January 2004

Towards more thorough data-driven segmentation in tourism - a tracking framework for exploring segment development

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

Towards more thorough data-driven segmentation in tourism - a tracking framework for exploring segment development

Abstract

Market segmentation has become a standard concept in tourism marketing. A priori and a posteriori (data-driven, post-hoc) segmentation approaches enjoy high popularity among both practitioners and researchers. In order to optimise 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 to understand 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 simple tracking framework is presented that allows testing of a posteriori segment developments over time on the basis of identical consecutive guest surveys. It comprises the following steps: (1) definition of the anchor period, (2) computation of a data-driven market segmentation solution, (3) characterisation of market segments, (4) assignment of data from other periods to the anchor segment 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 methods applied at each step and – through validation of explorative findings by means of repetition – allows insight into market structure from multiple perspectives.

Disciplines

Business | Social and Behavioral Sciences

Publication Details

This chapter was originally published as: Dolnicar, S, Towards more thorough data-driven segmentation in tourism - a tracking framework for exploring segment development, in Crouch, GI, Perdue, RR, Timmermans, HJP, et al, (eds.), Consumer Psychology of Tourism, Hospitality and Leisure, 2004, CAB International, New York, 245-252. Book information available here.

Towards more thorough data-driven segmentation in tourism

-

a tracking framework for exploring segment development

Sara Dolničar

School of Management, Marketing & Employment Relations

University of Wollongong, Wollongong, NSW 2522 Australia

Telephone: (+61 2) 4221 3862, Fax: (+61 2) 4221 4154

sara_dolnicar@uow.edu.au

Abstract

Market segmentation has become a standard concept in tourism marketing. *A priori* and *a posteriori* (data-driven, post-hoc) segmentation approaches enjoy high popularity among both practitioners and researchers. In order to optimise 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 to understand 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 simple tracking framework is presented that allows testing of *a posteriori* segment developments over time on the basis of identical consecutive guest surveys. It comprises the following steps: (1) definition of the anchor period, (2) computation of a data-driven market segmentation solution, (3) characterisation of market segments, (4) assignment of data from other periods to the anchor segment 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 methods applied at each step and – through validation of explorative findings by means of repetition – allows insight into market structure from multiple perspectives.

Introduction

Market segmentation has become a standard concept in strategic marketing. Alongside the wide use of *a priori* (Mazanec, 2000) segmentation approaches, splitting individuals on the basis of predefined criteria, data-driven (or *a posteriori*, Mazanec 2000) approaches have gained increased popularity in the past decades among researchers attempting to derive market segments on the basis of survey information using various segmentation bases, as e.g. behavioural or psychometric information (Lilien & Rangaswamy, 1998; Middleton, 1988; Myers, 1996; Smith, 1995). While it is common to monitor the development of *a priori* segments over time (e.g. development of tourism from certain countries of origin, different age groups, families etc.), data-driven approaches are typically conducted at one point of time only. This might be due to the fact that data-driven segmentation *per se* is an exploratory concept and studying multiple time periods in such an exploratory manner (when even the one period case is full of possible pitfalls, Baumann, 2000; Ketchen & Shook, 1996; Punj & Stewart, 1983) would further increase complexity and opacity and decrease reliability of results. The fact, however, that tracking *a posteriori* segments is extremely uncommon is strongly supported by the literature survey in the field of business studies in general conducted by Baumann (2000) and the subsequent analysis limited to tourism segmentation studies (Dolničar, 2002). Among 47 data-driven segmentation publications within the field of tourism not a single study reports on investigations over time.

The advantages of being able to trace *a posteriori* segment trends in the marketplace include (1) validation of single data-driven segment solutions that are used to build an entire marketing plan, (2) increased insight into the changes encountered in the marketplace, (3) provision of a sound basis for forecasting, (4) the possibility of regular evaluation as to whether the segment(s) targeted should be held on to or switched, (5) 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 paper is to suggest a framework for tracking *a posteriori* market segment trends over years based on guest surveys conducted at different points in time based on the same sampling method and the same questionnaire.

The tracking framework

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

1. **Definition of the anchor period**

   Data-driven market segmentation is no deterministic concept. The typically explorative nature of *a posteriori* segmentation has to be accounted for in this concept by choosing an anchor period for analysis. The anchor period is used as the starting point of the investigation. Any period of survey data available can be utilised: using the first period allows statements about the development that has taken place to date, while using the last period gives insight into how the present market situation has developed. Clearly, if data for all periods is available and the tracking framework is used for explorative *ex post* investigation only, the definition of an anchor period would not be necessary. If, however, ongoing monitoring is intended, the anchor period choice is inevitable.

2. **Computation of a data-driven market segmentation solution using data from the anchor period**

   Using the guest survey data set from the anchor period, a segment solution is derived. Any method that partitions the multidimensional data set in an appropriate manner can be used at this stage. The result is a number of market segments with each respondent assigned membership to one of the segments.

3. **Characterisation of market segments**

   Based on the answers provided by segment members to the segmentation base (Wedel & Kamakura, 1998), the groups of tourists are described in detail. Furthermore, relevant background variables (demographic, socio-economic, behavioural etc.) are studied for each segment. Using discriminant analysis at this stage, for example, can help to validate the existence of heterogeneous segments with regard to this background information.

4. **Assignment of data from other periods to the anchor segment solution**

   The guest survey data from the remaining periods is matched with the segment solution derived from the anchor year. This is achieved by extracting the centroids from the anchor solution; these function as representants or prototypes for the segments. The answer patterns of respondents from the remaining periods are then assigned to the closest corresponding prototypes. The result are frequency distributions of segment assignments for each period of time.

5. **Testing of distribution changes**

   The distributions of respondents over segments are compared over the various time periods applying Chi square tests based on contingency tables, including year and cluster membership information. 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 for the same segments over years.

7. **Validation of results**

Validation of this stepwise procedure is of utmost importance because data-driven segmentation is an exploratory tool by itself and potentially renders a million different solutions, one of which is then chosen at random or by comparing the usefulness of different solutions. By including multiple periods of time, another dimension of possible influence is included and this makes it even more dangerous to base the entire market structure interpretation on one single run of analysis. Basically, repetition is a useful tool. Repetition can be conducted with different numbers of clusters, different algorithms or different anchor years. By comparing solutions and time changes a picture emerges from the exploratory approach that allows conclusions to be made about the reliability of findings.

This stepwise 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, socio-economic, tourist behaviour, benefits sought etc.). Second, the choice of the anchor year allows a wide variety of explorative approaches (as well as ongoing monitoring of *a posteriori* segment development). Insights into segment structure development can thus be derived from 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.

In addition to the fundamental aim of this procedure, it might as well be used to determine the optimal segment solution in the first place. For this purpose it is recommended – as in the case of validation - to perform the entire procedure a number of times, with changing anchor years, varying partitioning algorithms and numbers of segments. Solutions with the highest stability would be favoured, unless distinct density segments can be identified in the data (Dolničar & Leisch, 2001).

Tracking development of activity based tourist segments in Austria – an illustration

Austrian National Guest Survey data from the summer seasons 1994 and 1997 are used to illustrate the tracking framework suggested. An activity-based segmentation is constructed with the first survey used as anchor year. The sample sizes amount to 7967 for the year 1994 and 6604 for 1997. Respondents were asked to state which leisure activities they engage in during their vacation. In the data used, "1" indicates that the activity was undertaken sometimes or often, whereas a "0" indicates either the fact that a respondent did not undertake that particular pastime or that he or she has not answered the question. Hence, the data set used is in binary format and includes 14571 respondents. Answers provided by each of the respondents with regard to 22 vacation activities are used as the basis of segmentation. In addition, a number of background variables are available, but the analysis of background variables is omitted in this empirical illustration, as inspection of the segmentation base renders sufficiently illustrative results.

Data from the first period of time is chosen as the anchor period. A self-organizing feature map functions as partitioning algorithm (SOFM, Kohonen, 1984, for applications in the field of *a posteriori* segmentation of guest surveys, see Mazanec, 1994 and 1999, Dolničar, 1997). A map with six prototypes is used (3 columns, 2 rows); starting points for the prototypes are chosen by drawing 100 points at random and picking the best solution as determined by the criterion of maximum between-segment variance and minimum inner-segment variance. The data set is presented to the neural network 200 times for learning purposes, with a decreasing amount of adaptation of both the prototype most similar to the answer pattern presented and the neighbouring prototypes in the grid. After this learning phase, in which the SOFM adapts prototype values to best mirror the data at hand both in terms of representation by six segments as well as with regard to topological arrangement along the grid, each data vector is

presented to the SOFM one more time, each respondent being assigned to the prototype best representing his or her vacation activity answer pattern.

The resulting segment profiles are provided in Figure 1 in an arrangement mirroring the SOFM grid. Each profile chart characterizes one of the segments that emerged from the partitioning step. The bars indicate the average tendency of the segment members to undertake the vacation activities listed and the line shows the total sample average. Segment # 1 could thus be described as “culture tourists”. Members of this group state that they have engaged in cultural activities of various kinds more often than the average summer tourist in Austria, especially the level of “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 (this is a typical Viennese restaurant)”. Segment # 2 is less distinct, showing interest in both cultural activities as well as sports. Segment # 3 is clearly sports-centred and segment # 5 represents the typical relaxed summer tourist who spends the days hiking and going for walks. The remaining two segments are not interpreted, as it cannot be validated which proportion of the respondents are “active in all respects” or “not interested in any activities”, as opposed to being mere answer tendencies that are concentrated in segments # 4 and # 6.

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

The contingency table in Table 1 describes changes taking place over the two years of the survey. The Pearson Chi-square renders a highly significant outcome at the 99.9% significance level. The major trends with regard to activity segment shifts from 1994 to 1997 are (1) the increase of the sports segment # 3, an increase 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 similar way as self organizing feature maps but does not force the prototypes into the predefined grid. Table 2 illustrates the relation between the two segmentation solutions. The culture segment is split up in segment TRN # 2 and TRN # 5 (cross-sections are pointed out in Table 2 with black frames around the cells), the main difference between these segments being the fact 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 the major discriminating variable being the amount of biking done during the vacation. 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 resulting from the ten-segment solution are very 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 10 segment

solution thus backs up segments # 1, # 3, # 4 and # 5 from the SOFM solution, as well as supporting the fact that the remaining group of tourists are not easily segmented in terms of vacation activities, as no stable representation can be determined.

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

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

---------- TABLE 3 ----------

All in all, validation of the initial six segment solution by repeating the tracking process for a ten-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 engaging in very few vacation activities (or the negative answer tendencies group) can be accepted as relatively stable and the same market trends concerning these groups of tourists result from both investigations. The remaining segments are not identified as stable and therefore should not be chosen as target segments without further investigation.

Conclusions

This paper has proposed a tracking framework for the tracing of trends in *a posteriori* segments over time. The stepwise framework includes the following stages: (1) definition of the anchor period, (2) computation of a data-driven market segmentation solution using data from the anchor period, (3) characterisation of market segments, (4) assignment of data from other periods to the anchor segment solution, (5) testing of distribution changes, (6) testing of changes in background variables, and (7) validation of results, with the last step being of central importance in reducing the influence of random factors in the final result. The framework is flexible regarding methodological approaches preferred at each step.

In order to apply the tracking framework the procedure requires multiple-period guest survey data based on an identical questionnaire and sampling strategy.

The advantages of being able to track *a posteriori* segment trends in the marketplace include (1) validation of single data-driven segment solutions on which an entire marketing plan can be built, (2) increased insight into the changes encountered in the marketplace, (3) provision of a sound basis for forecasting, (4) the possibility of regular evaluation as to whether the segment(s) targeted should be maintained or replaced, (5) only temporary reduction of dimensionality during the segment assignment procedure (no compression of the item information), and (6) applicability to typical multi-period data in tourism. (It is not necessary for the same individuals to be questioned. Such a requirement would be difficult to achieve in the case of touristic guest surveys.).

The main limitation of the concept is the insecurity arising from sampling over consecutive periods of time. It cannot automatically be assumed that multiple surveys based on representative samples are sufficient to exclude intervening variable effects that may distort the results. This problem is especially crucial in the case of surveys that rely strongly on weighting cases due to sampling restrictions. But this limitation clearly is not caused by the

tracking framework proposed but affects any analysis of data sets derived in typical guest survey manner over multiple periods of time.

Further work is needed to extend the data-driven segment tracking tool in a three way data situation, thus building upon the perceptions based market segmentation concept as first introduced by Dolničar, Grabler & Mazanec (1999) and comprehensively described in Mazanec & Strasser (2000). The relevance of this extension is founded on the increasing importance of brand and destination image studies. Extending the tracking tool in such a way would enable simultaneous trend tracking of *a posteriori* segmentation, positioning and competition in the marketplace.

Acknowledgement

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).

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., Grabler, K., and 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, pp. 39-62.

Dolničar, S. and Leisch, F. (2001) *Knowing what you get – A conceptual clustering framework for increased transparency of market segmentation studies*. Abstract Proceedings of the Marketing Science Conference in Wiesbaden, pp. 88-89.

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

Ketchen D. J. jr. and Shook, C.L. (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.

Lilien, G.L. and Rangaswamy, A. (1998) *Marketing Engineering*. Reading: Addison-Wesley.

Martinetz, Th. and Schulten, K. (1994) Topology Representing Networks, *Neural Networks* 7(5), 507-522.

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

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

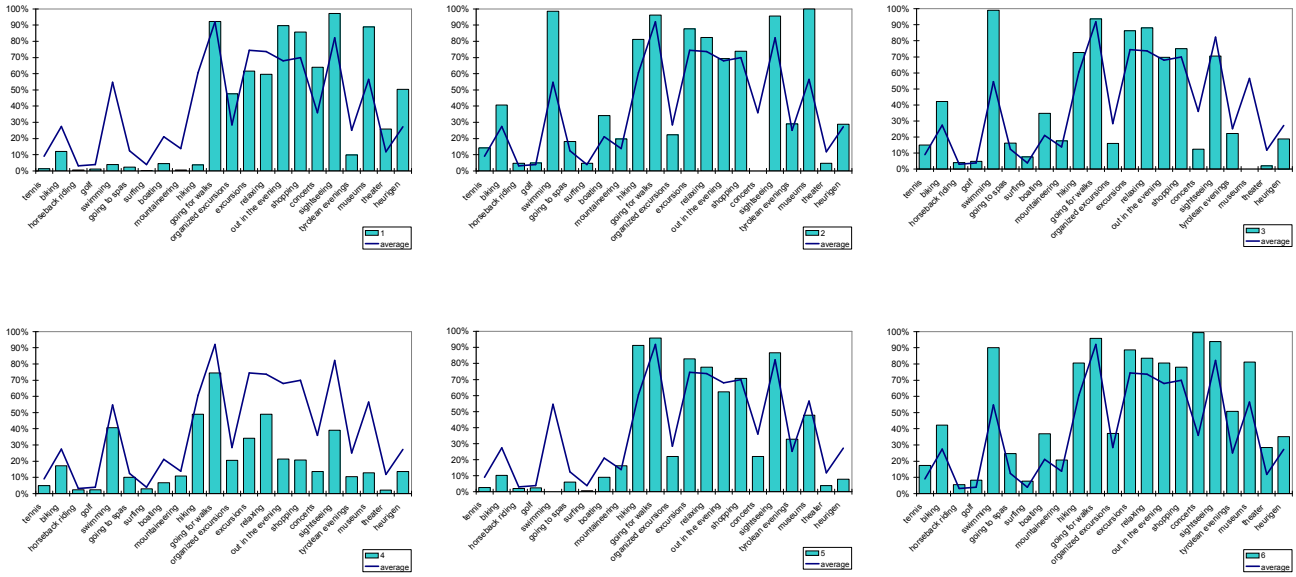

Table 1: Contingency table for segment size comparison 1994 and 1997 (SOFM based)

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: Cross-tabulation of the six-segment SOFM and the ten-segment TRN 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*

Table 3: Contingency table for segment size comparison 1994 and 1997 (TRN based)

Segment		1994	1997	Total
1	number of respondents	622	556	1178
	% within the year	8%	8%	8%
2	number of respondents	983	880	1863
	% within the year	12%	13%	13%
3	number of respondents	946	937	1883
	% within the year	**12%**	**14%**	13%
4	number of respondents	723	340	1063
	% within the year	**9%**	**5%**	7%
5	Number of respondents	839	588	1427
	% within the year	**11%**	**9%**	10%
6	number of respondents	735	681	1416
	% within the year	9%	10%	10%
7	number of respondents	864	638	1502
	% within the year	**11%**	**8%**	10%
8	number of respondents	523	306	829
	% within the year	**7%**	**5%**	6%
9	number of respondents	921	977	1898
	% within the year	**12%**	**15%**	13%
10	number of respondents	811	701	1512
	% within the year	10%	11%	10%
Total	number of respondents	7967	6604	14571
	% within the year	100%	100%	100%