

Using segment level stability to select target segments in data-driven market segmentation studies

Sara Dolnicar¹  · Friedrich Leisch²

Published online: 1 March 2017
© Springer Science+Business Media New York 2017

Abstract Market segmentation is widely used by industry to select the most promising target segment. Most organisations are interested in finding one or a small number of target segments to focus on. Yet, traditional criteria used to select a segmentation solution assess the global quality of the segmentation solution. This approach comes at the risk of selecting a segmentation solution with good overall quality criteria which, however, does not contain groups of consumers representing particularly attractive target segments. The approach we propose helps managers to identify segmentation solutions containing attractive individual segments (e.g., more profitable), irrespective of the quality of the global segmentation solution. We demonstrate the functioning of the newly proposed criteria using two empirical data sets. The new criteria prove to be able to identify segmentation solutions containing individual attractive segments which are not detected using traditional quality criteria for the overall segmentation solution.

Keywords Market segmentation · Niche segments · Cluster analysis · Bootstrap

✉ Sara Dolnicar
s.dolnicar@uq.edu.au

Friedrich Leisch
friedrich.leisch@boku.ac.at

¹ UQ Business School, The University of Queensland, Brisbane, Queensland 4072, Australia

² Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences, Peter-Jordan-Straße 82, 1190, Vienna, Austria

1 Introduction

Market segmentation is a critical building block of strategic marketing (Iacobucci 2013) and “essential for marketing success” (Lilien and Rangaswamy 2003, p. 61). Conceptually, there are two possible approaches to market segmentation. Segments can be defined by using *one single* segmentation variable. For example, profitability can be used to split existing customers into a high, medium, and low profit potential segments. These three market segments can then be profiled using descriptor variables such as benefits sought from the product, socio-demographics, or media behavior. This approach has been referred to as *a priori* (Myers and Tauber 1977; Mazanec 2000; Wedel and Kamakura 2000), convenience-group (Lilien and Rangaswamy 2003), or commonsense (Dolnicar 2004) segmentation.

Alternatively, multiple segmentation variables can be used. For example, benefits people seek when buying food in a fast food restaurant (save time, save money, keep kids happy, ...) may have been collected in a survey. The full set of benefits is used to extract market segments. As opposed to segmentations based on one variable, it is therefore not known in advance what the defining features of each of the market segments may be. Once the segments have been extracted from the data, they also need to be profiled in detail using descriptor variables, just like the high, medium, and low profit potential segments in the previous example. This approach where multiple segmentation variables are used is referred to as *a posteriori* (Mazanec 2000), response-based (Myers and Tauber 1977), post-hoc (Wedel and Kamakura 2000) or data-driven market segmentation (Dolnicar 2004). Throughout the manuscript, we will use the terms commonsense and data-driven segmentation because they are most intuitive in terms of what each of those concepts means.

When commonsense segmentation is conducted, it is typically obvious from the start which the most attractive target market will be. If profitability is used as the segmentation criterion, the high profit potential segment is undoubtedly the most attractive and should be chosen as the target segment.

When data-driven market segmentation is conducted, however, the decision which market segment to choose as the target segment is not at all obvious. The state of the art approach to data-driven market segmentation involves the following steps (Wedel and Kamakura 2000; Lilien and Rangaswamy 2003): First, a managerial decision is made about which set of variables will be used as segmentation variables. Second, these segmentation variables are collected, frequently by means of a survey study. Third, the empirical data forms the basis of extracting market segments. A wide range of distance- or model-based methods is available to achieve this. At this stage, it is common that segmentation solutions for a range of numbers of segments are calculated to determine which of these global market segmentation solutions (each containing multiple segments) performs best on statistical criteria. The best performing solution is selected. Next, all the market segments contained in this particular market segmentation solution are described in detail using both the segmentation variables and additional descriptive variables. Finally, based on this information, a target segment is selected using criteria such as how similar segment members are to

one another, how distinct the segment is with respect to the segmentation variables, whether it is large enough, whether it matches the firm's strengths, whether it is identifiable, and whether it can be reached with the tools of the marketing mix (Wedel and Kamakura 2000; Lilien and Rangaswamy 2003; McDonald and Dunbar 2012).

This state of the art approach is prone to making one critical mistake: selecting a *global* market segmentation solution which does not contain the most attractive individual segment or segments. This can happen because the statistical criteria used to select the *global* market segmentation solution are not aimed at identifying the most attractive individual market segments contained in the global solution.

This problem is exacerbated by the fact that data-driven market segmentation analysis—irrespective of the algorithm used—leads to different results if repeated (Dolnicar and Leisch 2010). As a consequence, experts in multivariate analysis for marketing research recommend to analyze data more than once with more than one algorithm. Iacobucci (2013, p. 15), for example, explicitly suggests to “Choose one of these algorithms (and play with more than one).” This advice reflects the fact that data-driven market segmentation analysis—whether it is done using distance based methods such as cluster analysis or model-based methods such as finite mixture models—is essentially exploratory in nature: these methods are nothing more than sophisticated fishing rods. But they can make mistakes. This paper presents an approach that prevents the mistake of choosing a global market segmentation solution that does not contain any attractive market segments from happening. Two illustrations with consumer data show how the most attractive target segments would indeed have gone undetected if segment level criteria had not been inspected. To stick with the fishing analogy: the proposed segment level stability measures proposed here make fishing less random. They ensure that no baby fish get caught when really we are after a big fat salmon. To date, no other approach has been proposed that can achieve this aim.

2 Traditional global criteria

Different indices can be used to assess the goodness of fit of the global market segmentation solution. Most cluster algorithms using Euclidean distance try to optimize a functional of between- and within- cluster sum of squares (Everitt et al. 2011; Kaufman and Rousseeuw 1990). Let \mathbf{T} be the total scatter matrix of a data set of size n in p dimensions, that is, the covariance matrix multiplied by $n - 1$. Let \mathbf{W} be the within-cluster scatter matrix, and \mathbf{B} be the between-cluster scatter matrix, such that $\mathbf{T} = \mathbf{W} + \mathbf{B}$. Examples for target functions of cluster algorithms include trace and determinant of \mathbf{W} , or trace of $\mathbf{B}\mathbf{W}^{-1}$ (Everitt 1974). The most common target is the sum of squares within clusters $SSW = \text{trace}(\mathbf{W})$. Some also consider the sum of squares between clusters $SSB = \text{trace}(\mathbf{B})$. To choose a specific number of clusters, one can either search for an elbow in the within sum of squares criterion, or use more refined criteria. The seminal paper by Milligan and Cooper (1985) lists over a dozen indices which can be used. Most search for minima,

maxima, or elbows in functionals of the above, such as SSW/k (Ball and Hall 1965), $[(SSB/(k-1))/(SSW/(n-k))]$ (Calinski and Harabasz 1974), or $\log(SSB/SSW)$ (Hartigan 1975). Calinski-Harabasz performs best in the simulations by Milligan and Cooper (1985). There are more recent additions to the list, but the classic indices are still most popular. For model-based clustering procedures, information criteria such as AIC or BIC relate the goodness of fit measured by the likelihood to the number of estimated parameters, which are a function of the number of clusters (Fraley and Raftery 1998).

Recent simulation studies have compared cluster indices on large simulation designs. Most focus on correct identification of the number of clusters (Chiang and Mirkin 2010; Steinley and Brusco 2011), others on validity or stability of cluster solutions (Brock et al. 2008; Steinley 2008; Vinh et al. 2010). Another line of research explores using resampling methods to evaluate cluster stability and choosing the right number of clusters (Dudoit and Fridlyand 2002; Grün and Leisch 2004; Lange et al. 2004; Tibshirani and Walther 2005). Stability of segmentation solutions has been proposed as a key evaluation criterion (Breckenridge 1989, 2000; Dolnicar and Leisch 2010; Putler and Krider 2012) where high levels of stability are interpreted as indicative of existing data structure, be it actual cluster structure or any other kind of data structure which enables similar clusters to be identified across repeated computations. Solutions with high global stability are preferable.

The difficulty with all these traditional criteria is that they evaluate the *global* segmentation solution, thus potentially discarding segmentation solutions that are globally suboptimal, but may contain the single most interesting segment for a particular organization.

To assess the current practice of evaluating alternative segmentation solutions, twenty-nine applied data-driven market segmentation studies published after 2006 across different disciplines were reviewed (reference list available upon request). Because applied market segmentation studies conducted by organizations are not accessible, all applied segmentation studies published in academic journals in the last decade which could be found were included. The search was very wide and included studies segmenting wine customers, tourists, households, green consumers, generation Y females, shoppers, mothers, university students, primary care patients, smokers and entrepreneurs. Studies which conducted data-driven market segmentation were included. A detailed inspection of the methodology indicates that, in all studies, the segmentation solution was chosen based on a global assessment. A segment level assessment was never undertaken before one specific segmentation solution was chosen. Seventeen percent of studies based the decision on one single cluster analysis; 21 % ran a hierarchical analysis, used the dendrogram to chose the number of clusters and then ran one run of a partitioning algorithm; 45 % ran one computation for a range of cluster numbers; ten percent reran computations both across and within certain numbers of clusters; and seven percent provided insufficient explanation. Criteria mentioned for the selection of the global segmentation solution include *overall* interpretability of the solution, *overall* distinctness of segments contained, size of *all* segments contained or a combination of statistical criteria and visual inspection of *all* segment profiles.

3 Segment level criteria

We propose two new criteria: *segment level stability within solutions with the same number of segments* (SLS_W) and *segment level stability across solutions with different numbers of segments* (SLS_A). Both have in common that the entity being evaluated is the segment, not the global segmentation solution. The following analogy illustrates the key benefit derived from those new criteria: traditional measures assess the quality of a haystack (the global overall segmentation solution containing a number of segments). But, as we will demonstrate, the nicest haystack may not contain the sharpest needles. The benefit of the newly proposed evaluation criteria is that they enable the assessment of needles within haystacks and thus allow data analysts and managers to focus on what really matters: finding one or a small number of good individual target segments.

3.1 Segment level stability across solutions with different numbers of segments (SLS_A)

SLS_A measures the persistence of a segment reoccurring across segmentation solutions with different numbers of clusters. Higher SLS_A values point to a higher likelihood of the segment representing a natural as opposed to an artificially constructed market segment (Dolnicar and Leisch 2010). Segment solutions containing one or more segments with high SLS_A should not be discarded.

Let P_1, P_2, \dots, P_m be a series of m partitions with numbers of clusters $k_1 < k_2 < \dots < k_m$. SLS_A can be quantified using an entropy measure (Shannon 1948). Entropy is defined as $-\sum p_j \log p_j$ and can be interpreted as the uncertainty in a discrete probability distribution p_1, \dots, p_k . In our case, the p_j are the percentages of data points any given segment in P_{i+1} obtains from each segment in P_i . Segments that have high entropy recruit a large number of members from different segments from the segmentation solution with fewer segments. In order to get a standardized measure for segment stability, entropy values are standardized by dividing them by the maximum possible entropy that would occur in the case of equal distribution of all segment members across all old segments; that is, $-\sum (1/k) \log(1/k) = \log(k)$. The SLS_A measure of stability is

$$SLS_A = 1 - \frac{\sum p_j \log p_j}{\log k}$$

As such, the values lie between a minimum (undesirable) SLS_A value of 0 and a maximum (desirable) value of 1.

An SLS_A plot is shown in Fig. 1. Each vertical column of circles represents one global segmentation solution; circles represent individual segments contained in those solutions. Lines between circles illustrate how many segment members stay in the same segment when more segments are extracted. High SLS_A segments are depicted by a thick, dark blue connecting line across solutions and no other lines branching off. Note that plotting SLS_A requires segment number relabelling; a relabelling algorithm is provided in the Appendix. Low SLS_A segments have many thin,

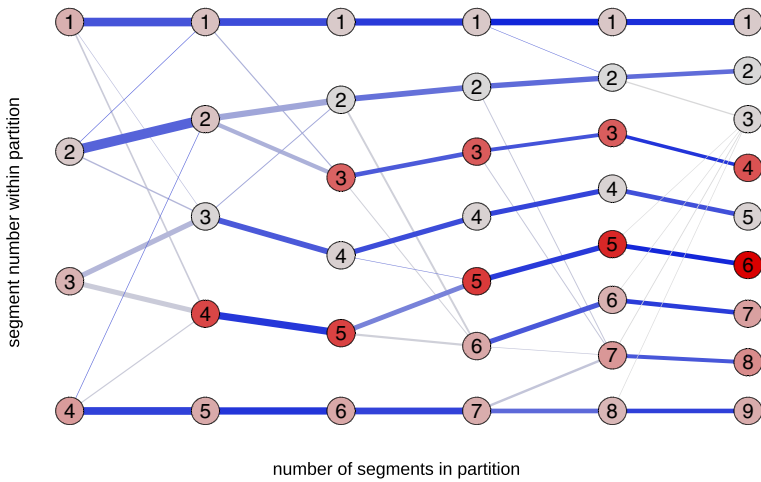


Fig. 1 Segment level stability for the guest survey data across solutions (SLS_A) with four to nine segments shown in columns, stability shown by the thickness of lines, and profitability reflected in the color of the nodes

light grey branches feeding into and running out of them. The thickness of the lines indicates the absolute number of segment members flowing into a segment; the color indicates the number of segments these members have been sourced from. Because one of the most critical selection criteria for a target segment in practice is that of segment profitability or future profit potential of a segment, the coloring of the circles in the SLS_A plot indicates the profitability or profit potential of each segment. The ability of SLS_A to identify naturally occurring market segments has been tested using artificial data sets with known structure.

3.2 Segment level stability within solutions with the same number of segments (SLS_W)

SLS_W measures how often—across multiple computations of the segmentation solution with the same number of clusters—a segment with the same key characteristics is identified. This criterion has been proposed as an evaluation criterion for *global* segmentation solutions (Dolnicar and Leisch 2010). Following Hennig (2007), we show how it can be applied at the segment level; technical details are provided in the Appendix. High SLS_W segments are attractive because they are likely to represent natural segments. Segmentation solutions containing high SLS_W segments should not be discarded.

To compute SLS_W , several bootstrap samples are drawn from the data set for each number of clusters of interest. Then, agreement between the original partition and each bootstrap partition is computed. Hennig (2007) defines maximum agreement as the stability of a segment in this bootstrap replica. SLS_W across all bootstrap replicates can be visualized using a boxplot as shown in Fig. 2. Maximum SLS_W is indicated by a horizontal line located at the top of the chart. Low SLS_W is shown by a low median reproducibility and/or a high level of dispersion around the median.

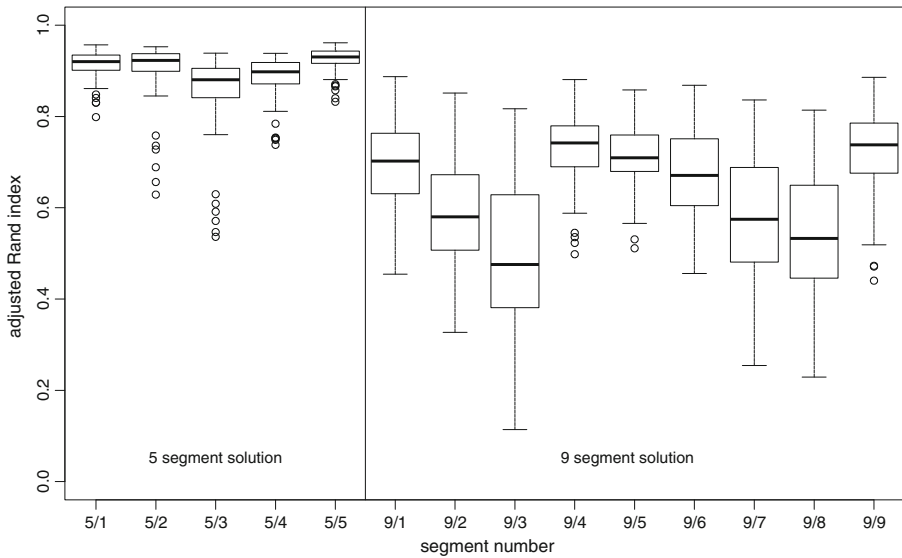


Fig. 2 Segment level stability (SLS_W) within the five and nine segment solutions for the guest survey data

The ability of SLS_W to identify naturally occurring market segments has been tested using artificial data sets with known structure.

4 Illustration with empirical data

4.1 Austrian national guest survey

Austrian National Guest Survey data from 1994 and 1997 from 11,378 tourists are used. The segmentation base contains 21 travel motives, such as “On holidays I want to rest and relax.” Answer options were as follows: applies to me greatly (1), mostly (2), slightly (3), and not at all (4). The motives are very distinct from each other, e.g., 14 principal components are needed to explain 80 % of the variance.

We compute segmentation solutions for between four and twenty segments using the k-means algorithm with 30 random starts for each number of clusters. All computations have been done in R (R Development Core Team 2016) using package flexclust (Leisch 2006). The traditional Calinski-Harabasz index gives no indication of a good choice of number of clusters at all; the index values decrease smoothly displaying no local minimum or elbow. The overall stability of the segmentation solutions for four to nine clusters leads to the conclusion that the five-segment solution should be chosen because it produces the highest median stability (0.84) across 100 replications.

Figure 1 plots SLS_A . Each column in this plot represents one segmentation solution ranging from four to nine clusters. At the far left is the four-cluster solution, at the far right the nine-cluster solution. As can be seen, segments 1 and 9 (in the

nine-cluster solution) have the highest SLS_A values, with their membership changing only marginally even when the number of segments doubles. Unfortunately, both these segments represent response styles rather than distinctly profiled segments and therefore cannot be considered attractive target segments. Response styles are systematic tendencies of responding to survey questions independent of question content which are consistent over time and across survey contexts (Paulhus 1991).

Using the segment numbering from the nine segment solution (right side of plot), segments 3, 7, and 8 have low SLS_A . They are artificially created as the number of segments increases. Segments 2, 4, 5, and 6 demonstrate reasonably high SLS_A . These segments emerge initially in the five- and six-segment solution and then reappear in all solutions containing higher numbers of segments. They are not response style segments and therefore represent potentially interesting target segments. Moving to the analysis of descriptor variables, segments 4 and 6 are of particular interest because they display the highest profitability (indicated by red color in the plot) as assessed by their daily expenditures during the holiday on which they have been surveyed. More specifically, members of segments 4 and 6 spend nearly twice as much money on holiday as the least profitable segment 2 (73 Euros and 80 Euros per person per day, respectively, as opposed to only 51 Euros in segment 2). This difference is highly statistically significant (Kruskal Wallis test p value $< 2.2e-16$). Members of segment 6 have another interesting feature: they are much more frequently than other tourists first time visitors to Austria. As such, they could be described as light users as opposed to heavy users who are regular visitors to Austria. Heavy versus light usage is a valuable variable in market segmentation. It can be used as the segmentation variable in commonsense segmentation resulting in light, medium, and heavy usage segments which can then be described in detail before targeting. They can also be used as descriptors in a data-driven market segmentation study as is the case in our example.

Figure 2 shows SLS_W for the travel motive data. The five segment solution is shown because it represents the segmentation solution recommended by the global stability criterion. The nine cluster solution is shown because it contains the segments with high SLS_A (segments 4 and 6). As can be seen from Fig. 4, all segments in the five-cluster solution are indeed highly stable, explaining why the global measure recommends this solution. However, this solution does not contain what later becomes the highly profitable, high SLS_A segment 4. Selecting this solution on the basis of global criteria would lead to segment 4 being irretrievably lost. Good haystack, no needle.

When inspecting the SLS_W for the nine-cluster solution, the two highly profitable, high SLS_A segments both emerge as high SLS_W as well. Segment 4 is the highest SLS_W segment (median 0.742), closely followed by segments 9 (0.738) and 5 (0.709). Segment 3 (which is not present in the five-cluster solution) has the lowest SLS_W . Segments 4, 5, and, to a lesser degree, 6, are not response-style segments, yet they display high SLS_W , thus warranting further assessment.

Who are segments 4 and 6? Segment 6 can be described as an adventure segment, and segment 4 as a health segment. This health segment is particularly interesting for the Austrian tourism context, which has many hot spring resorts and for which this segment is excellently suited. Members of this segment, as opposed to other

segments, are slightly older, spend significantly more money when on vacation and engage in a different set of activities: they are more into relaxing and swimming, less into hiking, sightseeing, going to museums, and riding bicycles. The health tourist segment represents an attractive target segment, which—using *overall* quality criteria for assessing segmentation solutions—would not have been detected.

4.2 Fast food restaurant image data

In this second illustration, image data about a fast food restaurant (Subway) which were collected in 2009 are used. A total of 1453 respondents assessed—using a binary response format—the following attributes: yummy, fattening, greasy, fast, cheap, tasty, expensive, healthy, disgusting, convenient, and spicy.

Segmentation solutions for three to nine segments using the k-means algorithm with 30 random starts for each number of clusters were generated. The traditional Calinski-Harabasz index again gives no indication of a good choice of number of clusters at all; the index values are almost constant over the whole range of clusters, any partition could be chosen. For this data set, assessing global segmentation solution stability leads to no recommendation whatsoever in terms of which solution to choose. The options are, then, to either randomly choose a solution or inspect segment level criteria.

Figure 3 plots segment level stability across numbers of clusters. To illustrate the versatility of this plot, the node color in this particular plot reflects the frequency of eating at subway (instead of profitability which was plotted for the guest survey data). This allows simultaneous inspection of stability across numbers of clusters and heavy versus light user segments. As can be seen in Fig. 3, segments 3, 6, and 9 in the right column (the nine cluster solution) are heavy users of subway. As can also be seen, segment 9 (bottom row in Fig. 3) is extremely stable across numbers of clusters. It

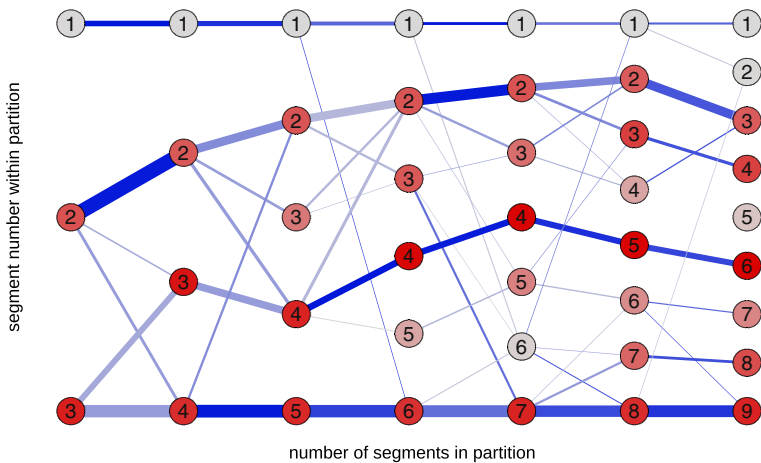


Fig. 3 Segment level stability for the fast food data across solutions (SLS_A) with three to nine segments shown in columns, stability shown by the thickness of lines, and user status (heavy versus light) reflected in the color of the nodes

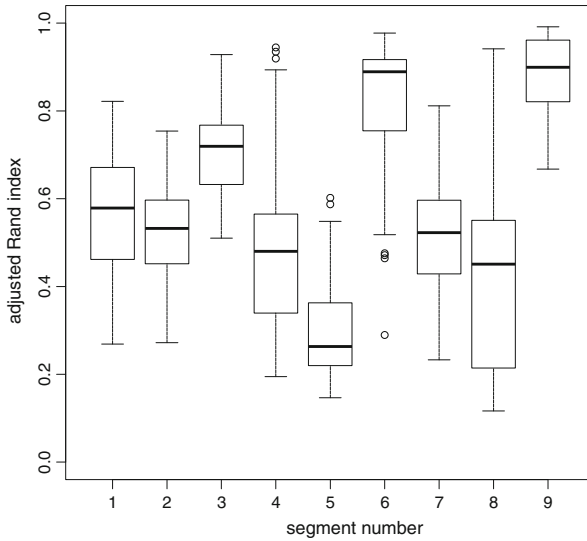


Fig. 4 Segment level stability (SLS_W) within the nine segment solution for the fast food data

first emerges in the four cluster solution and then remains virtually unchanged until the nine cluster solution. Segment 6 first emerges in the five cluster solution and also stays practically unchanged until the nine cluster solution. Based on the insights gained from Fig. 3, both market segments 6 and 9 represent very attractive candidates for target segments.

In terms of stability within cluster numbers, Fig. 4 shows that segments 6 and 9 also outperform all other segments in this criterion which both have a median stability level of 0.89.

Inspecting the profile of these segments reveals that members of segment 9 perceive Subway—more so than the other segments—as yummy, fast, cheap, tasty, healthy, and convenient, but *not as* fattening, greasy, expensive, disgusting, and *spicy*. Segment 6 largely shares this perception, with the one exception that they *do perceive Subway as spicy*. Both segments are attractive market segments. They could both be targeted in a differentiated marketing strategy. Alternatively, Subway could choose to position itself in a certain “spiciness position” and focus on one of those segments only.

Importantly, had the global four segment solution been chosen initially, segment 6 would have gone unnoticed. The ability to inspect segment level stability has been crucial in being able to detect its existence.

5 Conclusions

Market segmentation has greatly contributed to understanding consumer behavior in the marketplace (Roberts 2000) and has therefore been widely adopted by industry.

Despite its popularity, some aspects of segmentation analysis that may be statistically satisfactory do not provide optimal market insights for marketing managers. One such case is the selection of a market segmentation solution based on *global* criteria assessing an overall segmentation solution instead of assessing the segments contained therein. This is despite the fact that most organizations require only one or a small number of well-chosen, attractive (e.g., profitable) target segments to ensure survival and competitive advantage. The statistical properties of the global segmentation solution have the potential of distracting data analysts away from alternative solutions with worse global criteria values but containing individual highly attractive market segments. The excitement about a beautiful haystack may leave needles unnoticed.

This paper presents and demonstrates the usefulness of two new assessment criteria aiming at identifying segmentation solutions which contain interesting segments rather than being globally optimal. Interesting segments demonstrate high SLS_A (segment level stability across segmentation solutions with different numbers of clusters) and high SLS_W (segment level stability across repeated computations with the same number of clusters; reproducibility; replicability). The key advantage of the proposed criteria is that they protect segmentation solutions containing one or only a few very attractive market segments from being prematurely discarded. The two new criteria help data analysts to spot great needles in ugly haystacks.

The effectiveness of the proposed approach has been demonstrated using two empirical data sets. In both cases, using traditional selection criteria for market segmentation solutions would have failed in guiding the data analyst to select a good segmentation solution, either because traditional criteria resulted in no recommendation at all or because they resulted in a suboptimal recommendation. Using the two proposed segment level criteria allowed more detailed insights into the nature of segments emerging across different segmentation solutions and, in so doing, pointed to particularly attractive market segments. Pinpointing those segments made it possible to select a good market segmentation solution for further profiling and selection of one or more target segments.

Acknowledgements We thank the Australian Research Council for contributing to the funding of this study (ARC, DP110101347). We also thank our research assistants Alexander Chapple and Aaron Eden for their assistance with literature searches. Special thanks to Martin Natter, Bettina Grun, Dominik Ernst, Christina Yassouridis, Homa Hajibaba, and Nazila Babakhani for their feedback on previous versions of the manuscript.

Appendix: Technical Appendix

Relabelling algorithm required for the calculation of pertinaciousness

For series of partitions we propose a new relabelling algorithm which makes it possible to track segments over partitions with different numbers of clusters. Let again P_1, P_2, \dots, P_m be a series of m partitions with numbers of clusters $k_1 < k_2 < \dots < k_m$.

Algorithm 1 Relabel series of partitions

1. Relabel partition P_1 using any heuristic, for example, mean value of all variables for k-means clustering or prior probability for finite mixtures. Set $i = 1$.
2. Calculate the $k_i \times k_{i+1}$ contingency table T of cluster memberships in P_i and P_{i+1} , set column names of T to $1, \dots, k_{i+1}$.
3. Find the k_i clusters in P_{i+1} that have maximal agreement with the clustering P_i by solving a linear sum assignment problem (LSAP) using the Hungarian method (Papadimitriou and Steiglitz 1982) on T . Use the LSAP solution to permute the columns of T such that the main diagonal of the matrix in the first k_i columns is maximal. Permute the column names together with the columns.
4. Insert the remaining $k_{i+1} - k_i$ to the right of their best match in P_i (largest number of common data points). If the element in row j is the maximum, insert between columns j and $j + 1$.
5. Relabel cluster numbers in P_{i+1} according to the permuted column names of T .
6. Repeat from 2 until $i = m - 1$.

Note that if $k_{i+1} = k_i + 1$, then only one column needs to be inserted in step 4. However, the algorithm also works for the more general case.

Calculating segment-wise rerun stability**Algorithm 2** Segment-wise rerun stability (Hennig 2007)

1. Compute partition P with k clusters C_1, \dots, C_k of original data set \mathcal{D} . Set $i = 1$.
2. Draw bootstrap training sample \mathcal{T}^i of size n by sampling with replacement from the original data set.
3. Compute partition P^i with k clusters C_1^i, \dots, C_k^i of bootstrap training sample \mathcal{T}^i .
4. For each original cluster C_1, \dots, C_k compute the maximum agreement with a cluster in the bootstrap partition measured by the Jaccard coefficient:

$$s_j^i = \max_{1 \leq j' \leq k} \frac{|C_j \cap C_{j'}^i|}{|C_j \cup C_{j'}^i|}, \quad 1 \leq j \leq k$$

5. Repeat from 2 for a pre-specified number of iterations b .
6. Compute median or mean stability values for each cluster

$$s_j = \frac{1}{b} \sum_{i=1}^b s_j^i, \quad 1 \leq j \leq k$$

or look directly at distribution of s_j^i ; for example, using boxplots.

References

- Ball, G.H., & Hall, D.J. (1965). *ISODATA, a novel method of data analysis and pattern classification* (Tech. Rep. NTIS No. AD 699616). Menlo Park, CA: Stanford Research Institute.
- Breckenridge, J.N. (1989). Replicating cluster analysis: method, consistency, and validity. *Multivariate Behavioral Research*, 24(2), 147–161.
- Breckenridge, J.N. (2000). Validating cluster analysis: consistent replication and symmetry. *Multivariate Behavioral Research*, 35(2), 261–285.
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). Clvalid: an r package for cluster validation. *Journal of Statistical Software*, 25(4), 1–22.
- Calinski, R.B., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3, 1–27.
- 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, 3–40.
- Dolnicar, S. (2004). Beyond “Commonsense segmentation” – a systematics of segmentation approaches in tourism. *Journal of Travel Research*, 42(3), 244–250.
- Dolnicar, S., & Leisch, F. (2010). Evaluation of structure and reproducibility of cluster solutions using the bootstrap. *Marketing Letters*, 21(1), 83–101.
- Dudoit, S., & Fridlyand, J. (2002). A prediction-based resampling method to estimate the number of clusters in a data set. *Genome Biology*, 3(7), 1–21.
- Everitt, B.S. (1974). *Cluster analysis*. London: Heinemann Educational Books.
- Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*, 5th edn. Chichester: Wiley.
- Fraley, C., & Raftery, A.E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41(8), 578–588.
- Grün, B., & Leisch, F. (2004). Bootstrapping finite mixture models. In J. Antoch (Ed.), *COMPSTAT 2004* (pp. 1115–22). Heidelberg: Physica.
- Hartigan, J.A. (1975). *Clustering algorithms*. New York, NY: Wiley.
- Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, 52, 258–271.
- Iacobucci, D. (2013). *Marketing models: multivariate statistics and marketing analytics*. Mason, OH: South-Western.
- Kaufman, L., & Rousseeuw, P.J. (1990). *Finding groups in data*. New York: Wiley.
- Lange, T., Roth, V., Braun, M.L., & Buhman, J.M. (2004). Stability-based validation of clustering solutions. *Neural Computation*, 16(6), 1299–323.
- Leisch, F. (2006). A toolbox for K-Centroids cluster analysis. *Computational Statistics and Data Analysis*, 51(2), 526–544.
- Lilien, G.L., & Rangaswamy, A. (2003). *Marketing engineering*, 2nd edn. Upper Saddle River: Pearson Education.
- Mazanec, J.A. (2000). Market segmentation. In J. Jafari (Ed.), *Encyclopedia of tourism*. London: Routledge.
- McDonald, M., & Dunbar, I. (2012). *Market segmentation: how to do it and how to profit from it*, 4th edn. Wiley: Hoboken.
- Milligan, G.W., & Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159–79.
- Myers, J.H., & Tauber, E. (1977). *Market structure analysis*. Chicago: American Marketing Association.
- Papadimitriou, C., & Steiglitz, K. (1982). *Combinatorial optimization: algorithms and complexity*. Prentice Hall: Englewood Cliffs.
- Paulhus, D.L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.
- Putler, D.S., & Krider, R.E. (2012). *Customer and business analytics: applied data mining for business decision making using R*. London: Chapman&Hall/CRC.
- R Development Core Team (2016). R: a language and environment for statistical computing r foundation for statistical computing. Vienna, Austria.

- Roberts, J. (2000). The intersection of modelling potential and practice. *International Journal of Research in Marketing*, *17*, 127–134.
- Shannon, C.E. (1948). A mathematical theory of communication. *The Bell Systems Technical Journal*, *27*, 379–423.
- Steinley, D. (2008). Stability analysis in K-Means clustering. *British Journal of Mathematical and Statistical Psychology*, *61*, 255–273.
- Steinley, D., & Brusco, M.J. (2011). Choosing the number of clusters in K-Means clustering. *Psychological Methods*, *16*(3), 285–297.
- Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, *14*(3), 511–28.
- Vinh, N.X., Epps, J., & Bailey, J. (2010). Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, *11*, 2837–2854.
- Wedel, M., & Kamakura, W.A. (2000). *Market segmentation conceptual and methodological foundations*, 2nd edn. Boston: Kluwer Academic Publishers.