EKONOMETRIA ECONOMETRICS 3(57) • 2017

ISSN 1507-3866
e-ISSN 2449-9994

Ewa Genge
University of Economics in Katowice
e-mail: ewa.genge@ue.katowice.pl

A COMPARISON OF DICHOTOMOUS IRT MODELS BASED ON CONTINUOUS AND DISCRETE LATENT TRAIT – POLISH HOUSEHOLDS' SAVING SKILLS

PORÓWNANIE DYCHOTOMICZNYCH MODELI IRT O CIĄGŁEJ I DYSKRETNEJ CESZE UKRYTEJ – UMIEJĘTNOŚĆ OSZCZĘDZANIA WŚRÓD POLSKICH GOSPODARSTW DOMOWYCH

DOI: 10.15611/ekt.2017.3.03
JEL Classification: C1, C59

Summary: Item response theory is considered to be one of the two trends in the methodological assessment of the reliability scale. Depending on the complexity of the adopted item parameterization, different types of IRT models for dichotomous items are defined. Most applications carried out in practice concern educational testing or `psychological research` and are based largely on the continuous assumption of the latent trait.The aim of this paper is to compare the estimation results of the discrete (formulated by the latent class approach) and continuous dichotomous IRT models in the analysis of Polish households' saving skills as well as to assess Poles' responses according to their ability to save money and the difficulty of the items (evaluation of the reliability of the item scale). All the computations and graphics in this paper are prepared using the `MultiLCIRT` and `ltm` packages of R.

Keywords: IRT theory, dichotomous IRT models, discrete and continuous latent trait.

Streszczenie: Teoria reakcji na pozycję (*item response theory*) zaliczana jest do jednego z dwóch nurtów teorii pomiaru znajdujących zastosowanie w analizie pozycji testowych. W literaturze najczęściej spotykane są zastosowania dychotomicznych modeli IRT w analizach testów edukacyjnych, badaniach psychologicznych czy marketingowych, w których zakłada się, że cecha ukryta pochodzi z rozkładu normalnego. Celem pracy będzie porównanie wspomnianego podejścia klasycznego ze współczesnymi dychotomicznymi modelami IRT o dyskretnej cesze ukrytej, na przykładzie danych dotyczących zdolności do oszczędzania w polskim społeczeństwie. Badania będą przeprowadzone z zastosowaniem pakietów `ltm` oraz `MultiLCIRT` programu R.

Słowa kluczowe: teoria IRT, dychotomiczne modele IRT, zmienna ukryta.

1. IRT models – introduction

Item response theory is considered to be one of the two trends in the methodological assessment of the reliability scale[1]. In item response theory (IRT), latent trait is usually measured by employing probabilistic models for responses to a set of items. Depending on the complexity of the adopted item parameterization, different types of IRT models for dichotomous items are defined.

One of the most prominent examples for such an approach (for dichotomous items) is the Rasch model, or the one-parameter logistic model [Rasch 1960] which captures the difficulty (or equivalently, easiness) of binary items and the respondent's trait level on a single common scale. According to this model, the probability to answer correctly an item depends on the respondent's ability level and item difficulty. The difficulty level is the only parameter describing the item, whereas the discrimination power is assumed to be constant across items. In the two-parameter logistic model (2PL) model, two parameters are used to describe each item corresponding to the difficulty level and the discrimination power.

Another relevant difference between IRT models, other than the complexity of the parameterization of the conditional probability of approval of an item is related to the formulation of the latent trait. We need to distinguish between the fixed-effects and random-effects approach. In the first case, every subject's latent trait level is included in the model as a fixed parameter that is estimated together with the item parameters or is somehow eliminated. In the second case, the latent trait level is considered as a realization of a random variable with a certain distribution in the population from which the observed sample has been drawn. This distribution may be continuous, typically normal, or discrete, giving rise to latent classes in the population [Bartolucci et al. 2016b, p. 66].

The main goal of the article is to compare different models in the framework of the random-effects approach under the normality assumption for the latent trait and under that of the discreteness of the latent trait in measuring money-saving skills.

2. IRT models for dichotomous data – continuous approach

We consider that the questionnaire is aimed to measure, for each individual, the level of a certain latent trait θ (ability)[2]. One of the most well-known IRT models for binary responses is the Rasch model [Rasch 1960].

Latent trait can be measured through a set of items ($j = 1, \dots m$) to which binary responses are given. Success in solving an item or agreeing with it is coded as "1", while "0" codes the opposite response.

---

[1] The second one is the theory of reliability from the perspective of the classical test theory (CTT).

[2] θ is realization of Θ. The extended notation to the case of test items measuring more than one dimension of the latent trait is presented in [Bartolucci et al. 2014; Genge 2016].

The model suggested by Rasch [1960] uses the person's ability θ and item's difficulty ϑ_j (j = 1, ... m) to model the response of each person (household) to item j:

$$p_j(\theta) = \frac{e^{\theta - \vartheta_j}}{1 + e^{\theta - \vartheta_j}}, \ j = 1, \ldots, m. \tag{1}$$

It is important to stress that both θ and ϑ_j are measured on the same scale and lie on $\Re$. The item difficulty represents the level of ability required to have a 50% probability of answering correctly (or wrongly) that item.

The second well-known model is the 2PL model [Birnbaum 1968], which generalizes the Rasch model (one-parameter logistic) by allowing items to vary not only in terms of their difficulty but also in terms of their ability to discriminate among individuals:

$$p_j(\theta) = \frac{e^{\alpha_j(\theta - \vartheta_j)}}{1 + e^{\alpha_j(\theta - \vartheta_j)}}, \ j = 1, \ldots, m, \tag{2}$$

where α_j is the discriminating parameter for item j, which measures the capacity of that item to distinguish between individuals with different ability levels. The discriminating parameter is typically assumed to be positive.

Applications of Rasch models (under continuous assumption for the latent trait) are described in a wide variety of sources, including: Baker [1985], Sagan [2002], Alagumalai et al. [2005], Bezruczko [2005], Panayides et al. [2010], Bond and Fox [2013], Christensen et al. [2013], Brzezińska [2016].

3. IRT models for dichotomous data – discrete approach

A crucial assumption characterizing the latent class IRT (LC-IRT) models concerns the discreteness of the random variable Θ, with support points $\xi_1, \ldots, \xi_u$ and weights $\pi_1, \ldots, \pi_u$. Each weight π_s ($s = 1, \ldots, u$) represents the probability that a subject belongs to class s:

$$\pi_s = p(\Theta = \xi_s), \tag{3}$$

assuming that $\sum_{s=1}^{u} \pi_s = 1$, $\pi_s \geq 0$, $s = 1, \ldots, u$.

For each subject, the manifest distribution of the response vector $\mathbf{X} = X_1, \ldots, X_m$ is given by

$$p(\mathbf{x}) = p(\mathbf{X} = \mathbf{x}) = \sum_{s=1}^{u} p(\mathbf{x} \mid \xi_s)\pi_s. \tag{4}$$

According to the assumption of local independence [Hambleton, Swaminathan 1985]:

$$p(\mathbf{x}\,|\,\xi_s)=\prod_{j=1}^{m} p_j(\xi_s)^{x_{ij}}\,[1-p_j(\xi_s)]^{1-x_{ij}}\,. \tag{5}$$

The conditional probabilities $p(X_j = x\,|\,\Theta=\xi_s)$ depend on the nature of the response variable. In the case of the binary variables the following two-parameters logistic (2PL) specification [Birnbaum 1968] may be adopted:

$$\log\frac{p(X_j \geq x\,|\,\theta)}{p(X_j < x\,|\,\theta)}=\alpha_j(\theta-\vartheta_j),\ j=1,\ldots,m,\ x=0,1, \tag{6}$$

where α_j and ϑ_j are item parameters corresponding to the discriminating power and the difficulty of the item *j*, X_j denotes the response variable for the *j*-th item of the questionnaire, with $j = 1, \ldots m$. A more parsimonious model is obtained by constraining all the discriminating parameters to be equal to one other, that is $\alpha_j = 1$ for all $j = 1, \ldots m$. In this way a Rasch type model [Rasch 1960] is specified.

As usual in the LC model, individuals do not differ within latent classes, as the same ability level ξ_s is assumed for all individuals in class *s*. Moreover, the item parameters are supposed to be constant across classes.

4. Parameter estimation

An important modeling issue is the parameter estimation. There are several approaches to the estimation of traditional dichotomous IRT models, namely conditional, full and marginal maximum likelihood have been developed under maximum likelihood estimation. The 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 the `ltm` [Rizopolous 2015] package of R applying marginal maximum likelihood estimation (MML) to estimate IRT models under the assumption of continuous distribution of the latent trait. Parameter estimation under MML assumes that objects represent a random sample from a population and their ability is normally distributed.

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

$$l_{cont}(\boldsymbol{\varphi})=\log p(\mathbf{x})=\log\int p(\mathbf{x}\,|\,\theta)f(\theta)d\theta, \tag{7}$$

where $f(\theta)$ denotes the density function of Θ which is common to all the subjects in the sample and $\boldsymbol{\varphi}$ is the vector containing all the free parameters of the model.

The parameters of the dichotomous LC-IRT class of models may be estimated by the discrete marginal maximum likelihood approach (MML-LC), making use of the EM algorithm [Dempster et al. 1977] using the `MultiLCIRT` [Bartolucci et al. 2016a] package of R. The model log-likelihood is then defined as:

$$l_{LC}(\boldsymbol{\varphi}) = \sum_{\mathbf{x}} n_{\mathbf{x}} \log[p(\mathbf{x})], \qquad (8)$$

where n_{x} is the frequency of the response configuration **x**, $p(\mathbf{x})$ is computed according to (4) and (5) as a function of $\boldsymbol{\varphi}$ and $\sum_{\mathbf{x}}$ is the sum extended to all the possible response configurations of **x** (see [Bartolucci 2007]).

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

5. Empirical analysis

The comparison analysis was illustrated on the Polish households' saving behaviour data set (a public data set available at *www.diagnoza.com*, see also [*Diagnoza społeczna 2015*]). The parameter estimation was performed using two R packages: `ltm` [Rizopolous 2015] particularly suitable for an MML estimation when the normal distribution of the latent trait is assumed and `MultiLCIRT` [Bartolucci et al. 2016a], suitable for estimating the Rasch model and the 2PL model under the assumption that the ability has a discrete distribution.

The data concern 12 binary response variables measured at the last year of the survey i.e. 2015. In total, there is complete information on n = 7399 households.

The following items 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 charges (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.

At the beginning of our analysis we fitted the original form of the Rasch model that assumes the known discrimination parameter fixed at one. After that, we investigated the two parameter logistic model (2PL model), which assumes a different discrimination parameter per item. Then, a crucial point in applying the discrete approach was the choice of the number of support points of the latent trait. The number of the latent classes (support points) was set to $s = 3$, in accordance with the previous studies carried out with the same data (see for details [Genge 2016]). Finally, both approaches based on continuous and discrete latent distribution of the latent trait were compared (based on AIC and BIC criteria, see Table 1).

In both approaches we came to the same conclusion on the basis of the MML method, that is the Rasch model, is too restrictive for these data and the inclusion of the discriminant indices in the models are unavoidable.

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

Model	LL	AIC	BIC
Rasch	–40438.74	80903.48	80993.30
2PL	–39937.9	79923.80	80089.62
LC-Rasch	–40165.15	80362.31	80472.85
LC-2PL	**–39853.52**	**79761.05**	**79947.59**

Source: own calculations in R.

The estimates of the parameters of the latent distribution for the discrete dichotomous IRT model (for LC-2PL model), chosen on the basis of likelihood ratio test as well as AIC and BIC information criteria is presented in Table 2.

Table 2. Prior probabilities and the ability levels for the LC-2PL model

	Cluster 1	Cluster 2	Cluster 3
π_s	0.704	0.246	0.050
ξ_s	–0.659	–0.188	0.313

Source: own calculations in R.

We can observe that the ability levels are increasing in order and the first class (with the lowest saving skills) has the highest prior probability (see Table 2).

However, in order to compare these results with those obtained previously under the normal distribution of the ability (for the 2PL model), the estimated support points and the item parameter estimates were standardized (see Figure 1 and Table 3).

The estimates of the item difficulty levels and discrimination parameters (for both approaches, i.e. under the normality (N) and discreteness (LC) assumption for the latent trait) together with the proportion (prop) of positive (yes) responses for each item is presented in Table 3.

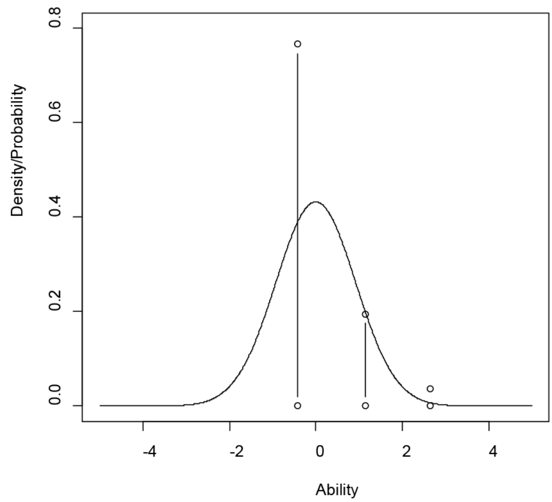

Fig. 1. Estimated distribution of the ability under the 2PL model with the assumption of normality (curve) and discreteness (bars) for the latent trait

Source: own calculations in R.

Table 3. The item parameter estimates for the 2PL (N) and LC-2PL (LC) models

Item	ϑ_j(LC)	ϑ_j(N)	α_j(LC)	α_j(N)	prop
X_1 (current needs)	1.807	**4.915**	*0.274*	*0.099*	0.380
X_2 (charges)	3.257	3.335	0.526	0.511	0.165
X_3 (food)	1.382	1.229	1.222	1.543	0.204
X_4 (home)	**5.909**	**4.769**	0.563	0.709	*0.041*
X_5 (renovation)	1.359	1.209	1.066	1.329	0.227
X_6 (treatment)	1.384	1.214	0.795	0.975	0.269
X_7 (rehabilitation)	2.859	2.223	0.964	1.339	0.089
X_8 (leisure)	1.159	1.045	**1.342**	**1.752**	0.226
X_9 (rainy day)	*-0.760*	*-0.803*	0.921	0.785	**0.635**
X_{10} (children's future)	2.048	1.689	0.812	1.053	0.186
X_{11} (old age)	0.975	0.905	0.885	1.054	0.313
X_{12} (business_develop.)) s_development)	**4.158**	**3.177**	0.773	1.059	*0.053*

Source: own calculations in R.

We observe some differences in the discriminant indices that are rather smaller under the MML method based on the normal distribution for the ability. However, a comparison among the items in terms of difficulty and discriminating power generally leads to the same results under both approaches. Leisure and current needs have the highest and the lowest discriminating power, respectively. The items X_4 (*home*) and X_{12} (*business development*) are considered to have the highest difficulty level of the item. In turn, the ninth item (*rainy day*) is consider to have the lowest difficulty level. There is one difference in the difficulty power of the items, i.e. in continuous approach item X_1 (current needs) is considered to have the strongest difficulty level (X_4 and X_{12} are in the second and third position).

As expected, there is a correspondence between the difficulty level and the proportion of correct (yes) responses especially in LC approach. The `rainy day` item has the highest success rate and then the lowest difficulty estimate. `Home` or `business_development` items have a lower success rate and then higher difficulty estimates.

6. Conclusions

IRT models are widely known in education, psychological and social sciences. However, the paper presents an application of item response models in economic analysis, which is relatively rare. One of the most popular dichotomous IRT models is the Rasch model (under the assumption of normal distribution of the latent trait) used to separate the ability of test respondents and the quality of the test. We have presented the extended dichotomous IRT models considering the assumption of the discreteness of the latent trait (giving rise to latent classes in the population).

We fitted and compared the results of the traditional and extended Rasch models as well as the two-parameter logistic models assuming the discrete and continuous distribution of the latent trait in measuring money-saving skills in Poland. We provided an illustration of latent saving skills and the performance on the items (for both approaches) as well.

Although we have received slightly better results (lower value of BIC and AIC) for the discrete approach (LC-2PL), on the basis on this real datasets analysis, we would like to stress that there is no dominating “good” model (approach) for item parameters estimation. The choice of the method should first of all depend on the research question and the context of the study. However it is worth to stress that the discrete dichotomous IRT (LC-IRT) models are more flexible in comparison with the traditional formulations of IRT models, often based on restrictive assumptions such as the normality of latent trait (explicitly introduced). Moreover, in the LC-IRT class of models, no specific assumption about the distribution of the latent trait is necessary since the latent class approach is adopted, in which the latent trait is represented by a random vector with a discrete distribution common to all subjects. In this way, subjects with a similar latent trait are assigned to the same latent class so as to detect homogeneous subpopulations of subjects (which may be useful in heterogeneous socio-economic data sets analyses).

The discrete IRT class of models allows also for introducing the assumption of multidimensionality of the latent trait allowing to take more than one latent trait into account at the same time (for ordinal polytomous responses as well) (see [Bacci et al. 2014; Genge 2016]).

Bibliography

Agresti A., 2002, *Categorical Data Analysis*, John Wiley&Sons, Hoboken, NJ.

Alagumalai S., Curtis D.D., Hungi N., 2005, *Our Experiences and Conclusion*, Springer Netherlands, pp. 343-346.

Akaike H., 1974, *A new look at statistical model identification*, IEEE Transactions on Automatic Control, 19, pp. 716-723.

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, 43, pp. 787-800.

Baker F.B., 1985, *The basic of item response theory*, College Park, MD: ERIC Clearinghouse on Assesment and Evaluation.

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

Bartolucci F., 2007, *A class of multidimensional IRT models for testing unidimensionality and clustering items*, Psychometrika, 72, pp. 141-157.

Bartolucci F., Bacci S., Gnaldi M., 2014, *MultiLCIRT: An R package for multidimensional latent class item response models*, Computational Statistics and Data Analysis, 71, pp. 971-985.

Bartolucci F., Bacci S., Gnaldi M., 2016a, *MultiLCIRT: Multidimensional latent class Item Response Theory models. R package version 2.10*, URL http://CRAN.R-project.org/package=MultiLCIRT.

Bartolucci F., Bacci S., Gnaldi M., 2016b, *Statistical Analysis of Questionnaires. A Unified Approach Based on R and Stata,* A Chapman &Hall/ CRC.

Birnbaum A., 1968, *Some latent trait models and their use in inferring an examinee's ability*, [in:] Lord F.M., Novick M.R., *Statistical Theories of Mental Test* Scores, Addison-Wesley, Reading, MA, pp. 395-479.

Bezruczko N., 2005, *Rasch measurement in health sciences*, Maple Grove, MN: Jam Press, Springer-Verlag, New York

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

Brzezińska J., 2016, *Modele IRT i modele Rascha w badaniach testowych*, [in:] Jajuga K., Walesiak M., Taksonomia 26. Klasyfikacja i analiza danych – teoria i zastosowania, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 426, Wrocław, pp. 49-57.

Christensen K.B., Kreiner S., Mesbah M., 2013, *Rasch Models in Health*, ISTE-Wiley, London – Hoboken.

Dempster A.P., Laird N.P., Rubin D.B., 1977, *Maximum likelihood for incomplete data via the EM algorithm (with discussion)*, Journal of the Royal Statistical Society, 39, ser. B, pp. 1-38.

Diagnoza społeczna 2015. Warunki i jakość życia Polaków (raport), Czapiński J., Panek T. (eds.), Rada Monitoringu Społecznego, Warszawa (2.02.2017).

Genge E., 2016, *Teoria reakcji na pozycję w podejściu modelowym w taksonomii*, Ekonometria, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 1(51), Wrocław, pp. 9-19.

Hambleton R.K., Swaminathan H., 1985, *Item Response Theory: Principles and Applications*, Kluwer Nijhoff, Boston.

Linacre J.M., 1998, *Understanding Rasch measurement: Estimation methods for Rasch measures*, Journal of Outcome Measurement, 3(4), pp. 382-405.

Mair P., Hatzinger R., 2007, *Extended Rasch Modeling: The eRm package for the application of IRT models in R*, Journal of Statistical Software, 20(9), pp. 1-20.

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

Panayides P., Robinson, C., Tymms P., 2010, *The assessment revolution that has passed England by: Rasch measurement*, British Educational Research Journal, 36(4), pp. 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.

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, 605, pp. 73-92.

Schwarz G., 1978, *Estimating the dimension of a model*, Annals of Statistics, 6, pp. 461-464.