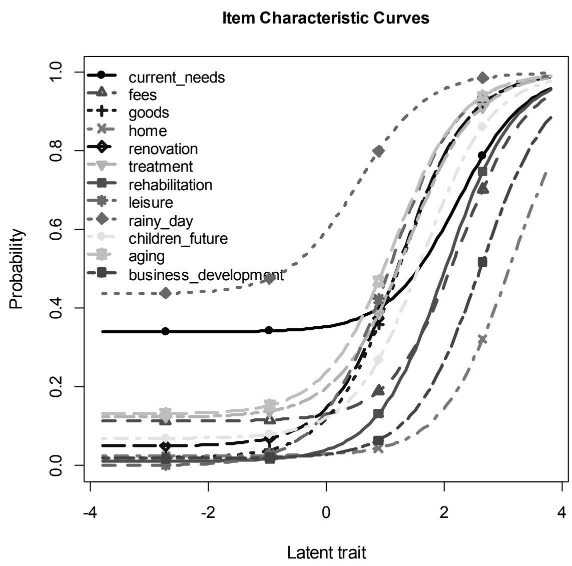

Figure 1. Item Characteristic Curves

Source: Own calculations in R.

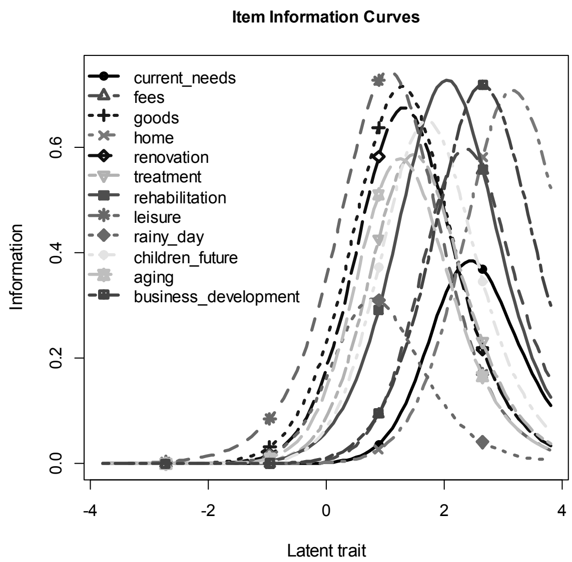

Figure 2. Item Information Curves

Source: Own calculations in R.

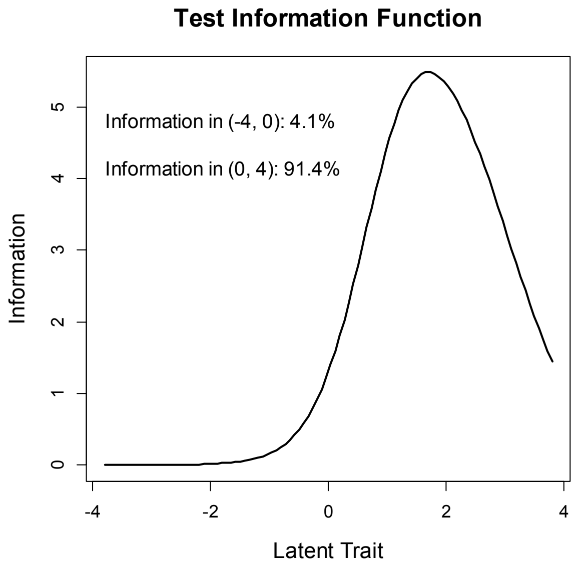

Figure 3. Test Information Function

Source: Own calculations in R.

From the Item Response Characteristic Curves (Figure 1), we observe that there is low probability of endorsing „yes" answer for relatively low latent trait levels, i.e. low saving skills (with exception of two items, i.e. *rainy day* and *current needs*). Respondents with higher saving skills are more likely to give positive answers for items aging, treatment, leisure and children future as well.

According to the Test Information Curve (Figure 3), we observe that the set of 12 items mainly provides over than 91% of the total information for high latent trait level, whereas the items that seems to distinguish respondents with lower ability levels is over 4%. Finally, the Item Information Curves (Figure 2) indicate that items X_9 and X_2 (*rainy day* and *current needs*) provide little information in the whole latent trait continuum. Whereas the item that seems to distinguish between respondents with higher ability levels is the forth (*home*) and twelfth ones (*business development*).

Conclusions

Latent variable models are increasingly becoming established in social research, particularly in the analysis of performance or attitude data in education, psychology and other fields. We presented the analysis of multivariate dichotomous data using latent variable models, under the Item Response Theory approach. We fitted and compared the results of the following models: Rasch, two-parameter logistic, Birnbaum's three-parameter models. We provided the graphical illustration, i.e. Item Characteristic, the Item Information and the Test Information Curves plots, describing the relationship between a latent saving skills and the performance on the items as well.

In future research we would like to analyze data presented above using the extended version of traditional IRT models allowing for multidimensionality and discreteness of latent traits. This class of models also allows for different parameterizations for the conditional distribution of the response variables given the latent traits, depending on both the type of link function and the constraints imposed on the discriminating and the difficulty item parameters [Bacci et al., 2014; Bartolucci et al., 2014].

References

Adams R., Wilson M., Wang W. (1997), *The Multidimensional Random Coefficients Multinomial Logit*, "Applied Psychological Measurement", No. 21, s. 1-24.

Agresti A. (1993), *Computing Conditional Maximum Likelihood Estimates for Generalized Rasch Models Using Simple Loglinear Models with Diagonals Parameters*, "Scandinavian Journal of Statistics", No. 20, s. 63-71.

Agresti A. (2002), *Categorical Data Analysis*, Wiley, New Jersey.

Akaike H. (1974), *A New Look at Statistical Model Identification*, "IEEE Transactions on Automatic Control", No. 19, s. 716-723.

Alagumalai S., Curtis D.D., Hungi N. (2005), *Our Experiences and Conclusion*, Springer, Dordrecht, The Netherlands.

Baker F., Kim S.H. (2004), *Item Response Theory*, Marcel Dekker, New York.

Bartholomew D.J. (2002), *Old and New Approaches to Latent Variable Modeling* [in:] G.A. Marcoulides, I. Moustaki (eds.), *Latent Variable and Latent Structure Models*, Quantitative Methodology Series: Methodology for Business and Management, Lawrence Erlbaum Associates, Mahwah, NJ, s. 1-14.

Bacci S., Bartolucci F., Gnaldi M. (2014), *A Class of Multidimensional Latent Class IRT Models for Ordinal Polytomous Item Responses*, "Communication in Statistics – Theory and Methods", No. 43, s. 787-800.

Bartolucci F., Bacci S., Gnaldi M. (2014), *MultiLCIRT: An R Package for Multidimensional Latent Class Item Response Models*, "Computational Statistics and Data Analysis", No. 71, s. 971-985.

Bezruczko N. (2005), *Rasch Measurement in Health Sciences*, Maple Grove, MN: Jam Press. Springer-Verlag, New York.

Birnbaum A. (1968), *Some Latent Trait Models and Their Use in Inferring an Examinee's Ability* [in:] F.M. Lord, M.R. Novick (eds.), *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, MA, s. 395-479.

Bond T.G., Fox C.M. (2013), *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Psychology Press, Hove, UK.

Czapiński J., Panek T. (red.) (2016), *Diagnoza społeczna 2015. Warunki i jakość życia Polaków (raport)*, Rada Monitoringu Społecznego, Warszawa.

Duncan O., Stenbeck M. (1987), *Are Likert Scales Unidimensional?* "Social Science Research", No. 16, s. 245-259.

Frick H., Strobl C., Leisch F., Zeileis A. (2012), *Flexible Rasch Mixture Models with Package Psychomix*, "Journal of Statistical Software", No. 48(7), s. 1-25.

Kelderman H. (1996), *Multidimensional Rasch Models for Partial-Credit Scoring*, "Applied Psychological Measurement", No. 20, s. 155-168.

Kelderman H., Rijkes J. (1994), *Loglinear Multidimensional IRT Models for Polytomously Scored Items*, "Psychometrika", No. 59(2), s. 149-176.

Lazarsfeld P.F., Henry N.W. (1968), *Latent Structure Analysis*, Houghton Mill, Boston, MA.

Likert R. (1932), *A Technique for the Measurement of Attitudes*, "Archives of Psychology", No. 140(22), s. 1-55.

Linacre J.M. (1998), *Understanding Rasch Measurement: Estimation Methods for Rasch Measures*, "Journal of Outcome Measurement", No. 3(4), s. 382-405.

Lord F.M., Novick M.R. (1968), *Statistical Theories of Mental Test Stores*, Addison-Wesley, Reading, MA.

Masters G. (1982), *A Rasch Model for Partial Credit Scoring*, "Psychometrika", No. 47, s. 149-174.

Mair P., Hatzinger R. (2007), *Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R*, "Journal of Statistical Software", No. 20(9), s. 1-20.

Martin A., Quinn K. (2006), *MCMCpack: Markov Chain Monte Carlo (MCMC) Package*, R package version 0.7-3, http://mcmcpack.wustl.edu/ (dostęp: 10.02.2015).

Panayides P., Robinson C., Tymms P. (2010), *The Assessment Revolution That Has Passed England by: Rasch Measurement*, "British Educational Research Journal", No. 36(4), s. 611-626.

Rasch G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Danish Institute for Educational Research, Copenhagen.

Rizopoulos D. (2015), *Latent Trait Models Under IRT*, https://cran.r-project.org/web/packages/ltm/ltm.pdf (dostęp: 21.01.2016).

Sagan A. (2002), *Zastosowanie wielowymiarowych skal czynnikowych i skal Rascha w badaniach marketingowych (na przykładzie oceny efektów komunikacyjnych reklamy)*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, nr 605, s. 73-92.

Schwarz G. (1978), *Estimating the Dimension of a Model*, "Annals of Statistics", No. 6, s. 461-464.

Van der Linden W., Hambleton R. (1997), *Handbook of Modern Item Response Theory*, Springer-Verlag, New York.

Wright B., Masters G. (1982), *Rating Scale Analysis*, Mesa Press, Boston.

DYCHOTOMICZNE MODELE IRT W BADANIU SKŁONNOŚCI DO OSZCZĘDZANIA POLSKICH GOSPODARSTW DOMOWYCH

Streszczenie: Ze względu na charakter zmiennych obserwowanych oraz zmiennych ukrytych w grupie modeli zmiennych ukrytych wyróżnić można: modele teorii reakcji na pozycję (modele IRT), modele klas ukrytych, analizę ukrytych profili oraz analizę czynnikową. W artykule przedstawiono dychotomiczne modele IRT, w których zakłada się, że zmienne obserwowane są dyskretne, a zmienna ukryta jest zmienną ciągłą.

Choć w literaturze najczęściej spotykane są zastosowania modeli IRT w analizach testów edukacyjnych, w artykule przedstawiony zostanie przykład ich zastosowania w badaniu społeczno-ekonomicznym. Celem artykułu jest dopasowanie najlepszego modelu IRT do analizowanego zbioru danych rzeczywistych, ocena tzw. parametrów skali oraz „ukrytej zdolności do oszczędzania" polskich gospodarstw domowych.

Słowa kluczowe: modele zmiennych ukrytych, zmienne dychotomiczne, teoria reakcji na pozycję.