

Using Multidimensional Item Response Theory Models to Explain Multi-Category Purchases

By Nadine Schröder

We apply multidimensional item response theory models (MIRT) to analyse multi-category purchase decisions. We further compare their performance to benchmark models by means of topic models. Estimation is based on two types of data sets. One contains only binary the other polytomous purchase decisions. We show that MIRT are superior w. r. t. our chosen benchmark models. In particular, MIRT are able to reveal intuitive latent traits that can be interpreted as characteristics of households relevant for multi-category purchase decisions. With the help of latent traits marketers are able to predict future purchase behaviour for various types of households. These information may guide shop managers for cross selling activities and product recommendations.

1. Introduction

The analysis of multi-category purchases has been of interest for marketers for a long time and still is of increasing importance today. Multi-category purchases show up wherever households are faced with a situation in which



Nadine Schröder is Assistant Professor of Marketing at the University of Regensburg, Universitätsstraße 31, 93040 Regensburg, Germany, Phone: +49 941 943 2274, Fax: +49 941 943 2828, E-Mail: Nadine.schroeder@wiwi.uni-regensburg.de.

Please note: The author would like to thank two anonymous referees and the editor of the journal for their helpful and constructive comments.

they are to pick one or various products out of a set of different products or product categories. Such situations are typical for visits to traditional stores like supermarkets or retail shops. The same applies however as well when considering online purchase behaviour where households are not only faced with different products or brands of one retailer but can usually choose across many vendors. For marketers it is not only important to understand what drives households in their process of decision but also how they can induce further purchases.

Previous research on multi-category purchases can be classified according to exploratory or explanatory market basket analyses (e. g., Boztuğ und Silberhorn 2006; Mild and Reutterer 2003). Market baskets consist of product (categories) which have been purchased by a household. The majority of authors use either multivariate probit (e. g., Duvvuri et al. 2007) or logit models (e. g., Dippold and Hruschka 2013) when dealing with an explanatory market basket analysis. These models can deal with correlated binary dependent variables, i. e., the observed choices of a household when deciding on purchasing a product from a certain product category. Bivariate relationships between categories are captured with a model coefficient in multivariate logit models and with the error correlation matrix in multivariate probit models respectively. Hruschka (2014) on the other hand introduces a new approach, namely topic models, in order to describe multi-category purchase behaviour. His approach admits that a market basket and hence an observed purchase decision may be linked to several topics, i. e. latent activities. In his study he examines multi-category purchase behaviour for 60 different product categories offered by a medium-sized supermarket. Purchases from these categories can be linked to ten different latent activities. Whereas one latent activity consists of, e. g., beverages, periodicals, and cigarettes on the one hand another one can be linked to baking as it contains product categories like baking ingredients, sugar, and flour etc. The idea of topic models is comparable to a factor analytic approach which also tries to detect latent factors that are responsible for observed behaviour (Böcker 1975). So far, only few studies have employed factor analytic models when dealing with multi-category purchases. Böcker (1975), e. g., uses a factor analysis to analyse multi-category purchase behaviour of households of a medium-sized household supply store. In total, he investigates purchases

from eight product categories that consist of, e. g., household goods, glassware, giftware, and electrical goods. Instead of using information on purchase incidences regarding the various product categories he focuses on sales in order to meet with prerequisites of a factor analysis. He finds four different factors that can be interpreted as sources for the purchases or latent traits of the households. He labels factor one that consists of household goods, glassware, and electrical goods as kitchen articles. Factor four on the other hand consists of only one product category, namely kitchen and garden furniture. In consequence Böcker (1975) identifies the factor as furniture. Apart from Böcker (1975) there are as well some other publications that use factor analytic structures in their models to investigate interdependencies between products. Hansen et al. (2006) use a logit model on a data set from a retail chain and analyse purchase behaviour among ten different product categories comprising food and non-food related products. The factor analytic structure is placed on the covariance matrix of the model coefficients. This way, the dependence between product categories is captured and factors are interpreted as household specific latent traits. The authors find that two factors best describe the relationships. They interpret one factor as an indicator for preferences of store brands and the remaining as indicator for price sensitivity.

The small number of factor analytic models w. r. t. multi-category purchases can be explained by data characteristics. According to Bartholomew et al. (2008, p. 178) factor analytic models assume metric scales for observed variables. However, in multi-category purchases a market basket displays which and how many products a household has purchased within various product categories during a transaction. Typically, such a basket contains a lot of zeros (for product categories where no purchase has taken place) and ones (for product categories with purchases). Hence, data are usually binary or polytomous if more than one product has been purchased per product category. As an example for the binary scenario, consider a supermarket that offers ten different product categories. If a household purchases at least one product from the first two of the ten product categories the respective row in the data set looks like this: (1,1,0,0,0,0,0,0,0,0). Thus, what would be needed is a factor analytic model that can deal with a binary or polytomous data set.

(Multidimensional) item response theory models ((M)IRT) stem from, e. g., educational and psychological testing situations (Reckase 2009, p. 57). They serve to analyse the relationship between a respondent's underlying latent trait and an item level stimulus (Chalmers 2012). These models are applied to data sets that collect responses of respondents to a set of questions, i. e., items. Responses can be binary (yes / no) or polytomous if a multiple choice question has been posed. MIRT estimate probabilities of answering a question correctly. Depending on the respondent's ability or latent trait the probability changes (Reckase 2009, p. 84). Today, many

different reasons for applying MIRT exist. These range from assessing construct validity to the analysis of attitudes (Embretson and Reise 2000, pp. 273, 307). As an example, Bartholomew et al. (2008, p. 226–228) use a data set from the British Social Attitudes Survey in 1991 in which respondents were asked ten questions on matters like whether homosexuals should hold public positions or whether homosexual couples should be allowed to adopt children. Based on the answers the authors are able to identify two latent traits that relate to public matters on the one hand and private behaviour on the other hand. Within the area of marketing research applications of MIRT have been limited apart from a few exceptions. For instance, de Jong et al. (2008) apply them for measuring extreme response styles which can be a problem in survey-based marketing research. Extreme response styles result when respondents either tend to only mark end points of a rating scale or avoid these answer categories completely. Based on an IRT the authors are able to study the origins of extreme response styles. To the best of the authors' knowledge MIRT have not been applied to analyse multi-category purchases.

The following example shows that data structure resulting from a survey is similar to that of multi-category purchases. Consider a questionnaire that consists of ten questions to which respondents can either answer "yes" or "no" (binary scenario). If a respondent agrees on the first two questions of such a survey, the answer can be captured in the same way as the decision of a household that purchases from the first two of ten different product categories. Due to this, (M)IRT are able to deal with market basket data as well. In applying MIRT for analysing multi-category purchase decisions of households we use their similarity to factor analytic models and perform a market basket analysis. This enables us to identify latent traits of households. Based on these results we predict purchase behaviour. When using MIRT to analyse multi-category purchase behaviour we assume that the multi-category purchase decision of a household is equivalent to a respondent who decides on answering questions of a questionnaire. Hence, in our multi-category purchase scenario we call an item a product category.

We add to the literature in the following way: We apply MIRT in two contexts of multi-category purchase decisions (a binary and a polytomous scenario) which to the best of our knowledge is the first time it has been applied in this area of research. We show that by MIRT's similarity to factor analytic models we are able to detect intuitive latent traits for both scenarios which can then be related to observable purchase decisions and can be used to predict multi-category purchase behaviour. We further assess how MIRT perform in comparison to topic models which also assume a latent trait behind the multi-category purchase behaviour. Finally, we show how marketers can benefit from knowing the latent traits of households by giving examples how future purchases can be induced.

In what follows, we introduce to (M)IRT and their link to factor analysis. We then describe our two models that are used for estimation. We conclude the section by briefly sketching the process of estimation and topic models which are used as our benchmark models. Section three informs about the data set. Results are discussed in section four. The final section provides information relevant for managerial decisions and concludes the article.

2. Method

2.1. Link to factor analysis

Since we apply MIRT as binary and polytomous factor analytic models respectively, we begin by addressing their similarities with the traditional factor analysis. According to Reckase (2009, p. 63) MIRT can be seen as a special case of, e. g. binary factor analysis. Traditional, i. e. linear factor analytic models assume manifest variables to be metric whereas in situations with categorical variables one speaks of IRT (e. g., Raykov and Calantone 2014). IRT usually assume unidimensionability, i. e., different manifest variables (answers to test items) are based on one latent trait. MIRT on the other hand allow more than one latent trait. This way, they may also serve as a factor analytic model that deals with binary or polytomous data. According to Bartholomew et al. (2008, pp. 212–213) the general linear factor analysis model can be written as

$$x_{ij} = d_j + \alpha_{j1}\theta_{1i} + \alpha_{j2}\theta_{2i} + \dots + \alpha_{jT}\theta_{Ti} + \varepsilon_{ij} \quad (1)$$

and assumes an unobservable decision process of household i on product category j . Whereas x_{ij} shows the observed purchase decision $\theta_i = (\theta_{1i}, \dots, \theta_{Ti})$ is defined as a vector of $t = 1, \dots, T$ latent traits or factors. $\alpha_j = (\alpha_{j1}, \dots, \alpha_{jT})$ is a vector of the corresponding unrotated factor loadings and d_j a constant term. Residuals ε_{ij} are assumed to be $N(0, \sigma^2)$. The decision of making a purchase in a certain product category can be explained by latent traits of households. Whereas x_{ij} are metric in the general linear factor model they are binary (or polytomous) when used w. r. t. (M)IRT. To calculate the probability for household i of purchasing from product category j a logit link function is employed. Thus, Eq. (1) changes into

$$\text{logit } \pi_{ij}(\theta) = d_j + \alpha_{j1}\theta_{1i} + \alpha_{j2}\theta_{2i} + \dots + \alpha_{jT}\theta_{Ti}. \quad (2)$$

Once α_{jt} are standardized according to Eq. (3)

$$\text{st}\alpha_{jt} = \frac{\alpha_{jt}}{\sqrt{1 + \sum_{t=1}^T \alpha_{jt}^2}} \quad (3)$$

they can be interpreted as correlations between latent traits and observed purchase decisions (Bartholomew et al. 2008, p. 225).

2.2. MIRT

Researchers usually use (M)IRT to analyse the relationship between a respondent's characteristic and the feature of test items (Reckase 2009, p. 68). They can as a next step be used to predict responses to items, i. e., product categories based on a person's characteristic (Lord 1980, p. 11). We further differentiate whether a household only decides whether or not to buy from a certain product category (binary scenario) or how often she or he buys from that product category (polytomous scenario).

2.2.1. Binary Scenario

We use a multidimensional extension of the two-parameter logistic model (MIRTbin). This model assumes that the observed purchase decision can be explained by a combination of latent traits of households. It belongs to compensatory models since a low level of a certain latent trait can be compensated for by a high level of another latent trait (Reckase 2009, p. 87). The probability of household i to purchase from product category j results from Eq. (2) and is given by

$$P(x_{ij} = 1 | \theta_i, \alpha_j, d_j) = \frac{1}{1 + \exp(-(\alpha_j'\theta_i + d_j))}. \quad (4)$$

In MIRT d_j is a product category intercept which can be derived by looking at the purchase probability of a product category being equal to .5. In such a case the exponent of \exp is zero. In traditional IRT it is called the difficulty parameter. However, for MIRT one can only assess the relative difficulty of a certain product category l which is given by $-d_j/\alpha_j$. The product category slopes $\alpha_j = (\alpha_{j1}, \dots, \alpha_{jT})$ are also called discrimination parameters. The larger the value (in absolute terms), the better that product category differentiates across households (Reckase 2009, pp. 89–90).

2.2.2. Polytomous scenario

In the polytomous scenario we estimate a multidimensional graded response model (MIRTpol) that consists of sequential MIRTbins. Eq. (5) shows the probability of household i for purchase frequency $c \in (0, 1, \dots, C_j - 1)$ w. r. t. product category j :

$$P(x_{ij} = c | \theta_i, \alpha_j, d_j) = P(x_{ij} \geq c | \theta_i, \alpha_j, d_j) - P(x_{ij} \geq c + 1 | \theta_i, \alpha_j, d_j). \quad (5)$$

The boundaries of response probabilities for model MIRTpol are:

$$\begin{aligned} P(x_{ij} \geq 0 | \theta_i, \alpha_j, d_j) &= 1, \\ P(x_{ij} \geq 1 | \theta_i, \alpha_j, d_j) &= \frac{1}{1 + \exp(-(\alpha_j'\theta_i + d_1))} \\ P(x_{ij} \geq 2 | \theta_i, \alpha_j, d_j) &= \frac{1}{1 + \exp(-(\alpha_j'\theta_i + d_2))} \\ &\dots \\ P(x_{ij} \geq C_j | \theta_i, \alpha_j, d_j) &= 0. \end{aligned} \quad (6)$$

Parameters of the model can be interpreted in the same manner as in the binary scenario. However, d_j now is a vector of $d_j = (d_{j1}, \dots, d_{j(C-1)})$ intercepts referring to product category j (Chalmers 2012).

2.3. Process of estimation

2.3.1. MIRT

We estimate our models with the MIRT software package (Chalmers 2012) in R (R Core Team 2016). We vary the amount of latent traits from one to five for MIRTbin and MIRTpol. Estimation is carried out via expectation-maximisation-(EM-)method (Bock and Aitken 1981). The E-step is performed with Gauss-Hermite quadrature. To account for the possible inaccuracy in high dimensional models a quasi-Monte Carlo EM (QMCEM)-estimation is used when dealing with models of at least four latent traits (Chalmers 2012). The algorithm of the EM-/QMCEM-procedure stops when parameter changes are less than .0001. Further information on the EM-/QMCEM-procedure can be found in, e. g. Bock et al. (1988) and Chalmers (2012).

2.3.2. Topic models

As stated in section one we do not only want to demonstrate that MIRT are able to describe multi-category purchase decisions. In addition, we want to assess how they perform with a benchmark model by means of a topic model. We choose a topic model because it has been successfully applied to analyse multi-category purchases. Like MIRT topic models are able to reduce purchases across many product categories into a smaller set of latent traits and are hence comparable to factor analytic models (Atkins et al. 2012). In this case topics can be interpreted in the same manner as latent traits. To the benefit of the reader we refer to topics as latent traits as well.

We estimate topic models based on a binary (TOPICbin) and on a polytomous scenario (TOPICpol) similar to our procedure regarding MIRT. Both types of topic models are based on latent Dirichlet allocations (LDA) which according to Hruschka (2014) prove to be superior over a correlated topic model. The LDA estimates the probability

$$p_{ij} = \sum_{t=1}^T \theta_{jt}^* \delta_{it} \quad (7)$$

It is the probability that market basket of household i contains product category j . θ_{jt}^* and δ_{it} both come from a Dirichlet distribution. θ_{jt}^* captures the importance of product category j for latent trait t and can hence be compared to θ_j in MIRT. δ_{it} on the other hand captures the importance of a latent trait t for a market basket of household i (Hruschka 2014). We vary the number of topics between two and five in analogy to the amount of latent traits of MIRTbin and MIRTpol. Estimation is carried out with a variational EM-algorithm as documented by Grün and Hornik (2011). Readers with further interests

in topic models are advised to consult Hruschka (2014) for an application in multi-category purchase behaviour or e. g., Blei (2012) for general text mining applications.

3. Data

We use an IRI data set as discussed by Bronnenberg et al. (2008)[1]. Our original data set of a single store in 2001 contains 8,531 weekly transactions of 1,237 households across 31 different product categories. The median number of purchased product categories per transaction is two. To achieve a higher median we aggregate purchases across households on a yearly basis, i. e. we sum in how many weeks w household i purchases at least one item from any of the product categories j , i. e. $\sum_{w=1}^{W_i} data_{ij}$. This results in a data set of $n = 1,237$ observations and in a median number of purchased product categories per transaction of five. We can use this data set to create two different data sets, i. e., a polytomous and a binary scenario.

3.1. Binary Scenario

In the binary scenario we only observe whether a household has made at least one purchase from a product category or not. We therefore only observe if $\sum_{w=1}^{W_i} data_{ij}$ for each household i and product category j is larger than zero. Hence our binary data set (Bdata) can be displayed as:

$$Bdata_{ij} = \begin{cases} 0, & \text{if } \sum_{w=1}^{W_i} data_{ij} = 0 \\ 1 & \text{else} \end{cases} \quad (8)$$

Column (3) in *Tab. 1* shows relative purchase frequencies of the respective product category according to Bdata. Thus, 66.69 % of the households made at least one purchase of milk during the observation period. As stated earlier, (M)IRT are typically applied to survey or test situations. Data set Bdata can thus be considered as answers (yes or no) of our 1,237 households to their purchase behaviour during a year. In particular, a household could have responded to questions of a type “have you made at least one purchase from product category j in the past year?”.

3.2. Polytomous Scenario

In the polytomous scenario we try to use the frequency of purchases of each household i in product category j , i. e., $\sum_{w=1}^{W_i} data_{ij}$ as far as possible. We therefore use the distribution of $\sum_{w=1}^{W_i} data_{ij}$, in particular third quartiles q_j , for each of the 31 product categories. These vary between zero and three. We use this information in order to construct the polytomous data set (Pdata) in the following way: if no purchase has taken place Pdata_{ij} is zero. Pdata_{ij} equals one if the third quartile of a product category is equal to zero and a purchase has taken place or alternatively if the third quartile is larger than zero and exactly one purchase has taken place, etc.

Categories	Abbreviation	Relative purchase frequencies (based on Bdata)	Maximum per product category <i>j</i> and its correspondence to the five-point-response-scale (based on Pdata)
Milk	Milk	.6669	4 – almost always
Carbonated beverages	Carbbev	.5578	3 – often
Salty snacks	Saltsnck	.5004	3 – often
Cold cereal	Coldcer	.4293	2 – sometimes
Soup	Soup	.4002	2 – sometimes
Toilet tissue	Toitisu	.3323	2 – sometimes
Spaghetti/Italian sauce	Spagsauc	.3104	2 – sometimes
Coffee	Coffee	.3072	2 – sometimes
Yogurt	Yogurt	.2967	2 – sometimes
Mayonnaise	Mayo	.2724	2 – sometimes
Margarine/spreads/ butter blends	Margbutr	.2700	2 – sometimes
Paper towels	Paptowl	.2700	2 – sometimes
Facial tissue	Factiss	.2668	2 – sometimes
Laundry detergent	Laundet	.2668	2 – sometimes
Peanut butter	Peanutr	.2223	1 – almost never
Mustard & ketchup	Mustketc	.2167	1 – almost never
Frankfurters	Hotdog	.2021	1 – almost never
Frozen dinners/ entrees	Fzdin	.1908	1 – almost never
Beer/ale/alcoholic cider	Beer	.1876	1 – almost never
Toothpaste	Toothpa	.1803	1 – almost never
Frozen pizza	Fzpizza	.1407	1 – almost never
Shampoo	Shamp	.1318	1 – almost never
Deodorant	Deod	.1253	1 – almost never
Household cleaner	Hhclean	.1075	1 – almost never
Blades	Blades	.0574	1 – almost never
Sugar substitutes	Sugarsub	.0428	1 – almost never
Cigarettes	Cigets	.0397	1 – almost never
Toothbrush	Toothbr	.0364	1 – almost never
Diapers	Diapers	.0259	1 – almost never
Photography supplies	Photo	.0251	1 – almost never
Razors	Razors	.0081	1 – almost never

Tab. 1: Descriptive statistics w.r.t. product categories

$$P_{data_{ij}} = \begin{cases} 0, \text{ if } \sum_{w=1}^{W_{ij}} data_{ij} = 0 \\ 1, \text{ if } (0 < \sum_{w=1}^{W_{ij}} data_{ij} \ \& \ q_j = 0) \text{ or } (0 < \sum_{w=1}^{W_{ij}} data_{ij} \leq 1 \ \& \ q_j > 0) \\ 2, \text{ if } (1 < \sum_{w=1}^{W_{ij}} data_{ij} \ \& \ q_j = 1) \text{ or } (1 < \sum_{w=1}^{W_{ij}} data_{ij} \leq 2 \ \& \ q_j > 1) \\ 3, \text{ if } (2 < \sum_{w=1}^{W_{ij}} data_{ij} \ \& \ q_j = 2) \text{ or } (2 < \sum_{w=1}^{W_{ij}} data_{ij} \leq 3 \ \& \ q_j > 2) \\ 4, \text{ if } (3 < \sum_{w=1}^{W_{ij}} data_{ij} \ \& \ q_j = 3) \end{cases} \quad (9)$$

This way, we keep some information on the frequency of purchases. This results in data values that vary between zero and four which can be interpreted as answers to questions which are based on a five-point-response-scale (see column 4 in Tab. 1). Here, the question could look like “how often have you made a purchase from product category *j* in the past year?”. The respective answer categories can be thought of as “almost always (4)” – “often (3)” – “sometimes (2)” – “almost never (1)” – “never (0)” (Cai 2010).

We set the minimum relative frequency of answer categories to 2 % both for the binary and the polytomous scenario. This results in omitting razors from further analyses.

3.3. Socio-demographic characteristics

In addition, our data set also contains various socio-demographic variables such as annual income in \$ or family sizes per household. The distribution of relevant variables is shown in Tab. 2. Of course, these information relate to our households independent of the binary or polytomous scenario.

4. Results

4.1. Model fit

4.1.1. (M)IRT vs. topic models

We evaluate the performance of our various types of models, i. e., MIRTbin, MIRTpol, TOPICbin, and TOPICpol, with the help of the Akaike Information Criterion (AIC) (Akaike 1973) that is calculated as follows

$$AIC = -2 * ll + 2 * p. \quad (10)$$

ll is the log-likelihood and *p* stands for the number of parameters and thus penalizes for complex models.

For MIRTbin and MIRTpol the log-likelihood is calculated with the help of an indicator function

Income in \$* in %	0 – 9,999	10,000– 11,999	12,000– 14,999	15,000– 19,999	20,000– 24,999	25,000– 34,999
	6.22	4.93	4.77	6.39	9.46	15.12
Income in \$* in % (cont.)	35,000– 44,999	45,000– 54,999	55,000– 64,999	65,000– 74,999	75,000– 99,999	100,000 -
	12.53	11.96	7.11	7.52	8.08	5.74
Family Size (in # of persons) in %	1	2	3	4	5	6
	24.98	40.34	14.79	13.18	4.61	2.10
Marital status* in %	Single	Married	Divorced	Widowed	Separated	
	11.08	59.66	10.51	14.87	2.59	

Note: * do not sum to 100 % due to N/A-values

Tab. 2: Distribution of socio-demographic variables

$$I_c(x_{ij}) = \begin{cases} 1, & \text{if } x_{ij} = c, \\ 0, & \text{else} \end{cases} \quad (11)$$

and can be written as

$$ll(x_{ij} | \theta_i, \alpha_j, d_j) = \sum_{i=1}^n \left[\sum_{j=1}^J \sum_{c=1}^{C-1} I_c(x_{ij}) * \ln P(x_{ij} = c | \theta_i, \alpha_j, d_j) \right] \quad (12)$$

according to (Chalmers 2012). The calculation of the log-likelihood of topic models is documented in Grün and Hornik (2011).

In addition to the AIC we also use its corrected form denoted by AICc. In small sample sizes in relation to the number of parameters AIC runs the risk of overfitting the data. This can be avoided by using a correction term which is based on Sugiura (1978) and results in

$$AICc = AIC + \frac{2p(p+1)}{n-p-1} \quad (13)$$

Burnham and Anderson (2002, p. 66) advocate using AICc instead of AIC when $n/P \ll 40$ where P stands for the number of the highest-dimensioned model.

By means of AIC and AICc we can determine the optimal number of latent traits per model. In addition, we can

decide whether MIRT or topic models are to be preferred. Results are shown in Tab. 3 and Tab. 4, respectively.

Regarding MIRT the best, i. e., minimum AIC and AICc are obtained for models consisting of four latent traits each. On the other hand, w. r. t. topic models only two latent traits are chosen.

Let us remind you that data sets Bdata and Pdata are different from each other and that therefore AIC- and AICc-values cannot be compared between these two scenarios. However, within Bdata (Pdata) comparisons are possible. We may therefore compare MIRTbin to TOPICbin (based on Tab. 3) and MIRTpol to TOPICpol (based on Tab. 4). It is obvious from both comparisons that MIRT outperform topic models due to much smaller AIC – and AICc-values. That is why we will discuss further results only based on MIRTbin(4) and MIRTpol(4), the respective MIRT with four latent traits for the binary and polytomous scenario, respectively.

4.1.2. Binary vs. polytomous model

MIRTbin(4) (MIRTpol(4)) needs 245 (481) iterations. We now take a closer look at the binary or the polytomous model. In general, both models deal with sparse

(M)IRTbin	AIC	AICc	TOPICbin	AIC	AICc
1 latent trait	29,103	29,109	1 latent trait	---	---
2 latent traits	29,027	29,041	2 latent traits	55,740	55,746
3 latent traits	29,003	29,028	3 latent traits	55,833	55,847
4 latent traits	28,961	28,999	4 latent traits	55,919	55,945
5 latent traits	29,018	29,072	5 latent traits	56,002	56,044

Note: Best model in bold.

Tab. 3: AIC- and AICc-values of binary (M)IRT and topic models

(M)IRTpol	AIC	AICc	TOPICpol	AIC	AICc
1 latent trait	37,719	37,730	1 latent trait	---	---
2 latent traits	37,610	37,630	2 latent traits	78,511	78,518
3 latent traits	37,570	37,603	3 latent traits	78,612	78,626
4 latent traits	37,470	37,519	4 latent traits	78,698	78,724
5 latent traits	37,532	37,599	5 latent traits	78,788	78,830

Note: Best model in bold.

Tab. 4: AIC- and AICc-values of polytomous (M)IRT and topic models

Tab. 5: Limited-information goodness-of-fit and proportion of explained variation

Models	M_2	Degrees of freedom	p -value	Proportion of explained variation
MIRTbin(4)	328.7401	321	.3710	51.5 %
MIRTPol(4)	336.1913	303	.0919	53.2 %

data. This is typically the case for a large number of product categories. That is why the usual goodness of fit statistics like X^2 are not applicable (e. g., Bartholomew et al. 2008, p. 219) and further evaluation of the two models is based on limited-information goodness-of-fit M_2 that is asymptotically X^2 -distributed and tests the null hypothesis that the model fits exactly (Cai and Hansen 2013; Maydeu-Olivares and Joe 2006). The results for both models, i. e. MIRTbin(4) and MIRTPol(4) are displayed in Tab. 5. For a significance level of .05 both models fail to reject the null hypothesis (p -value > .05) and thus support the choice based on AIC and AICc. However, MIRTPol(4) leads to a slightly larger proportion of explained variation which is why we prefer MIRTPol(4) over MIRTbin(3). This is in accordance with Ostini and Nering (2006, pp. 7–8) who state that polytomous items contain more information than binary items. The reason for this small difference can be explained with 16 out of 30 product categories having only two observed answering categories, i. e., “never” and “almost never”. Hence, these product categories are treated in the same manner as in the binary scenario.

4.2. Interpretation of latent traits

Consistent with Böcker (1975) and Hansen et al. (2006) we now proceed by interpreting and hence labeling the resulting latent traits. Factor loadings, i. e., standardized product category slopes α_j according to Eq. (3) are calculated via the “oblimin”-rotation which is an oblique rotation (Bernaards and Jennrich 2005). According to Bartholomew et al. (2008, p. 189) oblique rotations sometimes yield latent traits that are easier to interpret. Tab. 6 shows the outcome of MIRTPol(4) with four latent traits. Product categories whose correlations in terms of absolute values are larger than .4 are considered valid descriptors of the respective latent trait consistent with Bartholomew et al. (2008, p. 197). Latent traits differ in whether they consist of food (latent traits one and four) or non-food (latent traits two and three) related product categories. Latent trait one consists of general groceries. The top three most dominant food categories of this latent trait are margarine/spreads/butter blends (.817), soup (.623), and sugar substitutes (.619). Latent trait four on the other hand contains food related product categories

	Latent trait 1 – groceries related	Latent trait 2 – tissue related	Latent trait 3 – drugstore related	Latent trait 4 – indulgent food related
Beer	-.007	.129	-.016	.493
Blades	-.026	.367	.373	.079
Carbbev	.131	.352	-.049	.561
Cigets	.082	-.202	-.033	.601
Coffee	.405	.277	.039	.075
Coldcer	.436	.146	.204	.118
Deod	.111	.317	.466	.038
Diapers	-.126	-.211	.729	.150
Factiss	.236	.587	-.007	.022
Fzdin	.321	.112	-.056	.218
Fzpizza	.269	-.004	.139	.420
Hhclean	.315	.282	.074	.052
Hotdog	.425	.075	.207	.104
Laundet	.297	.261	.349	.028
Margbutr	.817	-.023	-.125	.045
Mayo	.467	.130	.107	.177
Milk	.487	.139	.123	.138
Mustketc	.577	.044	.182	.100
Paptowl	.106	.750	.043	-.004
Peanbutr	.472	.059	.202	.105
Photo	-.157	.576	.076	.266
Saltsnck	.133	.139	.194	.523
Shamp	.071	.408	.364	.048
Soup	.623	.177	.028	.046
Spagsauc	.508	-.009	.179	.189
Sugarsub	.619	-.027	-.027	-.156
Toitisu	.141	.722	.012	.103
Toothbr	.083	.080	.807	-.043
Toothpa	.198	.338	.477	-.105
Yogurt	.226	.021	.474	.078

Tab. 6: Latent traits and rotated factor loadings for MIRTPol(4)

Note: Correlations which are larger than .4 in absolute terms in bold.

Latent trait 1 – groceries related	Latent trait 2 – tissue related	Latent trait 3 – drugstore related	Latent trait 4 – indulgent food related
Mustketc	Toitisu	Toothbr	Cigets
Peanbutr	Paptowl	Diapers	Carbbev
Yogurt	Factiss	Blades	Saltsnck
Soup	Photo	Toothpa	
Sugarsub	Shamp		
Margbutr	Deod		
Coldcer			
Coffee			
Milk			
Spagsauc			
Hotdog			
Mayo			

Note: Sorted in descending order

Tab. 7: Latent traits for MIRTbin(4)

which show some indulgence character. The top three most product categories are cigarettes (.601), carbonated beverages (.561), and salty snacks (.523). Latent traits two and three on the other hand contain non-food related product categories. Top three product categories of latent trait two are paper towels (.750), toilet tissue (.722), and facial tissues (.587). It can hence be described as a tissue-related latent trait. Latent trait three is mainly described through toothbrush (.807), diapers (.729), and toothpaste (.477) and can thus be considered as drugstore-related.

Latent traits for MIRTbin(4) can be labeled in the same manner. Tab. 7 shows the corresponding latent traits with the respective product categories if we keep a threshold of .4 in absolute terms w. r. t. $st\alpha_{ji}$.

5. Managerial Relevance, conclusion, and limitations

We use this section to show how results of MIRT can be used for further analyses and managerial implications. For the ease of clarity we base our managerial implications on MIRTpol(4). Of course, the following results can be obtained in the same manner if MIRTbin(4) had been chosen.

5.1. Further analyses based on MIRTpol(4)

Let us remind you, that purchases are driven by a combination of latent traits. Knowing the extent of latent traits per households does not only help to describe characteristics of a household (Reckase 2009, p. 11) but can also inform the marketer about expected purchases. Hence, ascribing actual values of these latent traits to households can support management decisions for several reasons. First of all, as can be seen from Eqns. (4) and (5) respectively, the extent of a latent trait has an influence on purchase probability of a product. By using

$$E(\text{purchase frequency}_j) = \sum_{c=1}^C c * P(x_{ij} = c | \theta_i, \alpha_j, d_j) \quad (14)$$

according to Reckase (2009, p. 104) marketers are able to estimate expected purchase frequencies for each prod-

uct category and household. Of course, companies can use these latent traits to further analyse their households and treat them with individual or segment-specific marketing tools. If available, other characteristics of households can be added to the analyses as well.

To demonstrate this, we calculate for all of our 1,237 households estimates of our four latent traits according to Chalmers (2012). These estimates are able to condense a household’s product category preferences into four different variables. In order to see how households’ latent traits differ we perform a *k*-means cluster analysis for which these four latent traits serve as input variables. We vary the number of segments, i. e., clusters from one to 10 and choose *k* = 3 which explains 80.6 % of the total sum of squares and divides households into three segments. According to Tab. 8 these three segments differ regarding size, socio-demographic attributes, and expected purchase frequency of product categories.

51.41 % of households fall into segment one. Households belonging to segment three on the other hand constitute 15.12 % of our sample. In addition, segments of households can also be described by socio-demographic variables. For the metric variable number of persons per household we use a one-way ANOVA with segment membership as factor. Results are significant with a *p*-value < .05. For the remaining categorical variables we perform a X^2 – test of independence. Marital status and annual income show a significant relationship with segment membership as supported by *p*-value < .05 and *p*-value < .07, respectively. Households belonging to segment three tend to have a larger annual income than households from segment one. In addition, the former have a higher number of persons per household. W. r. t. the marital status almost 69 % of households from segment three are married as opposed to 59.18 % (57.23 %) of households from segment two (one).

Tab. 8 contains the five most frequently expected purchased product categories per segment. The three most frequently expected purchased product categories are the same for all three average households, namely milk, car-

bonated beverages, and salty snacks. However, an average household of segment one almost never purchases from these categories. Regarding average households from segment two it is observable that the top five categories stem from a mixture of latent traits one and four. For an average household of segment three the tissue-related latent trait is observable as well.

5.2. Managerial relevance

5.2.1. Implications based on the general composition of the latent traits

Marketers can use results of MIRT_{pol}(4) in several ways. First of all, by observing the contents of the different latent traits they see which product categories are frequently purchased together. This is of relevance for cross selling strategies. When a shop registers at the check-out which products have been bought by a household it could print recommendations or even discounts for further product categories of the same latent trait on the receipt. Take for instance the market basket of a household that has purchased, e. g., coffee or coldcer. As both product categories relate to latent trait one marketers could advertise or reduce prices for other product categories of this latent trait, e. g., milk or margbutr.

Marketers can however also use the composition of the latent traits for instore arrangements that support cross selling strategies. This could involve placing products from respective categories of a latent trait in close proximity within the shop. A company that is putting cross selling to the extremes is the franchise company “Kochhaus”. In such shops, customers can buy groceries. However, products are not sorted according to different cate-

gories but are arranged to recipes instead. A customer wishing to buy food for a certain recipe would find the respective products combined on one shelf (Teuber und Gehrman 2010). Latent traits that do not consist of too many different product categories could guide a shop manager in presenting products that are not grouped by recipes but by latent traits. Hence, a shelf which is related to, e. g., latent trait four could combine products from the categories beer, carbbev, saltsnck, fzipizza, and dummys of cigets (to apply with legal restrictions).

Alternatively, if it is not possible to present the above mentioned product categories next to each other a shop manager can put cards on the shelves addressing the other product category. If we consider again latent trait four these cards could be placed in the alcoholic section asking “have you thought of salty snacks”?

5.2.2. Implications based on segment-specific expected purchase behaviour

Let us give an example how a marketer may use the information contained in Tab. 8. Based on a household’s latent trait different expected purchase frequencies for various product categories may result. Households from segment three have the largest expected purchase frequencies for various product categories members from segment one the smallest. Depending on the segment, marketing tools should differ. While the main task for households belonging to segment one should involve in becoming households more active the main goal for segment three can be to induce further purchases. The strategy for segment two members lies somewhat in the middle.

	Segment 1	Segment 2	Segment 3
Size (in %)	51.41	33.47	15.12
Top three categories of significant socio-demographic variables: number of households in %			
Family size (in number of persons)	2: 41.19 1: 26.10 3: 14.31	2: 37.20 1: 27.78 3: 15.70	2: 44.39 4: 16.58 1: 14.97
Marital status	Married: 57.23 Widowed: 14.31 Single: 12.89	Married: 59.18 Widowed: 17.15 Divorced: 10.63	Married: 68.98 Widowed: 11.76 Single: 8.02
Annual income in \$	25,000-34,999: 14.62 35,000-44,999: 12.74 20,000-24,999: 10.85	25,000-34,999: 16.43 45,000-54,999: 12.80 35,000-44,999: 11.84	45,000-54,999: 18.18 25,000-34,999: 13.90 35,000-44,999: 13.37
Five most frequently expected purchased product categories	Milk (.6197 – almost never) Carbbev (.2392 - never) Saltsnck (.2347 - never) Coldcer (.1722 – never) Yogurt (.1195 – never)	Milk (2.3831 - sometimes) Carbbev (1.7917 - sometimes) Saltsnck (1.4974 – almost never) Coldcer (1.0672 – almost never) Soup (1.0015 – almost never)	Milk (3.8900 - almost always) Carbbev (2.9701 - often) Saltsnck (2.9056 - often) Toitisu (1.9788 - sometimes) Soup (1.9504 - sometimes)

Tab. 8: Descriptive statistics per segment

If a shop manager has more detailed information on households like, e. g., sociodemographic attributes she or he can combine this pool of information with latent traits. Due to this households can be targeted directly depending their latent traits and further characteristics.

The main goal w. r. t. households from segment one should be to increase purchase frequencies. The marketer could use the top five expected product categories, namely milk, carbbev, saltsnck, coldcer, and yogurt as recommendations when contacting these households. Of course, it is possible to proceed in the same manner with households belonging to segment two. As the order of the top four product categories is the same for households in both segments shop managers might be tempted to recommend in the same manner. However, the company should use further external information available on households as well. Whereas family size and marital status are comparable to households from segment two households from segment one have less annual income available. Hence, one option could be to target households from segment one not only with the respective product categories but to include, e. g., price promotions. This way, the marketer accounts for the latent traits of these households combined with their income situation.

We conclude by showing how a marketer might react towards households from segment three. If the goal of the marketer is to increase volumes of usually bought product categories one might recommend the top five categories at reduced prices to the households. If on the other hand households should be induced to buy from other categories as well the company could recommend product categories that are, e. g., on rank six to ten according to their latent traits.

5.3. Conclusions and limitations

Our research reveals that MIRT, a concept that has a long history in psychological and educational testing, has only few applications in marketing studies. In particular, it has never been used to explain multi-category purchase behaviour. We show that MIRT are able to model multi-category purchase behaviour. Based on two types of data sets (a binary and a polytomous scenario) these models have a better model fit than corresponding topic models. Furthermore, we are able to show that MIRT can display latent traits, i. e. characteristics of a household that explain the observed purchase decision. In our data sets, households show four different traits that reflect purchase behaviour of 30 different product categories. Both models, i. e., MIRTbin(4) and MIRTpol(4), identified two food- and two non-food related latent traits whereas the latter was able to explain a slightly larger percentage of the variation in the data set. Due to this, we chose MIRTpol(4) over MIRTbin(4). While we would generally advise to use the model that is able to explain the data more accurately we have shown in *Tab. 7* that MIRTbin(4) is equally well suited in identifying the house-

hold's latent traits. Hence, if a researcher is not that fortunate of having a polytomous data set MIRT will still be able to have reliable results. Finally, our investigation based on MIRTpol(4) shows how purchase behaviour differs depending on the latent trait and further characteristics of a household. This knowledge may support the marketer to stimulate further purchases.

We see our study as a first attempt to show that MIRT are suited for analysing multi-category purchase decisions. Of course, our procedure is not free of limitations which constitute interesting avenues for further research. One disadvantage is the aggregation of household transactions onto a yearly basis. This way, we are, e. g., not able to analyse marketing actions such as price reductions, etc. that are performed during certain weeks of the observation period. Furthermore, MIRT could be adapted to investigate how unobserved heterogeneity may influence latent traits.

Notes

[1] Analyses are the responsibility of the authors, not IRI.

References

- Atkins, D. C., Rubin, T. N., Steyvers, M., Doeden, M. A., Baucum, B. R., & Christensen, A. (2012). Topic models: A Novel Method for Modeling Couple and Family Text Data. *Journal of Family Psychology*, 26(5), 816–827.
- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory*, Budapest: Akademiai Kiado, 267–281.
- Bartholomew, D. J., Steele, F., Moustaki, I. & Galbraith, J. I. (2008). *Analysis of Multivariate Social Science Data*, 2nd Ed., Boca Raton et al.: CRC Press.
- Bernaards, C. A., & Jennrich, R. I. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement*, 65(5), 676–696.
- Blei, D. (2012). Probabilistic Topic Models. *Communications of the ACM*, 55(4), 77–84.
- Bock, R. D., & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46(4), 443–459.
- Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-Information Item Factor Analysis. *Applied Psychological Measurement*, 12(3), 261–280.
- Böcker, F. (1975). Die Analyse des Kaufverbunds – ein Ansatz zur bedarfsorientierten Warentypologie. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung ZFBF*, 27(1), 290–306.
- Boztuğ, Y., & Silberhorn, N. (2006). Modellierungsansätze in der Warenkorbanalyse im Überblick. *Journal für Betriebswirtschaft*, 56(2), 105–128.
- Bronnenberg, B. J., Kruger, M. W., & Mela, C. F. (2008). Database Paper – The IRI Marketing Data Set. *Marketing Science*, 27(4), 745–748.
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodal Inference, a Practical Information-Theoretic Approach*, 2nd Ed., New York et al.: Springer.
- Cai, L. (2010). High-Dimensional Exploratory Item Factor Analysis by a Metropolis-Hastings Robbins-Monro Algorithm. *Psychometrika*, 75(1), 33–57.
- Cai, L., & Hansen, M. (2013). Limited-Information Goodness-of-Fit Testing of Hierarchical Item Factor Models. *British Journal of Mathematical and Statistical Psychology*, 66(2), 245–276.

- Chalmers, R. P. (2012). Mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1–29.
- De Jong, M., Steenkamp, J.-B. E. M., Fox, J. -P., & Baumgartner, H. (2008). Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation. *Journal of Marketing Research*, 45(1), 105–115.
- Dippold, K., & Hruschka, H. (2013). A Model of Heterogeneous Multicategory Choice for Market Basket Analysis. *Review of Marketing Science*, 11(1), 1–31.
- Duvvuri, S. D., Ansari, A., & Gupta, S. (2007). Consumers Price Sensitivities Across Complementary Categories. *Management Science*, 53(12), 1933–1945.
- Embretson, S., & Reise, S. P. (2000). *Item Response Theory for Psychologists*, Mahwah, NJ et al.: Erlbaum.
- Grün, B., & Hornik, K. (2011). Topicmodels: an R Package for Fitting Topic Models. *Journal of Statistical Software*, 40(13), 1–30.
- Hansen, K, Singh, V., & Chintagunta, P. (2006). Understanding Store-Brand Purchase Behavior Across Categories. *Marketing Science*, 25(1), 75–90.
- Hruschka, H. (2014). Linking Multi-Category Purchases to Latent Activities of Shoppers: Analysing Market Baskets by Topic Models. *Marketing ZFP – Journal of Research and Management*, 36(4), 268–274.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*, Hillsdale, NJ: Erlbaum.
- Maydeu-Olivares, A., & Joe, H. (2006). Limited Information Goodness-of-Fit Testing in Multidimensional Contingency Tables. *Psychometrika*, 71(4), 713–732.
- Mild, A., & Reutterer, T. (2003). An Improved Collaborative Filtering Approach for Predicting Cross-Category Purchases Based on Binary Market Basket Data. *Journal of Retailing and Consumer Services*, 10(3), 123–133.
- Ostini, R., & Nering, M. L. (2006). *Polytomous Item Response Theory Models*. Thousand Oaks, CA: Sage.
- R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raykov, T., & Calantone, R. J. J. (2014). The Utility of Item Response Modeling in Marketing Research. *Journal of the Academy of Marketing Science*, 42(4), 337–360.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*, New York et al.: Springer.
- Sugiura, N. (1978). Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections. *Communications in Statistics – Theory and Methods*, 7(1), 13–26.
- Teuber, K., & Gehrman, L. (2010). Wer im Kochhaus einkauft, kocht niemals zu viel, <https://www.welt.de/lifestyle/essen-trinken/article9238226/Wer-im-Kochhaus-einkauft-kocht-niemals-zu-viel.html>, (accessed 27.01.2017)

Keywords

Multidimensional Item Response Theory, Multi-Category Purchase Behaviour, Market Basket Analysis.

MARKETING

ZFP – Journal of Research and Management

Editor-in-Chief: Prof. Dr. Bernhard Swoboda, Chair for Marketing and Retailing, University of Trier, Universitätsring 15, D-54296 Trier, Phone: +49 651 201 3050, Fax: +49 651 201 4165, Email: editor@marketing-zfp.de

Senior Editors: Prof. Dr. Heribert Gierl, University of Augsburg, Prof. Dr. Andrea Gröppel-Klein, Saarland University, Prof. Dr. Lutz Hildebrandt, Humboldt-University of Berlin, Prof. Dr. Hans Mühlbacher, International University of Monaco, Prof. Dr. Henrik Sattler, University of Hamburg, Prof. Dr. Udo Wagner, University of Vienna.

Manuscripts: We ask all authors who would like to submit a paper to send this paper to the editor-in-chief Bernhard Swoboda via mail: editor@marketing-zfp.de. Neither the publisher nor the editors assume any liability for unsolicited manuscripts. Unsolicited manuscripts will only be returned if accompanied by return postage. The acceptance of a contribution has to be in writing.

Copyright: Upon acceptance for publication the author transfers to the C.H.BECK the exclusive copyright of his

or her contribution for the duration of the copyright as laid down by law. The copyright covers the exclusive right and licence to reproduce, publish, distribute and archive the article in all forms and media of expression now known or developed in the future, including reprints, translations, photographic reproductions, microform, electronic form (offline and online) or any other reproduction of similar nature. The author's second window right after the expiry of 12 months after first publication, as laid down in article 38(4) German Copyright Law, remains unaffected. All articles published in this journal are protected by copyright law. Without first obtaining permission from the publisher, no material published in this journal may be reproduced, distributed, performed or displayed publicly, or made accessible or stored in electronic databases or reproduced, distributed or utilized electronically, outside the narrow limitations of copyright law.

Publisher: C.H.BECK oHG, Wilhelmstr. 9, 80801 München; postal address: P.O.Box 40 03 40, 80703 München; phone: +49 89 38189 0; fax: +49 89 38189 398, Bank account: Postbank München IBAN: DE82 7001 0080 0006 2298 02, BIC: PBNKDEFFXXX.

The publisher is a *offene Handelsgesellschaft* under German law with Dr. Hans Dieter Beck and Dr. h.c. Wolfgang Beck as partners.

Subscription: An annual subscription to the journal comprises four issues.

Subscription rates 2017: € 219 (VAT incl.) annual subscription rate, discounted rate for students (related subject, proof needed) € 135 (VAT incl.), campus licence € 349 (VAT incl.). Single Issue: € 61,50 (VAT incl.), shipping charges have to be added to the rates. Subscription and rate include print issue and a licence for the online archive. The components cannot be cancelled separately. Complaints about copies not received must be lodged within 6 weeks starting at the end of the quarter.

Subscription service: Please order with either the publisher or any book shop.

CustomerServiceCenter: Phone: +49 89 38189 750, Fax: +49 89 38189 358, E-Mail: kundenservice@beck.de

Cancellation: The subscription may be cancelled in writing 6 weeks before the end of a calendar year.

Citation: Marketing ZFP – Journal of Research and Management, number of volume(number of issue), year, page.

Typesetting: FotoSatz Pfeifer GmbH, 82152 Krailling.

Printing: Kessler Druck und Medien GmbH & Co. KG, Michael-Schäffer-Straße 1, 86399 Bobingen.