Food Quality and Preference 34 (2014) 70–78

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Stability of market segmentation with cluster analysis – A methodological approach

Henriette Müller *, Ulrich Hamm

Faculty of Organic Agricultural Sciences, University of Kassel, Germany

ARTICLE INFO

Article history:
Received 8 April 2013
Received in revised form 16 December 2013
Accepted 16 December 2013
Available online 7 January 2014

Keywords:
Market segmentation
Cluster analysis
Reliability
Validity
Stability over time
Panel survey

ABSTRACT

Market segmentation is a very popular marketing tool. In the food sector, the characteristics of different consumer attitudes and consumption habits are often used as the basis for segmentation. However, the success of a target-oriented marketing approach to selected groups of consumers depends on the results of the methodology applied. So far, relatively little attention has been paid to the reliability of the analysis used for attitude-based market segmentation, to the validity or internal stability of results or to the dynamic stability over time with regard to number, size and properties of the segments.

In our study, we used data from a panel of more than 10,000 German households. The participants were segmented using a statement battery and the application of cluster analysis. In order to ensure an internally stable cluster solution, our focus was on the analytical and technical process of decision making when clustering a large dataset. A combination of various statistical measures was applied in order to enable objective decision making in the determination of the optimal number of clusters. The dynamic stability of the resulting segments was determined by confirmatory cluster analyses using data from the same individuals in three subsequent years.

The results of the analyses show that neither the internal nor the dynamic stability of market segments should be taken for granted. Therefore, marketers face the challenge of designing segment-specific marketing strategies in a way that allows changes in consumer preferences to be integrated.

1. Introduction

Market segmentation is a very popular marketing technique and its benefits are emphasised in every textbook on marketing research. One option for the food market is segmentation based on consumer attitudes and stated consumption habits, which are measured on a given scale via the evaluation of a battery of statements. Using these data, it is possible to identify market segments according to groups of consumers that are as homogeneous as possible in their attitudes and consumption habits, and which differ from other groups. Such segments can be used to select particular markets, and these are then targeted with specific marketing measures according to their characteristics (Aaker, Kumar, & Day, 2007).

However, if market changes cannot be reliably predicted, strategies for segmentation and the identification of target groups can only be successful when the segments remain stable over time. For market segmentation based on statement batteries, the 'time' stability of statement ratings is therefore a basic requirement. Stability is defined in terms of a consistent answer to a repeated, identical question (Batista-Foguet & Saris, 1997). Stability of attitudes and of attitude-based market segmentation is usually assumed *per se* (Hoek, Gendall, & Esslemont, 1996).

Nevertheless, while there has been a significant amount of research regarding consumer attitudes and their stability, rather less attention has been paid to the reliability of the methodology applied in attitude-based market segmentation, or to the validity of results and the stability over time of the number, size and properties of segments (Wedel & Kamakura, 2000).

In this context, cluster analysis is a commonly applied procedure for segmenting customers. In cluster analysis, consumers are sorted into relatively homogenous groups according to chosen criteria, so that consumers placed together in the same group, or cluster, show more similarities with each other than with those placed in other clusters. When referring to the stability of a cluster solution, it is necessary to differentiate between 'internal' and 'dynamic' stability (Wedel & Kamakura, 2000). The internal stability, or validity, of a cluster solution describes the potential for replicating segmentation results within the same or similar dataset. It is evaluated using, for example, split samples' procedures or variations of clustering methodology (e.g., Bouguessa, Wang, & Sun, 2006; Dubes & Jain, 1979; Hennig, 2007; Kuncheva & Vetrov, 2006; Levine & Domany, 2001; Meinshausen & Bühlmann, 2010;

* Corresponding author. Address: Faculty of Organic Agricultural Sciences, University of Kassel, Nordbahnhofstrasse 1a, D-37213 Witzenhausen, Germany. Tel.: +49 5542 98 1651.

E-mail address: h.mueller@uni-kassel.de (H. Müller).

0950-3293/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodqual.2013.12.004

Schellinck & Fenwick, 1981; Volkovich, Barzily, & Morozensky, 2008; Wu, Chen, Xiong, & Xie, 2009). According to Lange, Roth, Braun, and Buhmann (2004) and Dolnicar and Leisch (2010), the reproducibility of a cluster solution depends mainly on whether the data contain natural or true segments, as opposed to segments that are constructed by the applied clustering method. Dolnicar (2002) examined various market segmentation studies based on cluster analyses and found that they focused mainly on the interpretability of results, while the validity of the cluster solution was usually ignored. Yet, in order to evaluate dynamic stability, internal stability is a necessary prerequisite.

Dynamic stability refers to the stability of a cluster solution over time and is based on data collected in different time periods. Research has been undertaken on the dynamic stability of market segments using various datasets and different methods (see also Blocker and Flint (2007) for an overview). Calantone and Sawyer (1978) found that benefit segments in a consumer mail panel were relatively stable over the course of 2 years, whereas the allocation of individual households to the segments tended to change. Yuspeh and Fein (1982) determined segment stability via a membership predictor using discriminant function analysis, which could not predict segment membership correctly even for core members of a segment after almost a year. Farley, Winer, and Lehmann (1987) found segments based on consumption patterns in panel data to be highly unstable over 3 years. Ailawadi, Gedenk, and Neslin (1999) analysed segment stability in choice experiments using different approaches and identified relatively stable large segments but rather unstable smaller ones. Dolnicar (2004) introduced a stepwise procedure for tracking a posteriori segments, and found both stable and unstable tourist segments based on binary data. Using latent segments' models, Fonseca and Cardoso (2007) analysed stability in segments of supermarket customers, revealing significant differences in their profiles and size after 3 years. Müller (2011) repeated a choice experiment with a different sample after 18 months and found segments to be relatively stable regarding size and properties, using a latent class analysis.

Within the multivariate techniques used for market segmentation, various forms of cluster analyses are among the most popular. However, the application of clustering methodology requires that the researcher takes several (more or less) subjective decisions (Tonks, 2009). One of the most crucial elements in the practical application of cluster analysis is the determination of the optimal number of clusters, which is most often undertaken subjectively by the researcher (Dolnicar, 2002).

As the literature regarding internal stability or validity shows, it remains difficult to determine even a single cluster solution with certainty and, as a result, very much depends on the methodology applied. Consequently, it is even more challenging to determine dynamic stability. Our research approach was therefore to establish, as far as was possible, an objective and internally stable anchor solution in order to be able to determine dynamic stability through subsequent comparative analyses. The available dataset offered considerable scope insofar as the research could be undertaken within the same sample, consisting of a large number of respondents who were given the same questionnaire over four consecutive years.

In this paper, the focus is on the analytical and technical process of decision making when applying cluster analyses to a large dataset containing a statement battery. A combination of various statistical measures is used for determining the optimal number of clusters, in order to enable objective decision making. To determine the dynamic stability of the segments established in the first year, confirmatory cluster analyses are performed on data from the same individuals in three subsequent years. Thus, in order to determine the overall stability (or validity) of market segments, it is necessary to analyse firstly, the extent to which consumers change their ratings of identical statements over time and, secondly, the degree to which such changes might alter the composition of clusters over the given period.

As a first step in the following analysis, the stability of identical statements is evaluated over a period of 4 years, and the statements are combined to factors using factor analysis. The second stage examines how distinctively market segments can be formed, and how definitely consumers can be assigned to a particular segment. Following this, the extent to which assignment to a specific segment remains stable is determined and, similarly, to what extent a regrouping of people takes place.

The results allow conclusions to be drawn about the informative value of market segments based on attitudes and consumption habits, and the usefulness of customer segmentation in the medium to long term.

2. Material and methods

The analyses are based on German household panel data from the market research institute GfK (Gesellschaft für Konsumforschung). The data consist of a battery of 68 statements on a rating scale from 1 (totally disagree) to 5 (totally agree). The statements contain information regarding attitudes toward food consumption as well as stated purchasing habits and were recorded annually from 2005 to 2008 in Germany. The annual mortality rate in the GfK household panel ranged between 21% and 29% during this period of time. Only households that participated continuously in the survey for all 4 years were included, resulting in a sample of 10,001 households. For each household, additional socio-demographic data were available.

To determine the stability of the statements over time, correlation and reliability analyses with Cronbach's α were used. The statements were combined to factors using factor analysis. An exploratory factor analysis using principal axes methodology was performed via the statistical programme SPSS 19, in order to formulate the model for the year 2005. Afterwards, confirmatory factor analyses using the programme LISREL were carried out for 2005, as well as for the years 2006–2008, and factor scores were calculated for each year. The data were transferred back to SPSS and then cluster analyses were undertaken.

2.1. Clustering methodology and internal stability

First, an exploratory cluster analysis was used to determine the optimal number of clusters in year 2005. For this, the hierarchical Ward method and partitioning *K*-Means analysis were employed. For this dataset, the potential for applying and interpreting hierarchical clustering methods was limited due to large sample size ($n = 10{,}001$). Generally, the researcher is confronted with a difficult decision when determining the optimal number of clusters because usual statistical programmes are able to calculate an unlimited number of possible cluster solutions, but scarcely provide guidance for selecting the best solution (Bacher, 2001). Whereas the visual representation of a hierarchical cluster structure based on a dendrogram can assist in the selection of the most appropriate cluster solution for smaller datasets, this is difficult with large samples. For large datasets, only statistical criteria can be considered. Based on a comparison of empirical results, an overview and assessment of 30 possible criteria for determining the optimal number of clusters was published in 1985 by Milligan & Cooper. Since then, there has been considerable research into this issue and several other methods have been developed (e.g., Albatineh & Niewiadomska-Bugaj, 2011; Ben-Hur, Elisseeff, & Guyon, 2002; Chiang & Mirkin, 2010; Fraboni & Saltstone, 1992; Krieger & Green, 1999; Krolak-Schwedt & Eckes, 1992; Milligan & Cooper, 1985;

Overall & Magee, 1992; Tibshirani, Walther, & Hastie, 2001; Tonidandel & Overall, 2004). So far, however, no method has become standard procedure. The optimal cluster solution for our study was therefore determined using a combination of methods, as suggested by Bacher, Pöge, and Wenzig (2010), who have developed a validation assessment for the determination of the formal adequacy of different cluster solutions. The quality of the selected solutions can then be assessed and compared according to further criteria.

In order to achieve this, *K*-Means analyses must be undertaken using different numbers of clusters. The assignment to clusters must be saved in every case and each solution, so that the *K*-Means analysis can be used in an exploratory way. The initial cluster centres are calculated by previous hierarchical cluster analyses using the Ward method. Schmidt and Hollensen (2006) proposed the comparison of at least 2–5, but better 2–10, cluster solutions. From these solutions, those that fulfil criteria of formal adequacy are chosen. Formal adequacy can be determined by the explained variance (ETA^2_K), the relative improvement compared to the previous solution (PRE_K) and the *F*-MAX statistics ($F\text{-MAX}_K$) (see Bacher et al., 2010).[1]

In the social sciences, it is not uncommon to find that several cluster solutions fulfil the required criteria and are therefore considered to be formally adequate (Bacher et al., 2010). These solutions can then be analysed further to determine the optimal cluster solution. Additional criteria are calculated and the cluster solutions are compared according to these criteria. Depending on the focus of the analysis, criteria can be chosen which may also be weighted differently (Bacher et al., 2010). Ideally, all criteria should point towards the same solution (Schendera, 2010). Besides the criteria PRE_K and $F\text{-MAX}_K$ which have already been mentioned, the number of clusters, the number of outliers and representatives, the Rand and the adjusted Rand Index, and Beale (1969) , *F*-type statistic can also be used (see Bacher et al., 2010).

If an 'optimal' solution is identified according to these criteria, the internal stability of the cluster solution can be evaluated. This can be done by splitting the dataset into two equal-sized sub-samples in which cluster analyses are performed separately. The results from the sub-samples should correspond to the original solution (Bacher et al., 2010). The internal stability of a cluster solution can be tested further by analysing whether the assignment of cases to clusters is independent of the initial cluster centres (Schendera, 2010). If no cluster centres are imported into the *K*-Means analysis, the SPSS programme uses the first cases in the dataset that are the farthest apart as initial cluster centres. The initial cluster centres therefore depend on the sorting of cases in the dataset. To test the stability of a solution, random variables are calculated which are used to sort the cases in the dataset. This is undertaken once in descending and once in ascending order. With each sorting, a *K*-Means analysis with *K* clusters is calculated. The agreement of the solutions can be measured in crosstabs using the Rand, the adjusted Rand Index and with Kappa. Kappa can be used for solutions with identical numbers of clusters (Schendera, 2010). The coding of cluster variables must be identical so that the diagonal is filled maximally. If this is not the case, the cluster variables must be recoded. According to Schendera (2010), Kappa values >0.74 show very good or excellent accordance, values from 0.60 to 0.74 satisfactory or good accordance.

2.2. Dynamic stability

To determine dynamic stability through cluster analyses, the cluster solution of 2005 was chosen as anchor point against which solutions for subsequent years could be compared. To determine changes in cluster structures and in the allocation of individuals to clusters over the years, confirmatory cluster analyses were conducted by *K*-means with fixed cluster centres. For this, the cluster centres of the anchor solution were imported and fixed, so that they were not adapted in the course of iteration. This ensured that the basic orientation of the clusters remained unchanged in terms of content.

In *K*-Means analysis, objects are allocated to clusters in a way that minimises the within clusters total variance. Nevertheless, when *K*-Means is applied in a confirmatory procedure with fixed centres, it is possible that the model fits the data only poorly (Bacher et al., 2010). To assess goodness of cluster fit for the solutions that emerge from this analysis, *F*-values can be used to describe the variance in each cluster in relation to the total variance of the sample (Backhaus, Erichson, Plinke, & Weiber, 2008).

In order to determine the similarities between cluster solutions relating to different time periods, pair-wise comparisons were undertaken in crosstabs, and the indices Kappa, Rand and adjusted Rand were computed. Also, as an overall measure of the reliability of cluster membership, Krippendorff's α was calculated (see Krippendorff, 2004).

3. Results

3.1. Statements

In order to analyse the 68 statements descriptively, means and standard deviations were calculated for each statement from 2005 to 2008. To determine the similarity of these statistical measures with regard to each statement over 4 years, the correlation between measures was also analysed. In addition, reliability was estimated using the reliability coefficient, Cronbach's α. The results of this analysis showed that correlations between measures were significant for all statements, whereby all correlation coefficients were greater than 0.3 and smaller than 0.8. For all statements, Cronbach's α was higher than 0.7, which is generally considered to be a sufficient value for this coefficient (Hair, Black, Babin, & Anderson, 2010).

The differences in the ratings of statements in the years 2006–2008 compared to those in 2005 are described in Table 1. In 2006, the 10,001 households rated on average 31.1 statements differently from 2005. In 2007 and 2008, an average of 33.9 and 34.9 ratings, respectively, were changed, amounting to about half of all statements. In terms of the actual number of statements in which ratings were changed, this ranged in all years from none at all to changes in almost every statement.

As the rating scale varied from 1 (totally disagree) to 5 (totally agree), the greatest possible difference between 2 years is four, which represents a complete change from "totally agree" to "totally disagree" (or vice versa). A change of two scale points stands for a change from agree to disagree (or vice versa) or a change from one extreme to the indifferent middle (or vice versa).

On average, over all statements, the majority of panellists showed a change in their ratings of >0.5–1.0 scale points compared to 2005. In all years, there were also people who changed the ratings by >1.0 scale points on average, which means that they differed in their ratings of all statements by at least one point on the scale, or evaluated some statements very differently by two or more scale points. In 2008, 10% of all people changed their ratings by more than one point on the scale on average. It is also interesting, however, that in all years there were panellists who did not change any of the 68 statement ratings compared to 2005.

3.2. Factor analysis

For the intended cluster analysis, it was necessary to eliminate correlations between the statements of the sample as far as possi-

[1] The criteria ETA^2, PRE and *F*-MAX can be calculated in SPSS with an appropriate syntax. A guideline was published by Bacher (2001).

Table 1
Differences in the rating of the statements in 2006–2008 compared to 2005.

		2005/2006	2005/2007	2005/2008
Number of differently rated statements	Mean	31.1	33.9	34.9
	Standard deviation	11.8	9.0	8.5
Percentage of interviewes	Mean difference in scale points			
	0	8.2%	2.1%	1.2%
	>0–0.5	17.6%	16.0%	14.1%
	>0.5–1.0	67.2%	72.8%	74.7%
	>1.0–1.5	6.6%	8.5%	9.3%
	>1.5–2.0	0.4%	0.5%	0.6%
	>2.0	0.1%	0.0%	0.1%

n = 10,001, 68 statements/year.

ble, and to reduce the number of measured variables to a manageable level. This was achieved using factor analysis. In order to be able to compare the results of the four consecutive years and to realise similar conditions for each year, the method of confirmatory factor analysis (CFA) was chosen.

To develop the model, an exploratory factor analysis (EFA) was conducted for the year 2005. In SPSS, a principal axes analysis was performed with the 68 statements, in which factors with Eigenvalues greater than 1 were extracted. In evaluating the model, all variables with a loading of less than 0.5 were excluded and the analysis was repeated; 38 variables remained in a 12 factor solution. In order to avoid factors with fewer than four variables, additional variables were added to such factors on a trial basis. For this, the variables used were those which had originally loaded onto the factor, but with loadings of less than 0.5. Six variables showed loadings greater than 0.4 and were added to the model, so that the final result contained 44 variables and 12 factors.

The model explained 50.3% of the total variance. The model, thus specified, was then entered into a CFA in LISREL. The results of the CFA showed that one more variable had to be excluded due to a loading of <0.5. The final result of the CFA for 2005 is displayed in Table 2.

According to this optimised model for 2005, CFAs were conducted with the same specification for the years 2006, 2007 and 2008. The model fit for the CFA in the 4 years is shown in Table 3.

While the degrees of freedom are the same in all years due to identical construction of the measurement models, the quality measures change. In all years, the chi-square value is very high and the p-value is very low, indicating a low quality model. However, according to Hair et al. (2010), the chi-square test is not a reliable criterion for the goodness of model fit when large samples are used. Given the present sample size (n = 10,001), the significance of the chi-square test is therefore limited. In contrast, the Root Mean Square Error of Approximation (RMSEA) is, in all years, below the critical value of 0.07 and the Comparative Fit Index (CFI) is above 0.90. Thus, it can be assumed that there is a good model fit (see Hair et al., 2010). Therefore, the respective factor scores were calculated for the 4 years for each panellist, and stored in the dataset. After transfer to SPSS, these were available for the cluster analysis.

3.3. Exploratory cluster analysis

Cluster analysis was undertaken using SPSS with the previously z-standardised factor scores for all cases (see Backhaus et al., 2008; Hair et al., 2010). The Ward method was used to determine an initial cluster solution and was carried out with squared Euclidean distances and all other default settings. Cluster membership was saved for solutions 1–15. In order to use the mean values of the factors in each of the cluster solutions as cluster centres for subsequent K-Means analysis, the cluster centres were saved in new files for the solutions with 2–15 clusters. It is not necessary to provide initial cluster centres for the 1-cluster solution.

The cluster centres were then read from Ward as the starting values for the K-Means analysis which was carried out for each number of clusters from 1 to 15. Cluster membership and the distance of every case to its cluster centre in each solution were saved in the dataset. The cluster solutions were examined according to the criteria described above and 214 outliers were determined. After elimination of the outliers, the 6-cluster solution turned out to be the best one. It was also plausible as regards interpretation of content. The 6-cluster solution was therefore chosen as anchor point for further analyses.

Tests of the internal stability of the 6-cluster solution were applied using split samples and random starting values. The resulting solutions were compared in crosstabs with the original 6-cluster solution. In all crosstabs, Kappa had values >0.8 and thus showed excellent agreement (Schendera, 2010). The Rand and the adjusted Rand Index also showed very good agreement of cluster solutions. The 6-cluster solution was therefore relatively independent of initial cluster centres, as well as of modifications of the sample, and could be considered to be sufficiently stable.

At this point, the outliers were inspected in order to reveal differences in relation to other cases in each of the respective clusters. Outliers were evident in every cluster and, when compared with the other cases, they were found to be similar with respect to some variables but considerably different as regards others. In themselves, however, they were not homogeneous groups within their respective clusters.

Factor scores for the clusters of the 6-cluster solution are described in Table 4, followed by a description of the clusters' properties and characteristics.

Cluster 1 – friends of foreign cuisine

People in Cluster 1 are particularly enthusiastic about foreign specialties. Accordingly, they attach no importance to the regional or German origin of their food. For them, cooking does not have to be fast or simple and they do not care about particular brands; they like to try new and unusual recipes.

Cluster 2 – sceptical conservatives

People in Cluster 2 are quite sceptical about anything new, but do not prefer small shops with personal service. They do not value specialties, new products, organic foods or dietary supplements. For them, cooking does not necessarily have to be healthy, but fast and simple. They like to stick to tried and tested recipes.

Cluster 3 – alternative gourmets

People in Cluster 3 prefer organic products. The regional or German origin of the food is important to them, but they also like foreign specialties. Fast, simple food is not their favourite, and

Table 2
Confirmatory factor analysis for 2005.

Factor	Statement	Factor loading
F1 health consciousness	In my household, I pay attention to a light and balanced diet	0.71**
	I pay strict attention to eating little fat	0.66**
	I pay strict attention to low salt	0.60**
	I pay strict attention to the sugar content of foods	0.59**
	I watch what I eat and drink, because I have to consider my health	0.63**
	The topic of food cholesterol plays a role when shopping	0.56**
	I avoid eating anything unhealthy	0.61**
F2 organic products	I would like to see a wider range of organic products in stores	0.91**
	I am willing to spend more money for organic products	0.87**
	When grocery shopping I prefer organic food	0.88**
	I would like more information about organic products	0.74**
F3 new products	I often have new products earlier than my friends	0.86**
	I already buy many items that other housewives do not know yet	0.82**
	I like to try out new products	0.70**
	I'm always on the lookout for new products that meet my needs more	0.64**
F4 regional origin	If I have a choice, I definitely buy food from Germany	0.83**
	For me food products from Germany are qualitatively the best	0.74**
	I do not care if my food comes from Germany or any other country	−0.70**
	I pay attention to purchasing food products from my region	0.64**
F5 foreign specialties	It's fun to try foreign specialties	0.86**
	I'm enthusiastic about specialties from other countries	0.69**
	I like to cook fancy foods and dishes	0.82**
F6 fast and simple cooking	The simpler the cooking the better I like it	0.80**
	I prefer to cook dishes that don't take too long	0.81**
	For cooking I take a lot of time	−0.64**
F7 brands	Food products from well-known brands are better than products with unknown names	0.83**
	Branded products are better than products with unknown names	0.70**
	I have no real confidence in food without a brand name	0.70**
F8 snacks	I often go eat something on the way	0.77**
	I often eat fast food (fast food, takeaways, pizza delivery service)	0.81**
	I eat mostly snacks	0.55**
F9 scepticism	New products are often more expensive than the old, but not better	0.72**
	I greatly distrust advertising statements	0.65**
	To buy brand new products often is a failure	0.52**
	I think that most products that continually enter the market are unnecessary	0.57**
F10 food supplements	I regularly use vitamin and mineral supplements to keep myself physically fit	0.72**
	A normal diet contains all the vital nutrients, no additional intake is necessary	−0.62**
	I like to buy food/drinks that contain added minerals and vitamins (ACE, calcium, etc.)	0.66**
	Multivitamin juices are an important addition to the diet	0.55**
F11 established and traditional dishes	I love to eat home cooking	0.59**
	I prefer to stick to established recipes when cooking	0.80**
F12 corner shops	I would not want to miss personal shopping service	0.65**
	I love the atmosphere of small shops and specialty stores	0.64**

** Significance at least 5% ($t \leqslant -1.96$, respectively $t \geqslant 1.96$) n = 10,001.

Table 3
Model fit for the confirmatory factor analysis from 2005 to 2008.

Criterion	2005	2006	2007	2008
Degrees of freedom	794	794	794	794
Chi-square	14337.80 (p = 0.0)	15516.49 (p = 0.0)	15364.97 (p = 0.0)	15441.99 (p = 0.0)
RMSEA (<0.07)	0.041	0.043	0.043	0.043
CFI ($\geqslant$0.90)	0.93	0.94	0.94	0.95

n = 10,001.

nutritional supplements are equally unnecessary for them. They prefer shopping in smaller shops with personal service.

Cluster 4 – health conscious traditionalists
People in Cluster 4 attach particular importance to regional origin and a healthy diet. For this, they rely on branded products. They appreciate home cooking and like to buy in smaller shops with personal service. They do not care for new products, foreign dishes or snacks and fast food.

Cluster 5 – fast food fans
People in Cluster 5 like to eat snacks or fast food and they appreciate quick and easy meals. They are receptive to new products. They do not care especially for healthy nutrition or regional food origin.

Cluster 6 – health conscious modernists
People in Cluster 6 focus on a healthy diet and like to make use of new products as well as food supplements. They are open-minded about novelties, organic products and foreign specialties.

3.4. Confirmatory cluster analyses for the years 2006–2008

In order to determine how the cluster structures found in 2005 changed during the years 2006, 2007 and 2008, as well as the extent of changes in the allocation to clusters, confirmatory cluster analyses were conducted. To ensure an identical case basis for all

Table 4
Factor scores in the clusters of 2005.

Factor	Cluster					
	1 n = 1467	2 n = 1877	3 n = 1609	4 n = 1707	5 n = 1446	6 n = 1681
Health consciousness	−0.60	−0.44	0.19	**0.83**	**−0.64**	0.56
Organic products	−0.69	**−0.73**	**0.74**	0.40	−0.25	0.51
New products	0.25	**−0.79**	−0.27	−0.39	0.55	**0.85**
Regional origin	**−1.01**	−0.21	0.49	**0.90**	−0.49	0.15
Foreign specialties	**0.84**	**−0.95**	0.56	−0.80	0.04	0.57
Fast and simple cooking	−0.62	0.65	**−0.69**	0.22	**0.69**	−0.34
Brands	**−0.69**	−0.55	−0.06	**0.83**	−0.10	0.52
Snacks	−0.12	−0.30	**−0.47**	**−0.47**	**1.50**	0.09
Scepticism	−0.12	**0.49**	0.46	0.27	−0.38	**−0.83**
Food supplements	−0.34	−0.47	**−0.59**	0.30	0.24	**0.87**
Established and traditional dishes	**−0.79**	**0.87**	−0.57	0.80	0.01	−0.56
Corner shops	−0.46	**−0.49**	**0.60**	0.50	−0.28	0.11

The highest and the lowest scores per factor are highlighted in bold print. n = 9787 (after elimination of 214 outliers).

years, the 214 outliers identified in 2005 were removed from the datasets in every year. Then the factor scores for the years 2006–2008 were z-standardised. As starting values for the K-Means analyses, the cluster centres of the final solution from 2005 were imported and fixed. The respective cluster membership was saved for the years 2006, 2007 and 2008.

In the cluster solutions of all years, the F-values for each cluster and for all factors were smaller than one. This means that all the clusters were more homogeneous, in themselves, than the total sample. However, in each year there were several factors where the F-value for one or more clusters was above 0.9. This means that the corresponding cluster was only a little more homogeneous in relation to this factor than the total sample. The average of the F-values of all factors and clusters was 0.659 in 2005. In subsequent years, this figure rose slightly so that a slight annual deterioration in the homogeneity of the cluster solutions could be observed.

The value of Krippendorff's α for the exploratory cluster solution of 2005 and the confirmatory solutions of 2006, 2007 and 2008 was 0.5307, with a 95% confidence interval between 0.4911 and 0.5686. The probability of reaching an α value of at least 0.5 was 94.2% (n = 9787, 4 years, 6 clusters).

The pair-wise comparisons show that the number of changes in cluster allocations increases over the years, as more and more people are allocated to clusters that are different from those of 2005 (Table 5). Therefore, the quality criteria Kappa, Rand Index and adjusted Rand Index decreased over the years. Nevertheless, accordance with the solution of 2005 was sufficient or even very good in every year, according to Kappa, Rand and the adjusted Rand Index. Over the course of the four years, 3689 people (37.7%) were always grouped into the same cluster; 4804 people (49.1%) were grouped into two different clusters, 1165 people (11.9%) into three different clusters and 129 people (1.3%) into four different clusters.

3.4.1. Changes in cluster content

The calculated factor scores were used for the clustering procedure and the description of clusters in 2005 because they clearly highlight the differences between panellists. These scores represent the relative evaluation made by each person in relation to each factor. Since the values are relative to the rating of other panellists in the dataset of the year, they are not suitable for a comparison of the absolute scores of individuals between years. In order to be able to describe the impact of changing attitudes and the subsequent regrouping of people on the content of clusters, the average scores of statements per factor and cluster were calculated for the years 2005 and 2008 (Table 6). To achieve this, the mean of the statement evaluations per factor was calculated. For this it was necessary to recode those statements with negative loadings, in order to capture the nature of agreement or disagreement per factor.

Overall, there were only minor changes in the characteristics of the clusters between the cluster solutions in 2005 and 2008. Although this was to be expected to some extent due to the confirmatory procedure with fixed cluster centres, it is an interesting outcome, considering that the membership of all clusters has changed considerably. In all clusters more than one third, and in some clusters almost one half, of the original members have switched to another cluster.

In all clusters, the approval of panellists regarding such factors as *organic products*, *new products, foreign specialties* and *food supplements* dropped, both among those who expressed little support in 2005, as well as those in stronger agreement. Conversely, agreement with respect to *regional origin*, *brands* and *established and traditional dishes* tended to increase. The evaluation of the factors *health consciousness, fast and simple cooking, scepticism, snacks* and *corner shops* changed only slightly in the clusters between 2005 and 2008.

For Cluster 1 (friends of foreign cuisine), the high value associated with foreign dishes was particularly characteristic in 2005 and although this had fallen slightly in 2008, it remained very high. The previously negative assessment of the regional origin of products and established and traditional dishes was shifted slightly towards a neutral rating. In Cluster 2 (sceptical conservatives), the conservative attitudes reflected in a rejection of organic products, dietary supplements and new products became even stronger. The regional origin of products retained its neutral value as before. In Cluster 3 (alternative gourmets), the importance of organic products fell slightly. The preference for regional origin and the rejection of dietary supplements was even more pronounced than before. In Cluster 4 (health-conscious traditionalists), the importance of regional origin which was particularly prominent in 2005, increased further. The rejection of new products and dietary supplements was also more pronounced in 2008 than was the case in 2005. For Cluster 5 (fast food fans), nothing changed much. The importance of regional origin shifted slightly to the neutral centre and branded products were evaluated slightly less negatively than before. For Cluster 6 (health conscious modernists), regional origin was now rated very positively. The evaluation of the statements about organic products and food supplements remained despite slight shifts in the neutral range (neither agree nor disagree).

The examination of cluster contents in different years reveals the general tendencies across all clusters, rather than changes in individual clusters. Overall, agreement with the statements about *regional origin* increased whereas agreement with those about *organic products* and *food supplements* dropped. However, these changes did not affect the general nature of the clusters, and neither the ratings that characterised respective clusters, nor those

Table 5
Comparison of the exploratory cluster solution in 2005 and the confirmatory solutions 2006 to 2008.

Exploratory cluster solution 2005	1	2	3	4	5	6	Total	Kappa	Rand index	Adj. Rand index
Confirmatory cluster solution 2006										
1	**906**	140	114	22	144	141	1467	0.557	0.810	0.319
2	145	**1258**	112	187	139	36	1877			
3	126	95	**967**	188	51	182	1609			
4	22	185	174	**1103**	69	154	1707			
5	161	123	49	45	**934**	134	1446			
6	131	50	170	169	149	**1012**	1681			
Total	1491	1851	1586	1714	1486	1659	9787			
Confirmatory cluster solution 2007										
1	**774**	193	151	13	172	164	1467	0.491	0.791	0.253
2	166	**1189**	119	194	168	41	1877			
3	124	119	**908**	183	62	213	1609			
4	14	208	182	**1038**	70	195	1707			
5	172	150	71	48	**809**	196	1446			
6	136	54	213	177	176	**925**	1681			
Total	1386	1913	1644	1653	1457	1734	9787			
Confirmatory cluster solution 2008										
1	**757**	188	163	19	158	182	1467	0.476	0.787	0.238
2	153	**1167**	111	206	183	57	1877			
3	148	109	**876**	192	70	214	1609			
4	23	235	179	**1027**	61	182	1707			
5	156	153	77	58	**792**	210	1446			
6	160	65	216	174	161	**905**	1681			
Total	1397	1917	1622	1676	1425	1750	9787			

The identical cluster classifications are highlighted in bold print $n = 9787$ (after elimination of 214 outliers).

Table 6
Average rating of the statements per factor in the clusters 2005 and changes in the confirmatory solution of 2008.

	C1		C2		C3		C4		C5		C6	
	Friends of foreign cuisine		Sceptical conservatives		Alternative gourmets		Health-conscious traditionalists		Fast food fans		Health-conscious modernists	
	2005	2008	2005	2008	2005	2008	2005	2008	2005	2008	2005	2008
Share of identical assignments[a]	51.6%		62.2%		54.4%		60.2%		54.8%		53.8%	
N factor	1467	1397	1877	1917	1609	1622	1707	1676	1446	1425	1681	1750
F1 health consciousness	2.5	−0.1	2.6	±0.0	3.1	−0.1	3.6	±0.0	2.5	±0.0	3.4	−0.1
F2 organic products	1.9	−0.2	1.9	−0.2	3.4	−0.1	3.1	−0.2	2.4	−0.1	3.2	−0.2
F3 new products	2.9	−0.1	2.0	−0.1	2.5	−0.1	2.4	−0.2	3.2	−0.1	3.4	−0.1
F4 regional origin	2.3	+0.2	2.9	+0.2	3.6	+0.2	3.9	+0.1	2.7	+0.2	3.3	+0.2
F5 foreign specialties	4.0	−0.1	2.4	−0.1	3.8	−0.1	2.6	−0.1	3.3	−0.1	3.8	−0.1
F6 fast and simple cooking	2.7	±0.0	3.7	−0.1	2.6	±0.0	3.3	−0.1	3.9	−0.1	2.9	±0.0
F7 brands	1.6	+0.1	1.7	+0.1	2.1	+0.1	2.8	+0.1	2.0	+0.2	2.5	+0.1
F8 snacks	1.6	−0.1	1.6	−0.1	1.4	±0.0	1.4	−0.1	2.9	±0.0	1.8	±0.0
F9 scepticism	3.4	±0.0	3.8	±0.0	3.8	±0.0	3.7	±0.0	3.3	±0.0	3.0	+0.1
F10 food supplements	2.2	−0.2	2.1	−0.2	1.9	−0.1	2.7	−0.2	2.6	−0.1	3.1	−0.2
F11 established and traditional dishes	2.6	+0.2	3.8	+0.1	2.8	±0.0	4.0	±0.0	3.3	±0.0	3.0	+0.1
F12 corner shops	3.0	±0.0	2.9	±0.0	3.8	±0.0	3.8	±0.0	3.1	±0.0	3.4	+0.1

[a] Share of individuals, who were assigned to the same cluster in 2005 and 2008 as percentage of the cases of the respective cluster in 2005. Rating scale of the statements from 1 (totally disagree) to 5 (totally agree) $n = 9787$ (after elimination of 214 outliers).

ratings that were previously neutral, changed significantly. This is partly due to the fixing of cluster centres, which did not allow drastic changes but, rather, caused a different allocation of people according to changes in their statement ratings.

4. Discussion and conclusions

The main point of market segmentation is to find and characterise particular consumer groups, and attain profitable customer segments. However, the needs of customers are not necessarily stable and can cause dynamic changes in the market. In this context, depending on objectives, segment instability can be both blessing and curse (Blocker & Flint, 2007). Obviously, it is easier for marketers to deal with segments that remain relatively stable or that undergo changes which are, at least, predictable. Nevertheless, since the development and implementation of relevant measures take time and money, an important condition for successful market segmentation is that the target groups identified will persist for a certain period of time. In order to draw conclusions about dynamic stability based on statements regarding food attitudes and consumption habits, both the stability of statements and the stability of market segments, which were formed on the basis of such statements, must be examined.

This research was based on surveys conducted within the scope of GfK household panel studies from 2005 to 2008 and, for the analyses, only those households that had participated in all 4 years of the survey were selected. Inevitably, this results in selectivity bias through which the sample becomes distorted. Although the extent of this distortion was measured according to socio-demographic criteria, it was not possible to analyse how much this influenced the stability of statements, as these data are not available for households leaving the panel.

Our results show that although the extent of alterations to the rating of statements increased over the years, the average rating of statements remained relatively stable. Furthermore, the statements were well-represented by the factor structure. Recently however, doubts have been raised about the common practice of applying factor analysis prior to cluster analysis, mainly because of loss of information arising from combining statements to factors. It has been shown that cluster analysis without previous factor analysis outperforms the factor–cluster procedure with respect to identifying the correct number of clusters in the data set and to recovering segmentation structure correctly. An alternative, if there are too many variables, is to eliminate the redundant ones from the data set before segmenting (Dolnicar & Grün, 2008, 2011). In the case of our research, however, we decided in favour of retaining the large number of variables available in a factor–cluster procedure rather than reducing the number of possibly redundant variables beforehand.

From the methodological viewpoint, market segmentation was to be achieved through the use of cluster analysis, the critical issue being the selection of a suitable cluster solution for 1 year, against which to compare solutions from other years. We used the formal validation procedure used by Bacher et al. (2010), through which a combination of different criteria is applied in order to ensure the greatest possible objectivity in the choice of cluster solution. Even though the final selection of the best solution remains ultimately subjective because both choice of criteria and their weighting are determined by the researcher, such criteria enable a more objective decision to be made with respect to number of clusters and should therefore be more commonly applied.

The results of the confirmatory cluster analyses indicate that with increasing distance from the reference solution (2005) over time, more and more people "move" into other clusters. In spite of the different composition of the clusters however, their basic structure could still be observed in later years and the main characteristics of the segments remained. It was therefore possible to assign consumers to a given set of segments with acceptable results.

The findings presented here show that the common practice of cluster analysis should be applied with caution. Neither the internal nor the dynamic stability of market segments should be taken for granted. Marketers need to design segment-specific marketing strategies that include opportunities for flexible adaption, so that changes in consumer preference, such as increasing interest in regional food production or a reduction in the importance attached to food supplements, can be properly integrated.

Acknowledgement

We gratefully acknowledge the helpful comments of two anonymous reviewers and we are thankful for the proofreading and copy-editing by Anne-Marie Sherwood (Aberystwyth, UK).

References

Aaker, D. A., Kumar, V., & Day, G. S. (2007). *Marketing research* (9th ed.). Hoboken: Wiley.

Ailawadi, K. L., Gedenk, K., & Neslin, S. A. (1999). Heterogeneity and purchase event feedback in choice models: An empirical analysis with implications for model building. *International Journal of Research in Marketing, 16*(3), 177–198.

Albatineh, A. N., & Niewiadomska-Bugaj, M. (2011). MCS: A method for finding the number of clusters. *Journal of Classification, 28*(2), 184–209.

Bacher, J. (2001). Teststatistiken zur Bestimmung der Clusterzahl für QUICK cluster. *ZA-Information, 48*, 71–97.

Bacher, J., Pöge, A., & Wenzig, K. (2010). *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren* (3rd ed.). Munich: Oldenbourg.

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2008). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (12th ed.). Berlin: Springer.

Batista-Foguet, J. M., & Saris, W. E. (1997). Tests of stability in attitude research. *Quality and Quantity, 31*(1), 269–285.

Beale, E. M. L. (1969). Euclidean cluster analysis. In *Bulletin of the international statistical institute: Proceedings of the 37th session* (*London*). Voorburg, Netherlands: International Statistical Institute. [book 2]

Ben-Hur, A., Elisseeff, A., & Guyon, I. (2002). A stability based method for discovering structure in clustered data. *Pacific Symposium on Biocomputing, 7*, 6–17.

Blocker, C. P., & Flint, D. J. (2007). Customer segments as moving targets: Integrating customer value dynamism into segment instability logic. *Industrial Marketing Management, 36*(6), 810–822.

Bouguessa, M., Wang, S., & Sun, H. (2006). An objective approach to cluster validation. *Pattern Recognition Letters, 27*(13), 1419–1430.

Calantone, R. J., & Sawyer, A. G. (1978). The stability of benefit segments. *Journal of Marketing Research, XV*, 395–404.

Chiang, M. M.-T., & Mirkin, B. (2010). Intelligent choice of the number of clusters in *K*-means clustering: An experimental study with different cluster spreads. *Journal of Classification, 27*(1), 3–40.

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel and Tourism Marketing, 12*(1), 1–22.

Dolnicar, S. (2004). Tracking data-driven market segments. *Tourism Analysis, 8*(2), 227–232.

Dolnicar, S., & Grün, B. (2008). Challenging "factor cluster segmentation". *Journal of Travel Research, 47*(1), 63–71.

Dolnicar, S. & Grün, B. (2011). Three good reasons NOT to use factor-cluster segmentation. In *CAUTHE 2011 : 21st CAUTHE national conference*. Adelaide, Australia.

Dolnicar, S., & Leisch, F. (2010). Evaluation of structure and reproducibility of cluster solutions using the bootstrap. *Marketing Letters, 21*(1), 83–101.

Dubes, R., & Jain, A. K. (1979). Validity studies in clustering methodologies. *Pattern Recognition, 11*(4), 235–254.

Farley, J. U., Winer, R. S., & Lehmann, D. R. (1987). Stability of membership in market segments identified with a disaggregate consumption model. *Journal of Business Research, 15*(4), 313–328.

Fonseca, J. R. S., & Cardoso, M. G. M. S. (2007). Supermarket customers segments stability. *Journal of Targeting, Measurement and Analysis for Marketing, 15*(4), 210–221.

Fraboni, M., & Saltstone, R. (1992). The Wais-R number-of-factors quandary: A cluster analytic approach to construct validation. *Educational and Psychological Measurement, 52*(3), 603–613.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Upper Saddle River: Pearson.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis, 52*(1), 258–271.

Hoek, J., Gendall, P., & Esslemont, D. (1996). Market segmentation: A search for the holy grail? *Journal of Marketing Practice. Applied Marketing Science, 2*(1), 25–34.

Krieger, A. M., & Green, P. E. (1999). A cautionary note on using internal cross validation to select the number of clusters. *Psychometrika, 64*(3), 341–353.

Krippendorff, K. (2004). Reliability in content analysis. Some common misconceptions and recommendations. *Human Communication Research, 30*(3), 411–433.

Krolak-Schwedt, S., & Eckes, T. (1992). A graph theoretic criterion for determining the number of clusters in a data set. *Multivariate Behavioral Research, 27*(4), 541–565.

Kuncheva, L., & Vetrov, D. (2006). Evaluation of stability of *k*-means cluster ensembles with respect to random initialization. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(11), 1798–1808.

Lange, T., Roth, V., Braun, M. L., & Buhmann, J. M. (2004). Stability-based validation of clustering solutions. *Neural Computation, 16*(6), 1299–1323.

Levine, E., & Domany, E. (2001). Resampling method for unsupervised estimation of cluster validity. *Neural Computation, 13*(11), 2573–2593.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72*(4), 417–473.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*(2), 159–179.

Müller, S. (2011). Stability of latent class segments over time: Results from a stated preference experiment. In *Proceedings of the second international choice modelling conference*. Oulton Hall: ICMC.

Overall, J. E., & Magee, K. N. (1992). Replication as a rule for determining the number of clusters in hierarchical cluster analysis. *Applied Psychological Measurement, 16*(2), 119–128.

Schellinck, T., & Fenwick, I. (1981). Testing the stability of market segmentation analysis. *Advances in Consumer Research, 8*, 723–727.

Schendera, C. F. G. (2010). *Clusteranalyse mit SPSS: Mit faktorenanalyse*. Munich: Oldenbourg.

Schmidt, M. J., & Hollensen, S. (2006). *Marketing research: An international approach*. Harlow: Prentice Hall.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B, 63*(2), 411–423.

Tonidandel, S., & Overall, J. E. (2004). Determining the number of clusters by sampling with replacement. *Psychological Methods, 9*(2), 238–249.

Tonks, D. G. (2009). Validity and the design of market segments. *Journal of Marketing Management, 25*(3–4), 341–356.

Volkovich, Z., Barzily, Z., & Morozensky, L. (2008). A statistical model of cluster stability. *Pattern Recognition, 41*(7), 2174–2188 [Retrieved 24.07.2013].

Wedel, M., & Kamakura, W. (2000). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Norwell: Kluwer Academic Publishers.

Wu, J., Chen, J., Xiong, H., & Xie, M. (2009). External validation measures for *K*-means clustering: A data distribution perspective. *Expert Systems with Applications, 36*(3), 6050–6061.

Yuspeh, S., & Fein, G. (1982). Can segments be born again? *Journal of Advertising Research, 22*, 13–22.