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 polynominal 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 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 inci-
dences regarding the various product categories he fo-
cuses 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 interdependen-
cies between products. Hansen et al. (2006) use a logit
model on a data set from a retail chain and analyse pur-
chase behaviour among ten different product categories
comprising food and non-food related products. The fac-
tor analytic structure is placed on the covariance matrix
of the model coefficients. This way, the dependence be-
tween product categories is captured and factors are in-
terpreted 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 sen-
sitivity.

The small number of factor analytic models w. r. t. multi-
category purchases can be explained by data characteris-
tics. According to Bartholomew et al. (2008, p. 178) fac-
tor analytic models assume metric scales for observed
variables. However, in multi-category purchases a mar-
ket basket displays which and how many products a
household has purchased within various product cate-
gories during a transaction. Typically, such a basket con-
tains a lot of zeros (for product categories where no pur-
chase has taken place) and ones (for product categories
with purchases). Hence, data are usually binary or poly-
tomous 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 ten different product categories the
respective row in the data set looks like this: (1,1,0,0,0,0,0,0,
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 underly-
ing 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 esti-
mate probabilities of answering a question correctly. De-
pending 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 atti-
tudes (Embrtsson 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 mat-
ters like whether homosexuals should hold public posi-
tions 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 mat-
ters 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 mea-
suring 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 catego-
ries 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 result-
ing 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 ana-
lysing 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 re-
results we predict purchase behaviour. When using MIRT
to analyse multi-category purchase behaviour we as-
sume 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 deci-
sions (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 similar-
ity to factor analytic models we are able to detect intu-
itive latent traits for both scenarios which can then be re-
lated 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-catego-
ry purchase behaviour. Finally, we show how marketers
can benefit from knowing the latent traits of households
by giving examples how future purchases can be in-
duced.
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 polychotomous 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 unidimensionality, i.e., different manifest variables (answers to test items) are based on one latent trait. MIRT on the other hand allow different manifest variables (answers to test items) to be metric whereas in situations with categorical variables one speaks of IRT (e.g., Raykov and Calantone 2014). IRT usually assume unidimensionality, 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 polychotomous data. According to Bartholomew et al. (2008, pp. 212–213) the general linear factor analysis model can be written as

\[ x_i = \theta_i + \alpha_i^1 \theta_i^1 + \alpha_i^2 \theta_i^2 + \ldots + \alpha_i^n \theta_i^n + \epsilon_i \]  

(1)

and assumes an unobservable decision process of household \( i \) on product category \( j \). Whereas \( x_i \) shows the observed purchase decision \( \theta_i = (\theta_i^1, \ldots, \theta_i^n) \) is defined as a vector of \( t = 1, \ldots, T \) latent traits or factors. \( \alpha_i = (\alpha_i^1, \ldots, \alpha_i^n) \) is a vector of the corresponding unrotated factor loadings and \( \epsilon_i \) a constant term. Residuals \( \epsilon_i \) 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 \( x_i \) are metric in the general linear factor model they are binary (or polychotomous) 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

\[ \logit \pi_{ij}(\theta) = d_j + \alpha_{ij} + \alpha_{i2} \theta_2 + \ldots + \alpha_{in} \theta_n \]  

(2)

Once \( \alpha_j \) are standardized according to Eq. (3)

\[ \text{st} \alpha_j = \frac{\alpha_j}{\sqrt{1 + \sum_{n=1}^{N} \alpha_j^2}} \]  

(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 characteristics (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 (polychotomous scenario).

2.2.1. Binary Scenario

We use a multidimensional extension of the two-parameter logistic model (MIRBin). 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

\[ P(x_{ij} = 1 | \theta_i, \alpha_j, d_j) = \frac{1}{1 + \exp(-\alpha_j \theta_i + d_j)} \]  

(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 \( \alpha_j = (\alpha_{j1}, \ldots, \alpha_{jn}) \) 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 polychotomous 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 \( c \in (0,1,...,C_j-1) \) w. r. t. product category \( j \):}

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

(5)

The boundaries of response probabilities for model MIRTPol are:

\[ P(x_{ij} \geq 0 | \theta_i, \alpha_j, d_j) = 1, \]
\[ P(x_{ij} \geq 1 | \theta_i, \alpha_j, d_j) = \frac{1}{1 + \exp(-\alpha_j \theta_i + d_j)} \]
\[ P(x_{ij} \geq 2 | \theta_i, \alpha_j, d_j) = \frac{1}{1 + \exp(-\alpha_j \theta_i + d_j)} \]
\[ \ldots \]
\[ P(x_{ij} \geq C_j | \theta_i, \alpha_j, d_j) = 0. \]
Parameters of the model can be interpreted in the same manner as in the binary scenario. However, \( d_j \) now is a vector of \( d_{j1}, \ldots, d_{jC_{j-1}} \) 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 polychotomous 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

\[
p_{ij} = \theta_{it} \cdot \delta_{jt}, \tag{7}
\]

It is the probability that market basket of household \( i \) contains product category \( j \). \( \theta_{it} \) and \( \delta_{jt} \) both come from a Dirichlet distribution. \( \theta_{it} \) captures the importance of product category \( j \) for latent trait \( t \) and can hence be compared to \( \theta_{it} \) in MIRT. \( \delta_{jt} \) 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. \( \sum_{w=1}^{w} \text{data}_{ij} \). 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 polychotomous 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 \( \sum_{w=1}^{w} \text{data}_{ij} \) for each household \( i \) and product category \( j \) is larger than zero. Hence our binary data set (Bdata) can be displayed as:

\[
\text{Bdata}_{ij} = \begin{cases} 
0, & \text{if } \sum_{w=1}^{w} \text{data}_{ij} = 0 \\
1, & \text{else} 
\end{cases} \tag{8}
\]

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 polychotomous scenario we try to use the frequency of purchases of each household \( i \) in product category \( j \), i. e., \( \sum_{w=1}^{w} \text{data}_{ij} \), as far as possible. We therefore use the distribution of \( \sum_{w=1}^{w} \text{data}_{ij} \), 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 polychotomous 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.
<table>
<thead>
<tr>
<th>Categories</th>
<th>Abbreviation</th>
<th>Relative purchase frequencies (based on Bdata)</th>
<th>Maximum per product category (j) and its correspondence to the five-point-response-scale (based on Pdata)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>Milk</td>
<td>.6669</td>
<td>4 – almost always</td>
</tr>
<tr>
<td>Carbonated beverages</td>
<td>Carbbev</td>
<td>.5578</td>
<td>3 – often</td>
</tr>
<tr>
<td>Salty snacks</td>
<td>Saltsnck</td>
<td>.5004</td>
<td>3 – often</td>
</tr>
<tr>
<td>Cold cereal</td>
<td>Coldcer</td>
<td>.4293</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Soup</td>
<td>Soup</td>
<td>.4002</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Toilet tissue</td>
<td>Toitisu</td>
<td>.3323</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Spaghetti/Italian sauce</td>
<td>Spagsauc</td>
<td>.3104</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Coffee</td>
<td>Coffee</td>
<td>.3072</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Yogurt</td>
<td>Yogurt</td>
<td>.2967</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Mayonnaise</td>
<td>Mayo</td>
<td>.2724</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Margarine/spreads/butter blends</td>
<td>Margbutr</td>
<td>.2700</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Paper towels</td>
<td>Paptowl</td>
<td>.2700</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Facial tissue</td>
<td>Factiss</td>
<td>.2668</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Laundry detergent</td>
<td>Laundet</td>
<td>.2668</td>
<td>2 – sometimes</td>
</tr>
<tr>
<td>Peanut butter</td>
<td>Peanbutr</td>
<td>.2223</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Mustard &amp; ketchup</td>
<td>Mustkete</td>
<td>.2167</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Frankfurters</td>
<td>Hotdog</td>
<td>.2021</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Frozen dinners/entrees</td>
<td>Fzdin</td>
<td>.1908</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Beer/ale/alkoholic cider</td>
<td>Beer</td>
<td>.1876</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Toothpaste</td>
<td>Toothpa</td>
<td>.1803</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Frozen pizza</td>
<td>Fzpizza</td>
<td>.1407</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Shampoo</td>
<td>Shamp</td>
<td>.1318</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Deodorant</td>
<td>Deod</td>
<td>.1253</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Household cleaner</td>
<td>Hhclean</td>
<td>.1075</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Blades</td>
<td>Blades</td>
<td>.0574</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Sugar substitutes</td>
<td>Sugarsub</td>
<td>.0428</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Cigarettes</td>
<td>Cigets</td>
<td>.0397</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Toothbrush</td>
<td>Toothbr</td>
<td>.0364</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Diapers</td>
<td>Diapers</td>
<td>.0259</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Photography supplies</td>
<td>Photo</td>
<td>.0251</td>
<td>1 – almost never</td>
</tr>
<tr>
<td>Razors</td>
<td>Razors</td>
<td>.0081</td>
<td>1 – almost never</td>
</tr>
</tbody>
</table>

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

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

\[
AIC = -2 \cdot \log{ll} + 2 \cdot p. \tag{10}
\]

\(ll\) 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.
<table>
<thead>
<tr>
<th>Income in $* in % (cont.)</th>
<th>35,000–</th>
<th>45,000–</th>
<th>55,000–</th>
<th>65,000–</th>
<th>75,000–</th>
<th>100,000–</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44,999</td>
<td>54,999</td>
<td>64,999</td>
<td>74,999</td>
<td>99,999</td>
<td></td>
</tr>
<tr>
<td>Family Size (in # of persons in %)</td>
<td>12.53</td>
<td>11.96</td>
<td>7.11</td>
<td>7.52</td>
<td>8.08</td>
<td>5.74</td>
</tr>
</tbody>
</table>

Marital status* in %

- Single: 11.08
- Married: 59.66
- Divorced: 10.51
- Widowed: 14.87
- Separated: 2.59

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

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

and can be written as

\[
ll(x_{ij} | \theta_i, \alpha_j, \delta_j) = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{c=1}^{C_j} I_c(x_{ij}) \cdot \ln P(x_{ij} = c | \theta_i, \alpha_j, \delta_j) \quad (12)
\]

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 Sugiyura (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 \( n/P << 40 \) 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...
Tab. 5: Limited-information goodness-of-fit and proportion of explained variation

<table>
<thead>
<tr>
<th>Models</th>
<th>$M_2$</th>
<th>Degrees of freedom</th>
<th>$p$-value</th>
<th>Proportion of explained variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIRTbin(4)</td>
<td>328.7401</td>
<td>321</td>
<td>.3710</td>
<td>51.5%</td>
</tr>
<tr>
<td>MIRTpol(4)</td>
<td>336.1913</td>
<td>303</td>
<td>.0919</td>
<td>53.2%</td>
</tr>
</tbody>
</table>

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

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 $\alpha_j$ 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.
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. $\alpha_i$.

### Tab. 7: Latent traits for MIRTbin(4)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mustketc</td>
<td>Toitisu</td>
<td>Toothbr</td>
<td>Cigets</td>
</tr>
<tr>
<td>Peanbutr</td>
<td>Paptowl</td>
<td>Diapers</td>
<td>Carbbev</td>
</tr>
<tr>
<td>Yogurt</td>
<td>Factiss</td>
<td>Blades</td>
<td>Saltsnck</td>
</tr>
<tr>
<td>Soup</td>
<td>Photo</td>
<td>Toothpa</td>
<td></td>
</tr>
<tr>
<td>Sugarsub</td>
<td>Shamp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margbutr</td>
<td>Deod</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coldcer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coffee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spagsauc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hotdog</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayo</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Sorted in descending order

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

$$E(\text{purchase frequency}) = \sum_{i=1}^{n} c^i P(x_i = c | \theta, \alpha_i, d)$$  \hspace{1cm} (14)

according to Reckase (2009, p. 104) marketers are able to estimate expected purchase frequencies for each product 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 $\chi^2$ – 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 MIRTpol(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 checkout 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 in-store 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 categories 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 Gehrmann 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, fpzpiza, and dum-mies of ciges (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.

<table>
<thead>
<tr>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (in %)</td>
<td>51.41</td>
<td>33.47</td>
</tr>
</tbody>
</table>

Top three categories of significant socio-demographic variables: number of households in %

<table>
<thead>
<tr>
<th>Family size (in number of persons)</th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2: 41.19</td>
<td>2: 37.20</td>
<td>2: 44.39</td>
<td></td>
</tr>
<tr>
<td>1: 26.10</td>
<td>1: 27.78</td>
<td>4: 16.58</td>
<td></td>
</tr>
<tr>
<td>3: 14.31</td>
<td>3: 15.70</td>
<td>1: 14.97</td>
<td></td>
</tr>
</tbody>
</table>

Marital status

| Married: 57.23 | Married: 59.18 | Married: 68.98 |
| Widowed: 14.31 | Widowed: 17.15 | Widowed: 11.76 |

Annual income in $

| 25,000-34,999: 14.62 | 25,000-34,999: 16.43 | 45,000-54,999: 18.18 |
| 35,000-44,999: 12.74 | 45,000-54,999: 12.80 | 25,000-34,999: 13.90 |
| 20,000-24,999: 10.85 | 35,000-44,999: 11.84 | 35,000-44,999: 13.37 |

Frequent expected purchased product categories

| Milk (.6197 – almost never) | Milk (.2381 - sometimes) | Milk (.3890 - almost always) |
| Carbev (.2392 - never)      | Carbev (.1791 - sometimes) | Carbev (.2970 - often) |
| Saltsnck (.2347 - never)    | Saltsnck (.14974 – almost never) | Saltsnck (.29056 - often) |
| Coldcer (.1722 – never)     | Coldcer (.10672 – almost never) | Toisu (.19788 - sometimes) |
| Yogurt (.1195 – never)      | Yogurt (.10015 – almost never) | Yogurt (.19504 - 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 on 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, carbeve, saltsnack, coldece, 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 polychotomous 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 household’s latent traits. Hence, if a researcher is not that fortunate of having a polychotomous 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


**Keywords**

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