Bayesian Methods for Conjoint Analysis-Based Predictions: Do We Still Need Latent Classes?

Daniel Baier

Abstract Recently, more and more Bayesian methods have been proposed for modeling heterogeneous preference structures of consumers (see, e.g., Allenby et al., J Mark Res 32:152–162, 1995, 35:384–389, 1998; Baier and Polasek, Stud Classif Data Anal Knowl Organ 22:413–421, 2003; Otter et al., Int J Res Mark 21(3):285–297, 2004). Comparisons have shown that these new methods compete well with the traditional ones where latent classes are used for this purpose (see Ramaswamy and Cohen (2007) Latent class models for conjoint analysis. In: Gustafsson A, Herrmann A, Huber (eds) Conjoint measurement – methods and applications, 4th edn. Springer, Berlin, pp 295–320) for an overview on these traditional methods). This applies especially when the prediction of choices among products is the main objective (e.g. Moore et al., Mark Lett 9(2):195–207, 1998; Andrews et al., J Mark Res 39:479–487, 2002a; 39:87–98, 2002b; Moore, Int J Res Mark 21:299–312, 2004; Karniouchina et al., Eur J Oper Res 19(1):340–348, 2009, with comparative results). However, the question is still open whether this superiority still holds when the latent class approach is combined with the Bayesian one. This paper responds to this question. Bayesian methods with and without latent classes are used for modeling heterogeneous preference structures of consumers and for predicting choices among competing products. The results show a clear superiority of the combined approach over the purely Bayesian one. It seems that we still need latent classes for conjoint analysis-based predictions.

D. Baier (✉)
Chair of Marketing and Innovation Management, Brandenburg University of Technology Cottbus, Postbox 101344, 03013 Cottbus, Germany
e-mail: daniel.baier@tu-cottbus.de

1 Introduction

Since many years conjoint analysis has proven to be a useful modeling approach when preference structures of consumers w.r.t. attributes and levels of competing products have to be modeled (see, e.g. Green and Rao 1971; Green et al. 2001; Baier and Brusch 2009). Preferential evaluations of sample products (attribute-level-combinations) are collected from sample consumers and for each consumer the relation between attribute-levels and preference values is modeled. Then, these individual models can be used for predicting choices of these consumers in different scenarios. Since in conjoint analysis typically the number of evaluations is low compared to the number of model parameters and many consumers show a similar preference structure, various approaches have been proposed that assume identical model parameters so that the ratio between evaluations and model parameters and – hopefully – the choice predictions using these model parameters can be improved.

Besides approaches that assume the same model parameters across all consumers especially latent class approaches have been proposed for this purpose (see Ramaswamy and Cohen 2007 for an overview on these traditional methods). Here, a division of the market into segments or (latent) classes with homogeneous preference structures is assumed and modeled by identical model parameters within a class. During the modeling step, the class-specific model parameters as well as the number and the size of the classes have to be estimated. Latent Class Metric Conjoint Analysis (shortly: LCMCA, DeSarbo et al. (1992)) is one of the most popular approaches of this kind. In the upper part of Fig. 1 a typical situation is given: The diagrams show a market with three market segments that differ w.r.t. to their preference for "high quality" and for "modern" products. Since the market seems to be clearly segmented, the sharing of evaluations within these segments could lead to an improvement of choice predictions.

Alternatively, recently, Hierarchical Bayesian procedures have been proposed for the same purpose (see, e.g. Allenby et al. 1995, 1998; Lenk et al. 1996). Here, no explicit market segmentation with identical model parameters within the segments is assumed. Instead, a common distribution of the model parameters is postulated for all consumers (first level model), which then is adjusted to individual consumers using their individual evaluations (second level model). Hierarchical Bayes Metric Conjoint Analysis (shortly: HB/MCA, Lenk et al. (1996)) is a popular approach of this kind. In the lower part of Fig. 1 a typical situation is given, where this approach is useful: The diagrams show a market obviously without segments. Consumers differ individually w.r.t. to their preference for "high quality" and for "modern" products, however, they cannot be grouped consistently into homogeneous segments. Market researchers call this situation the "water melon problem" (see, e.g. Sentis and Li 2002): Each dividing up into segments seems to be arbitrarily, so the sharing of evaluations within segments should lead to no improvement of choice predictions. Recently, many comparison studies have shown, that these Hierarchical Bayes approaches seem to compete well with the traditional latent class approaches w.r.t. criteria like model fit or predictive validity (see Table 1 for an overview on comparison studies and their results).
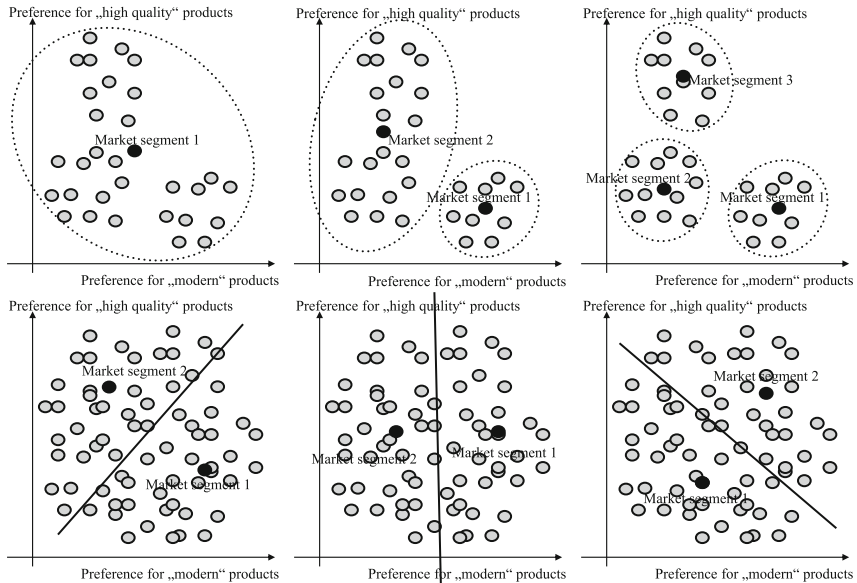
Fig. 1 A market with three market segments (*upper part*) and one without obvious segments (*lower part*); *grey points* indicate individual preferences, *black points* mean preferences when grouping the individuals; the *lines* are used to indicate the allocation of individuals to groups; in the lower – unsegmented – part their exists no obvious grouping

Table 1 Segmentation gains for conjoint analysis-based choice predictions: an overview

Reference	Segm. bases	Criteria	Result
Lenk et al. (1996)	2 real	Pred. validity (FCH, RMSE)	No segm. gains
Allenby et al. (1998)	3 real	Pred. validity (FCH, RMSE)	No segm. gains
Moore et al. (1998)	2 real	Pred. validity (FCH, RMSE)	No segm. gains
Andrews et al. (2002a)	Simulated	Model fit (Pearson, Kendall),Pred. validity (FCH, RMSE)	No sign. diff.
Andrews et al. (2002b)	Simulated	Model fit (Pearson, Kendall),Pred. validity (FCH, RMSE)	No sign. diff.
Gensler (2003)	Simulated	Model fit (Pearson, Kendall),Pred. validity (FCH, RMSE)	No sign. diff.
Moore (2004)	1 real	Pred. validity (FCH, RMSE)	No segm. gains
Karniouchina et al. (2009)	1 real	Pred. validity (FCH, RMSE)	No segm. gains

Across all studies, the assumption of market segments leads to no or only few segmentation gains (i.e. no significant differences w.r.t. model fit or predictive validity) and one could draw the conclusion that we don't need latent classes for conjoint analysis-based choice predictions. However, up to now, it is not clear

whether this is also true for a combination of Hierarchical Bayes and Latent Class approaches. For this reason, we compare in this paper a version of such combined approaches, Hierarchical Bayes Latent Class Metric Conjoint Analysis (HB/LCMCA), with HB/MCA, a purely Bayesian one. Since HB/MCA is a special case of HB/LCMCA (with only one latent class) the introduction of HB/LCMCA in chapter "The Randomized Greedy Modularity Clustering Algorithm and the Core Groups Graph Clustering Scheme" suffices. In chapter "Comparison of Two Distribution Valued Dissimilarities and Its Application for Symbolic Clustering" a Monte Carlo design is developed which is used to compare HB/MCA and HB/LCMCA. The paper closes with conclusions and outlook in chapter "Pairwise Data Clustering Accompanied by Validation and Visualisation".

2   Hierarchical Bayes Latent Class Metric Conjoint Analysis

In the following a combination of Hierarchical Bayes and Latent Class approaches for conjoint analysis-based choice prediction is introduced for answering the research question. The HB/LCMCA approach follows the HB/MCA approach in Lenk et al. (1996), but uses similar modeling assumptions as in DeSarbo et al. (1992) for the latent class part of the model and as in Baier and Polasek (2003) for the distributional assumptions. HB/LCMCA contains HB/MCA as a special case (with only one latent class). As in Lenk et al. (1996) the preferential evaluations are modeled as the addition of corresponding partworths (preferential evaluations of attribute-levels).

2.1   The Data, the Model, and the Model Parameters

Let $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^m$ describe observed preferential evaluations from n consumers ($i = 1, \ldots, n$) w.r.t. to m products ($j = 1, \ldots, m$). y_{ij} denotes the observed preference value of consumer i w.r.t. product j. As an example, these preference values could come from a response scale with values -5 ("I totally dislike this product.") to $+5$ ("I totally like this product."). $\mathbf{X} \in \mathbb{R}^{m \times p}$ denotes the characterization of the m products using p variables. As an example, cars could be characterized by attributes like price, performance, weight, and so on. For estimating the effects of the different attributes on the consumer's preference evaluations, one uses a set of products that reflects possible attribute-levels (e.g. a "low" and a "high" price) in an adequate way, using, e.g., factorial designs w.r.t. to nominal scaled attributes. In this case for $\mathbf{X}$ dummy coded variables are used instead of the original (possibly nominal) attributes.

The observed evaluations are assumed to come from the following model

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad \text{for } i = 1, \ldots, n \quad \text{with } \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \tag{1}$$

with $\mathbf{I}$ as the identity matrix, σ^2 as an error variance parameter, and individual part-worths $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ coming from T latent classes ($t = 1, \ldots, T$) with class-specific partworths $\boldsymbol{\mu}_t \in \mathbb{R}^p$ and class-specific (positive definite) variance/covariance matrices $\mathbf{H}_t \in \mathbb{R}^{p \times p}$:

$$\boldsymbol{\beta}_i \sim \begin{cases} N(\boldsymbol{\mu}_1, \mathbf{H}_1) & \text{if } C_i = 1, \\ \vdots \\ N(\boldsymbol{\mu}_T, \mathbf{H}_T) & \text{if } C_i = T, \end{cases} \qquad i = 1, \ldots, n. \tag{2}$$

$\mathbf{C} = (C_1, \ldots, C_n)$ indicates the (latent) classes to which the consumers belong with $C_i \in \{1, \ldots, T\}$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_T)$ reflects the (related) size of the classes ($\eta_t = \sum_{i=1}^n 1_{\{C_i = t\}} / T$).

2.2 The Bayesian Estimation Procedure

For estimating the model parameters ($\boldsymbol{\eta}, \mathbf{C}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T, \mathbf{H}_1, \ldots, \mathbf{H}_T, \sigma^2$), Bayesian procedures provide a mathematically tractable way that combines prior information about the model parameters with the likelihood function of the observed data. The result of this combination, the posterior distribution of the model parameters, depends on the modeling assumptions and the assumed prior distributions of the model parameters. It can be derived using iterative Gibbs sampling steps as explained in the following. We use variables with one asterisk ("$*$", e.g., $\mathbf{a}_*$) to denote describing variables of an a priori distribution (prior information) and two asterisks ("$**$", e.g., $\mathbf{a}_{**}$) to denote describing variables of a posterior distribution of the model parameters. Note that the describing variables of the a priori distributions and initial values for the model parameters have to be set before estimation whereas the describing variables of the posterior distributions have derived values allowing iteratively to draw values from the posterior distributions resulting in empirical distributions of all model parameters. We use repeatedly the following five steps:

1. Sample the class indicators $\mathbf{C}$ using the likelihood l of the normal distribution

$$p(C_i = t | \boldsymbol{\eta}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T, \mathbf{H}_1, \ldots, \mathbf{H}_T, \sigma^2, \mathbf{y}_i) \propto l(\mathbf{y}_i | \mathbf{X}\boldsymbol{\mu}_t, \mathbf{X}\mathbf{H}_t\mathbf{X}' + \sigma^2\mathbf{I})\eta_t$$

    (The consumer is allocated to the class that reflects her/his evaluations best.).
2. Sample the class sizes $\boldsymbol{\eta}$ from

$$p(\boldsymbol{\eta}|\mathbf{C}) \propto \text{Di}(e_{1**}, \ldots, e_{T**}) \text{ with } e_{1**} = e_{1*} + n_1, \ldots, e_{T*} + n_T, n_t = \sum_{i=1}^n 1_{\{C_i = t\}}$$

    (Di($e_1, \ldots, e_T$) represents the Dirichlet distribution with concentration variables $e_1, \ldots, e_T$. The variables of the a priori distribution are set to 1: $e_{t*} = 1 \ \forall \ t$.).

3. Sample the class-specific partworths $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T$ from

$$p((\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_T')'|\mathbf{C}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \sigma^2, \mathbf{y}_1, \ldots, \mathbf{y}_n) \propto N(\mathbf{a}_{**}, \mathbf{A}_{**})$$

$$\text{with } \mathbf{Z}_i = (\mathbf{X}1_{\{C_i=1\}}, \ldots, \mathbf{X}1_{\{C_i=T\}}), \mathbf{V}_i = \mathbf{X}(\mathbf{H}_{C_i}^{-1})\mathbf{X}' + \sigma^2\mathbf{I},$$

$$\mathbf{A}_{**} = (\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{V}_i^{-1}\mathbf{Z}_i + \mathbf{A}_*^{-1})^{-1}, \quad \mathbf{a}_{**} = \mathbf{A}_{**}(\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{V}_i^{-1}\mathbf{y}_i + \mathbf{A}_*^{-1}\mathbf{a}_*)$$

(Due to known problems with slow convergence, the class-specific partworths are sampled simultaneously. The class-specific partworths are stacked, $\mathbf{a}_{**}$ and $\mathbf{A}_{**}$ are the mean and the blocked variance/covariance matrix of the corresponding posterior distribution. The variables of the a priori distribution, $\mathbf{a}_*$ and $\mathbf{A}_*$, are set to be non-informative, alternatively, they could be used as in Baier and Polasek (2003) to constrain the partworths. The $\mathbf{Z}_i$ and $\mathbf{V}_i$ matrices are used to allocate the individual evaluations to the corresponding class.).

4. Sample the individual partworths $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ from

$$p(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n|\mathbf{C}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T, \mathbf{H}_1, \ldots, \mathbf{H}_T, \sigma^2, \mathbf{y}_1, \ldots, \mathbf{y}_n) \text{ using } \boldsymbol{\beta}_i \sim N(\mathbf{b}_{i**}, \mathbf{B}_{i**})$$

$$\text{with } \mathbf{B}_{i**} = (\mathbf{X}'\mathbf{X}/\sigma^2 + \mathbf{H}_{C_i}^{-1})^{-1} \text{ and } \mathbf{b}_{i**} = \boldsymbol{\mu}_{C_i} + \mathbf{B}_{i**}\mathbf{X}'\mathbf{y}_i/\sigma^2 + \mathbf{H}_{C_i}^{-1}\boldsymbol{\mu}_{C_i}.$$

(The posterior distribution of the partworths for individual i with describing variables $\mathbf{b}_{i**}$ and $\mathbf{B}_{i**}$ combines the information from the corresponding class-specific partworths with the observed preferential evaluations of individual i.)

5. Sample the variance/covariance model parameters $\mathbf{H}_1, \ldots, \mathbf{H}_T, \sigma^2$ from

$$p(\mathbf{H}_1, \ldots, \mathbf{H}_T, \sigma^2|\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n, \mathbf{C}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T, \mathbf{y}_1, \ldots, \mathbf{y}_n) \text{ using }$$

$$\mathbf{H}_t \sim IW(w_{t**}, \mathbf{W}_{t**}) \text{ with }$$

$$w_{t**} = w_{t*} + 0.5\sum_{i=1}^{n} 1_{\{C_i=t\}}, \mathbf{W}_{t**} = \mathbf{W}_{t*} + 0.5\sum_{i=1}^{n}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_t)(\boldsymbol{\beta}_i - \boldsymbol{\mu}_t)'1_{\{C_i=t\}} \text{ and}$$

$$\sigma^2 \sim IG(g_{**}, G_{**}) \text{ with } g_{**} = g_* + \frac{nm}{2}, G_{**} = G_* + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{X}\boldsymbol{\beta}_i - \mathbf{y}_i)'(\mathbf{X}\boldsymbol{\beta}_i - \mathbf{y}_i).$$

(IW stands for the Inverse Wishart distribution, IG for the Inverse Gamma distribution. Both distributions are used to model the a priori and the posterior distributions of the variance/covariance model parameters. We use similar settings for the a priori distributions as in Baier and Polasek (2003).)

As usual in Bayesian research, the posterior distribution of the model parameters are empirical distributions which collect the draws of the iterative Gibbs steps. Each empirical distribution consists typically of 1,000–2,000 draws, the "first" draws

(e.g. the first 200 draws) are typically discarded due to the need of a so-called "burn-in phase" during estimation.

When latent classes have to be modeled in Bayesian research, often the so-called "relabeling problem" occurs: From a statistical point of view the "labels" of the classes (their number $1,\dots,T$) provide no information. For one draw of all model parameters, changing the numbers of two or more classes makes no difference ("unidentifiability problem"). However, during the iterative process over 1,000 or more draws, such changes (due to algorithmic indeterminacy) lead to bad results w.r.t. the empirical distributions. Therefore, usually, in step 2 a relabeling is enforced that – after drawing the segment sizes – ensures that the class 1 has the smallest size, 2 the second smallest and so on. Alternatively, the relabeling could take place in step 3 w.r.t. class-specific partworths by ensuring that the importance of, e.g., attribute 1 is highest for class 1, second highest for class 2, and so on.

2.3 Model Fit and Predictive Validity

Once the posterior distribution of the parameters is available one can control model fit or predictive validity in various ways. So, w.r.t. model fit, the preferential evaluations w.r.t. to the estimation sample of evaluations could be compared with the corresponding predictions using Pearson's correlation coefficient. W.r.t. predictive validity one uses the possibility that the model can also be used to predict preferential evaluations w.r.t. modified sets of products (scenarios) by changing m and $\mathbf{X}$ accordingly. One collects additional preferential evaluations w.r.t. to so-called hold-out products and compares this evaluations with predictions of the model using criteria like the Root Mean Squared Error (RMSE) which stands for the deviation between the observed and predicted preferential evaluations or the first choice which stands for the percentage of predictions where the "best" holdout product w.r.t to the observed and predicted evaluations is the same.

3 Monte Carlo Comparison of HB/MCA and HB/LCMCA

In order to decide whether one still needs latent classes for conjoint analysis-based choice predictions a comprehensive Monte Carlo analysis was performed to compare the purely Bayesian approach (HB/MCA) with the combination of the Bayesian approach and latent class modeling. One should keep in mind that HB/MCA is the HB/LCMCA version with only one latent class ($T = 1$), so, w.r.t. model fit there should be a superiority of the combined over the purely Bayesian approach. However, the question is, whether this also holds w.r.t. predictive validity.

3.1   Design of the Monte Carlo Study

In total, 1,350 datasets were generated, using 50 replications w.r.t. 3 dataset generation factors with 3 possible levels each (forming $3 \times 3 \times 3 \times 50 = 1,350$ datasets). Each generated dataset describes a conjoint experiment for estimating the preferences of 300 consumers w.r.t to products characterized by 8 two-level attributes. The simulated conjoint task for each consumer was to evaluate a set of 16 products whose dummy coded descriptions w.r.t. the 8 two-level attributes were generated using a Plackett and Burman (1946) factorial design (with 16 rows and 8 columns). Also, a set of 8 additional products was used to generate additional preferential evaluations from each consumer for checking the predictive validity. The first 16 products form the estimation set, the last 8 products the holdout set of products.

A "true" preference structure of the consumers was assumed that could come – according to the first dataset generation factor ("Heterogeneity between segments") – from a market with one, two, or three segments. The market with only one segment is used as a proxy for an unsegmented market, the markets with two or three segments as proxies for segmented markets. As in other simulation studies, the means of the "true" segment-specific partworths were randomly drawn from the $[-1, 1]$ uniform distribution. All in all the following three dataset generation factors were used:

- **Heterogeneity between segments (unsegmented or not segmented market):** For a third of the datasets (level "low" for factor "heterogeneity between segments"), it was assumed that there is no segment-specific preference structure, i.e. all "true" individual partworths are drawn from one (normal) distribution (one market segment). For the other datasets (levels "medium" and "high"), it was assumed that there is a segment-specific preference structure, i.e. all "true" individual partworths are drawn from two ("medium") or three ("high") different (normal) distributions (two or three market segments). The size of these market segments was predefined as 100 % (in the case of one market segment, 300 consumers), 50 and 50 % in the case of two market segments (each segment contains 150 consumers) resp. 50, 30 and 20 % in the case of three market segments (containing 150, 90 and 60 consumers).
- **Heterogeneity within segments (segment-specific distributions of individual partworths):** For all datasets it was assumed that the individual partworths are drawn from normal distributions around the mean of their corresponding segment-specific partworths (drawn from a uniform distribution as described above). The variance/covariance matrix of these normal distributions was assumed to be diagonal with identical values σ^2 in the diagonal. For a third of the datasets these diagonal values (and consequently the heterogeneity within segments) were assumed to be "low ($\sigma = 0.1$)", for another third "medium ($\sigma = 0.25$)", and for another third "high ($\sigma = 0.5$)".
- **Disturbance (additive preference value error in data collection):** Additionally, as in other studies, a measurement error was introduced for the simulated

Table 2 Model fit across the datasets in the Monte Carlo analysis

		HB/MCA		HB/LCMCA	
Factor	Level	Corr($\mathbf{y}_i$)	Corr($\boldsymbol{\beta}_i$)	Corr($\mathbf{y}_i$)	Corr($\boldsymbol{\beta}_i$)
Heterogeneity	Low (450 datasets)	**0.744**	**0.954**	0.742	0.951
Between	Medium (450 datasets)	0.535	0.698	**0.685*****	**0.817*****
Components	High (450 datasets)	0.456	0.576	**0.655*****	**0.767*****
Heterogeneity	Low (450 datasets)	0.594	0.763	**0.724*****	**0.872*****
Within	Medium (450 datasets)	0.612	0.784	**0.731*****	**0.884*****
Components	High (450 datasets)	0.529	0.682	**0.627*****	**0.779*****
Disturbance	Low (450 datasets)	0.746	0.752	**0.930*****	**0.936*****
(Data error)	Medium (450 datasets)	0.686	0.759	**0.847*****	**0.925*****
	High (450 datasets)	0.303	**0.718***	**0.305**	0.674
Total (1,350 datasets)		0.578	0.743	**0.694*****	**0.845*****

* Significant differences at $\alpha = 0.1$; **at $\alpha = 0.01$; ***at $\alpha = 0.001$

data collection step. The calculated preference values for each product using the generated "true" individual partworths were superimposed by a normally distributed additive error (see model formulation in Sect. 2.1) with a "low ($\sigma = 0.4$)", "medium ($\sigma = 1$)" or "high ($\sigma = 2$)" standard deviation.

For each possible factor-level-combination – a total of $3 \times 3 \times 3 = 27$ combinations was possible – the dataset generation was repeated 50 times (full factorial design with 50 repetitions). As a result each dataset comprised conjoint evaluations from 300 consumers with respect to 16 products for estimation (using – as above mentioned – a Plackett and Burman (1946) factorial design) and 8 randomly generated holdout products for checking the predictive validity. It should be mentioned that – besides transforming the generated preferential evaluations into a Likert scale – the dataset generation process reflects the model formulation quite good (as usual, see the simulation studies in Table 1).

The HB/MCA and HB/LCMCA procedures were used with non-informative priors in order not to distort the estimation results by information outside the available collected data w.r.t. the 16 products. The number of segments (T) was predefined according to the HB/MCA ($T = 1$) or HB/LCMCA ($T = 2, 3$) procedure. For all estimations, 1,000 Gibbs iterations with 200 burn-ins proved to be sufficient for convergence. For HB/LCMCA relabeling w.r.t. to the class size (label order equals size order) was used.

3.2 Results w.r.t. Model Fit

For checking the model fit, mean Pearson correlation coefficients between true and estimated individual preference values for products (Corr($\mathbf{y}_i$)) as well as mean Pearson correlation coefficients between true and estimated individual partworths (Corr($\boldsymbol{\beta}_i$)) were calculated. Table 2 shows aggregated results (mean values w.r.t. to

Table 3 Predictive validity across the datasets in the Monte Carlo analysis

		HB/MCA		HB/LCMCA	
Factor	Level	First choice	RMSE	First choice	RMSE
Heterogeneity	Low (450 datasets)	**0.775**	**0.060**	0.770	0.062
Between	Medium (450 datasets)	0.567	0.121	**0.702******	**0.066*****
Components	High (450 datasets)	0.491	0.132	**0.673*****	**0.063*****
Heterogeneity	Low (450 datasets)	0.640	0.099	**0.757*****	**0.056*****
Within	Medium (450 datasets)	0.631	0.099	**0.761*****	**0.052*****
Components	High (450 datasets)	0.563	0.116	**0.628*****	**0.083*****
Disturbance	Low (450 datasets)	0.687	0.087	**0.870*****	**0.029*****
(Data error)	Medium (450 datasets)	0.677	0.088	**0.798*****	**0.041*****
	High (450 datasets)	0.469	0.139	**0.478**	**0.120**
Total (1,350 datasets)		0.611	0.105	**0.715*****	**0.063*****

[*] Significant differences at $\alpha = 0.1$; **at $\alpha = 0.01$; ***at $\alpha = 0.001$

the Pearson correlation coefficients) across all datasets with one factor-level fixed ($3 \times 3 \times 50 = 450$ datasets) and across all datasets ($3 \times 3 \times 3 \times 50 = 1,350$ datasets).

For each factor-level combination of the Monte Carlo analysis these values were calculated and compared between HB/MCA and HB/LCMCA. The results are convincing: If a segment-specific structure is in the data, the segment-free HB/MCA is outperformed by the segment-specific HB/LCMCA procedure. Overall the superiority can clearly be seen.

3.3   Results w.r.t. Predictive Validity

In a similar way, the predictive validity was checked. For the eight holdout products and each consumer, preference values were calculated from the estimated individual partworths and compared to the preference values that were derived from the "true" partworths. As criteria for the comparison the so-called first choice hit rate (first choice) and mean root mean squared error (RMSE) were calculated. First choice hit indicates for a consumer whether her/his preference values from the estimated and from the "true" partworths are maximum for the same holdout product, the first choice hit rate is the share of consumers where a first choice hit occurs. RMSE compares also the preference values from the estimated and from the "true" partworths but more according to their absolute values.

Table 3 shows (again) aggregated results (mean values w.r.t. to the first choice hit rate and RMSE) across all datasets with one factor-level fixed ($3 \times 3 \times 50 = 450$ datasets) and across all datasets ($3 \times 3 \times 3 \times 50 = 1,350$ datasets). Again, the results are convincing: If a segment-specific structure is in the data, the segment-free HB/MCA is outperformed by the segment-specific HB/LCMCA procedure. Overall the superiority of the combined approach can clearly be seen.

4   Conclusions and Outlook

The comparison in this paper clearly shows that we still need latent classes for conjoint analysis-based predictions even if we use Bayesian procedures for parameter estimation. HB/LCMCA was clearly superior to HB/MCA w.r.t. model fit and predictive validity, especially in cases when markets are segmented. However, these results are only based on a rather small number of datasets (1,350 datasets) generated synthetically and therefore no real data. More research in this field needs to be done, especially with a larger set of conjoint data from real markets.

References

Allenby GM, Arora N, Ginter JL (1995) Incorporating prior knowledge into the analysis of conjoint studies. J Mark Res 32:152–162

Allenby GM, Arora N, Ginter JL (1998) On the heterogeneity of demand. J Mark Res 35:384–389

Andrews RL, Ainslie A, Currim IS (2002a) An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. J Mark Res 39:479–487

Andrews RL, Ainslie A, Currim IS (2002b) Hierarchical Bayes versus finite mixture conjoint analysis models: a comparison of fit, prediction, and partworth recovery. J Mark Res 39:87–98

Baier D, Brusch M (2009) Conjointanalyse: Methoden – Anwendungen – Praxisbeispiele. Springer, Berlin

Baier D, Polasek W (2003) Market simulation using Bayesian procedures in conjoint analysis. Stud Classif Data Anal Knowl Organ 22:413–421

DeSarbo WS, Wedel M, Vriens M, Ramaswamy V (1992) Latent class metric conjoint analysis. Mark Lett 3(3):273–288

Gensler S (2003) Heterogenität in der Präferenzanalyse. Ein Vergleich von hierarchischen Bayes-Modellen und Finite-Mixture-Modellen. Gabler, Wiesbaden

Green PE, Rao VR (1971) Conjoint measurement for quantifying judgmental data. J Mark Res 8(3):355–363

Green PE, Krieger AM, Wind Y (2001) Thirty years of conjoint analysis: reflections and prospects. Interface 31(3):S56–S73

Karniouchina EV, Moore WL, Rhee BVD, Vermad R (2009) Issues in the use of ratings-based versus choice-based conjoint analysis in operations management research. Eur J Oper Res 19(1):340–348

Lenk PJ, DeSarbo WS, Green PE, Young MR (1996) Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. Mark Sci 15(2):173–191

Moore WM (2004) A cross-validity comparison of rating-based and choice-based conjoint analysis models. Int J Res Mark 21:299–312

Moore WL, Gray-Lee J, Louviere JJ (1998) A cross-validity comparison of conjoint analysis and choice models at different levels of aggregation. Mark Lett 9(2):195–207

Otter T, Tüchler R, Frühwirth-Schnatter S (2004) Capturing consumer heterogeneity in metric conjoint analysis using Bayesian mixture models. Int J Res Mark 21(3):285–297

Plackett RL, Burman JP (1946) The design of optimum multifactorial experiments. Biometrika 33:305–325

Ramaswamy V, Cohen SH (2007) Latent class models for conjoint analysis. In: Gustafsson A, Herrmann A, Huber (eds) Conjoint measurement – methods and applications, 4th edn. Springer, Berlin, pp 295–320

Sentis K, Li L (2002) One size fits all or custom tailored: which HB fits better? In: Proceedings of the sawtooth software conference, Sequim, pp 167–175, Sept 2001