DE GRUYTER OPEN

Folia Oeconomica Stetinensia
DOI: 10.1515/foli-2016-0032

WYDZIAŁ NAUK EKONOMICZNYCH I ZARZĄDZANIA
UNIWERSYTET SZCZECIŃSKI

LATENT VARIABLE MODELLING AND ITEM RESPONSE THEORY ANALYSES IN MARKETING RESEARCH

Justyna Brzezińska, Ph.D.

University of Economics in Katowice
Faculty of Finance and Insurance
Department of Economic and Financial Analysis
1 Maja 50, 40-287 Katowice, Poland
e-mail: justyna.brzezinska@ue.katowice.pl

Received 6 March 2016, Accepted 6 October 2016

Abstract

Item Response Theory (IRT) is a modern statistical method using latent variables designed to model the interaction between a subject's ability and the item level stimuli (difficulty, guessing). Item responses are treated as the outcome (dependent) variables, and the examinee's ability and the items' characteristics are the latent predictor (independent) variables. IRT models the relationship between a respondent's trait (ability, attitude) and the pattern of item responses. Thus, the estimation of individual latent traits can differ even for two individuals with the same total scores. IRT scores can yield additional benefits and this will be discussed in detail. In this paper theory and application with R software with the use of packages designed for modelling IRT will be presented.

Keywords: latent class analysis, latent variables, item response theory models, survey discrete survey response data, R software

JEL classification: C25, C51, C59

Introduction

Latent variable analysis has become a very popular modern statistical method and it is widely used in psychology, education, marketing and survey research. By the latent variable model we mean any model including unobserved random variables which can alternatively be thought of as random parameters. Examples include factor, item response, latent class, structural equation, mixed effects and frailty models (Rabe-Hesketh, Skrondal, 2008). Latent variable models now have a wide range of applications, especially in the presence of repeated observations, longitudinal (panel) data, and multilevel data. Latent variables are variables that cannot be measured directly but are rather inferred (through a mathematical model) from other variables that are observed (directly measured).

Latent variable analysis is a statistical method that had its roots in psychometric and education research at the beginning of the 20th century. It has a relatively long history, dating back from the measure of general intelligence by a common factor analysis to the modern statistical methods known today (Jöreskog, 1973; Keesling, 1972; Wiley, 1973). It is possible to involve latent variables in almost all kinds of regression models, where all additive error terms in regression models are latent variables because they cannot be measured nor observed directly.

One of the methods belonging to the latent class analysis is the Item Response Theory (IRT). It has a set of latent variable techniques designed to model interaction between a subject's ability and the item level stimuli (difficulty, guessing, etc.). The focus is on the pattern of responses rather than on composite or total score variables and the linear regression theory. The classical test theory (CTT) was the most popular and dominant approach till 1953 when Frederic Lord published his thesis focused on the latent trait theory. While CTT models test outcomes based on the linear relationship between true and observed score (observed score is defined as true score + error), ITR models the probability of a response pattern of an examinee as a function of the person's ability and the characteristics of the items in a test survey. Interest in Lord's work (Lord, 1953) spread quickly, and later succeeded in Allen Birnbaum's work on logistic test models (Birnbaum, 1968). In 1960 George Rasch published his book proposing several models for item responses, later Baker (Baker, 1961) proposed the comparison between logistic and normal give functions, while Lord and Novick (1968), as well as Wright (968) worked on dichotomous models. Through the 1970's and 1980's, a new group of scholars surfaced including Samejima (1969), Andrich (1978), Andersen (1977, 1980), Hambleton and Swaminathan (1985), Wright and Stone (1979), and Harris (1989). Nowadays, with the use of high-speed computers and computer software, the application of latent class models can be found in the publications of: Garrett and Zeger (2000), Vermunt (2000), Hagenaars and McCutcheon (2002), Galimberti and Soffritti (2006), Yang (2006), Collins and Lanza (2010).

Latent class analysis may be applied in a wide area of scientific fields such as psychology, behavioural science, genetics and the evaluation of diagnostic tests. In this paper we review the latent variable method and item response theory models (IRT models) in marketing research based on a survey. We also present the use of packages designed for latent variable modelling in R software.

1. Latent variable modelling

Latent variable models (Bartholomew, Knott, 1999; Skrondal, Rabe-Hesketh, 2004) constitute a general class of models suitable for the analysis of multivariate data. In principle, latent variable models are multivariate regressions models that link continuous or categorical responses to unobserved covariates. The basic assumption and objectives of latent variable modelling can be summarized as follows (Bartholomew, Steele, Moustaki, Galibraith, 2002):

1. A small set of variables is assumed to explain the interrelationships in a set of observed response variables. This is known as the conditional independence assumption, which postulates that the response variables are independent given the latent variables.
2. Unobserved variables, which cannot be measured by conventional means, can be quantified by assuming latent variables.
3. Latent variable modelling is also used to assign scores to sample units in the latent dimensions based on their responses. This score, also known as factor score, is a numerical value that indicates a person's relative spacing or standing on a latent variable. Factor scores may be used either to classify subjects or in the place of the original variables in a regression analysis, provided that the meaningful variation in the original data has not been lost.

Latent variables can be included in statistical models based on latent variables with different goals: representing the effect of unobservable covariates (factors) and then accounting for the unobserved heterogeneity between subjects (latent variables are used to represent the effect of these unobservable factors), accounting for measurement errors (the latent variables represent the "true" outcomes and the manifest variables represent their "disturbed" versions), and summarizing different measurements of the same (directly) unobservable characteristics (e.g., quality-of-life), so that sample units may be easily ordered/classified on the basis of these traits (represented by the latent variables). Latent variable models have now a wide range of applications, especially in the presence of repeated observations, longitudinal (panel) data, and multilevel data.

2. Item Response Theory models

The Item Response Theory (IRT) is a statistical method that distinguishes the latent trait (ability) of a participant from the difficulty of a set of items with well-correlated response patterns. Item Response Theory models were formally presented by Lawley (1943) who introduced IRT as a measurement theory and later developed by Rasch (1960) to measure ability and to devise tests for the military. Fumiko Samejima (1969) developed graded response models in IRT for polytomous IRT models that deal with Likert-scale data and other tests with ordered multiple response options for each item (DeMars, 2010). Gerhard Fischer extended Rasch's binary or dichotomous model so as to handle with polytomous data in the linear logistic latent trait model. The Item Response Theory (IRT) considers a class of latent variable models that link mainly dichotomous and polytomous manifest (response) variables to a single latent variable (Bock, 1997; van der Linden, Hambletown, 1997; Baker, Kim, 2004). The wide area application of item response theory models can be mostly found in educational and psychological testing. In such surveys the researcher is interested in the measurement of examinees' ability with the use of a test consisting of several questions.

The main goal of the IRT method is to provide a framework for evaluating how well assessments work, and how well individual items on assessments work. The most common application of IRT is in education, where psychometricians use it for developing and designing exams, maintaining banks of items for exams, and equating the difficulties of items for successive versions of exams (for example, to allow comparisons between results over time) (Hambleton, 1991).

IRT models are treated as latent trait models due to their use of latent variables. The latent variables are used to emphasize that discrete item responses are taken to be observable manifestations of hypothesized traits, constructs, or attributes, not directly measured nor observed. Statistical models based on latent variables were developed in the field of psychology, education and sociology. Item response theory models can be used in assessments and evaluation research to explain how respondents or participants of the survey respond to items (questions). IRT assumes that examinees respond to an item according to their ability and the items difficulty. IRT models are built and based on the fundamental that the probability of a subject's certain reaction to a stimulus can be described as a function characterising the subject's location on a latent trait plus one or more parameters characterising the stimulus (Fox, 2007).

IRT consists of a set of models that describe the interactions between a person and the test items. Persons may possess different traits and instruments may be assigned to measure more than one trait and these models are referred to as unidimensional IRT. In an educational testing situation in which n individuals answer I questions for items. For j = 1, ..., n and i = 1, ..., I, let Y_{ij} be random variables associated with the response of individual j to item i. These respondents may be binary (correct or incorrect answer) or may be discrete with a number of categories. Let Ω_Y denote the set of possible values of the Y_{ij}, assumed to be identical for each item in the test. Let θ_j denote the latent trait of ability for individual j, and let η_i denote a set of parameters that will be used to model item (question) characteristics. Different IRT models arise from different sets of possible responses Ω_Y and different functional forms assumed to describe the probabilities with the Y_{ij} assuming those values, namely:

$$P\left(Y_{ij} = y \middle| \theta_j, \eta_i\right) = f\left(y \middle| \theta_j, \eta_i\right); \quad y \in \Omega_Y \tag{1}$$

The item parameters η_i may include four distinct types of parameters: a discrimination parameter a_i, a difficulty parameter b_i, a guessing parameter c_i, and a carelessness parameter d_i. 1 the parameter IRT model is the Rasch model for dichotomous items defined as (Rasch, 1960):

$$P\left(y_{ij} = 1 \middle| \theta_j, b_i\right) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)} \tag{2}$$

where b_i ($-\infty < b_i < \infty$) is difficulty (location, threshold) parameter. The sign of expression $\theta_j - b_i$ in any particular instance indicates the probable outcome the person-item interaction. If $\theta_j - b_i > 0$ then the most probable outcome is a correct response. If $\theta_j - b_i < 0$ then the most likely outcome is an incorrect response.

Another IRT model is the 2-parameter Birnbaum model defined as follows:

$$P\left(y_{ij} = 1 \middle| \theta_j, a_i, b_i\right) = \frac{\exp a_i(\theta_j - b_i)}{1 + \exp a_i(\theta_j - b_i)} \tag{3}$$

where a_i ($-\infty < a_i < \infty$) is a discrimination (slope) parameter. This parameter is related to how rapidly the probability in equation (3) changes with the changes in ability θ_j.

Another 3-parameter Birnbaum IRT model:

$$P\left(y_{ij} = 1 \middle| \theta_j, a_i, b_i, c_i\right) = c_i + (1 - c_i)\frac{\exp a_i(\theta_j - b_i)}{1 + \exp a_i(\theta_j - b_i)} \tag{4}$$

where c_i ($0 \le c_i \le 1$) is a guessing parameter.

There is also the 4-parameters model defined as:

$$P\left(y_{ij} = 1 \middle| \theta_j, a_i, b_i, c_i, d_i\right) = c_i + (d_i - c_i)\frac{\exp a_i(\theta_j - b_i)}{1 + \exp a_i(\theta_j - b_i)} \tag{5}$$

with the carelessness parameter d_i ($0 \leq c_i < d_i \leq 1$), but this model is not used very often in survey research due to its complexity and number of parameters included.

The most popular application of IRT models can be found in education and psychological testing in which the researcher measures examinees' ability using a test consisting of several different items. The IRT can be characterized by a large number of advantages in comparison to the well-known and established classical test theory (CTT). IRT models yield invariant item and latent trait estimates (within a linear transformation), standard errors conditional on trait level, and trait estimates anchored to item content. Such models also facilitate the evaluation of differential item functioning, inclusion of items with different response formats in the same scale, and assessment of person fit and is ideally suited for implementing computer adaptive testing.

3. Application of latent class models in R

The availability of computer software and statistical packages for the analysis of latent class models and IRT is rising nowadays due to technological development. Researchers can choose among different statistical software, commercial, as well as free of charge. For IRT modelling in R software we can use libraries such as: `ltm`, `eRm`, `mlirt`, `gpcm`, `MCMCpack`, `mirt` and `lme4`.

In this paper we present the use of latent class analysis using the `values` dataset available in the `poLCA` library in R. Survey responses and which are based on dichotomous data from 216 respondents giving answers to four questions (A, B, C, D) which measure the tendencies among "universalistic" or "particularistic" values (Goodman, 1974; Stouffer, Toby, 1951). Data for the analysis are presented in the form of a data frame consisting of 216 observations on 4 variables (1 denoting the "particularistic" values response and 2 denoting the "universalistic" values response). For latent class analysis we use the `poLCA` function.

We built the following models without covariates: M_0: log-linear independence model, M_1: two-class latent class model, M_2: three-class latent class model.

Table 1. Goodness of fit statistics for latent class models

Model	G^2	χ^2	AIC	BIC	Estimated class population shares
M_0	81.084	104.107	1,095.300	1,108.801	1
M_1	2.719	2.719	1,026.935	1,057.313	0.2792; 0.7208
M_2	0.387	0.423	1,034.602	1,081.856	0.6713; 0.1944; 0.1343

Source: own calculations in R.

Using the `poLCA` package we obtained a number of goodness of fit statistics. We can see that the minimum AIC and BIC criteria both indicate that the M_1 model which is a two latent class with AIC equal 1,026.935 and BIC equal 1,057.313. For this model the estimated class population shares are 0.2792 and 0.7208 respectively.

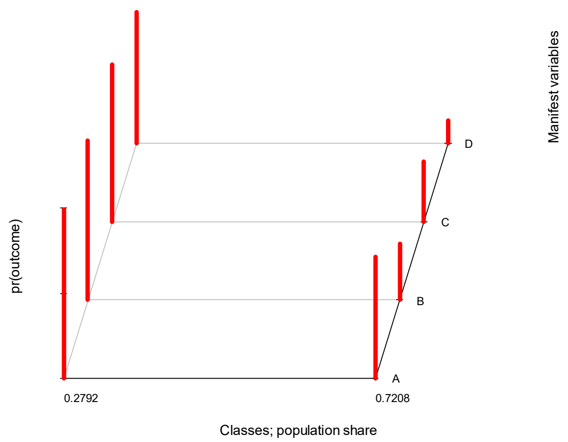

Figure 1. Estimation of the two-class latent class models

Source: own calculations in R.

Looking at Figure 1 we can see that the red bars represent the conditional probabilities, by latent class of being labelled A through D by each of the four values. The greater the bar is, then closer it is to the 1 conditional probabilities of a positive rating.

The R package `poLCA` presented in the paper provides a user friendly and easy to use framework that can be applied for the estimation of latent class models, as well as latent class regression models for the analysis of multivariate categorical data. The `poLCA` package is based on the expectation-maximization (EM algorithm) and Newton-Raphson algorithms to find the maximum likelihood estimates of the parameters of the latent class. It also includes a tool for observing the iterative parameter estimation process and presenting the results graphically.

4. Application of Item Response Theory in R

In this part of the paper we present the use of Rasch IRT models in R with the use of the `ltm` package. As an example we use the `LSAT` (Law School Administration Test) based on a survey of 100 respondents answering 5 test questions (Bock, Lieberman, 1970).

Table 2. Difficulty, SE and probability of giving a correct answer for the Rasch model

Item	Difficulty parameter	SE	Discrimination parameter	$P(x = 1\|z = 0)$
1	2.87	0.129	1	0.946
2	1.06	0.082	1	0.743
3	0.26	0.077	1	0.564
4	1.39	0.087	1	0.800
5	2.22	0.105	1	0.902

Source: own calculations in R.

Looking at the information criteria we can see that AIC is 4,956.108, BIC is 4,980.646. We can also see that out of all items, the most difficult is item 1 with the difficulty parameter 2.87 and the easiest one is item 3 with the difficulty parameter equal 0.26.

Secondly, we can build an unconstrained Rasch model. Results of the estimation are presented in Table 3.

Table 3. Difficulty, SE for an unconstrained Rasch model

Item	Coefficient	SE
1	–3.615	0.327
2	–1.322	0.142
3	–0.318	0.098
4	–1.739	0.169
5	–2.780	0.251
Discrimination parameter	0.755	0.069

Source: own calculations in R.

For an unconstrained Rasch model AIC is 4,945.875, BIC is 4,975.322 and the discrimination parameter is equal 0.755.

To compare both Rasch models we can use the ANOVA function. The results are in Table 4.

Table 4. Likelihood ratio table

Item	AIC	BIC	Log-likelihood	LRT	p-value
Constrained Rasch model	4,956.11	4,980.65	–2,473.05		
Uncinstrained Rasch model	4,945.88	4,975.32	–2,466.94	12.23	<0.001

Source: own calculations in R.

The LRT value indicates that an unconstrained model fits better than a constrained model. For this model we present the Item Characteristic Curve (ICC), and Item Information Curve (IIC) (Figure 2).

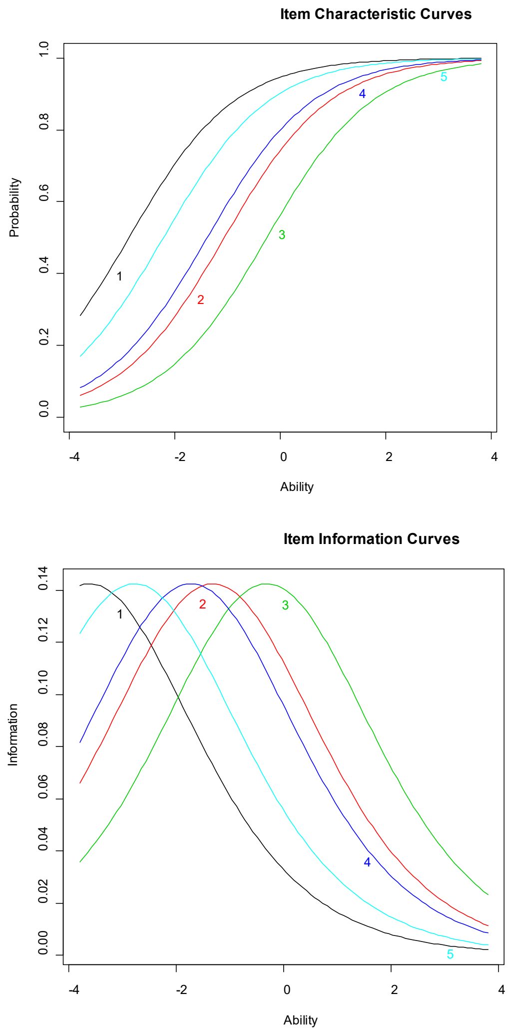

Figure 2. Item Characteristic Curve (ICC), and Item Information Curve (IIC)

Source: own calculations in R.

The Item characteristic curve describes the relationship between a latent ability and the performance on a test item. Each item corresponds with the information functions, which shows the relationship between the information task and the level of knowledge. In Figure 2 we can see that the position in the research provides information for respondents with low ability. Information for ability level ranges from –4 to 0 and covers almost 60% of the total information. The position of distinguishing respondents with higher levels of ability is only 1/3.

Conclusions

Latent variable modelling comprises an important set of techniques for a broad range of fields, especially in the presence of repeated observations, longitudinal (panel) data, and multilevel data. In this paper we have presented the advantages of a latent variable analysis including Item Response Modelling. A latent variable is a variable which is not directly observable and is assumed to affect the response variables (manifest variables).

In this paper we have presented latent variable modelling and item response theory models using R. We focused on parameter estimates and the graphical presentation of the study. For latent variables we used the `poLCA` package. We built and compared some latent variables models and with the use of information criteria and we indicated one best-fitting model. We also presented graphically the probability estimates for a selected model. In the item response theory analysis we showed the use of the `ltm` package. We constructed an unconstrained and constrained Rasch model and we presented an item characteristic curve, and item information curve for the best-fitting model. The conducted analysis shows the application of methods with the use of latent variables, together with their advantages and possible outcomes. Finally it should encourage a wide range of scientists to use such an analysis in survey research.

References

Andersen, E.B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*, 69–81.

Andersen, E.B. (1980). *Discrete statistical models with social science applications*. Amsterdam: North-Holland.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573.

Baker, F.B. (1961). Empirical comparison of item parameters based on the logistic and normal functions. *Psychometrika*, *36*, 239–246.

Baker, F., Kim, S.H. (2004). *Item Response Theory*. 2nd edition. New York: Marcel Dekker.

Bartholomew, D., Knott, M. (1999). *Latent Variable Models and Factor Analysis*. 2nd edition. London: Arnold.

Bartholomew, D., Steele, F., Moustaki, I., Galbraith, J. (2002). *The analysis and interpretation of multivariate data for social scientist*. London: Chapman & Hall.

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

Bock, R.D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practices*, *16*, 21–33.

Bock, R., Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items, *Psychometrika*, *35*, 179–197.

Collins, L.M, Lanza, S.T. (2010). *Latent class and latent transition analysis for the social, behavioral, and health sciences*. New York: Wiley.

De Boeck, P., Wilson, M. (2004). *Explanatory Item Response Models: A Generalized and Nonlinear Approach*. New York: Springer-Verlag.

Fox, J.P. (2007). Multilevel irt modelling in practice with the package mlirt. *Journal of Statistical Software*, *20* (5). Available at: http://www.jstatsoft.org.

Galimberti, G., Soffritti, G. (2006). Identifying multiple cluster structures through latent class models. In: M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nürnberger, W. Gaul (eds.), *From data and information analysis to knowledge engineering* (pp. 174–181). Berlin: Springer.

Garrett, E.S., Zeger, S.L. (2000). Latent class model diagnosis. *Biometrics*, *56*, 1055–1067.

Goodman, L.A. (1974). Exploratory Latent-Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika*, *61* (2), 215–231.

Hagenaars, J.A., McCutcheon, A.L. (2002). *Applied latent class analysis*. Cambridge: Cambridge University Press.

Hambleton, R.K. (1991). *Fundamentals of item response theory* (Vol. 2). Sage publications.

Hambleton, R.K., Swaminathan, H. (1985). *Item response theory: Principles and Applications*. Boston, MA: Kluwer Academic Publishers.

Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice*, *8*, 35–41.

Joreskog, K.G. (1973). A general method for estimating a linear structural equation system. In: A.S. Goldberger, O.D. Duncan (eds.), *Structural equation models in the social sciences* (pp. 5–112). New York: Seminar Press.

Keesling, J.W. (1972). *Maximum Likelihood Approach to Causal Analysis*. Unpublished Ph.D. Dissertation, University of Chicago.

Lawley, D.N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Statistical Society of Edinburgh*, *61*, 273–287.

Lord, F.M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, *13*, 517–548.

Lord, F.M., Novick, M.R. (1968). *Statistical Theories of Mental Test Scores (with contributions by A. Birnbaum)*. Reading, MA: Addison-Wesley.

Rabe-Hesketh, S., Skrondal, A. (2008). Classical latent variable models for medical research. *Statistical Methods in Medical Research*, *17*, 5–32, Sage.

Rasch, G. (1960). Probabilistic models for some intelligence and achievement tests. *Copenhagen: Danish Institute for Educational Research*.

Samejima, F. (1969). Calibration of latent ability using response pattern of graded score. *Psychometrika Monograph Supplement*, *17*.

Skrondal, A., Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman & Hall.

Stouffer, S.A., Toby, J. (1951). Role conflict and personality. *American Journal of Sociology*, *56*, 395–406.

van der Linden, W.J., Hambleton, R.K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.

Vermunt, J.K., Magidson, J. (2000). *Latent GOLD's User's Guide*. Boston: Statistical Innovations Inc.

Wiley, D.E. (1973). The Identification Problem for Structural Equation Models with unmeasured Variables. In: A.S. Goldberger, O.D. Duncan (eds.), *Structural Equation Models in the Social Sciences* (pp. 69–83). New York: Academic Press.

Wright, B.D. (1968). Sample-free test calibration and person measurement. In: *Proceedings of the 1967 Invitational Conference on Testing Problems* (pp. 85–101). Princeton, NJ: Educational Testing Service.

Wright, B.D., Stone, M.H. (1979). *Best test design*. Chicago: MESA Press.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. C*omputational Statistics & Data Analysis*, *50*, 1090–1104.