Journal of Classification 30:306-337 (2013)
DOI: 10.1007/s00357-013-9141-8

Constrained Multilevel Latent Class Models for the Analysis of Three-Way Three-Mode Binary Data

Michel Meulders

HUBrussel, Belgium and KU Leuven, Belgium

Francis Tuerlinckx

KU Leuven, Belgium

Wolf Vanpaemel

KU Leuven, Belgium

Abstract: Probabilistic feature models (PFMs) can be used to explain binary rater judgements about the associations between two types of elements (e.g., objects and attributes) on the basis of binary latent features. In particular, to explain observed object-attribute associations PFMs assume that respondents classify both objects and attributes with respect to a, usually small, number of binary latent features, and that the observed object-attribute association is derived as a specific mapping of these classifications. Standard PFMs assume that the object-attribute association probability is the same according to all respondents, and that all observations are statistically independent. As both assumptions may be unrealistic, a multilevel latent class extension of PFMs is proposed which allows objects and/or attribute parameters to be different across latent rater classes, and which allows to model dependencies between associations with a common object (attribute) by assuming that the link between features and objects (attributes) is fixed across judgements. Formal relationships with existing multilevel latent class models for binary three-way data are described. As an illustration, the models are used to study rater differences in product perception and to investigate individual differences in the situational determinants of anger-related behavior.

Keywords: Multilevel latent class model; Latent feature; Three-way data.

We thank Jeroen Vermunt for helping us with the specification of multilevel PFMs in the syntax module of Latent Gold.

Corresponding Author's Address: M. Meulders, Centre for Modelling and Simulation, HUBrussel, Warmoesberg 26, 1000 Brussels, Belgium, email: Michel.Meulders@hubrussel.be

Published online: 20 October 2013

1. Introduction

Binary three-way data are often used to study substantive phenomena in different areas of applied research. In personality psychology, such data may for instance be used to study the types of behavior that are displayed by a certain person in a certain situation (Kuppens, Van Mechelen, and Meulders 2004; Vansteelandt and Van Mechelen 1998). When studying the competitive structure analysis of products in marketing research, data on the attributes that are assigned to different products by different consumers can be used to derive a perceptual mapping of products and attributes (Torres and Bijmolt 2009). In social network analysis, one may for instance study the social structure of a group by asking each member of the group to judge the friendship ties between all (pairs of) members of the group (Kumbasar, Kimball, and Batchelder 1994) or study simultaneously different types of relations between group members (Gonzalez, Tuerlinckx, and De Boeck 2009). In psychiatric diagnosis, one may study the structure of psychiatric syndromes by analyzing which symptoms are ascribed to certain patients by a number of clinicians (Van Mechelen and De Boeck 1990).

Maris, De Boeck and Van Mechelen (1996) introduced probabilistic feature models (PFMs) as a method to model binary three-way data. PFMs have been used to analyze data from several substantive domains, such as marketing (Candel and Maris 1997), cross-cultural psychology (Meulders, De Boeck, Van Mechelen, Gelman, and Maris 2001), personality assessment (Meulders, De Boeck, Kuppens, and Van Mechelen 2002; Meulders, De Boeck, and Van Mechelen 2003), psychiatric diagnosis (Maris et al. 1996), and the facial perception of emotions (Meulders, De Boeck, Van Mechelen, and Gelman 2005).

To explain binary rater judgements about the associations between objects and attributes, PFMs assume that respondents classify both the objects and the attributes with respect to a number of unobserved binary variables (further denoted as 'latent features'), and that the observed object-attribute associations are derived as a predetermined mapping of these classifications. Similar to the continuous latent factors in a factor analysis, latent features can be considered as higher-order properties of objects/attributes which provide a reduced description of those objects/attributes. Using binary instead of continuous latent variables to describe objects/attributes may be interesting for several reasons: First, in contrast to models with continuous latent factors, models with binary latent factors do not have to rely on specific distributional assumptions for the latent factors. Second, the classification of objects and attributes in terms of binary latent features can be used to categorize objects and attributes, which may enhance the interpretation of the analysis.

The PFM introduced by Maris et al. (1996) makes two assumptions that are often too restrictive in applications: First, the model assumes that the parameters used to model the classification of objects and attributes in terms of latent features are the same for all respondents. This assumption implies that the probability of an object-attribute association is the same for all respondents, which may be unrealistic in applications. Second, the model assumes that the classification of objects/attributes in terms of latent features is renewed each time a respondent makes a new judgement. This assumption implies that all object-attribute associations are stochastically independent, which is often contradicted by the data (Meulders et al. 2003).

In this paper, we will describe a multilevel latent class extension of PFMs (further denoted as multilevel LC-PFM), which allows respondent heterogeneity (i.e., object and/or attribute parameters can be different across latent rater classes) and dependencies between observations (i.e., by assuming that the link between features and objects/attributes is fixed across judgements). Besides improving the statistical fit of the model, a key advantage of the proposed multilevel latent class extension is that it allows studying rater differences in the categorization of objects and/or attributes. The study of rater heterogeneity is an important substantive goal in many applications.

In the following sections we will first formally describe PFMs. Second, we describe model extensions which allow to model respondent heterogeneity in object- or attribute parameters and which allow to capture dependencies between observations. Third, we discuss related latent class models for the analysis of binary three-way data. Fourth, we discuss the estimation of multilevel LC-PFMs with the standard latent class software 'Latent Gold' (Vermunt and Magidson 2005; 2008). Fifth, we discuss model selection and model checking. Sixth, we illustrate the models with applications on product perception and anger-related behavior.

2. Probabilistic Feature Models

In order to explain binary associations between elements, PFMs assume a three-fold process. To clarify this process we consider the case of consumers who judge whether or not certain car models have certain predefined attributes (e.g 'safe', 'fun to drive with', 'powerful', 'stylish', 'rough', 'spacious', 'exclusive', 'practical', 'expensive', 'comfortable', 'versatile', 'economic', etc.). Let D_{ijk} denote binary observed variables that equal 1 if product j ($j = 1, \ldots, J$) has attribute k ($k = 1, \ldots, K$) according to rater i ($i = 1, \ldots, I$), and 0 otherwise. In order to explain the observed judgements, PFMs make the following assumptions:

1. There exist F binary latent features for both products and attributes.

2. It is assumed that certain binary latent features representing properties of a car model are activated when a rater judges whether or not a specific car model has a certain attribute. For instance, when judging a 'Renault Espace' the latent features 'family car' and 'pleasure' might be activated, whereas when judging a 'Porsche Cayenne' the latent features 'outdoor' and 'pleasure' might be activated. Formally, the activation of latent features is described with binary latent variables X_{ki}^{jf} $(f = 1, \ldots, F)$ that equal 1 if product j activates feature f when rater i judges the product-attribute pair (j, k), and 0 otherwise. Furthermore, PFMs assume that the activation of latent features during product perception is a stochastic process. More specifically, it is assumed that $X_{ki}^{jf} \sim \text{Bernoulli}(\sigma_{jf})$.
3. It is assumed that when a product-attribute pair is being judged, latent features activated by the product may or may not be linked to the attribute. For instance, the latent feature 'family car' may be linked to attributes such as 'safe', 'spacious', 'practical', 'comfortable', 'versatile', 'economic'; the latent feature 'pleasure' may be linked to attributes such as 'fun to drive with', 'powerful', 'stylish', 'exclusive', 'expensive', 'comfortable'; and the latent feature 'outdoor' may be linked to attributes such as 'safe', 'powerful', 'rough', 'expensive', and 'versatile'. Formally, the link between latent features and attributes is described with binary latent variables Y_{ji}^{kf} $(f = 1, \ldots, F)$ that equal 1 if attribute k is linked to feature f when consumer i judges product-attribute pair (j, k), and 0 otherwise. PFMs assume, like in the case of the products, that the link between latent features and attributes is stochastic. In particular, it assumed that $Y_{ji}^{kf} \sim \text{Bernoulli}(\rho_{kf})$.
4. It is assumed that the observed binary association D_{ijk} between a product j and an attribute k is obtained as a predetermined mapping of the latent features that are activated when the product is being judged and of the latent features that are linked to the attribute. Formally, $D_{ijk} = C(X_{ki}^{j1}, \ldots, X_{ki}^{jF}, Y_{ji}^{k1}, \ldots, Y_{ji}^{kF})$ with $C(\cdot)$ indicating a mapping rule. Most often deterministic mapping rules are used (Maris et al. 1996) but probabilistic mapping rules have been considered as well (see e.g., Meulders, De Boeck, and Van Mechelen 2001). Maris et al. (1996) describe several types of mapping rules for PFMs. In this paper we will consider the *disjunctive* rule which reads that the product is associated to the attribute if the product activates at least one feature which is also linked to the attribute [1]. This may be formally expressed as follows:

1. Another mapping rule is the conjunctive one: An attribute is assigned to a product when the product activates *all* the features that are linked to the attribute (see e.g., Meulders et al. 2005).

$$D_{ijk} = 1 \iff \exists f : X_{ki}^{jf} = Y_{ji}^{kf} = 1. \tag{1}$$

From both the distributional assumptions of the latent variables X and Y and the mapping rule $C(\cdot)$, it follows that the probability of a product-attribute association π_{jk} is a nonlinear function of the product parameters (σ) and the attribute parameters (ρ), more specifically:

$$\begin{aligned}
\pi_{jk} &= P(D_{ijk} = 1|\boldsymbol{\sigma}, \boldsymbol{\rho}) \\
&= \sum_{x_{ki}^{j1}} \cdots \sum_{x_{ki}^{jF}} \sum_{y_{ji}^{k1}} \cdots \sum_{y_{ji}^{kF}} P(D_{ijk} = 1|\mathbf{x}_{ki}^{j}, \mathbf{y}_{ji}^{k}) p(\mathbf{x}_{ki}^{j}|\boldsymbol{\sigma}_j) p(\mathbf{y}_{ji}^{k}|\boldsymbol{\rho}_k) \\
&= 1 - \prod_f (1 - \sigma_{jf}\rho_{kf}).
\end{aligned} \tag{2}$$

with $p(\mathbf{x}_{ki}^{j}|\boldsymbol{\sigma}_j)$ and $p(\mathbf{y}_{ji}^{k}|\boldsymbol{\rho}_k)$ products of Bernoulli distributed variables and with $P(D_{ijk} = 1|\mathbf{x}_{ki}^{j}, \mathbf{y}_{ji}^{k})$ fixed 0/1 probabilities that follow from the mapping rule. The observed variables D_{ijk} are statistically independent because they are based on independent realizations of X and Y. As a result, the likelihood of the PFM equals:

$$p(\mathbf{d}|\boldsymbol{\sigma}, \boldsymbol{\rho}) = \prod_i \prod_j \prod_k {\pi_{jk}}^{d_{ijk}} (1 - \pi_{jk})^{1 - d_{ijk}}. \tag{3}$$

The model described by (2) and (3) can be considered as a latent class model with $2F$ binary latent variables at the product-attribute-respondent combination level (i.e. the level of the data points d_{ijk}).

3. Multilevel Latent Class Extensions of PFMs

In this section, we first describe a multilevel LC-PFM which allows to model person heterogeneity in object and/or attribute parameters. Second, we further extend the models with different types of stochastic assumptions to account for statistical dependencies between product-attribute pairs with a common product or attribute. Third, we discuss how the proposed multilevel LC-PFMs relate to other hierarchical mixture models for the analysis of binary three-way data.

3.1 Respondent Heterogeneity in the Product-Attribute Association Probability

As can be seen in (2), the PFM makes the assumption that the association probability π_{jk} is the same for all respondents. To model respondent heterogeneity in the association probability, we propose a multilevel latent class extension of the PFM in which respondents are assigned to one of T

classes, and product parameters (σ_{jf}) and/or attribute parameters (ρ_{kf}) may depend on latent respondent classes. The classification of a respondent i is modeled by a T-element vector $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iT})$ which has $Z_{it} = 1$ if consumer i belongs to class t, and $Z_{it} = 0$ otherwise. The latent class sizes ξ_t indicate the probability that a respondent belongs to class t with $\sum_t \xi_t = 1$. The likelihood of the model equals:

$$p(\mathbf{d}|\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) = \prod_i \sum_t \xi_t \prod_j \prod_k \pi_{jkt}{}^{d_{ijk}} (1 - \pi_{jkt})^{1-d_{ijk}}. \tag{4}$$

Using a disjunctive rule (see (1)), the conditional probability of an observed association between product j and attribute k for a respondent belonging to latent class t can be expressed as follows:

$$\begin{aligned} \pi_{jkt} &= P(D_{ijk} = 1|Z_{it} = 1, \boldsymbol{\sigma}, \boldsymbol{\rho}) \\ &= \sum_{x_{ki}^{j1}} \cdots \sum_{x_{ki}^{jF}} \sum_{y_{ji}^{k1}} \cdots \sum_{y_{ji}^{kF}} P(D_{ijk} = 1|\mathbf{x}_{ki}^j, \mathbf{y}_{ji}^k) \\ &\quad p(\mathbf{x}_{ki}^j|Z_{it} = 1, \boldsymbol{\sigma}_{jt}) p(\mathbf{y}_{ji}^k|Z_{it} = 1, \boldsymbol{\rho}_{kt}) \qquad (5) \\ &= 1 - \prod_f (1 - \sigma_{jft}\rho_{kft}). \qquad (6) \end{aligned}$$

The model described by (4) and (5) can be considered as a multilevel LC model with a T-class latent variable at the respondent level and $2F$ binary latent variables at the product-attribute-respondent level. Note that the model in (6) assumes that both the product parameters and the attribute parameters depend on the latent respondent class. This assumption means that product-attribute associations can be explained by a different set of latent features in each of the latent respondent classes. Often, it will be of substantive interest to have a common latent feature interpretation across respondent classes, which can be achieved by using a model in which only product parameters or attribute parameters vary across latent classes, that is,

$$\pi_{jkt} = 1 - \prod_f (1 - \sigma_{jft}\rho_{kf}) \tag{7}$$

or

$$\pi_{jkt} = 1 - \prod_f (1 - \sigma_{jf}\rho_{kft}). \tag{8}$$

Note that the model in (7) is best suited to obtain a common feature interpretation across latent classes because it is natural to derive the interpretation of the latent features from the attribute parameters. In addition, this model allows a study of rater differences in product perception.

3.2 Dependencies Between Product-Attribute Associations with a Common Product or Attribute

The model in (4) assumes that all the judgements of a respondent are statistically independent conditional on the latent class membership. This assumption may be unrealistic, as one may expect product-attribute associations with a certain product or with a certain attribute to be more dependent than product-attribute associations that do not have a product or attribute in common. To overcome this limitation we may further extend the models along the same lines as described by Maris (1999) or Meulders, De Boeck, and Van Mechelen (2003). More specifically, in order to model dependencies between product-attribute associations with the same product, one may assume that the pattern of latent features assigned to a product is not refreshed at each new judgement but that each respondent assigns a fixed pattern of latent features to each product. In the same way, dependencies between product-attribute associations with the same attribute can be modelled by assuming that each respondent assigns a fixed pattern of latent features to each attribute. For the multilevel LC-PFM with fixed product features per respondent, the likelihood equals:

$$\begin{aligned} p(\mathbf{d}|\boldsymbol{\sigma},\boldsymbol{\rho},\boldsymbol{\xi}) &= \prod_i \sum_t \xi_t \prod_j \sum_{\mathbf{x}_i^j} \left[\prod_k p(d_{ijk}|\mathbf{x}_i^j, Z_{it}=1, \boldsymbol{\rho}_{kt}) \right] \\ &\quad p(\mathbf{x}_i^j|Z_{it}=1,\boldsymbol{\sigma}_{jt}) \,, \end{aligned} \tag{9}$$

with

$$p(\mathbf{x}_i^j|Z_{it}=1,\boldsymbol{\sigma}_{jt}) = \prod_f (\sigma_{jft})^{x_i^{jf}} (1-\sigma_{jft})^{1-x_i^{jf}}$$

and, using a disjunctive mapping rule:

$$\begin{aligned} &p(d_{ijk}|\mathbf{x}_i^j, Z_{it}=1, \boldsymbol{\rho}_{kt}) \\ &\quad = \sum_{\mathbf{y}_k^{ji}} p(d_{ijk}|\mathbf{x}_i^j, \mathbf{y}_{ji}^k) p(\mathbf{y}_{ji}^k|Z_{it}=1,\boldsymbol{\rho}_{kt}) \\ &\quad = [1-\prod_f (1-x_i^{jf}\rho_{kft})]^{d_{ijk}} [\prod_f (1-x_i^{jf}\rho_{kft})]^{1-d_{ijk}}. \end{aligned}$$

The model in (9) can be considered as a multilevel LC model with a T-class latent variable at the respondent level, F binary latent variables (X) at the product-respondent level, and F binary latent variables (Y) at the product-attribute-respondent level. The model involves a double classification: First, to model dependencies between product-attribute associations with the same product, for each product, respondents are classified in 2^F latent *clusters*

which assign a specific pattern of latent features to the product. Second, to capture respondent heterogeneity in the product- or attribute parameters, respondents are classified in T *classes*. Note that the likelihood of a model with fixed attribute features per respondent is analogous as it can be obtained by switching the role of products and attributes in (9).

3.3 Combining Different Types of Heterogeneity- and Dependency Assumptions

The three types of heterogeneity assumptions (i.e. class-specific product parameters, class-specific attribute parameters and class-specific product- and attribute parameters) can now be combined with the three types of dependency assumptions (i.e. independence, fixed latent product features per respondent, fixed latent attribute features per respondent) so that nine different types of multilevel LC-PFMs can be specified.

4. Related Latent Class Models for the Analysis of Three-Way Binary Data

The model described by (4) and (6) can be considered as a constrained latent class model for the $J \times K$ product-attribute associations in which the conditional probabilities π_{jkt} are constrained to be a nonlinear function of $J \times F$ product parameters and $K \times F$ attribute parameters. Note that using an unconstrained T-class model to analyze associations between the $J \times K$ binary variables could be problematic if $J \times K$ is large because the number of conditional probabilities in such a model rapidly increases with the number of latent classes, leading to overparameterized models that may fail to fit the data (Formann 1992).

Vermunt (2007) considered a hierarchical mixture model for the analysis of three-way data that, like the LC-PFM, involves a double classification: First, it is assumed that respondents belong to one of T classes with Z_{it} equal to 1 if respondent i belongs to class t, and 0 otherwise. The latent class sizes ξ_t indicate the probability that a respondent belongs to class t with $\sum_t \xi_t = 1$. Second, it is assumed that, when judging product j, respondents of t belong to one of S clusters with G_{ijs} equal to 1 if respondent i belongs to cluster s when judging product j, and 0 otherwise. The latent cluster sizes $P(G_{ijs} = 1|Z_{it} = 1) = \theta_{s|t}$ indicate the conditional probability of belonging to cluster s given that one belongs to class t with $\sum_s \theta_{s|t} = 1$. The likelihood of the model equals:

$$p(\mathbf{d}|\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \prod_i \sum_t \xi_t \prod_j \sum_s \theta_{s|t} \prod_k p(d_{ijk}|G_{ijs} = 1, Z_{it} = 1, \boldsymbol{\beta}_k), \tag{10}$$

with $\boldsymbol{\beta}_k$ regression parameters that are used to model the conditional probability of a product-attribute association between product j and attribute k given that the respondent belongs to t and s. To model this conditional probability one could for instance assume:

$$P(D_{ijk} = 1|G_{ijs} = 1, Z_{it} = 1, \boldsymbol{\beta}_k) = \frac{\exp(\beta_{0k} + \beta_{sk}^{att} + \beta_{jk}^{prod})}{1 + \exp(\beta_{0k} + \beta_{sk}^{att} + \beta_{jk}^{prod})}. \tag{11}$$

In other words, it is assumed that the probability to assign a particular attribute to a(ny) product differs among the S clusters, and that it is more likely to assign certain attributes to certain products, but that this effect is equally strong for all clusters. In (11), it is also assumed that the regression parameters β_{0k}, β_{sk}^{att} and β_{jk}^{prod} are equal across T classes of Z. However, the latter assumption could be relaxed if needed. The model in (10), which is a variant of the multilevel LC model proposed by Vermunt (2003), can be considered as a multilevel LC model with a T-class latent variable at the respondent level and a S-cluster latent variable at the product-respondent level.

The multilevel LC-PFM with T classes at the respondent level and with F fixed latent product features in (9) is closely related to the hierarchical mixture model with T classes at the respondent level and $S(= 2^F)$ clusters at the product-respondent level in (10) as both models are multilevel LC models that involve a double classification. However, there are also some notable differences: First, in the hierarchical mixture model, the marginal distribution of G is saturated, whereas in the multilevel LC-PFM the marginal distribution of $X^1, \ldots, X^F$ is modelled as a product of Bernoulli distributions. Second, the conditional probability of a product-attribute association given the latent class membership is modelled differently in both models: In the multilevel LC-PFM, the conditional product-attribute association probability $p(d_{ijk}|\mathbf{x}_i^j, Z_{it} = 1, \boldsymbol{\rho}_{kt})$ is a nonlinear function of product- and attribute parameters which follows directly from the specific pattern of features ascribed to a product, the strength of the feature-attribute links in class t, and the specific mapping rule used by the model. In contrast, in the hierarchical mixture model the logit of the conditional product-attribute association probability $p(d_{ijk}|G_{ijs} = 1, Z_{it} = 1, \boldsymbol{\beta}_k)$ is modelled as a linear function of product, attribute, and product-attribute parameters that can be made class- or cluster-specific.

Finally, hierarchical mixture models with a double classification structure have also been applied to other contexts: Vicari and Alfò (2010) used a hierarchical mixture model to derive a joint partition of customers and products based on two-way discrete choice data. Jordan and Jacobs (1994) and Titsias and Likas (2002) discussed hierarchical mixtures of experts for

supervised classification. Li (2005) presented a model-based approach to clustering in which the density of each cluster is modelled with a mixture of normals. However, as these models do not apply to three-way binary data they will not be further considered.

5. Estimation

For parameter estimation of multilevel LC-PFMs one may use an EM-algorithm (Dempster, Laird, and Rubin 1977) to compute maximum likelihood estimates because the complete-data loglikelihood has a simple structure:

$$p(\mathbf{d}, \mathbf{x}, \mathbf{y}, \mathbf{z}|\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \propto p(\mathbf{d}|\mathbf{x}, \mathbf{y}, \mathbf{z})p(\mathbf{x}|\mathbf{z}, \boldsymbol{\sigma})p(\mathbf{y}|\mathbf{z}, \boldsymbol{\rho})p(\mathbf{z}|\boldsymbol{\xi}).$$

However, as maximum likelihood estimates will not always exist within the interior of the parameter space, it is convenient to specify a concave prior distribution on the parameters as this will guarantee the existence of posterior mode estimates in the interior of the parameter space. For a multilevel LC-PFM in which both product- and attribute parameters depend on the latent respondent class, a conjugate prior distribution can be specified as follows:

$$\begin{aligned} p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \quad = \quad & \prod_j \prod_f \prod_t \text{Beta}(\sigma_{jft}|1 + \frac{\alpha_\sigma}{JT}, 1 + \frac{\beta_\sigma}{JT}) \\ & \prod_k \prod_f \prod_t \text{Beta}(\rho_{kft}|1 + \frac{\alpha_\rho}{KT}, 1 + \frac{\beta_\rho}{KT}) \\ & \times \text{Dir}(\boldsymbol{\xi}|1 + \frac{\gamma_1}{T}, \ldots, 1 + \frac{\gamma_T}{T}). \end{aligned} \tag{12}$$

A concave prior distribution can be specified by using values larger than 0 for the constants $\alpha_\sigma, \beta_\sigma, \alpha_\rho, \beta_\rho$ and $\gamma_1, \ldots, \gamma_T$ in (12). Note that in (12), the constants $\alpha_\sigma, \beta_\sigma, \alpha_\rho, \beta_\rho$ and $\gamma_1, \ldots, \gamma_T$ are weighted (with $1/JT$, $1/KT$ and $1/T$) to obtain exactly the same prior as in Latent Gold (Vermunt and Magidson 2005; 2008).

All the presented multilevel LC-PFMs, can be estimated using the syntax module of the standard latent class software Latent Gold (version 4.5). To illustrate the syntax for estimating multilevel LC-PFMs with Latent Gold, we consider the analysis of binary judgements of respondents on the associations between 10 products and 5 attributes. Table 1 shows how the data should be structured for the analysis: The observed judgements should be structured in one column (D), and the variables 'respondent', 'product' and 'attribute' should be added to further describe each record. Note that, the data should be sorted by respondent if a latent classification at the respondent

Table 1. Structure of data on product-attribute associations for analysis with Latent Gold 4.5

respondent	product	attribute	D
1	1	1	1
1	1	2	0
1	1	3	1
1	1	4	1
1	1	5	0
1	2	1	1
1	2	2	0
1	2	3	0
1	2	4	0
1	2	5	1
2	1	1	0
...			

level is used, and that, within respondents, the data should be sorted by product (attribute) if a latent classification at the product (attribute)-respondent level is used. Appendix A describes the syntax code that can be used to estimate a 2-feature 2-class ($T = 2, F = 2$) disjunctive multilevel LC-PFM with fixed latent product features (i.e. $d_{ijk} = C(\mathbf{x}_i^j, \mathbf{y}_{ji}^k)$) and heterogeneity in only product parameters (i.e. $\pi_{jkt} = f(\boldsymbol{\sigma}_{jt}, \boldsymbol{\rho}_k)$). Adapting the syntax code to estimate other multilevel LC-PFMs is straightforward.

As can be seen in the specified syntax code (see lines 18-41 of Appendix A), Latent Gold evaluates the association probability $P(D_{ijk} = 1|Z_{it} = 1, \boldsymbol{\sigma}, \boldsymbol{\rho})$ of multilevel LC-PFMs by summing the joint probability $P(D_{ijk} = 1, \mathbf{x}, \mathbf{y}|Z_{it} = 1, \boldsymbol{\sigma}, \boldsymbol{\rho})$ over all latent data patterns $(\mathbf{x}, \mathbf{y})$ (see for instance (5)). As the number of latent data patterns $(= 2^{JF+KF})$ rapidly increases with the number of latent features, estimation of multilevel LC-PFMs with many latent features $(F > 5)$ may become slow (although it remains possible using Latent Gold). In order to solve this problem we have implemented an EM algorithm in which the association probabilities are directly computed using the analytical results. Appendix B describes the derivation of an EM-algorithm for a disjunctive multilevel LC-PFM with fixed product features (i.e. $d_{ijk} = C(\mathbf{x}_i^j, \mathbf{y}_{ji}^k)$) and heterogeneity in only product parameters (i.e. $\pi_{jkt} = f(\boldsymbol{\sigma}_{jt}, \boldsymbol{\rho}_k)$). For all the models proposed in the paper, EM-algorithms were derived. The results of the implemented algorithms (parameter estimates, loglikelihood) for applications presented in the paper were identical to the results obtained with Latent Gold.

6. Model Comparison and Model Checking

To choose the number of latent features and the number of classes of multilevel LC-PFMs, one may use information criteria such as the Akaike

Information Criterion (AIC, Akaike 1973; 1974) or the Bayesian Information Criterion (BIC, Schwarz 1978). Both measures are obtained by summing a badness-of-fit term (i.e. $-2\ln[p(\mathbf{d}|\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\xi}})]$) and a penalty term, depending on the number of free parameters k. For AIC and BIC the penalty terms equal $2k$ and $k\log(I)$, with k the number of free parameters and I the sample size. With PFMs, the sample size equals the number of raters I rather than the total number of observed judgements $I \times J \times K$. As indicated by Miloslavsky and van der Laan (2003), the approximations used in AIC and BIC are also based on regularity conditions that do not hold in the mixture case. However, as both criteria are easy to compute when the posterior mode has been located and as they perform well if the sample size is sufficiently large (Meulders et al. 2003; Miloslavsky and van der Laan 2003) they are often used in latent class analysis. Furthermore, as in contrast to AIC, BIC has been shown to be a consistent estimate of the number of components in a mixture (Keribin 2000), we will use this criterion for model comparison.

Once a specific multilevel LC-PFM has been selected, it is important to check whether this model can capture key substantive aspects of the data. A first aspect of interest is whether observed frequencies of product-attribute associations ($\sum_i d_{ijk}$), can be captured by the model. A simple descriptive analysis to check this is to compute the correlation between observed and expected frequencies.

A second aspect of interest is whether observed rater differences can be captured by the model. To check this, we will compare observed and expected correlations (computed across raters) between all pairs of product-attribute associations (j, k) and (j', k'). To evaluate whether observed correlations can be fitted by the model we use a parametric bootstrap procedure (Efron and Tibshirani 1993) to simulate the reference distributions of the correlations under the model, and we check whether observed correlations are within the corresponding simulated 99% confidence interval (CI).

6.1 Example 1: Situational Determinants of Anger-Related Behavior

In the last decade the focus of personality research has shifted from identifying stable personality traits (which explain behavior across situations) to identifying person types with stable situation-behavior relations (Mischel and Shoda 1998; Mischel, Shoda, and Mendoza-Denton 2002). As will be shown in this section, multilevel latent class PFMs are especially well-suited to meet the latter challenge: More specifically, the model allows classifying behaviors in terms of more basic 'behavior types' or 'behavioral strategies', and it allows classifying situations with respect to the behavior types that are considered appropriate in these situations. On top of that, the model allows identifying person types with distinct situation-behavior

profiles by taking into account individual differences in the classification of situations.

6.1.1 Data

To illustrate multilevel LC-PFMs we analyze binary judgements of 101 first-year psychology students who indicated whether or not they would display each of 8 anger-related behaviors when being angry at someone in each of 6 situations (see also, Kuppens, Van Mechelen, and Meulders 2004; Meulders et al. 2002; Vermunt 2007). The 8 behaviors consist of 4 pairs of reactions that reflect a particular strategy to deal with situations in which one is angry at someone, namely, (1) fighting (fly off the handle, quarrel), (2) fleeing (leave, avoid), (3) emotional sharing (pour out one's heart, tell one's story), and (4) making up (make up, clear up the matter). The six situations are constructed from two factors with three levels: (1) the extent to which one likes the instigator of anger (like, dislike, unfamiliar), and (2) the status of the instigator of anger (higher, lower, equal). Each situation is presented as one level of a factor, without specifying a level for the other factor.

6.1.2 Analysis

To explain associations between situations and behaviors we use Latent Gold to fit disjunctive multilevel LC-PFMs with one up to five latent features ($f = 1, \ldots, 5$), one up to 5 latent classes ($t = 1, \ldots, 5$), three types of heterogeneity assumptions (i.e. class-specific situation parameters, class-specific behavior parameters, class-specific situation- and behavior parameters) and three types of dependency assumptions (independence, fixed latent situation features per respondent, fixed latent behavior features per respondent). For each of the resulting 225 $(= 5 \times 5 \times 3 \times 3)$ models, the posterior mode(s) were estimated using 20 runs with random starting points, and the solution with the highest posterior density was selected.

For each of the five models with lowest BIC, it was checked to what extent the model can capture dependencies between pairs of situation-behavior judgements. We used a parametric bootstrap procedure based on 2000 replicated datasets to simulate the reference distributions of all correlations $\text{cor}_i(d_{ijk}, d_{ij'k'})$ under the model, and we compute the proportion of observed correlations that are outside the corresponding 99% CI. As one may expect extra strong dependencies between situation-behavior pairs with a common situation or with a common behavior, we also computed, for each of these types, the proportion of observed correlations that lie outside the 99% CI.

Table 2 lists for the five models with lowest BIC, the BIC values and the proportion of all observed correlations that are outside the 99% CI. The results of the analysis indicate a remarkable resemblance between the five

Table 2. Number of features (F), number of latent classes (T), elements with a fixed feature-classification (fixed classif.), elements with class-specific parameters (heterogeneity), loglikelihood (LL), number of parameters (k), BIC value, proportion of all observed correlations outside the 99% CI (p_{all}), proportion of observed correlations between situation-behavior pairs with a common situation outside the 99% CI (p_{sit}), proportion of observed correlations between situation-behavior pairs with a common behavior outside the 99% CI (p_{behav}), and correlation between observed and expected frequencies in the $J \times K$ table ($cor_{J\times K}$) for the five models with lowest BIC applied to the situation-behavior data.

F	T	fixed classif.	heterogeneity	LL	k	BIC	p_{all}	p_{sit}	p_{behav}	$cor_{J\times K}$
4	1	situation	situation	-2510	56	5279	.05	.20	.12	.99
4	2	situation	situation	-2459	81	5291	.03	.14	.08	.99
5	1	situation	situation	-2492	70	5307	.06	.20	.11	.99
4	3	situation	situation	-2424	106	5337	.03	.08	.08	.99
5	2	situation	situation	-2444	101	5355	.03	.11	.08	.99

best models according to BIC as all of these models assume 4 or 5 latent features, class-specific situation parameters, and a fixed classification of situations per respondent. Table 2 further shows that the model with the lowest BIC value $(F = 4, T = 1)$ does not fully capture the observed dependencies between situation-behavior pairs as about 5% of the residual correlations are significantly different from 0 at the 1% level. Also, the model does not fully capture dependencies between situation-behavior pairs with a common situation or a common behavior as 20% and 12% of these types of correlations are outside the 99% CI.

Of all the models in Table 2, model $(F = 4, T = 3)$ seems to have the best performance to fit dependencies between situation-behavior pairs (i.e. 97% of all observed correlations and 92% of the correlations with a common situation or behavior are in the 99% CI). Therefore, we will further discuss this model. Note that this model is also very similar to model $(F = 4, T = 2)$ because the situation-behavior profiles of the latter model are very similar to those of the first two classes of model $(F = 4, T = 3)$.

6.1.3 Interpretation of the Selected Model

Table 3 lists the estimated behavior parameters of the selected model. As can be seen in Table 3, it turns out that each of the four extracted features shows especially strong links with behaviors that reflect a specific type of reaction. In particular, F_1, F_2, F_3 and F_4 respectively have strong links with the reaction types 'emotional sharing' (pour out one's hart (.85), tell one's story (.91)), 'making up' (make up (.96), clear up the matter (.87)), 'fighting' (fly off the handle (.74), quarrel (.49)) and 'fleeing' (leave (.75), avoid (.85)).

Table 3. Estimated behavior parameters for a four-feature model (F=4) with class-specific situation parameters (T=3) and fixed situation-features per respondent. The highest behavior parameters are displayed in bold.

behavior	F_1	F_2	F_3	F_4
fly off the handle	.00	.16	**.74**	.00
quarrel	.04	.33	**.49**	.01
leave	.02	.00	.01	**.75**
avoid	.08	.01	.01	**.85**
pour out one's hart	**.85**	.04	.01	.08
tell one's story	**.91**	.00	.15	.12
make up	.00	**.96**	.07	.06
clear up the matter	.07	**.87**	.06	.00

Figures 1 to 4 visualize for each feature the estimated class-specific situation parameters and the corresponding 95% CIs. The class-specific situation parameter σ_{jft} can be interpreted as the probability that persons in latent class t consider the reaction type associated to latent feature f to be appropriate in situation j. Note that, the three latent classes contain 32%, 50%, and 18% of the respondents, respectively.

As can be seen in Figures 1 to 4, in some cases situational characteristics are very strong leaving little room for individual differences in the displayed behavior. For instance, for all latent classes, 'making up' is very likely when being angry at a liked person, or when being angry at a person of lower or equal status and it is very unlikely when being angry at someone disliked or unfamiliar. Also, for all respondents, emotional sharing is very likely when being angry at a disliked person or when being angry at a person of equal status.

However, in most cases different person types will react in a different way to different types of situations. In particular, as can be seen in Figures 3 and 4 the first two latent classes can be considered as an approach and an avoidance class: In all situations, except when being angry at someone of equal status, respondents of class 2 are more likely to flight (i.e. avoid a confrontation) than respondents of class 1. In contrast, the estimated propensity to fight (i.e. approach) is generally higher for respondents of class 1 than for respondents of class 2. However, the difference in propensity to fight between the two classes is only significant when one is angry at a liked person, a person of higher status, or an unfamiliar person. Furthermore, respondents of the third latent class are very likely to fight and/or make up when being angry at a liked person or when being angry at a person of lower or equal status, and they are very likely to flight (and not make up) when being angry at someone disliked, unfamiliar or of higher status. Finally, in contrast to

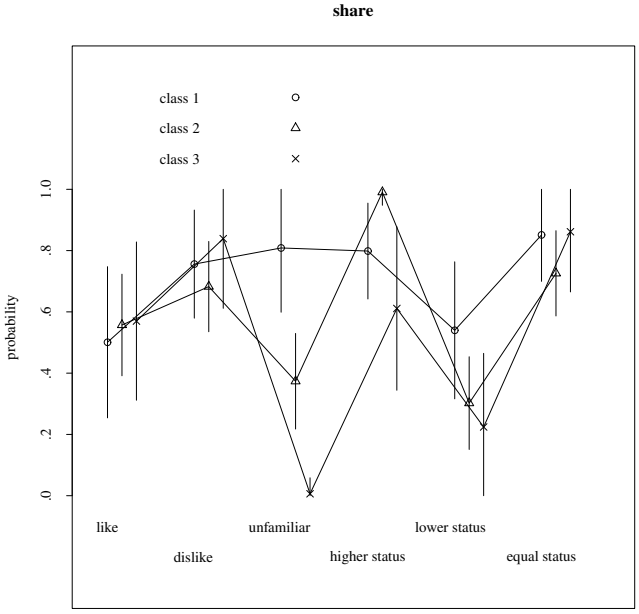

Figure 1. Propensity to share emotions (pour out one's heart, tell one's story) when being angry at someone in a specific situation.

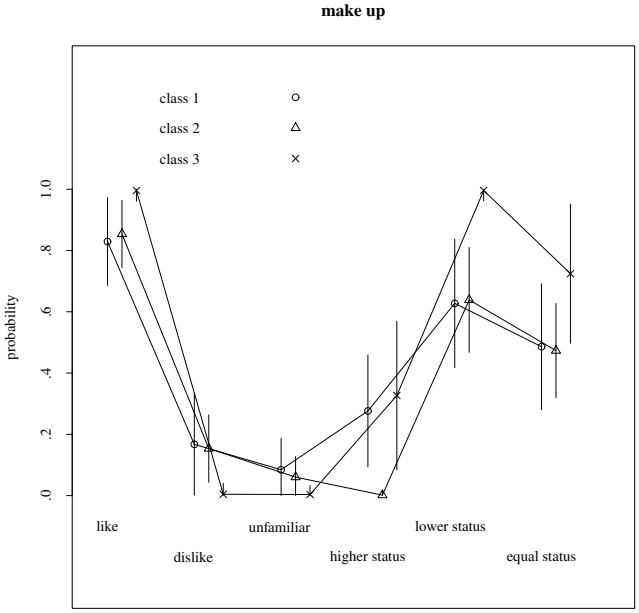

Figure 2. Propensity to make up (make up, clear up the matter) when being angry at someone in a specific situation.

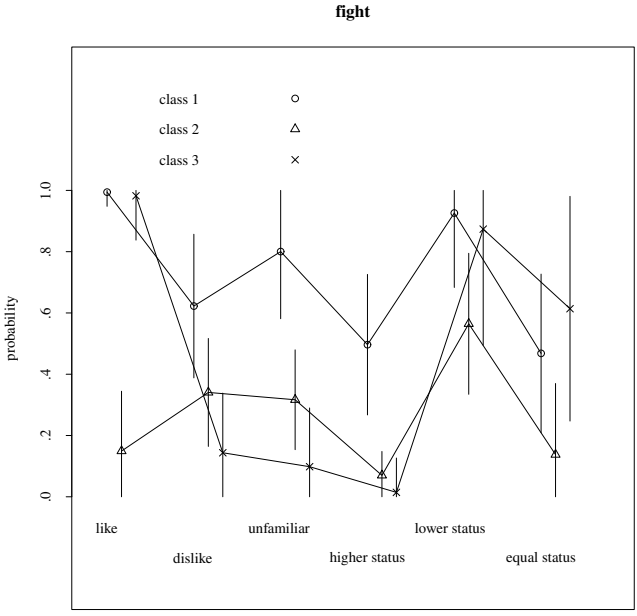

Figure 3. Propensity to fight (fly off the handle, quarrel) when being angry at someone in a specific situation.

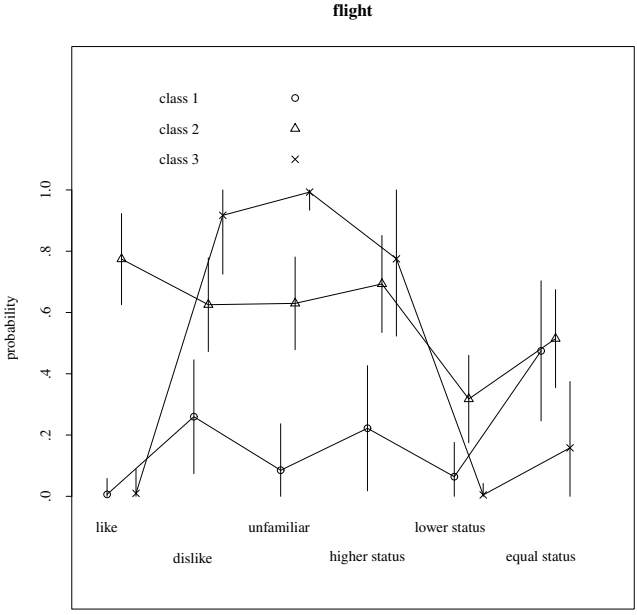

Figure 4. Propensity to flight (leave, avoid) when being angry at someone in a specific situation.

respondents of the other classes, respondents of class 3 are very unlikely to share emotions when being angry at a person they are unfamiliar with.

6.1.4 Comparison with Related Models

The fit of the selected 4-feature 3-class multilevel LC-PFM (with BIC equal to 5337) can be compared with several related models using the BIC criterion. First, standard T-class LC models for the $J \times K$ situation-behavior pairs with 1 up to 4 latent classes were fitted. The 2-class model has the lowest BIC value, namely 5928. Second, we fitted hierarchical mixture models defined by (10) and (11) (with j indicating situations and k indicating behaviors) assuming 1 up to 4 latent classes and 1 up to 4 latent clusters. The 3-class 4-cluster model has the lowest BIC value, namely 5324. However, as shown by Vermunt (2007), the fit of this model can be considerably improved by including direct effects to model the dependency between behavior pairs of the same reaction type (i.e. fight, flight, share, and make-up), yielding a BIC value of 5111. In sum, we may conclude that the selected multilevel LC-PFM has a better fit in terms of BIC than a standard latent class model, that it has a slightly higher BIC value than a standard hierarchical mixture model, but that it is outperformed by a hierarchical mixture model that includes direct effects between behaviors of the same type.

Although the multilevel LC-PFM has a lower fit than a standard hierarchical mixture model, it offers some substantive advantages: First, in contrast to the hierarchical mixture model, the multilevel LC-PFM provides a reduction of behaviors (in terms of the involved reaction style) and of situations (in terms of the reaction styles that are appropriate in the situation). Second, in line with the objectives of personality research, the multilevel LC-PFM allows to identify person types with specific situation-behavior relations. In contrast, the hierarchical mixture model assumes that person types only differ in the behaviors they will typically select. Moreover, the model assumes that certain behaviors are more likely to occur in certain situations, but that this effect is the same for all person types (see Vermunt 2007).

6.2 Example 2: Analysis of Rater Differences in Product Perception

Strategic marketing decisions about (re)positioning products are often based on information about the competitive structure of the products in a certain product space. Such competitive structure analysis typically involves using a perceptual mapping technique to derive a dimension-based spatial configuration of products (DeSarbo, Grewal, and Scott 2008). However, as the metric and dimensional assumptions at the basis of geometric representations may be doubtful, non-spatial approaches based on the categorization of products were proposed as an alternative (for an overview of non-spatial

techniques, see DeSarbo, Manrai, and Manrai 1993). As shown by Candel and Maris (1997), PFMs are well-suited to derive a categorization-based perceptual mapping from binary product-attribute associations.

In this application, we will use multilevel LC-PFMs with class-specific product parameters to model rater differences in product perception. Investigating differences in product perception among raters may be important for two reasons: First, if rater classes have a different product perception (e.g., in the sense that they assign different feature patterns to products), the derived competitive structure for the products will also be class-specific, and as a result, decisions based on the competitive structure should account for these rater differences in product perception. Second, an external preference analysis, which involves modelling product preference on the basis of product coordinates obtained with a perceptual mapping (i.e. product-feature probabilities), will be suboptimal if rater differences in product perception are ignored.

6.2.1 Data

To illustrate using multilevel LC-PFMs for the analysis of rater differences in product perception, we collected and analyzed binary judgements of 78 respondents who indicated for each of 27 attributes whether they apply to each of 14 car models.

6.2.2 Analysis

Disjunctive multilevel LC-PFMs with one up to seven latent features ($f = 1, \ldots, 7$), one up to 5 latent classes ($t = 1, \ldots, 5$), class-specific product parameters (see (7)), and three types of dependency assumptions (independence, fixed latent product features per respondent, fixed latent attribute features per respondent) were estimated using an EM-algorithm. Attribute parameters were constrained to be equal across latent classes (see (7)) to guarantee a common feature interpretation for all respondents. For each of the $105\ (= 7 \times 5 \times 3)$ models, the posterior mode(s) were estimated using 20 runs with random starting points, and the solution with the highest posterior density was selected.

Table 4 lists for the five best models (according to BIC) how well dependencies between product-attribute pairs can be fitted and how well product-attribute frequencies in the $J \times K$ table can be predicted by the model. As can be seen, the model with the lowest BIC (model $(F = 7, T = 3)$ with fixed latent attribute features) also fits the dependencies between product-attribute associations best: 98% of all correlations between product-attribute pairs are in the 99% CI. Also, 94% of the correlations between product-attribute pairs with a common product and 97% of the correlations between product-attribute pairs with a common attribute are in the 99% CI.

Table 4. Number of features (F), number of latent classes (T), elements with a fixed feature-classification (fixed classif.), loglikelihood (LL), number of parameters (k), BIC value, proportion of all observed correlations outside the 99% CI (p_{all}), proportion of observed correlations between product-attribute pairs with a common product outside the 99% CI (p_{prod}), proportion of observed correlations between product-attribute pairs with a common attribute outside the 99% CI (p_{att}), and correlation between observed and expected frequencies in the $J \times K$ table ($cor_{J\times K}$) for the five models with lowest BIC applied to the car data. Only models assuming class-specific product-parameters were considered.

F	T	fixed classif.	LL	k	BIC	p_{all}	p_{prod}	p_{att}	$cor_{J\times K}$
7	3	attribute	-11960	485	26033	.02	.06	.03	.96
7	2	attribute	-12183	386	26049	.03	.08	.03	.96
7	2	product	-12221	386	26124	.03	.03	.12	.95
6	2	attribute	-12359	331	26159	.03	.09	.04	.95
7	1	product	-12466	288	26183	.06	.04	.21	.96

Note that, because of the specific double classification of respondents involved, models with fixed latent product features per respondent can better fit correlations between attributes with a certain product than correlations between products with a certain attribute, and that in the same way, models with fixed latent attribute features per respondent can always better fit correlations between products with a certain attribute than correlations between attributes with a certain product. Finally, Table 4 shows that models with 6 or 7 features all fit the product-attribute frequencies in the $J \times K$ table rather well as the correlation between observed and expected frequencies ranges from .95 to .96. As model $(F = 7, T = 3)$ with fixed latent attribute features has the best overall performance we will further discuss this model.

6.2.3 Interpretation of the Selected Model

Tables 5 and 6 present estimates of attribute- and product parameters for the selected model. Note that the three classes presented in Table 6, contain respectively 27%, 18% and 55% of the raters. As can be seen in the tables the extracted latent features have a meaningful interpretation: Features 1 and 7 are similar in that both are most likely ascribed to small (i.e., agile with city-focus) popular cars. However, these features also have a different meaning as F_1 refers to economical cars with a good price-quality ratio (e.g., 'Opel Corsa'), and F_7 refers to attractive cars which are original and have a nice design (e.g., 'Mini Cooper'). Feature 2 is ascribed to cars that are considered to be reliable, safe, and versatile (e.g., 'Volvo V50'). Features 3 and 6 are similar in that both are ascribed to luxurious, powerful and attractive cars which are regarded as a status symbol. However, F_6 is ascribed to cars that are especially suited for outdoor activities (e.g. 'BMW X5') whereas

Table 5. Estimated attribute parameters for a seven-feature model ($F = 7$) with class-specific product parameters ($T = 3$) and fixed latent attribute features per respondent. Probabilities larger than or equal to .50 are displayed in bold.

attribute	F_1	F_2	F_3	F_4	F_5	F_6	F_7
Economical	**.88**	.18	.00	.02	**.75**	.00	.08
Agile	**.68**	.13	.20	.00	.00	.13	**.98**
Environmentally friendly	.32	.12	.03	.01	**.89**	.00	.12
Reliable	.35	**.99**	**.83**	.05	.19	.06	.10
Practical	**.83**	**.54**	.08	**.57**	.00	.00	.45
Family Oriented	.00	.36	.15	**1.00**	.32	.00	.00
Versatile	.02	**.69**	.00	**.51**	.07	.27	.00
Good price-quality ratio	**.90**	.24	.04	.02	.33	.09	.06
Luxurious	.00	.05	**.98**	.08	.01	**.82**	.12
Safe	.20	**.88**	**.70**	.15	.02	.04	.00
Sporty	.04	.00	**.65**	.00	.00	**1.00**	**.56**
Attractive	.14	.13	**.74**	.01	.03	**.56**	**.87**
Comfortable	.09	**.50**	**.96**	**.60**	.00	.14	.04
Powerful	.00	.05	**.88**	.06	.00	**.91**	.00
Status symbol	.00	.00	**.92**	.01	.12	**.62**	.40
Technically advanced	.02	.02	**.73**	.00	**.71**	.26	.00
Sustainable	.15	**.88**	.40	.04	.29	.03	.05
Original	.00	.00	.03	.00	.26	.27	**.86**
Nice design	.07	.08	**.53**	.02	.06	**.51**	**.86**
Value for the money	**.65**	.29	.05	.00	.09	.11	.02
High trade-in value	.00	.08	**.82**	.00	.00	.08	.03
Exclusive	.03	.02	.17	.00	.07	.30	.40
Popular	**.81**	.12	.25	.11	.06	.00	**.57**
Outdoor	.00	.01	.05	.24	.00	**.85**	.02
Green	.01	.06	.00	.02	**.61**	.00	.09
City focus	**.79**	.06	.00	.00	.37	.00	**.84**
Workmanship	.00	.09	**.65**	.02	.25	.19	.02

F_3 is ascribed to comfortable high-quality (i.e., safe, reliable, workmanship, technically advanced) cars (e.g. 'Mercedes C-class'). Finally, feature 4 is ascribed to comfortable, family-oriented (i.e. practical, versatile) cars (e.g. 'Renault Espace'), and feature 5 is ascribed to technically advanced, environmentally friendly (i.e. green) cars such as the 'Toyota Prius' which uses hybrid drive technology to minimize gas consumption and CO_2 emissions.

6.2.4 Differences in Product Perception

An ANOVA decomposition of the (logit of the) product parameters in Table 6 as a function of the factors product (P), feature (F), class (C) and their two-way interactions $P \times F$, $P \times C$, and $F \times C$ shows that respectively 3%, 8%, 17%, 53%, 1% and 3% of the total variation in the product

Table 6. Product parameters for a seven-feature model ($F = 7$) with class-specific product parameters ($T = 3$) and fixed latent attribute features per respondent.

class	car model	F_1	F_2	F_3	F_4	F_5	F_6	F_7
1	Volkswagen Golf	.34	.74	.44	.06	.09	.16	.47
2	Volkswagen Golf	.00	.09	.12	.02	.00	.00	.06
3	Volkswagen Golf	.57	.43	.21	.00	.06	.00	.02
1	Opel Corsa	.89	.19	.02	.10	.07	.00	.14
2	Opel Corsa	.26	.00	.00	.00	.00	.00	.03
3	Opel Corsa	.60	.10	.00	.07	.00	.00	.06
1	Nissan Qashgai	.11	.38	.11	.65	.07	.62	.08
2	Nissan Qashgai	.01	.02	.00	.19	.02	.14	.04
3	Nissan Qashgai	.10	.19	.00	.41	.01	.27	.00
1	Toyota Prius	.09	.28	.02	.55	.87	.04	.06
2	Toyota Prius	.00	.01	.01	.07	.36	.04	.00
3	Toyota Prius	.00	.13	.13	.17	.77	.00	.07
1	BMW X5	.00	.51	.81	.44	.00	.88	.01
2	BMW X5	.00	.00	.12	.05	.00	.25	.00
3	BMW X5	.00	.08	.38	.45	.00	.50	.00
1	Volvo V50	.10	.81	.41	.57	.22	.17	.00
2	Volvo V50	.00	.13	.04	.15	.00	.04	.00
3	Volvo V50	.08	.62	.10	.59	.00	.07	.00
1	Renault Espace	.25	.40	.03	1.0	.12	.02	.08
2	Renault Espace	.01	.00	.00	.65	.00	.00	.03
3	Renault Espace	.14	.13	.00	.85	.00	.05	.11
1	Citroen C4 Picasso	.40	.26	.00	.90	.13	.01	.20
2	Citroen C4 Picasso	.03	.00	.00	.42	.00	.00	.04
3	Citroen C4 Picasso	.25	.04	.00	.86	.00	.02	.02
1	Ford Focus Cmax	.35	.57	.06	.75	.00	.05	.16
2	Ford Focus Cmax	.11	.00	.00	.18	.00	.03	.00
3	Ford Focus Cmax	.23	.18	.02	.61	.00	.00	.06
1	Mercedes C-class	.00	.49	.96	.08	.00	.01	.18
2	Mercedes C-class	.00	.00	.20	.01	.00	.11	.00
3	Mercedes C-class	.02	.14	.73	.00	.00	.02	.05
1	Fiat 500	.53	.03	.00	.11	.18	.03	.65
2	Fiat 500	.04	.00	.00	.00	.00	.00	.42
3	Fiat 500	.29	.00	.00	.00	.19	.00	.44
1	Audi A4	.00	.76	.79	.26	.00	.17	.37
2	Audi A4	.00	.00	.23	.07	.00	.09	.00
3	Audi A4	.13	.47	.65	.12	.00	.03	.05
1	Mini Cooper	.00	.21	.48	.00	.00	.12	.94
2	Mini Cooper	.00	.00	.00	.00	.00	.07	.33
3	Mini Cooper	.02	.03	.14	.00	.00	.15	.73
1	Mazda MX5	.11	.00	.25	.00	.08	.44	.30
2	Mazda MX5	.00	.01	.00	.00	.00	.33	.00
3	Mazda MX5	.02	.00	.14	.01	.00	.42	.21

parameters is explained by these effects, and that 15% of the total variation is due to error or a three-way interaction between the factors. Furthermore, all the effects, except the $P \times C$ interaction are significant at the 0.001 level.

The strong $P \times F$ interaction, which reflects the tendency to assign specific features to specific products, indicates a strong psychological basis for a common product-categorization in terms of latent features by raters. However, there is also evidence for class-specific effects. First, the moderate main-effect of class provides evidence for class-specific distributions of the product parameters. More specifically, the averages of the product parameters in the three classes equal 0.27 0.06 and 0.16. This class-effect may be interpreted as a response-style effect (Van Rosmalen, Van Herk, and Groenen 2010), as it can be explained as a tendency of classes to generally pick more or less attributes per product when completing the survey. Second, further inspection of the product parameters also shows that the classes can have a different perception of specific car models. For instance, in class 3, the 'Volkswagen Golf' is regarded as a popular city car with a good price-quality ratio (probability of .57 for F_1) and not as an attractive, original city car with a nice design (probability of .02 for F_7), whereas in class 1 this model is more likely to be regarded as a reliable safe car (probability of .74 for F_2), which is attractive, original and has a nice design (probability of .47 for F_7), and less as a popular car with a good price-quality ratio (probability of .34 for F_1). As a second example, we see that the 'Toyota Prius' is primarily regarded by class 3 as an environmentally friendly, economical, green car (probability of .77 for F_5 and low probabilities for the other latent features), whereas class 1 perceives this car not only as 'green' (probability of .87 on F_5), but also as a comfortable family oriented car (probability of .55 for F_4).

6.2.5 Comparison with Related Models

Unconstrained latent class models for the $J \times K$ product-attribute pairs with 1 up to 4 classes were fitted using Latent Gold. The 2-class model has the lowest BIC, namely 28074. In addition, hierarchical mixture models defined by (10) and (11) (with j indicating products and with k indicating attributes, or with j indicating attributes and k indicating products) assuming 1 up to 9 latent classes and assuming 1 up to 9 latent clusters were fitted. The 7-class 8-cluster model (with 7 classes at the respondent level and 8 clusters at the attribute-respondent level) turns out to have the lowest BIC, namely 24951. Hence, the selected multilevel LC-PFM (with BIC equal to 26033) fits better than a standard unconstrained latent class model, but it is outperformed by the hierarchical mixture model.

Although the selected multilevel LC-PFM fits less well than the hierarchical mixture model in the present example, it still remains an interesting

alternative for several reasons: First, in contrast to the hierarchical mixture model, the multilevel LC-PFM yields a reduced description of products and attributes in terms of binary latent features. As a result, parameter interpretation is often easier than with the hierarchical mixture model, especially when many products and attributes are involved. Second, the multilevel LC-PFM is well-suited to detect latent classes with a different product perception (i.e. in terms of the latent features they assign to products). In contrast, the hierarchical mixture model does not include cluster-specific product-attribute relations. That is, it assumes that respondents of different latent clusters differ in the attributes they typically pick, and that certain attributes are more likely to be assigned to certain products, but that the latter effect is the same across latent clusters.

7. Discussion

In this paper, we presented multilevel latent class extensions of PFMs to extract latent features underlying binary judgements of raters about object-attribute associations. Specifically, PFMs with different types of statistical dependence assumptions (independence, fixed object features, fixed attribute features) were extended in a multilevel latent class framework so that object and/or attribute parameters can be made class-specific. The proposed multilevel latent class extensions turn out to be advantageous for several reasons:

First, as illustrated in the examples, modelling the heterogeneity in object- or attribute parameters can be interesting from a substantive point of view. For instance in the study on the situational determinants of anger-related behavior the heterogeneity in the situation parameters reveals a small number of clearly interpretable respondent classes who tend to react differently in certain anger-related situations. On the other hand, in the example on car perception, introducing heterogeneity in the product-parameters allows to identify rater classes who adopt a different response style (i.e. assign generally more or less attributes to a car when completing the survey), and who also have a different perception of specific car models in terms of latent features.

Second, as shown in the examples, modelling heterogeneity in the object-or attribute parameters may help to better fit the dependencies between object-attribute pairs and as a result to better capture related respondent differences. More specifically, where the dependence assumption of fixed object/attribute features especially helps to fit correlations between pairs with a common object/attribute (see also Meulders et al. 2003), introducing heterogeneity in the object/attribute parameters further improves to fit the correlations between pairs with a common attribute/object.

A comparison of multilevel LC-PFMs with standard hierarchical mixture models for three-way data (Vermunt 2007) shows that the latter are

likely to outperform the developed multilevel LC-PFMs in terms of model fit, but that multilevel LC-PFMs may still fit important substantive aspects of the data rather well. Furthermore, when modelling complex data (i.e. with many objects and many attributes as in the example on car perception), hierarchical mixture models may become complex (i.e. including many latent classes and latent clusters) and therefore hard to interpret. In contrast, multilevel LC-PFMs aim for a reduction of the attributes in terms of a small number of latent features. In as far as the extracted latent features have a simple interpretation, interpreting the object parameters in terms of the extracted features is also straightforward.

On a practical level, it is useful to note that the presented multilevel LC-PFMs can be estimated with the syntax module of Latent Gold. This enhances the application of the models by other researchers, and also opens the door to further model extensions, for instance by including person, object- or attribute covariates in the model.

Appendix A: Latent Gold 4.5 Syntax Code for the Estimation of Multilevel LC-PFMs

A.1 Syntax code

```
model
title fixedproductfeature-heterogeneousproductparameters-T2-F2;
options
algorithm tolerance=1e-010 emtolerance=0.01 emiterations=250 nriterations=50;
startvalues seed=0 sets=20 tolerance=1e-008 iterations=50;
bayes categorical=1 variances=1 latent=2 poisson=1;
montecarlo seed=0 replicates=500 tolerance=1e-008;
quadrature nodes=10;
missing excludeall;
output estimatedvalues parameters=last standarderrors probmeans
=posterior profile bivariateresiduals;
variables
groupid respondent;
caseid product;
dependent D nominal;
independent product nominal, attribute nominal;
latent Z group nominal 2, X1 nominal 2, X2 nominal 2, Y1 nominal 2
dynamic, Y2 nominal 2 dynamic;
```

17. equations
18. Z < -1 ;
19. X1 $< -1|$ product Z;
20. Y1 $< -1|$ attribute;
21. X2 $< -1|$ product Z;
22. Y2 $< -1|$ attribute;
23. D $< -(b \sim wei)$ 1| X1 X2 Y1 Y2;
24. b={
25. 1 0 // X1=0 X2=0 Y1=0 Y2=0
26. 1 0 // X1=0 X2=0 Y1=0 Y2=1
27. 1 0 // X1=0 X2=0 Y1=1 Y2=0
28. 1 0 // X1=0 X2=0 Y1=1 Y2=1
29. 1 0 // X1=0 X2=1 Y1=0 Y2=0
30. 0 1 // X1=0 X2=1 Y1=0 Y2=1
31. 1 0 // X1=0 X2=1 Y1=1 Y2=0
32. 0 1 // X1=0 X2=1 Y1=1 Y2=1
33. 1 0 // X1=1 X2=0 Y1=0 Y2=0
34. 1 0 // X1=1 X2=0 Y1=0 Y2=1
35. 0 1 // X1=1 X2=0 Y1=1 Y2=0
36. 0 1 // X1=1 X2=0 Y1=1 Y2=1
37. 1 0 // X1=1 X2=1 Y1=0 Y2=0
38. 0 1 // X1=1 X2=1 Y1=0 Y2=1
39. 0 1 // X1=1 X2=1 Y1=1 Y2=0
40. 0 1 // X1=1 X2=1 Y1=1 Y2=1
41. };
42. end model

A.2 Comments on the Latent Gold Syntax Code

Lines 3-10 describe the specific options used for the estimation (for further details see Vermunt and Magidson 2008). In line 6, the specification 'latent=2' indicates that we actually specify $\alpha_\sigma = \beta_\sigma = \alpha_\rho = \beta_\rho = 1$ and $\gamma_1 =, \ldots, = \gamma_T = 2$ for the prior distribution in (12). Lines 11-16 describe per type of variable, the variables that are involved in the analysis. In lines 12 and 13 it is indicated that respondents should be classified both at the respondent level (=group level) and at the product-respondent level (=case level). In line 16 it is indicated that the latent variable Z is used to

classify respondents in 2 classes, and that the binary latent variables $X1$ and $X2$ are used to classify respondents at the product-respondent level in $2^F (= 4)$ clusters. In addition, the keyword 'dynamic' indicates that the binary latent variables $Y1$ and $Y1$ are used to classify respondents at the level of individual observations. Lines 18-23 describe the relationships between the variables that are involved in the analysis. In particular, lines 19 and 21 indicate that the distributions of $X1$ and $X2$ are specific for all combinations of class- and product levels. Lines 20 and 22 indicate that the distributions of $Y1$ and $Y2$ are attribute-specific. Line 23 states that the observations D are a deterministic function of the latent variables $X1$, $X2$, $Y1$ and $Y2$. Lines 25-40 list the conditional probabilities $P(D = 0|\mathbf{x}, \mathbf{y})$ and $P(D = 1|\mathbf{x}, \mathbf{y})$ for all latent data patterns $(\mathbf{x}, \mathbf{y})$. Note that in lines 25-40, the parts after the double slash are comments. Concerning the order of the latent variables in the pattern, Latent Gold will first consider the latent variables at the case-level (i.e. $X1$ and $X2$) and then the latent variables at the level of individual observations (i.e. $Y1$ and $Y2$).

Appendix B: Derivation of the EM-Algorithm

For the disjunctive multilevel LC-PFM with fixed product features (i.e. $d_{ijk} = C(\mathbf{x}_i^j, \mathbf{y}_{ji}^k)$) and heterogeneity in only product parameters (i.e. $\pi_{jkt} = f(\boldsymbol{\sigma}_{jt}, \boldsymbol{\rho}_k)$), the complete-data likelihood is proportional to:

$$\prod_i \prod_t \left\{ \xi_t \prod_j \prod_f (\sigma_{jft})^{x_i^{jf}} (1-\sigma_{jft})^{1-x_i^{jf}} \prod_k \prod_f \prod_j \right.$$
$$\left. \left[(\rho_{kf})^{y_{ji}^{kf}} (1-\rho_{kf})^{1-y_{ji}^{kf}} \right]^{x_i^{jf}} \left[(0)^{y_{ji}^{kf}} (1)^{1-y_{ji}^{kf}} \right]^{1-x_i^{jf}} \right\}^{z_{it}} .$$

Using a conjugate prior distribution as in (12), except for a constant, the logarithm of the complete-data posterior equals:

$$\begin{aligned} = \quad & \sum_i \sum_t (\frac{2}{T} + z_{it}) \ln(\xi_t) \\ & + \sum_j \sum_f \sum_t [\frac{1}{JT} + \sum_i x_i^{jf} z_{it}] \ln(\sigma_{jft}) \\ & + [\frac{1}{JT} + \sum_i (1 - x_i^{jf}) z_{it}] \ln(1 - \sigma_{jft}) \\ & + \sum_k \sum_f [\frac{1}{K} + \sum_i \sum_j \sum_t x_i^{jf} y_{ji}^{kf} z_{it}] \ln(\rho_{kf}) \end{aligned}$$

$$+[\frac{1}{K} + \sum_i \sum_j \sum_t x_i^{jf}(1 - y_{ji}^{kf})z_{it}]\ln(1 - \rho_{kf}).$$

To estimate the parameters with the EM algorithm, in subsequent iterations, we have to maximize the complete-data posterior as function of the parameters (M-step), and replace the complete data statistics by their conditional expected values given the observed data, and the current value of the parameters (E-step). In iteration $(m + 1)$, one may update the parameters as follows:

$$\hat{\sigma}_{jft}^{(m+1)} = \frac{\frac{1}{JT} + E(\sum_i x_i^{jf} z_{it}|\mathbf{d}, \boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^m)}{\frac{2}{JT} + E(\sum_i z_{it}|\mathbf{d}, \boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^m)} \tag{13}$$

$$\hat{\rho}_{kf}^{(m+1)} = \frac{\frac{1}{K} + E(\sum_i \sum_j \sum_t x_i^{jf} y_{ji}^{kf} z_{it}|\mathbf{d}, \boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^m)}{\frac{2}{K} + E(\sum_i \sum_j \sum_t x_i^{jf} z_{it}|\mathbf{d}, \boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^m)} \tag{14}$$

$$\hat{\xi}_t^{(m+1)} = \frac{\frac{2}{T} + E(\sum_i z_{it}|\mathbf{d}, \boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^m)}{I + 2}. \tag{15}$$

To update the parameters, one has to compute the conditional expected values in (13), (14) and (15). The expected value in the numerator of (15) and the denominator of (13) can be derived as follows:

$$E(\sum_i z_{it}|\mathbf{d}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) = \sum_i P(Z_{it} = 1|\mathbf{d}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}).$$

The posterior probability $P(Z_{it} = 1|\mathbf{d}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})$ can be computed as follows:

$$\begin{aligned}
&= \frac{P(Z_{it} = 1|\boldsymbol{\xi})p(\mathbf{d}_i|Z_{it} = 1, \boldsymbol{\sigma}, \boldsymbol{\rho})}{p(\mathbf{d}_i|\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})} \\
&\propto \xi_t[\prod_j p(\mathbf{d}_{ij}|Z_{it} = 1, \boldsymbol{\sigma}, \boldsymbol{\rho})] \\
&\propto \xi_t \prod_j \{\sum_{\mathbf{x}_i^j} p(\mathbf{d}_{ij}|\mathbf{x}_i^j, \boldsymbol{\rho})p(\mathbf{x}_i^j|Z_{it} = 1, \boldsymbol{\sigma}_j)\} \\
&\propto \xi_t \prod_j \{\sum_{\mathbf{x}_i^j} [\prod_k p(d_{ijk}|\mathbf{x}_i^j, \boldsymbol{\rho}_k)]p(\mathbf{x}_i^j|Z_{it} = 1, \boldsymbol{\sigma}_j)\}
\end{aligned}$$

with

$$p(d_{ijk}|\mathbf{x}_i^j, \boldsymbol{\rho}_k) = [1 - \prod_f (1 - x_i^{jf}\rho_{kf})]^{d_{ijk}}[\prod_f (1 - x_i^{jf}\rho_{kf})]^{1-d_{ijk}}$$

and

$$p(\mathbf{x}_i^j|Z_{it}=1,\boldsymbol{\sigma}_{jt}) = \prod_f (\sigma_{jft})^{x_i^{jf}} (1-\sigma_{jft})^{1-x_i^{jf}}.$$

The expected value in the numerator of (13) can be computed as follows:

$$\begin{aligned}
&\sum_i E(x_i^{jf} z_{it}|\mathbf{d}_i,\boldsymbol{\sigma},\boldsymbol{\rho},\boldsymbol{\xi}) \\
&\quad = \sum_i P(Z_{it}=1, X_i^{jf}=1|\mathbf{d}_i,\boldsymbol{\sigma},\boldsymbol{\rho},\boldsymbol{\xi}) \\
&\quad = \sum_i P(Z_{it}=1|\mathbf{d}_i,\boldsymbol{\sigma},\boldsymbol{\rho},\boldsymbol{\xi}) P(X_i^{jf}=1|Z_{it}=1,\mathbf{d}_{ij},\boldsymbol{\sigma},\boldsymbol{\rho}) \\
&\quad = \sum_i P(Z_{it}=1|\mathbf{d}_i,\boldsymbol{\sigma},\boldsymbol{\rho},\boldsymbol{\xi}) \sum_{\mathbf{x}_i^j} x_i^{jf} P(\mathbf{X}_i^j=\mathbf{x}_i^j|Z_{it}=1,\mathbf{d}_{ij},\boldsymbol{\sigma},\boldsymbol{\rho})
\end{aligned}$$

with

$$\begin{aligned}
&P(\mathbf{X}_i^j=\mathbf{x}_i^j|Z_{it}=1,\mathbf{d}_{ij},\boldsymbol{\sigma},\boldsymbol{\rho}) \\
&\qquad \propto \prod_k p(d_{ijk}|\mathbf{x}_i^j, Z_{it}=1,\boldsymbol{\rho}_k) p(\mathbf{x}_i^j|Z_{it}=1,\boldsymbol{\sigma}_j).
\end{aligned}$$

The expected value in the numerator of (14) can be computed as:

$$\begin{aligned}
&E(\sum_i \sum_j \sum_t x_i^{jf} y_{ji}^{kf} z_{it}|\mathbf{d}_i,\boldsymbol{\sigma},\boldsymbol{\rho}) \\
&= \sum_i \sum_j \sum_t P(X_i^{jf}=1, Y_{ji}^{kf}=1, Z_{it}=1|\mathbf{d}_i,\boldsymbol{\sigma},\boldsymbol{\rho}) \\
&= \sum_i \sum_t P(Z_{it}=1|\mathbf{d}_i,\boldsymbol{\sigma},\boldsymbol{\rho},\boldsymbol{\xi}) \sum_j P(X_i^{jf}=1|\mathbf{d}_{ij}, Z_{it}=1,\boldsymbol{\sigma},\boldsymbol{\rho}) \\
&\qquad \times P(Y_{ji}^{kf}=1|X_i^{jf}=1, Z_{it}=1, d_{ijk},\boldsymbol{\rho})
\end{aligned}$$

with

$$\begin{aligned}
&P(Y_{ji}^{kf}=1|X_i^{jf}=1, Z_{it}=1, D_{ijk}=1) \\
&\quad = \frac{p(D_{ijk}=1|X_i^{jf}=1, Y_{ji}^{kf}=1, Z_{it}=1) P(Y_{ji}^{kf}=1|\boldsymbol{\rho}_k)}{P(D_{ijk}=1|X_i^{jf}=1, Z_{it}=1)} \\
&\quad = \frac{\rho_{kf}}{1-\prod_f (1-x_i^{jf}\rho_{kf})}
\end{aligned}$$

and with

$$P(Y_{ji}^{kf}=1|X_i^{jf}=1, Z_{it}=1, D_{ijk}=0) = 0.$$

The expected value in the denominator of (14) can be computed as:

$$
\begin{aligned}
& E(\sum_i \sum_j \sum_t x_i^{jf} z_{it} | \mathbf{d}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \\
&= \sum_i \sum_j \sum_t P(X_i^{jf} = 1, Z_{it} = 1 | \mathbf{d}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \\
&= \sum_i \sum_t P(Z_{it} = 1 | \mathbf{d}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \sum_j P(X_i^{jf} = 1 | \mathbf{d}_i, Z_{it} = 1, \boldsymbol{\sigma}, \boldsymbol{\rho}).
\end{aligned}
$$

References

AKAIKE, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle", in *Second International Symposium on Information Theory*, eds. B.N. Petrov and F. Csaki, Budapest: Academiai Kiado, pp. 271–283.

AKAIKE, H. (1974), "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control, 19*, 716–723.

CANDEL, M.J.J.M., and MARIS, E. (1997), "Perceptual Analysis of Two-Way Two-Mode Frequency Data: Probability Matrix Decomposition and Two Alternatives", *International Journal of Research in Marketing, 14*, 321-339.

DEMPSTER, A.P., LAIRD, N.M., and RUBIN, D.B. (1977), "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm (With Discussion)", *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

DESARBO, W.S., GREWAL, R., and SCOTT, C.J. (2008), "A Clusterwise Bilinear Multidimensional Scaling Methodology for Simultaneous Segmentation and Positioning Analyses", *Journal of Marketing Research, 45*, 280-292.

DESARBO, W.S., MANRAI, A.K., and MANRAI, L.A. (1993), "Non-Spatial Tree Models for the Assessment of Competitive Market Structures", in *Marketing*, eds. J. Eliashberg and G.L. Lilien, North-Holland: Elsevier Science, pp. 193–258.

EFRON, B., and TIBSHIRANI , R.J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

FORMANN, A.K. (1992), "Linear Logistic Latent Class Analysis for Polytomous Data", *Journal of the American Statistical Association, 87,* 476–486.

GONZÁLEZ, J., TUERLINCKX, F., and DE BOECK, P. (2009), "Analyzing Structural Relations in Multivariate Dyadic Binary Data", *Applied Multivariate Research, 13*, 77–92.

JORDAN, M.I., and JACOBS, R.A. (1994), "Hierarchical Mixtures of Experts and the EM Algorithm", *Neural Computation, 6*, 181–214.

KERIBIN, C. (2000), "Consistent Estimation of the Order of Mixture Models", *Sankhya Series A, 62*, 49-66.

KUPPENS, P., VAN MECHELEN, I., and MEULDERS, M. (2004), "Every Cloud Has a Silver Lining: Interpersonal and Individual Differences Determinants of Anger-Related Behaviors", *Personality and Social Psychology Bulletin, 30*, 1550–1564.

KUMBASAR, E., KIMBALL, R.A., and BATCHELDER, W.H. (1994), "Systematic Biases in Social Perception", *American Journal of Sociology, 100*, 477–505.

LI, J. (2005), "Clustering Based on a Multi-Layer Mixture Model", *Journal of Computational and Graphical Statistics, 14*, 547–568.

MARIS, E. (1999), "Estimating Multiple Classification Latent Class Models", *Psychometrika, 64*, 187-212.

MARIS, E., DE BOECK, P., and VAN MECHELEN, I. (1996), "Probability Matrix Decomposition Models", *Psychometrika, 61*, 7-29.

MEULDERS, M., DE BOECK, P., KUPPENS, P., and VAN MECHELEN, I. (2002), "Constrained Latent Class Analysis of Three-Way Three-Mode Data", *Journal of Classification, 19*, 277-302.

MEULDERS, M., De BOECK, P., and VAN MECHELEN, I. (2001), "Probability Matrix Decomposition Models and Main-Effects Generalized Linear Models for the Analysis of Replicated Binary Associations", *Computational Statistics and Data Analysis, 38*, 217-233.

MEULDERS, M., DE BOECK, P., and VAN MECHELEN, I. (2003), "A Taxonomy of Latent Structure Assumptions for Probability Matrix Decomposition Models", *Psychometrika, 68*, 61-77.

MEULDERS, M., DE BOECK, P., VAN MECHELEN, I., and GELMAN, A. (2005), "Probabilistic Feature Analysis of Facial Perception of Emotions", *Applied Statistics, 54*, 781-793.

MEULDERS, M., DE BOECK, P., VAN MECHELEN, I., GELMAN, A., and MARIS, E. (2001), "Bayesian Inference with Probability Matrix Decomposition Models", *Journal of Educational and Behavioral Statistics, 26*, 153-179.

MILOSLAVSKY, M., and VAN DER LAAN, M.J. (2003), "Fitting of Mixtures with Unspecified Number of Components Using Cross Validation Distance Estimate", *Computational Statistics and Data Analysis, 41*, 413-428.

MISCHEL, W., SHODA, Y., and MENDOZA-DENTON, R. (2002), "Situation-Behavior Profiles as a Locus of Consistency in Personality", *Current Directions in Psychological Science, 11*, 50–54

MISCHEL, W., and SHODA, Y. (1998), "Reconciling Processing Dynamics and Personality Dispositions", *Annual Review of Psychology, 49*, 229–258.

SCHWARZ, G. (1978), "Estimating the Dimensions of a Model", *Annals of Statistics, 6*, 461-464.

TITSIAS, M.K., and LIKAS, A. (2002), "Mixture of Experts Classification Using a Hierarchical Mixture Model", *Neural Computation, 14*, 2221–2244.

TORRES, A., and BIJMOLT, T.H.A. (2009), "Assessing Brand Image Through Communalities and Asymmetries in Brand-to-Attribute and Attribute-to-Brand Associations", *European Journal of Operational Research, 195*, 628-640.

VAN MECHELEN, I., and DE BOECK, P. (1990), "Projection of a Binary Criterion into a Model of Hierarchical Classes", *Psychometrika, 55*, 677–694.

VANSTEELANDT, K., and VAN MECHELEN, I. (1998), "Individual Differences in Situation-Behavior Profiles: A Triple Typology Model", *Journal of Personality and Social Psychology, 75*, 751–765.

VAN ROSMALEN, J., VAN HERK, H., and GROENEN, P.J.F. (2010), "Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model", *Journal of Marketing Research, 47*, 157–172.

VERMUNT, J.K. (2003), "Multilevel Latent Class Models", *Sociological Methodology, 33*, 213-239.

VERMUNT, J.K. (2007), "A Hierarchical Mixture Model for Clustering Three-Way Data Sets", *Computational Statistics and Data Analysis, 51*, 5368-5376.

VERMUNT, J.K., and MAGIDSON, J. (2005), *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Belmont MA: Statistical Innovations Inc.

VERMUNT J.K., and MAGIDSON, J. (2008), *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*, Belmont MA: Statistical Innovations Inc.

VICARI, D., and ALFÒ, M. (2010), "Clustering Discrete Choice Data", in *Proceedings in Computational Statistics 2010*, eds. Y. Lechevallier and G. Saporta, Berlin: Springer-Verlag, pp. 369–378.