Latent profile analysis of human values: What is the optimal number of clusters?

Mikkel N. Schmidt, Daniel Seddig, Eldad Davidov, Morten Mørup,
Jan Michael Bauer, and Fumiko Kano Glückstad

Abstract

Latent Profile Analysis (LPA) is a method to extract homogeneous clusters characterized by a common response profile. Previous works employing LPA to human value segmentation tend to select a small number of moderately homogeneous clusters based on model selection criteria such as AIC, BIC and Entropy. The question is whether a small number of clusters is all that can be gleaned from the data. While some studies have carefully compared different statistical model selection criteria, there is currently no established criteria to assess if an increased number of clusters generates meaningful theoretical insights. This article examines the content and meaningfulness of the clusters extracted using two algorithms: Variational Bayesian LPA and Maximum Likelihood LPA. For both methods, our results point towards eight as the optimal number of clusters for characterizing distinctive Schwartz value typologies that generate meaningful insights and predict several external variables.

1 Introduction

Latent Class Analysis (LCA) (McCutcheon, 1987) is a method to classify respondents into latent types based on their response patterns on a set of variables. In contrast to conventional heuristic techniques such as k-means (Bock, 2007) and hierarchical clustering (Ward, 1963), LCA is based on a finite latent variable mixture model (Berlin, Williams, & Parra, 2014; Eid, Langeheine, & Diener, 2003; Lanza, Flaherty, & Collins, 2003; G. J. McLachlan, Lee, & Rathnayake, 2019; Petersen, Qualter, & Humphrey, 2019). In the case where the observed variables are continuous, LCA is referred to as Latent Profile Analysis (LPA). A general rule in latent variable mixture modelling is that the larger the number of clusters, the more homogeneous the respective clusters will be. However, many of the works employing LCA (Davidov, Yang-Hansen, Gustafsson, Schmidt, & Bamberg, 2006; Eid et al., 2003; Moors & Vermunt, 2007; Reinecke & Seddig, 2011; Rudnev, Magun, & Schmidt, 2014) determined that the appropriate number of clusters is in the range of three to five on average, based on model selection criterion such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and Entropy. Although these criteria are generally considered as reliable, they do not always agree, and researchers tend to settle on a small number of moderately homogeneous clusters. A question that remains to be answered is whether this small number of clusters is the only solution that can be gleaned from the data, or whether a larger number of theoretically and empirically meaningful clusters is concealed within the data that may advance our understanding of the pattern of clusters in populations.

The purpose of LPA is to reduce the data to a reasonable number of classes, meaningful both in terms of statistical saliency and theoretical interpretability, and in our work we define the term "optimal number of clusters" as a number that is both statistically justifiable and substantively meaningful. Nylund, Asparouhov, and Muthén (2007) conducted a thorough study to compare existing model selection criteria for LCA and demonstrated that the Bootstrap Likelihood Ratio Test (BLRT) (G. McLachlan & Peel, 2000) was the best performing criterion to determine an optimal number of clusters of the mixture models including the Maximum Likelihood (ML) LCA. However, Nylund et al. (2007) also pointed out that BLRT is not commonly implemented in mixture modeling software, generally requires substantial computation time and sometimes leads incorrect p-value estimation. Nylund-Gibson and Choi (2018) further demonstrated that the BLRT and the Vuong-Lo-Mendell-Rubin (VLMR) likelihood ratio test (Lo, Mendell, & Rubin, 2001) perform well in determining an optimal number of clusters. In addition, Spurk, Hirschi, Wang, Valero, and Kauffeld (2020) reviewed the best practice for selecting an optimal number of clusters extracted by LPA. Spurk et al. (2020) also highlighted the importance of assessing whether increasing the number of clusters lead to new typological profiles that generate meaningful insights (Berlin et al., 2014; Vermunt & Magidson, 2002). Finally, Albers, Mørup, Schmidt, and Glückstad (2020, p.4) recently demonstrated that "the Bayesian LCA can be used in a predictive approach to model order selection, in which an appropriate number of clusters is identified by optimizing the predictive performance on held-out data." These previous works indicate that the determination of an optimal number of clusters is still a challenging issue, and that approaches suitable for determining a number of cluster of the mixture models are underexplored.

LCA and LPA are often used by psychologists and sociologists to classify individuals into specific personality types (Glückstad, Schmidt, & Mørup, 2017; Magun, Rudnev, & Schmidt, 2016; Moors & Vermunt, 2007; Petersen et al., 2019; Spurk et al., 2020). Some of these works (Magun et al., 2016; Rudnev et al., 2014) used the Portrait Values Questionnaire (PVQ-21) of the Schwartz theory of ten basic human values (Schwartz, 2006, 2012) to conduct a typological analysis. For example, Rudnev et al. (2014) implemented an extended version of LPA that adds a method factor loaded on the 21 PVQ items (Schwartz et al., 2012) and investigated the invariance of the Schwartz value typologies across the 4th-, 5th-, and 6th rounds of the European Social Survey (ESS). Similarly, Magun et al. (2016) analyzed 29 European countries from the 4th round of the ESS using a factor mixture model (Muthén, 2008). Both analyses used AIC, BIC, Entropy and VLMR as model selection criteria and identified five clusters characterized with typologies representing only part of the abstract value orientations of the Schwartz ten value factors (Schwartz, 1992; Schwartz et al., 2012).

The majority of these previous works use maximum likelihood LPA (ML-LPA) that estimates "conditional means and variances of the continuous indicators" (Nylund-Gibson & Choi, 2018, p.454). In this work we also consider a so-called, variational Bayesian approach to statistical inference in the LPA model (VB-LPA) that optimizes a criterion called the Evidence Lower Bound (ELBO), a lower bound on the marginal likelihood. This is of particular interest because the ELBO itself is a statistically consistent criterion for model order selection. Second, we implement several cluster evaluation metrics that can be used to assess quality of clusters extracted from a wide range of clustering algorithms, not only the aforementioned LCA/LPA but also other heuristic models such as k-means. For comparing qualities of the clusters extracted by VB-LPA and ML-LPA, we demonstrate how to identify an optimal number of clusters extracted by the two
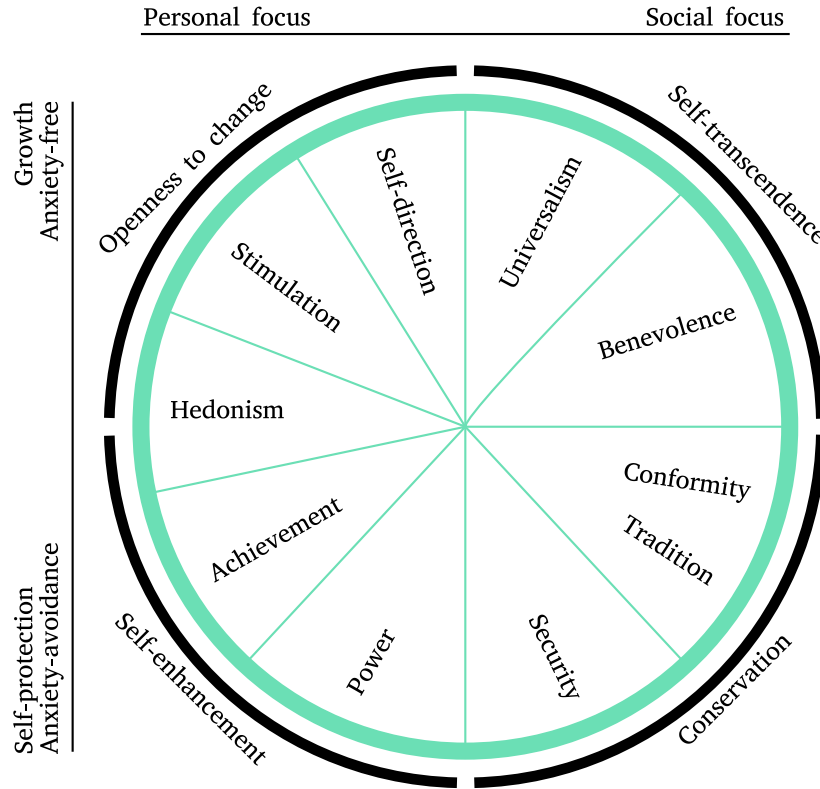
Figure 1: Ten basic human values of Schwartz organised in a circular system

approaches. To demonstrate the usability of the VB-LPA and the cluster evaluation metrics, we statistically evaluate typologies of the European population based on the 21 PVQ items (Schwartz, 2012) included in the 8th round of the European Social Survey (ESS8) dataset (Jowell, Roberts, Fitzgerald, & Eva, 2007). Accordingly, the next section reviews the Schwartz theory of ten basic human values, followed by the brief description of data, the methods applied to the current work, the analyses and the results. We finalize with a discussion of our findings.

1.1 Schwartz theory of ten basic human values

Schwartz (1992, 2006) proposed a theory summarising "ten motivationally distinct value orientations" organised as a circular structure (see Fig. 1). Each value has a unique motivation underlying it. Closer values in the value circle are expected to share similar motivations, and opposing values are expected to have contradictory motivations. The values can be summarized into four higher-order dimensions: *self-transcendence* consisting of "universalism" and "benevolence" that opposes *self-enhancement* consisting of "power" and "achievement"; *conservation* consisting of "tradition", "conformity" and "security" that opposes *openness to change* consisting of "stimulation" and "self-direction"; and "hedonism" lies between *openness to change* and *self-enhancement*. As shown in Fig. 1, these four higher-order values are further classified into the dimensional dichotomies of (i) *social-focus* vs. *personal-focus*; and (ii) *growth* vs. *self-protection*.

The Schwartz value theory has been widely applied to studies in various disciplines, not only in psychology and sociology (Beierlein, Kuntz, & Davidov, 2016; Prince-Gibson & Schwartz, 1998; Sagiv & Schwartz, 1995; Schwartz et al., 2012; Schwartz & Rubel-Lifschitz, 2009), but also the areas of marketing, tourism, and to assess ethical behaviours

(Collins, Steg, & Koning, 2007; de groot & Steg, 2008; Fritzsche & Oz, 2007; Hofer, Chasiotis, & Campos, 2006; Krystallis, Vassallo, & Chryssohoidis, 2012; Lee, Soutar, Daly, & Louviere, 2011; Leong, 2008; Lönnqvist, Jasinskaja-Lahti, & Verkasalo, 2011; Schultz et al., 2005).

2   Methods

2.1   Data

In this study we analyzed the secondary data compiled from the 8th round of the ESS (Jowell et al., 2007). From the original sample consisting of $44\,387$ respondents, $3\,094$ respondents with incomplete responses to the PVQ items were excluded. The PVQ21 consists of 21 questions responded to on a 6-point-Likert scale. To adjust for individual response styles, previous studies (Rudnev et al., 2014) employed a procedure such as a centering (Smith & Schwartz, 1997) and a method factor (Magun et al., 2016; Vermunt, 2010). For the demonstration of the VB-LPA and the cluster evaluation metrics, the current study used the Confirmatory Factor Analysis (CFA) to compute factor scores representing the Schwartz ten basic values for the respective respondents ($N=41\,293$) from the raw Likert scores of the 21 PVQ questions, following the instructions given in Schwartz (2014). The factor scores for the ten values provided the raw input for the LPA (ML-LPA and VB-LPA). Syntax, fit summary and parameter estimates of the CFA are found in the Supplemental materials.

2.2   Variational Bayesian Latent Profile Analysis (VB-LPA)

Assuming that the dataset is organized in a way that the observations (i.e. respondents of the ESS8) are the rows and the indicators (i.e. Schwartz' value indicators) are the columns of a matrix containing the CFA scores for each respondent, VB-LPA aims to group the rows to form homogeneous blocks. VB-LPA is specified according to the following generative model:

$$\pi \sim \text{Dirichlet}(\alpha) \qquad \text{Probability of belonging to each cluster.} \qquad (1)$$
$$z_n \sim \text{Categorical}(\pi) \qquad \text{Cluster membership for each respondent.} \qquad (2)$$
$$\mu_{km} \sim \text{Normal}(0, \lambda_0) \qquad \text{Mean value of each cluster profile.} \qquad (3)$$
$$\lambda_{km} \sim \text{Gamma}(a, b) \qquad \text{Precision (inverse variance) of each cluster profile.} \qquad (4)$$
$$A_{nm} \sim \text{Normal}(\mu_{z_n m}, \lambda_{z_n m}^{-1}) \qquad \text{Response for each respondent and each indicator.} \qquad (5)$$

where n, m, and k refer to each respondent, question, and latent profile respectively.

The model posits that there exists K profiles, each characterized by a mean value and a precision for each indicator, and that each respondent belongs to one of these clusters with some prior probability. Bayesian inference in the model involves estimating the posterior distribution of all the model parameters, conditioned on the observed data matrix A. Since exact Bayesian inference is intractable in this model, the Bayesian posterior distributions must be approximated. Let all model parameters be denoted by

$$\theta = \left\{ \{\pi_k | k \in 1...K\}, \{z_n | n \in 1...N\}, \{\mu_{km}, \lambda_{km} | k \in 1...K, m \in 1...M\} \right\}. \qquad (6)$$

In the variational Bayesian (VB) approach to approximate Bayesian inference, we assume that the posterior can be well represented by a factorized probability distribution

$$q(\theta) = q(\pi) \left(\prod_{n=1}^{N} q(z_n) \right) \left(\prod_{k=1}^{K} \prod_{m=1}^{M} q(\mu_{km}) q(\lambda_{km}) \right) \tag{7}$$

To fit the model, we optimize the ELBO given by

$$\text{ELBO} = \mathbb{E}_{q(\theta)} \left[\log \frac{p(A|\theta)p(\theta)}{q(\theta)} \right], \tag{8}$$

where $p(A|\theta)$ and $p(\theta)$ are the likelihood and prior as defined by the generative model. This corresponds to minimizing the Kullback-Leibler divergence between the true posterior and the variational approximation. The ELBO is optimized iteratively using closed form fixed-point updates for each term in the factorization, while keeping all other terms fixed.

Benefits of VB include the ability to explicitly monitor convergence as well as ease of interpretation accounting for uncertainty according to the inferred factorized distribution, although the assumed factorized distribution is only an (optimistic) approximation of the uncertainty and prone to local minima issues in the optimization.

2.3 Extraction of clusters

2.3.1 ML-LPA

Using the Mplus software implementation (Muthen & Shedden, 1999; Muthén & Muthén, 2017), we carried out the ML-LPA for solutions in the range from $K=3$ to $K=20$ with the default settings of equal variances across classes and zero covariance between the indicators for each class. To mitigate the potential local maxima issues, Mplus recommends that "the best log-likelihood value should be replicated in at least two final-stage solutions" (Spurk et al., 2020); to ensure this, we used 500 random restarts for $3 \leq K \leq 7$, 1 500 restarts for $8 \leq K \leq 17$, and 2 000 restarts for $18 \leq K \leq 20$. We computed fit performance scores including AIC, BIC, and Entropy, and we further computed VLMR- (Lo et al., 2001) and BLRT scores (G. McLachlan & Peel, 2000) that calculate p-values indicating whether the fit of K classes significantly increases from $K-1$ classes. The syntax for the Mplus analysis is provided in the Supplemental materials.

2.3.2 VB-LPA

Using our own software implementation[1] we carried out the VB-LPA for models in the range from $K=3$ to $K=20$ with hyperparameters fixed at $\alpha = a = b = \lambda_0 = 1$. The iterative fixed point update of the parameters were run until the relative improvement of the ELBO of a full cycle of updates of all parameters was less than 10^{-7} or until completion of 1 000 full update cycles. To mitigate effects of local sub-optimal solutions, we repeated the procedure 100 times using different random initializations and selected the solution with the highest ELBO as optimal. We verified that in all cases, the best solution was replicated to confirm that 100 repeats was adequate.

---

[1]Software implementation of the VB-LPA is available in Python in the supplement.

2.4 Cluster evaluation metrics

To evaluate the extracted clusters, we consider several generic metrics.

2.4.1 Maximum profile correlation

When the number of clusters is low, we expect the extracted profiles to be clearly distinct, but as the number of clusters increases, we expect that the extracted profiles will gradually become more similar. To assess this, we examined the maximum Pearson correlation coefficient between cluster profiles: The correlation coefficient will be equal to 1 when two cluster profiles are identical up to scaling and translation.

2.4.2 Within- and between-cluster variance

In general, a clustering solution is good when the observations in each cluster are very similar while the clusters are very dissimilar. To assess this we examined the average within-cluster variance

$$S = \frac{1}{K} \sum_{k=1}^{K} S_k, \qquad S_k = \sqrt{\frac{1}{N_k} \sum_{n|z_n=k} |a_n - \mu_k|^2}, \tag{9}$$

and the variance between the cluster profiles

$$M = \frac{1}{K(K-1)} \sum_{k=1}^{K} \sum_{\ell=k+1}^{K} M_{k\ell}, \qquad M_{k\ell} = \sqrt{|\mu_k - \mu_\ell|^2}. \tag{10}$$

2.4.3 Davies-Bouldin score

The Davies-Bouldin (DB) score is a measure of the trade-off between the within and between cluster variance, defined as the average similarity of each cluster to its closest neighbor

$$\text{DB} = \frac{1}{K} \sum_{k=1}^{K} \max_{\ell \neq k} R_{k\ell}, \tag{11}$$

where the similarity measure is defined as the ratio of within and between cluster spread

$$R_{k\ell} = \frac{S_k S_\ell}{M_{k\ell}}. \tag{12}$$

2.4.4 Entropy

Entropy, as defined in Mplus, is a statistic used to measure the uncertainty of the assignment of observations to cluster. More precisely *Entropy* is an inverted relative measure of entropy defined as

$$\text{Entropy}(z) = 1 - \frac{1}{N} \sum_{n=1}^{N} \frac{H(z_n)}{\log K}, \tag{13}$$

where the normalization $\log K$ is the maximum entropy obtained for a uniform distribution, and $H(z)$ is the information entropy of the categorical distribution,

$$H(z) = -\sum_{k=1}^{K} p_k \log p_k. \tag{14}$$

Here, p_k denotes the probability that the observation z is assigned to cluster k. The *Entropy* is in the range 0 to 1 and attains its maximum when there is no uncertainty in the cluster assignment.

2.5   External validation

External variables, that are expected to correlate with the value-based cluster assignment, can be used to validate the clustering and aid in determining the most appropriate number of clusters. To measure the degree of statistical relation between the cluster assignment and external categorical variables we used adjusted normalized mutual information (AMI) defined as

$$\text{AMI}(X,Y) = \frac{\text{MI}(X,Y) - \mathbb{E}[\text{MI}(X,Y)]}{\max(\text{H}(X), \text{H}(Y)) - \mathbb{E}[\text{MI}(X,Y)]}. \tag{15}$$

Here MI and H are the mutual information and information entropy defined as

$$\text{MI}(X,Y) = \sum_{x \in X} \sum_{y \in Y} p_{xy} \log \frac{p_{xy}}{p_x p_y}, \qquad \text{H}(X) = -\sum_{x \in X} p_x \log p_x, \tag{16}$$

where p_x is the fraction of observations that fall in category x in the label X, and p_{xy} is the fraction of observations that fall in categories x and y in labels X and Y respectively.

Mutual information measures the degree of statistical association between random variables and quantifies the amount of information (measured in nats) that the variables have in common. The AMI adjusts the mutual information for association due to chance, and scales the measure so that AMI=1 corresponds to perfect agreement and AMI=0 is the expected agreement due to chance when there is no association. Since mutual information takes the uncertainty related to the estimated cluster assignments into account, we find it well suited to measure the statistical relation between cluster assignments and external variables.

3   Results

3.1   Model fit based on the existing model selection criteria

3.1.1   ML-LPA fit summary

Table 1 shows the fit summary of the best log-likelihood solutions replicated at least twice. Both AIC and BIC scores continuously decrease, indicating that more than 20 clusters is appropriate. The entropy scores increase to K=16 where it levels off. According to the VLMR test, the first non-significant increase in model order is at K=9, indicating that a model with $K = 8$ should be examined further. The BLRT test required substantial computational time, which all resulted in P= 0.0000. In addition, for the range between K=18 and K=20, some (out of five) bootstrap draws had both a smaller LRT value than the observed LRT value and not a replicated best log-likelihood value.

Components	AIC	BIC	Entropy	VLMR	BLRT
3	671730	672092	0.865	—	
4	616549	617006	0.885	0.000	0.000
5	591249	591801	0.885	0.000	0.000
6	563286	563933	0.895	0.000	0.000
7	539388	540130	0.897	0.000	0.000
8	522803	523640	0.898	0.000	0.000
9	508036	508968	0.900	0.628	0.000
10	481011	482133	0.904	0.094	0.000
11	478958	480080	0.904	0.000	0.000
12	467310	468527	0.906	0.032	0.000
13	457601	458913	0.907	0.151	0.000
14	448117	449523	0.908	0.044	0.000
15	438810	440311	0.911	0.047	0.000
16	429312	430908	0.913	0.047	0.000
17	421874	423565	0.914	0.567	0.000
18	414963	416749	0.913	0.287	0.000
19	407574	409455	0.914	0.194	0.000
20	401787	403762	0.913	0.141	0.000

Table 1: Fit summary of the ML-LPA: AIC, BIC, Entropy, VLMR and BLRT.

3.1.2 VB-LPA fit performance

The model fit in terms of the highest attained ELBO is shown in Fig. 2. The ELBO increases continuously with a higher number of clusters, indicating that there is statistical support in the data for more than $K = 20$ clusters.

3.2 Cluster evaluation based on the proposed metrics

3.2.1 Profile correlation

The maximum Pearson correlation between cluster profiles is shown in Fig. 3(a). For ML-LPA at K=7 and for VB-LPA at K=6, the maximum correlation steeply increases from less than 0.7 to more than 0.9, indicating that above this threshold the methods discover two latent profiles which are close to identical up to a scaling and translation, which is also evident in the plot of the profiles in Fig. 4.

3.2.2 Within- and between-cluster variance

The variance within and between clusters is shown in Fig. 3 (b). As the number of clusters increases, the clusters become more dispersed (higher between-cluster variance) and contain fewer, more similar observations (lower within-cluster variance) as could be expected. The between-cluster variance is significantly lower for the VB-LPA, presumably because of the regularizing effect of the Bayesian estimation. The curves do not exhibit any conspicuous notches or plateaus which warrant further examination.
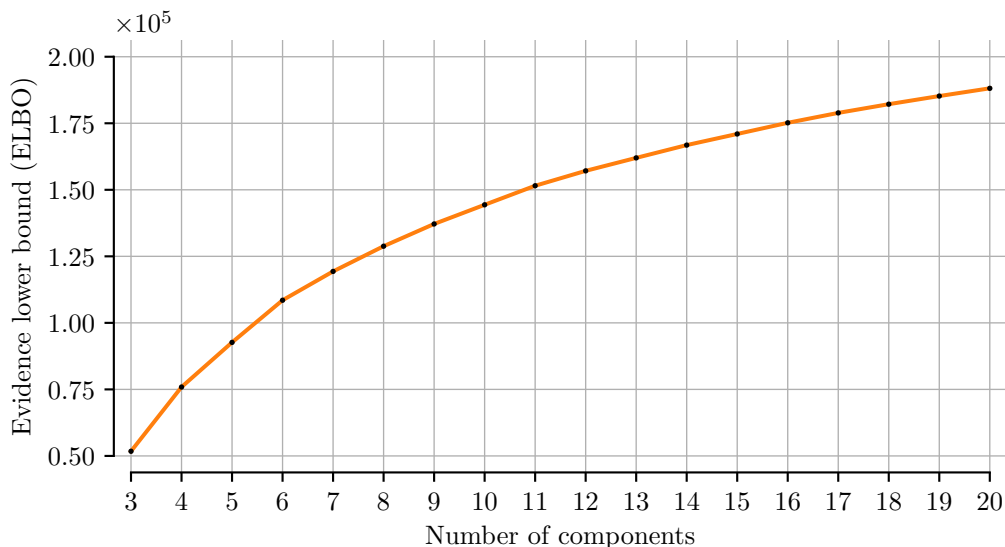
Figure 2: ELBO scores for the VB-LPA.

3.2.3 Davies-Bouldin score

The minima of the DB score shown in Fig. 3 (c) indicate model orders that achieve the best trade-off between within- and between-cluster variance. For both models, the best DB score is attained at K=8. For the ML-LPA, model orders K=4,9,20 and 5 are close contenders, but for the VB-LPA, K=8 is optimal with a large margin to the second-best solution at K=7.

3.2.4 Entropy

The certainty with which the observations are assigned to their respective clusters, as measured by the Entropy, is shown in Fig. 3 (d). The Entropy curves for the two methods are similar, with Entropy for VB-LPA consistently around a percentage point higher than ML-LPA. Models with low Entropy should be rejected because the assignment of observations to clusters should be unequivocal in order to interpret the latent factors. ML-LPA and VB-LPA reach an Entropy score above 0.9 around K=9 and K=6, respectively. Both methods yield a large increase in Entropy for K=6.

3.3 Value profiles

The results of cluster evaluation indicate that K=8 seems to be one of the optimal numbers of clusters for the two LPA techniques, although the Entropy score of K=8 for ML-LPA did not reach the level of 0.9. Fig. 4 illustrates the fitted model profiles (i.e. value typologies) for $K = 3 \ldots 8$ [2]. To aid the visualization, the profiles are color matched by greedy minimization of squared error, which enables us to inspect what type of clusters emerged when the number of clusters increased from K=3 to K=8. Generally, the typologies of the clusters shown are substantially similar between ML-LPA and VB-LPA (see "Overlap between ML-LPA8 and VB-LPA8" in Supplemental Materials). At the starting point, K=3 consists of three clusters characterized as (i) positive to all values

---

[2]Specific Schwartz typologies are provided in Appendix 2 of supplemental materials
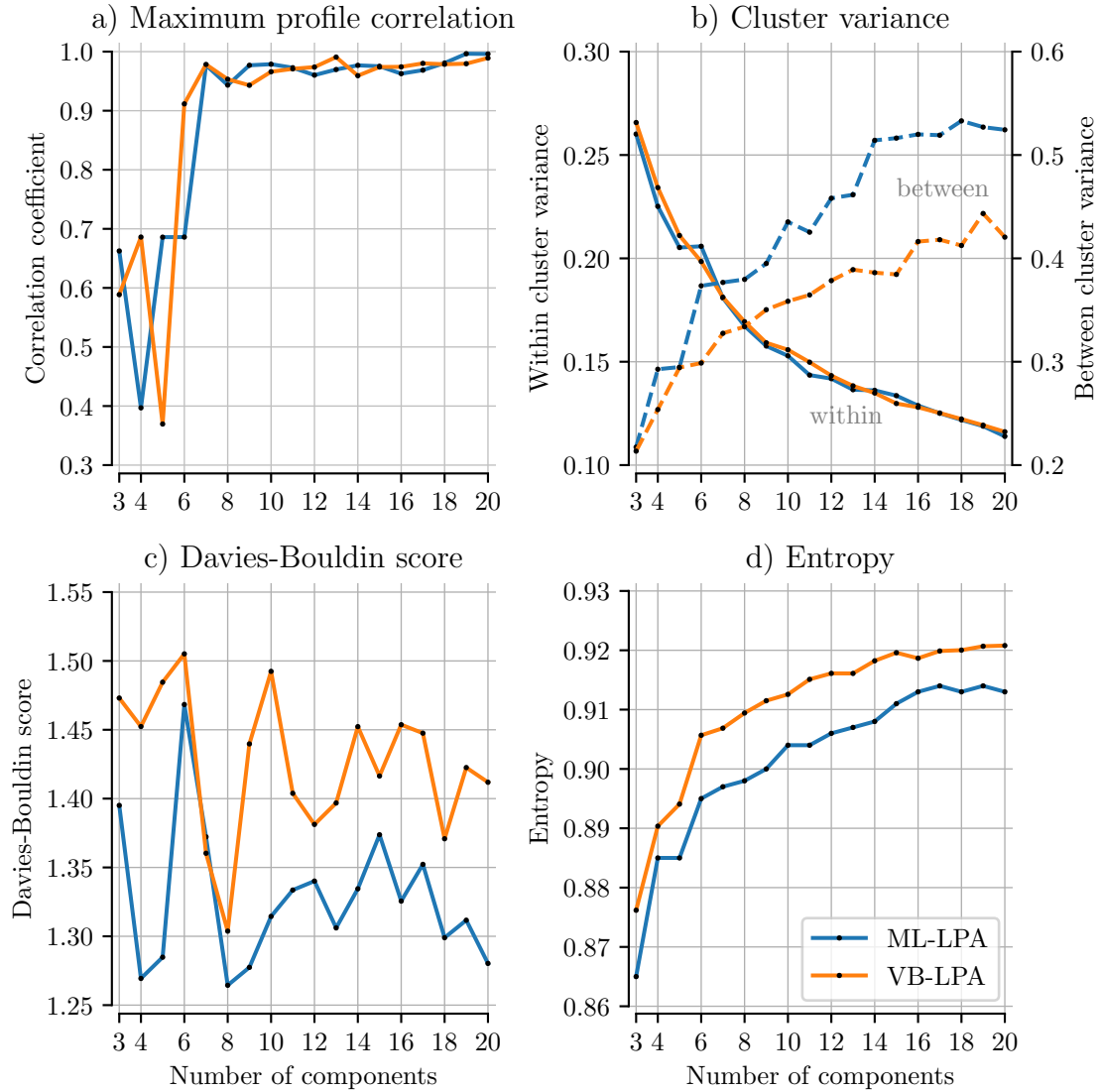
Figure 3: Cluster evaluation metrics. (a) Maximum correlation coefficient between any two profiles for different number of clusters. (b) Variance within and between clusters. (c) Davies-Bouldin score: Average similarity of each cluster with its nearest neighbor cluster. (d) Entropy (inverted relative entropy) of cluster assignments.

but especially positive to *achievement*; (ii) positive only to *social-focus*; and (iii) negative to all values but especially negative to *openness to change*. When the number of clusters increased to $K=5$, the pattern of the second cluster was slightly modified to (iii) negative to all values but especially negative to *power*, and two additional clusters, (iv) positive only to *personal-focus* and (v) positive only to *growth* were introduced. In the case of ML-LPA, the members of the two new clusters emerged in the $K=5$ solution were originated from the clusters (i) positive to all values and (iii) negative all values (see Supplemental Materials). When the number of cluster further increased to $K=7$, the second cluster (ii) only positive to *social-focus* split into two: (ii) moderately positive only to *social-focus* and (vi) strongly positive only to *social-focus*. In addition, a new cluster (vii) moderately positive only to *personal-focus* was introduced. Finally, when $K=8$ was reached, the fifth cluster, (v) positive only to *growth* split into (v) strongly positive only to *growth* and (viii) moderately positive only to *growth*. Importantly, although $K=3$ was dominated by two clusters characterized as (i) positive to all values and (iii) negative to all values, the increased number of clusters reduced the sizes of these two clusters, and subsequently revealed additional value typologies meaningfully related to the Schwartz circular value structure.

Finally, Fig. 5 depicts histograms of the cluster sizes for $K = 3 \ldots 20$. Clusters obtained using VB-LPA are considerably more uniform in size; for example, for $K \geq 14$, the smallest cluster obtained using ML-LPA covers 1% or less of the observations, whereas the smallest cluster in VB-LPA never covers less than 2% of the observations.

3.4 External validation

Plots of AMI between cluster assignments and nine external variables are shown in Fig. 6. In all cases, AMI increases from the $K=3$ solution and most either peak or level out around $K=5$–10. Two variables (b,h) have a global maximum at $K=8$–9 for both methods. In the ML-LPA for several variables (a,b,c,e,f,i), AMI exhibits a local peak at $K=5$, but this effect is absent in the VB-LPA. In the VB-LPA for several variables (b,c,e,f,g,h,i), AMI levels off at $K=8$–9.

4   Discussion and Conclusion

This article presented the VB-LPA algorithm and several statistical criteria inspired by Spurk et al. (2020) which can be used to evaluate content of the clusters and select an optimal number of clusters for various clustering techniques. Our work first reviewed the performance fit of the models obtained from the two LPA techniques. One of the criteria suggested by Nylund et al. (2007), i.e. the VLMR likelihood ratio test (Lo et al., 2001), selected $K=8$ as an optimal number of clusters. On the other hand, widely used criteria such as the AIC the BIC and the Entropy for the ML-LPA as well as the ELBO criterion for the VB-LPA were not able to identify a specific optimal number within the range of $K = 3 \ldots 20$. The criteria evaluating the maximum profile correlation, cluster variance and entropy indicated that, when the number of clusters is between 3 and 5, typological structures were not clearly identified. This may be due to the characteristics of the particular dataset that is rather homogeneous. However, when the number of clusters reached the level of $K = 6 \ldots 8$, contents of the clusters started to stabilize. The AMI analysis further suggested that when the number of clusters reached the level of
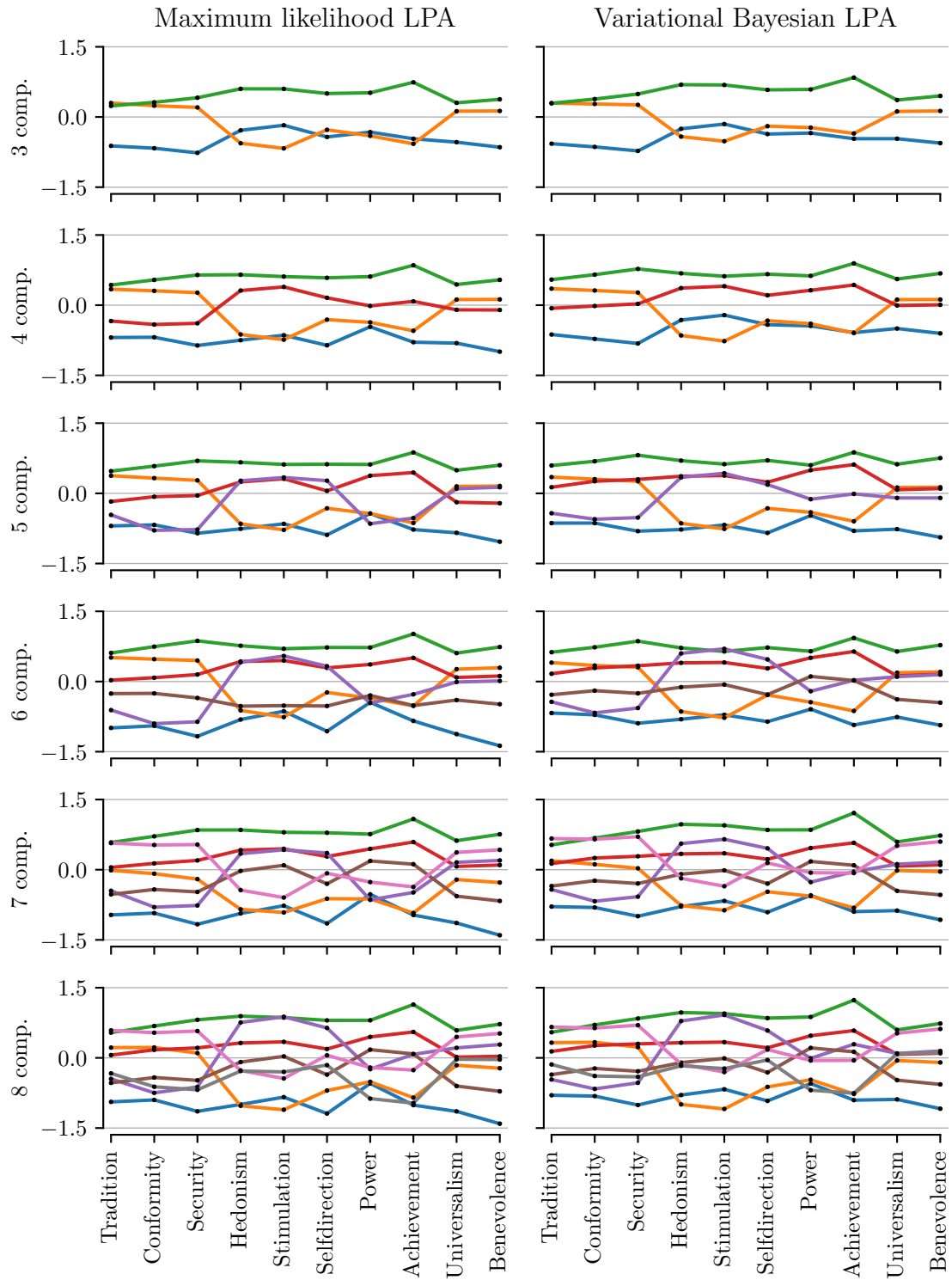
Figure 4: Latent profiles estimated using ML-LPA (left) and VB-LPA (right).
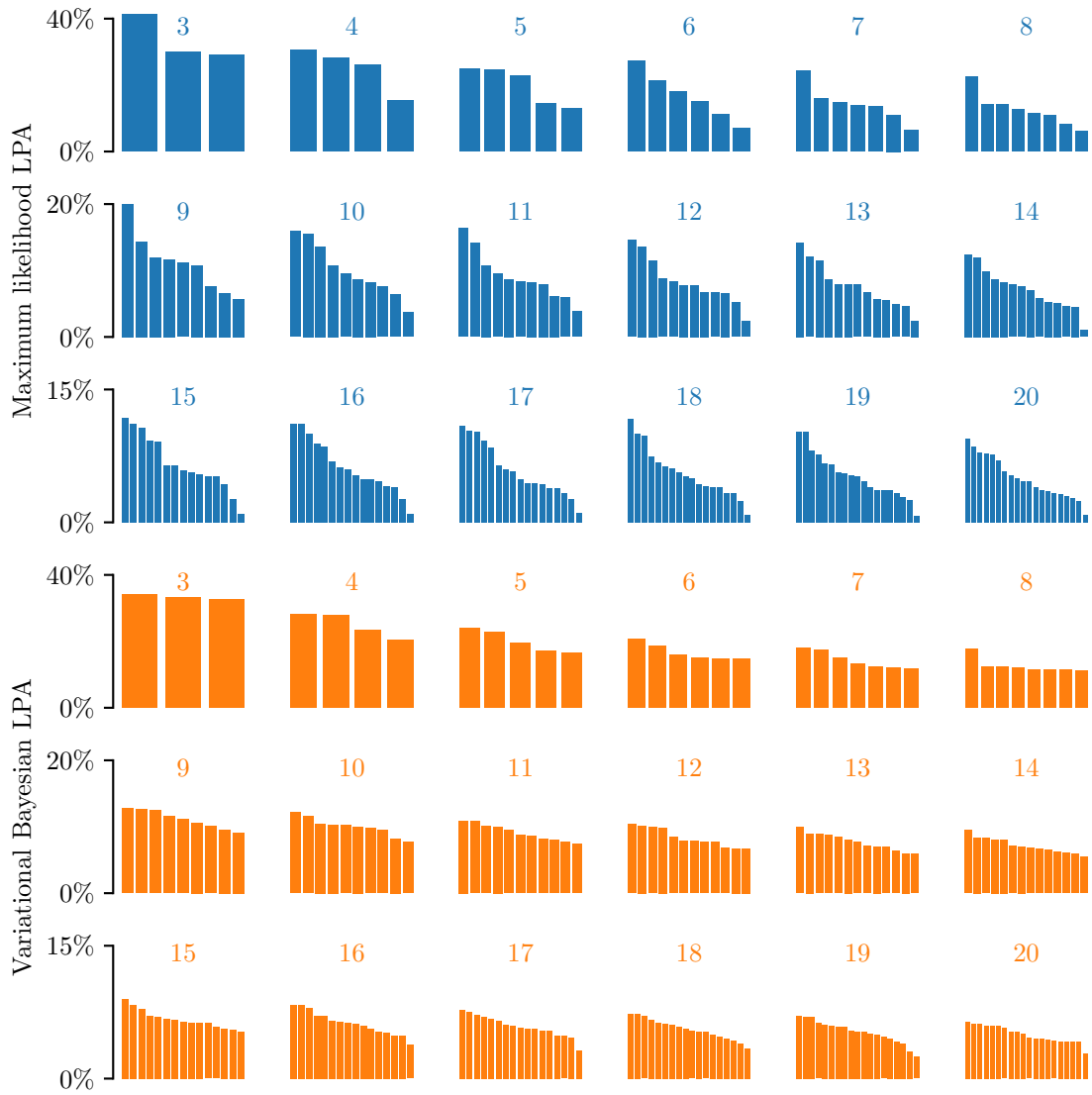
Figure 5: Cluster sizes in percent for ML-LPA (top three rows) and VB-LPA (bottom three rows).
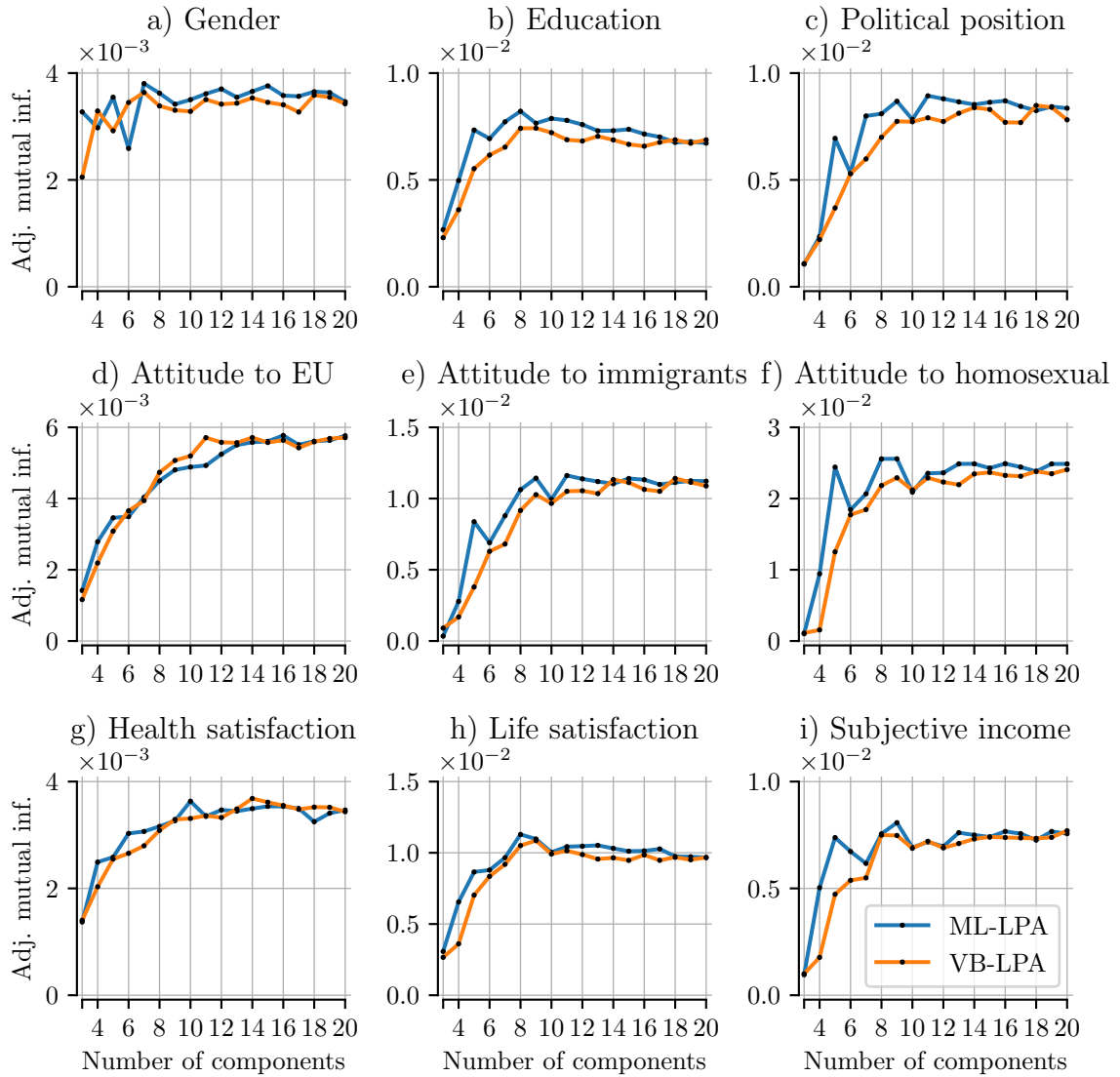
Figure 6: AMI between clustering and external variables.

$K = 8 \ldots 9$, the external variables measuring attitudes and subjective opinions started to level out. On the other hand, the AMI scores for the demographic variables leveled out when the number of cluster reached at $K = 5$.

The qualitative evaluation of the value profiles demonstrated that, for instance, the all positive or the all negative clusters identified in $K=3$ contain potentially meaningful clusters such as *strong growth*, *moderate growth*, *strong personal-focus*, *moderate personal-focus* that appeared when the number of clusters increased to $K=8$. The AMI and the qualitative inspection summarized in Supplemental Materials confirmed that these additional profiles identified in $K=8$ provided additional meaningful insights. Our findings therefore encourage researchers to evaluate if additional profiles obtained from the increased number of clusters provide meaningful insights in relation with the external variables researchers intend to investigate, before they decide on an optimal number of clusters, as Spurk et al. (2020) suggested. The AMI criterion can be especially useful in this respect for evaluating the abilities of the clustering solutions to predict external variables when researchers decide on an optimal number of clusters obtained from not only LPA or LCA but also from other clustering techniques such as k-means and hierarchical clustering techniques.

As for the comparison between the ML-LPA and the VB-LPA[3], our study identified several differences in the two algorithms. First, the sizes of clusters identified by the ML-LPA were varied, while the clusters identified by the VB-LPA were regulated. Second, the ML-LPA achieved higher between-cluster variance scores and lower Davies-Bouldin scores compared to the VB-LPA. This indicates that the regulated size distribution achieved in VB-LPA might have influenced the scores of the lower between-cluster variance. Third, the entropy scores for the VB-LPA were generally higher than those for the ML-LPA. These differences might be explained by the assumption of equal variances between classes in the ML-LPA, where the VB-LPA assumes unequal variances. Finally, in terms of the computational time, the ML-LPA required the random starts up to $2\,000$ times for the higher number of clusters, while the VB-LPA required only 100 random starts for the range of $K = 3 \ldots 20$.

A limitation of our study is that the comparisons of the ML-LPA and VB-LPA were not based on a simulation study but only on empirical data instead. The comparisons using the PVQ items based on the cluster evaluation metrics, however, enabled us to identify an optimal number of clusters that are statistically justifiable and substantively meaningful for interpreting the Schwartz theory. Another noteworthy point is that the ML-LPA based on the method factor presented in Magun et al. (2016); Vermunt (2010) as well as a test run of ML-LPA using the centering (Smith & Schwartz, 1997) scores of the PVQ items both resulted in a smaller number of clusters. The different optimal numbers of clusters identified from different data-sets may be due to the fact that the method factor and the centering techniques are usually used to eliminate the response biases, whereas the CFA scores inherently distinguished several Schwartz typologies with either positive or negative response styles. Some recent works (See, Klimstra, van den Bergh, Sijtsema, & Denissen, 2021) consider a hierarchical tree typology that distinguishes high- and low response styles of pathological personality traits. Hence, it is debatable whether the response styles should be eliminated within the clustering process. This clearly indicates that there is room for further investigation on various data pre-processing procedures. Finally, our study did not consider a covariance structure among the indicators, which can represent the underlying dimensionality of the Schwartz values in addition to classifi-

---

[3]A summary can be found in the supplemental materials.

cation. Thus, future research comparing ML-LPA and VB-LPA may consider the factor mixture model (Muthén, 2008), which can be implemented in the Variational Bayesian algorithm and the frequentist ML approach.

References

Albers, K. J., Mørup, M., Schmidt, M. N., & Glückstad, F. K. (2020, September). Predictive evaluation of human value segmentations. *The Journal of Mathematical Sociology*, 1–28. doi: 10.1080/0022250X.2020.1811277

Beierlein, C., Kuntz, A., & Davidov, E. (2016). Universalism, conservation and attitudes toward minority groups. *Social Science Research*, *58*, 68–79. doi: 10.1016/j.ssresearch.2016.02.002

Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An introduction to latent variable mixture modeling (Part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, *39*(2), 174–187. doi: 10.1093/jpepsy/jst084

Bock, H.-H. (2007). Clustering Methods: A History of k-Means Algorithms.. doi: 10.1007/978-3-540-73560-1_15

Collins, C. M., Steg, L., & Koning, M. A. (2007). Customers' values, beliefs on sustainable corporate performance, and buying behavior. *Psychology and Marketing*. doi: 10.1002/mar.20173

Davidov, E., Yang-Hansen, K., Gustafsson, J. E., Schmidt, P., & Bamberg, S. (2006). Does money matter? A theory-driven growth mixture model to explain travel-mode choice with experimental data. *Methodology*, *2*(3), 124–134. doi: 10.1027/1614-2241.2.3.124

de groot, J., & Steg, L. (2008). Value Orientations to. *Environment and Behavior*, *40*(3), 330–354. doi: 10.1177/0013916506297831

Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis: A primer. *Journal of Cross-Cultural Psychology*, *34*(2), 195–210. doi: 10.1177/0022022102250427

Fritzsche, D. J., & Oz, E. (2007). Personal values' influence on the ethical dimension of decision making. *Journal of Business Ethics*, *75*(4), 335–343. doi: 10.1007/s10551-006-9256-5

Glückstad, F. K., Schmidt, M. N., & Mørup, M. (2017, 1). Examination of Heterogeneous Societies: Identifying Subpopulations by Contrasting Cultures. *Journal of Cross-Cultural Psychology*, *48*(1), 39–57. doi: 10.1177/0022022116672346

Hofer, J., Chasiotis, A., & Campos, D. (2006). Congruence between social values and implicit motives: Effects on life satisfaction across three cultures. *European Journal of Personality*. doi: 10.1002/per.590

Jowell, R., Roberts, C., Fitzgerald, R., & Eva, G. (2007). *Measuring attitudes cross-nationally: Lessons from the European Social Survey*. Sage.

Krystallis, A., Vassallo, M., & Chryssohoidis, G. (2012). The usefulness of Schwartz's 'Values Theory' in understanding consumer behaviour towards differentiated products. *Journal of Marketing Management*, *28*(11-12), 1438–1463. doi: 10.1080/0267257X.2012.715091

Lanza, S. T., Flaherty, B. P., & Collins, L. M. (2003). Latent Class and Latent Transition Analysis. In *Handbook of psychology.* doi: 10.1002/0471264385.wei0226

Lee, J. A., Soutar, G. N., Daly, T. M., & Louviere, J. J. (2011). Schwartz values clusters in the United States and China. *Journal of Cross-Cultural Psychology*, *42*(2), 234–252. doi: 10.1177/0022022110396867

Leong, C. H. (2008). A multilevel research framework for the analyses of attitudes toward immigrants. *International Journal of Intercultural Relations*. doi: 10.1016/j.ijintrel.2007.10.002

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*. doi: 10.1093/biomet/88.3.767

Lönnqvist, J. E., Jasinskaja-Lahti, I., & Verkasalo, M. (2011). Personal values before and after migration: A longitudinal case study on value change in ingrian-finnish migrants. *Social Psychological and Personality Science*, *2*(6), 584–591. doi: 10.1177/1948550611402362

Magun, V., Rudnev, M., & Schmidt, P. (2016). Within- and between-country value diversity in Europe: A typological approach. *European Sociological Review*, *32*(2), 189–202. doi: 10.1093/esr/jcv080

McCutcheon, A. L. (1987). Latent Class Analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-064.

McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York, NY: John Wiley & sons.

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*. doi: 10.1146/annurev-statistics-031017-100325

Moors, G., & Vermunt, J. (2007). Heterogeneity in post-materialist value priorities. Evidence from a latent class discrete choice approach. *European Sociological Review*, *23*(5), 631–648.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 1–24). Charlotte, NC: Information Age Publishing.

Muthen, B., & Shedden, K. (1999). Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics*, *55*(2), 463–469.

Muthén, L. K., & Muthén, B. O. (2017). Mplus User's Guide. Eighth Edition. *Los Angeles, CA: Muthén & Muthén*. doi: 10.1111/j.1600-0447.2011.01711.x

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, *14*(4), 535–569.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*. doi: 10.1037/tps0000176

Petersen, K. J., Qualter, P., & Humphrey, N. (2019). The application of latent class analysis for investigating population child mental health: A systematic review. *Frontiers in Psychology*, *10*(MAY). doi: 10.3389/fpsyg.2019.01214

Prince-Gibson, E., & Schwartz, S. H. (1998). Value Priorities and Gender. *Social Psychology Quarterly*, *61*(1), 49. doi: 10.2307/2787057

Reinecke, J., & Seddig, D. (2011). Growth mixture models in longitudinal research. *AStA Advances in Statistical Analysis*, *95*(4), 415–434. doi: 10.1007/s10182-011-0171-4

Rudnev, M., Magun, V., & Schmidt, P. (2014). The Stability Of The Value Typology Of Europeans: Testing Invariance With Confirmatory Latent Class Analysis. *Higher School of Economics Research Paper No. WP BRP*.

Sagiv, L., & Schwartz, S. H. (1995). Value Priorities and Readiness for Out-Group Social Contact. *Journal of Personality and Social Psychology*, *69*(3), 437–448. doi: 10.1037/0022-3514.69.3.437

Schultz, P. W., Gouveia, V. V., Cameron, L. D., Tankha, G., Schmuck, P., & Franěk, M. (2005). Values and their relationship to environmental concern and conservation behavior. *Journal of Cross-Cultural Psychology*, *36*(4), 457–475. doi: 10.1177/0022022105275962

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*. doi: 10.1016/S0065-2601(08)60281-6

Schwartz, S. H. (2006). Basic human values: An overview. *Jerusalem Hebrew University*, *48*, 49–71. doi: 10.1080/026999499377664

Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*. doi: 10.9707/2307-0919.1116

Schwartz, S. H. (2014). *Instructions for computing scores for the 10 human values and using them in analyses.* doi: 10.1007/s13398-014-0173-7.2

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., ... Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663–688. doi: 10.1037/a0029393

Schwartz, S. H., & Rubel-Lifschitz, T. (2009). Cross-National Variation in the Size of Sex Differences in Values: Effects of Gender Equality. *Journal of Personality and Social Psychology*. doi: 10.1037/a0015546

See, A. Y., Klimstra, T. A., van den Bergh, M., Sijtsema, J. J., & Denissen, J. J. A. (2021). An empirically derived hierarchical tree typology of DSM-5 pathological personality traits in adolescence. *Psychological Assessment*, *33*(6), 499–510. doi: 10.1037/pas0001001

Smith, P. B., & Schwartz, S. H. (1997). Values. *Handbook of cross-cultural psychology, 2nd ed., vol.*, *3*, 77–118.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*(May). doi: 10.1016/j.jvb.2020.103445

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. doi: 10.1093/pan/mpq025

Vermunt, J. K., & Magidson, J. (2002). Applied latent class analysis. In . A. M. In J. Hagenaars (Ed.), (chap. Latent cla). Cambridge:: Cambridge University Press.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236–244.