2004

Tracking data-driven market segments

Sara Dolnicar
University of Wollongong, s.dolnicar@uq.edu.au

Publication Details

This article was originally published as: Dolnicar, S, Tracking Data-Driven Market Segments, Tourism Analysis, 2004, 8(2-4), 227-232. The journal homepage can be found here.

Tracking data-driven market segments

Abstract
Market segmentation has become a standard concept in tourism marketing. A priori and a posteriori (data-driven) segmentation approaches enjoy high popularity among both practitioners and researchers. In order to optimize the market segmentation strategy it is not only necessary to identify relevant market segments, describe them, evaluate the match between corporate or destination strengths and segment needs but also, to determine how segments develop over time. This knowledge is typically accounted for when a priori segments are used. In the case of a posteriori segments, however, such trend tracking is neglected. In this paper a tracking framework is presented that allows testing of a posteriori segment developments over time on the basis of identical consecutive guest surveys. The framework is flexible with regard to methods applied at each step and – through validation of explorative findings by means of repetition – allows insight into market structure from multiple perspectives.

Keywords
market segmentation, market structure analysis, change monitoring

Publication Details
This article was originally published as: Dolnicar, S, Tracking Data-Driven Market Segments, Tourism Analysis, 2004, 8(2-4), 227-232. The journal homepage can be found here.

Tracking data-driven market segments

Sara Dolničar[1]

School of Management, Marketing & Employment Relations, University of Wollongong,

Wollongong, NSW 2522 Australia, sara_dolnicar@uow.edu.au

Abstract

Market segmentation has become a standard concept in tourism marketing. *A priori* and *a posteriori* (data-driven) segmentation approaches enjoy high popularity among both practitioners and researchers. In order to optimize the market segmentation strategy it is not only necessary to identify relevant market segments, describe them, evaluate the match between corporate or destination strengths and segment needs but also, to determine how segments develop over time. This knowledge is typically accounted for when *a priori* segments are used. In the case of *a posteriori* segments, however, such trend tracking is neglected.

In this paper a tracking framework is presented that allows testing of *a posteriori* segment developments over time on the basis of identical consecutive guest surveys. The framework is flexible with regard to methods applied at each step and – through validation of explorative findings by means of repetition – allows insight into market structure from multiple perspectives.

Keywords:

market segmentation, market structure analysis, change monitoring

---

[1] This study was partially conducted during Sara Dolničar's appointment as assistant professor at the Institute for Tourism and Leisure Studies at the Vienna University of Economics and Business Administration and supported by the FWF grant 010 (Adaptive Information Systems and Modeling in Economics and Management Science).

Market segmentation has become a standard concept in strategic marketing. Alongside the wide use of *a priori* (Mazanec, 2000) segmentation, data-driven (or *a posteriori*, Mazanec, 2000) approaches have increasingly gained popularity among researchers attempting to derive segments on the basis of survey information using various bases, as e.g. behavioural or psychometric variables (Middleton, 1988; Myers, 1996; Smith, 1995). While it is common to monitor the development of *a priori* segments over time (e.g. geographical segments), data-driven approaches are typically conducted at one point of time only (Baumann, 2000). This might be due to the fact that data-driven segmentation *per se* is an exploratory concept and studying multiple time periods in such a manner (when even the one period case is tricky: Baumann, 2000; Dolničar, 2002; Ketchen & Shook, 1996; Punj & Stewart, 1983) would further increase complexity and decrease reliability of results.

The advantages of tracing *a posteriori* segment trends in the marketplace include (1) validation of single data-driven segment solutions that are used for marketing planning, (2) increased insight into the changes encountered in the marketplace, (3) provision of a sound basis for forecasting, (4) the possibility of regular evaluation to determine whether the segment(s) targeted should be held on to or switched, (5) only temporary reduction of dimensionality during the segment assignment procedure, and (6) applicability to typical multi-period data in tourism (non-panel format).

The aim of this article is to suggest a framework for tracking *a posteriori* segments over years based on consecutive guest surveys utilising the same sampling method and questionnaire.

The tracking framework

The tracking framework suggested is a stepwise process including seven steps.

1. Definition of the anchor period: The exploratory nature of *a posteriori* segmentation is accounted for by defining an anchor period for analysis, which serves as a starting point for investigation. If data for all periods is available and the tracking framework is used for *ex post* investigation only, the definition of an anchor period would not be necessary. If, however, ongoing monitoring is aimed at, the anchor period choice is inevitable.

2. Computation of a data-driven market segmentation solution: Using the guest survey data set from the anchor period, a segment solution is derived. Any appropriate partitioning method can be used. The result is a number of segments with membership assignments for respondents.

3. Characterisation of market segments: Tourist groups are described using answers to questions chosen as the segmentation base (Wedel & Kamakura, 1998). Furthermore, relevant background variables (demographic, socioeconomic, behavioural etc.) are able to be studied. For example, using discriminant analysis at this stage can help to validate the existence of heterogeneous segments with regard to the background information.

4. Assignment of data from other periods to the anchor segment solution: Using the centroids determined in step 2, the answer patterns of the respondents from the remaining periods of time are assigned to the most similar centroids. At this stage, frequency distributions of segment assignments for each period can be investigated.

5. Testing of distribution changes: These distributions are compared applying Chi square tests to contingency tables, including year and cluster membership. Bonferroni correction of significance values is necessary if more that two periods are studied. At this stage it is possible to determine whether there are any significant trends in *a posteriori* segments over time.

6. (Testing of changes in background variables): In addition, qualitative changes can be studied by investigating differences in background variables of the segments over years.

7. Validation of results: Validation is of utmost importance due to the exploratory nature of data-driven segmentation that can potentially render a million different solutions. By including multiple periods of time, another dimension of possible influence is included, making it even more dangerous to base interpretation on a single run of analysis. Repetition is a useful approach for validating results. It can be conducted with different numbers of clusters, algorithms and anchor years. By comparing solutions and time changes a picture emerges from the exploratory approach that allows conclusions regarding the reliability of findings.

The tracking framework is flexible in many respects: First, any kind of data can be used that is appropriate for traditional *a posteriori* segmentation (multi-dimensional data on demographics, socioeconomics, tourist behaviour, benefits sought, etc.). Second, the choice of the anchor year allows a wide variety of explorative approaches (as well as ongoing monitoring). Insights into segment structure development can thus be gained using various perspectives. Third, any algorithm that results in a partitioning of answer patterns of respondents can be applied, as long as each respondent is assigned to a segment deterministically. Finally, any background variables can be chosen for validation and detailed description of the results.

Tracking activity based tourist segments in Austria – an illustration

Austrian National Guest Survey data from the summer seasons 1994 and 1997 are used. Sample sizes amount to 7967 for the year 1994 and 6604 for 1997. Respondents were asked to state which vacation activities they engage in. In the data set, "1" indicates that the activity was undertaken sometimes or often, whereas a "0" indicates both the fact that a respondent indicated that they do not undertake that pastime or that the question was not answered. Hence, the data set is binary and includes 14571 respondents who answered with regard to 22 vacation activities. In addition, a number of background variables are available, but these are not analysed because the segmentation base renders sufficiently illustrative results in this empirical example.

1994 is chosen as the anchor period. A self organizing feature map (SOFM, Kohonen, 1984, for segmentation applications see Mazanec, 1994 and 1999, Dolničar, 1997) functions as partitioning algorithm. A map with six prototypes is used (3 columns, 2 rows), the starting points for the prototypes are chosen by randomly drawing 100 points and picking the best solution with regard to the criterion of maximum between-segment variance and minimum inner-segment variance. The data is presented to the network 200 times for learning purposes, with decreasing adaptation to the answer pattern by both the most similar prototype to the answer pattern and the neighbouring prototypes in the grid. After the learning phase, in which the SOFM adapts prototype values to best

mirror the data, each data vector is presented to the SOFM once more, with each respondent assigned to the prototype best representing his or her vacation activity answer pattern.

The resulting segment profiles are provided in Figure 1, arranged to mirror the SOFM grid. Each chart describes one segment. The bars indicate average agreement of the segment members, the line shows the total sample average. Segment # 1 can thus be described as “culture tourists”. They state that they have engaged in cultural activities more often than the average summer tourist, especially as their activity level with regard to “going to concerts”, “sightseeing”, “going out in the evening”, “shopping”, “going to the theatre”, “going to museums and exhibitions” and “spending the evening at a Heurigen (a typical Viennese restaurant)” is above average. Segment # 2 is less distinct, showing interest in cultural activities and sports, segment # 3 is clearly sports-centred and segment # 5 is composed of the relaxed summer tourists who spend the days hiking and going for walks. The remaining two segments are not interpreted, as it is not clear which proportion of the respondents are “active in all respects” or “not interested in any activities” as opposed to being mere answer tendencies.

---------- FIGURE 1 ----------

In order to investigate changes over years a contingency table (Table 1) is constructed. The Pearson Chi-square renders significant outcome at the 99.9% level. The major trends from 1994 to 1997 are (1) an increase of the sports segment # 3 and in hiking tourists (# 5) and a dramatic decrease in respondents stating either that they engage in all activities or have a positive answer tendency.

---------- TABLE 1 ----------

The results are validated by rerunning the entire process for 10 segments with a different algorithm (topology representing networks as introduced by Martinetz & Schulten, 1994) that functions in a manner similar to SOFMs but does not force the data into the predefined grid. Table 2 illustrates the relationship between the solutions. The culture segment is split up in two segments TRN # 2 and TRN # 5 (cross-sections are pointed out in Table 2 with black frames), the main difference between these segments being that no member of TRN # 2 participates in organized excursions, whereas all

TRN # 5 members do. Similarly, the sports-oriented segment is split up in TRN # 3 and TRN # 6 with “biking” representing the major discriminating variable. The hiking tourists remain very stable (matching TRN # 10) although the number of segments was almost doubled. The same is true for segment # 4: the potential negative answer tendency segment is represented by prototype TRN #10. The remaining segments are difficult to interpret. The cross-tabulation (significant Chi square test at the 99.9% level) shows that segments 2 and 6 in the SOFM solution are both split up among the four new segments. The TRN solution backs the existence of segments # 1, # 3, # 4 and # 5 from the SOFM solution, as well as supporting the fact that the remaining group of tourists is not easily segmented in terms of vacation activities, as no stable representation can be arrived at.

---------- TABLE 2 ----------

From the change tracking perspective over consecutive survey years, results from the six-segment solution are supported with regard to the increase in the sports activity segment, the hiking segment and the decrease of segment SOFM # 6 (TRN # 4, # 7 and # 8).

All in all, the validation of the initial six segment solution by repeating the tracking process for a 10 segment solution with a different partitioning algorithm revealed which of the findings based on the SOFM analysis can be built on for planning marketing action and which findings have to be used with caution: The “culture segment”, the “sports segment”, the “hiking segment” and the group that engaged in very few vacation activities engaged in (or the negative answer tendencies group) can be revealed in a stable manner and the same market trends concerning these groups of tourists result from both investigations. The remaining segments are not identified in a stable manner and therefore should not be chosen as target segments without further investigation.

Conclusions

A tracking framework for *a posteriori* segments is proposed. The stepwise procedure includes (1) definition of the anchor period, (2) computation of a data-driven segmentation using anchor period data, (3) characterisation of segments, (4) assignment of data from other periods to the anchor solution, (5) testing of distribution changes, (6) testing of changes in background variables, and (7)

validation of results. The framework is flexible with regard to techniques used in the single steps. However, multiple-period survey data based on the same questionnaire and sampling strategy is required. The advantages of being able to track *a posteriori* segment trends in the marketplace include (1) validation of single data-driven segment solutions, (2) increased insight into the changes encountered in the marketplace, (3) provision of a stronger forecasting basis, (4) the possibility of continuous evaluation of the segmentation strategy, (5) only temporary dimensionality reduction during partitioning (no compression of the item information), and (6) applicability to typical multi-period data in tourism (no panel format).

The main limitation is insecurity arising from sampling at consecutive periods of time. It cannot be assumed that multiple surveys based on representative samples are sufficient to exclude distortion through intervening effects. This is especially crucial when surveys rely strongly on weighting. However, this limitation is not caused by the tracking framework but affects any analysis of guest survey data from multiple periods. Future work looks at extending the tracking tool to the three-way data situation (building on the PBMS concept introduced by Dolničar, Grabler & Mazanec, 1999, and comprehensively described in Mazanec & Strasser, 2000).

References

Baumann, R. (2000). *Marktsegmentierung in den Sozial- und Wirtschaftswissenschaften: eine Metaanalyse der Zielsetzungen und Zugänge*. Diploma thesis at Vienna University of Economics and Management Science. Vienna.

Dolničar, S. (1997). *Erstellung einer psychographischen Taxonomie von Osterreich-Urlaubern unter Anwendung neuronaler Netzwerkverfahren*. Vienna: Service Fachverlag.

Dolničar, S. (2002). Review of Data-Driven Market Segmentation in Tourism, *Journal of Travel & Tourism Marketing,* 12, 1-22.

Dolničar, S., Grabler, K., Mazanec, J. A. (1999). A Tale of Three Cities: Perceptual Charting for Analyzing Destination Images, in: A.G.Woodside et al. (eds.), *Consumer Psychology of Tourism*, Hospitality, and Leisure, New York: CAB International, 39-62.

Ketchen D. J. jr. & C.L. Shook (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal,* 17, 441-458.

Kohonen, T. (1984). *Self Organisation and Associative Memory.* New York: Springer-Verlag.

Mazanec, J. and Strasser, H. (2000). *A Nonparametric Approach to Perceptions-Based Market Segmentation: Foundations.* Springer: Berlin.

Mazanec, J. (1994). Image Measurement with Self-Organizing Maps: A Tentative Application to Austrian Tour Operators. *Revue de Tourisme,* 49, 9-18.

Mazanec, J. (1999). Simultaneous Positioning and Segmentation Analysis with Topologically Ordered Feature Maps: A Tour Operator Example. *Journal of Retailing and Consumer Services* (Special Issue on Marketing Applications of Neural Networks), 6, 219-235.

Mazanec, J. (2000). Market Segmentation. In: Jafari, J. (ed.), *Encyclopedia of Tourism*. London: Routledge.

Middleton, V.T.C. (1988). *Marketing in Travel & Tourism*. Oxford: Heinemann.

Myers, J.H. (1996). *Segmentation and Positioning for Strategic Marketing Decisions*. Chicago: American Marketing Association.

Punj, G. & D.W. Stewart (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research,* 20, 134-148.

Smith, S.L.J. (1995). *Tourism Analysis*. Harlow: Longman.

Wedel, M. & W. Kamakura (1998) *Market Segmentation - Conceptual and Methodological Foundations*. Boston: Kluwer Academic Publishers.

Figure 1: Segment profiles based on a 6 prototype SOFM solution for 1994

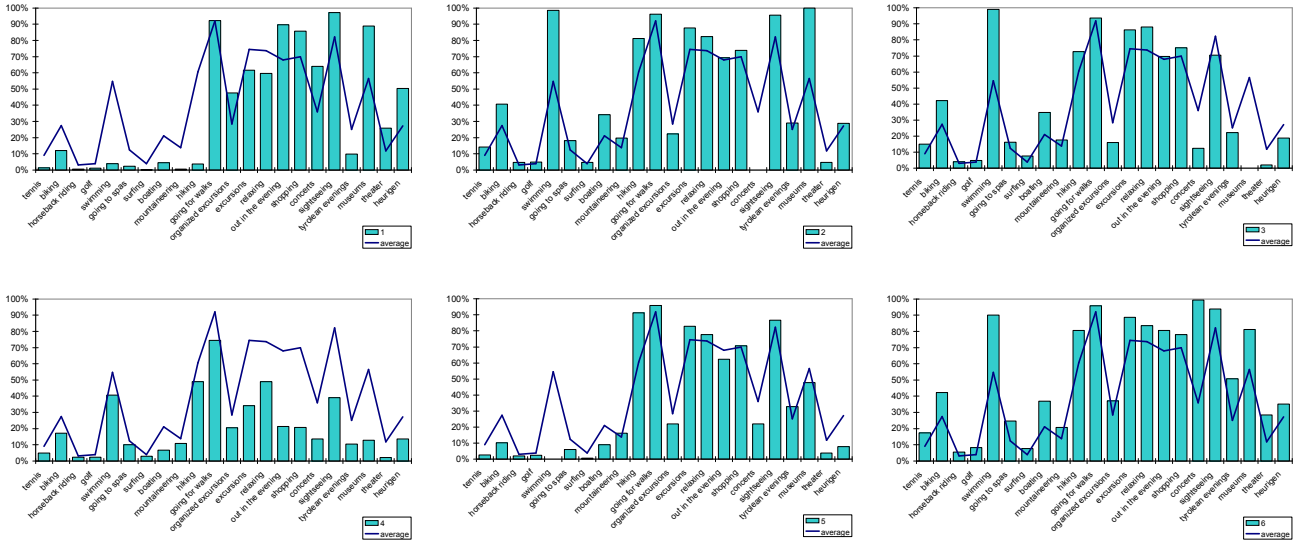

Table 1: Contingency table for segment size comparison 1994 and 1997 (SOFM based)

Segment		**1994**	**1997**	***Total***
1 (culture)	number of respondents	1672	1337	*3009*
	% within the year	21%	20%	*21%*
2 (culture & sports)	number of respondents	1321	1116	*2437*
	% within the year	17%	17%	*17%*
3 (sports)	number of respondents	1499	1474	*2973*
	% within the year	**19%**	**22%**	*20%*
4 (nothing, answer tend.)	number of respondents	1034	848	*1882*
	% within the year	13%	13%	*13%*
5 (hiking)	Number of respondents	1241	1233	*2474*
	% within the year	**16%**	**19%**	*17%*
6 (all, answer tend.)	number of respondents	1200	596	*1796*
	% within the year	**15%**	**9%**	*12%*
Total	*number of respondents*	*7967*	*6604*	*14571*
	% within the year	*100%*	*100%*	*100%*

Table 2: Crosstabulation of the 6 segment SOFM and the 10 segment TRN segment solution

			TRN solution with 10 segments										
			1	2	3	4	5	6	7	8	9	10	tot.
SOFM with 6 segm.	1	# members		893		16	522	3		1			*1435*
		% SOFM		62%		1%	36%						
	2	# members	874	10		131	8	3	1311	151	8	1	*2497*
		% SOFM	35%			5%			53%	6%			
	3	# members	1		2116	51	4	1495		69	12	5	*3753*
		% SOFM			56%	1%		40%		2%			
	4	# members		20	249	22	4	106	19		91	1724	*2235*
		% SOFM		1%	11%	1%		5%	1%		4%	77%	
	5	# members	36	69	1	412	31	60		9	2355	88	*3061*
		% SOFM	1%	2%		13%	1%	2%			77%	3%	
	6	# members	244	26	33	574	18	80	246	531			*1752*
		% SOFM	14%	1%	2%	33%	1%	5%	14%	30%			
	tot.	*# members*	*1155*	*1018*	*2399*	*1206*	*587*	*1747*	*1576*	*761*	*2466*	*1818*	*14733*