Mikulić, J., Prebežac, D. (2009), "Developing service improvement strategies under consideration of multicollinearity and asymmetries in loyalty intentions – A study from the airline industry", in: Stauss, B., Brown, S.W., Edvardsson, B., Johnston, R. (Eds.): QUIS 11: Moving Forward with Service Quality: Proceedings of the QUIS 11 - Services Conference. Wolfsburg, Germany, pp. 414-423.

DEVELOPING SERVICE IMPROVEMENT STRATEGIES UNDER CONSIDERATION OF MULTICOLLINEARITY AND ASYMMETRIES IN LOYALTY INTENTIONS– A STUDY FROM THE AIRLINE INDUSTRY

Josip Mikulić and Darko Prebežac

University of Zagreb, Faculty of Economics and Business, Croatia

ABSTRACT

The aim of this paper is to introduce a regression-based analytical framework for developing service improvement strategies which account for asymmetric effects in customer satisfaction and loyalty. A hierarchical research design is applied to minimize the risk of multicollinearity. The high managerial value of the framework is demonstrated in a case study on airline passenger satisfaction. A four-dimensional importance-performance analysis is used to derive improvement-priorities of the main components of airline passenger services, whereas several determinance-asymmetry analyses are used to derive priorities of the service attribute forming the service components.

INTRODUCTION

The essence of customer satisfaction (CS) research, through the lens of a service manager, is to identify areas of good and poor service performance, as well as more and less important service elements, in order to allocate resources in improvement strategies. Better performance increases CS, which is a necessary precondition for achieving high levels of customer loyalty (CL) in competitive environments (Heskett *et al.*, 1994), whereas stronger CL significantly boosts the customer lifetime-value for a company, why it should be regarded a key metric of long-term financial performance.

This marketing paradigm, called the service-profit chain, has several implications for the CS management. On the one hand, service firms should benchmark themselves against their competitors, because competitors in a market strongly influence customer choice and/or judgment of a service. Benchmarking reveals significant performance shortcomings, which need to be resolved to keep a competitive position in the market. On the other hand, managers further need to know, how improvements of particular service elements will be related to changes in CS, and how these changes in CS will consequently be related to changes in CL. Accordingly, service-elements have to be assigned weights based on their impact on the customer's choice and/or evaluation of the service, before prioritizing them for improvement. However, weighting service elements is quite a complex task, since there is growing evidence that an element's impact on the customer's choice and/or evaluation is performance-dependent (e.g. Füller *et al.*, 2006), which exhibits the need for some kind of dynamic weighting procedure. Another major issue plaguing CS researchers is the problem of multicollinearity, which calls into question the applicability of commonly used weighting procedures based on multiple regression analysis (MRA). Both these issues—i.e. (i) asymmetries in CS/CL and (ii) multicollinearity in CS data will be discussed in the following two sections of this paper and guidelines for dealing with them will be put forward. Based on insights from the discussion, a new analytical framework for developing service improvement strategies will be proposed, which is demonstrated in a case study from the airline industry.

ASYMMETRIES IN CUSTOMER SATISFACTION RESEARCH

In the marketing literature, there are two opposed viewpoints regarding the issue of asymmetry in the links of the service-profit chain, particularly in the links between attribute-performance, CS and CL. On the one hand, there is a growing number of studies that hypothesize the existence of both positive and negative asymmetry—i.e. some product/service attributes have a larger impact on CS/CL when performance is low than when it is high, whereas for some attributes it is the other way round (e.g. Brandt, 1987; Matzler *et al.*, 2003). Most of these authors refer to Herzberg's Motivator-Hygiene theory (Herzberg, 1958), the Kano model (Kano *et al*., 1984) and/or the three-factor theory of CS. On the other hand, there is a smaller number of influential studies in which it is hypothesized that negative performance generally has a larger impact on CS than positive performance, whereby authors frequently draw an analogy to Kahnemann and Tversky's *prospect theory* (1979). For instance, Mittal *et al.* (1998) conclude that negative attribute-performance has a larger impact on overall CS (OCS) and repurchase intentions than positive performance, in both a product (automobiles) and service context (healthcare). This finding was confirmed in several other studies (e.g. Mittal and Baldasare, 1996), but one should be cautious when considering generalizing it. The hypothesis of negative performance exceeding the impact of positive performance on CS might proof true in many cases when *salient* service attributes are subject to analysis, but for *facets* of a service this must not always hold. Consider the following example. A diverse offer of movies on a continental flight (a service-facet) might have a strong positive impact on a passenger's satisfaction. Conversely, a poor choice (or absence) of in-flight movies might not affect the passenger's satisfaction in a negative way at all, because he usually does not watch movies when travelling shorter distances. Considering this example, it does not seem reasonable to assume that negative performance generally has a larger impact on CS than positive performance. On the other hand, any salient service attribute, such as flight-safety or on-time performance, would be likely to have a stronger negative impact on the passenger's satisfaction when performing low, than a positive impact when performing high.

The managerial implication of such dynamics in attribute-impact is that improvement priorities should not be based on the assumption of linear and symmetric relationships between the links of the service-profit chain. Accordingly, besides assessing the average impact of attributes on CS/CL (e.g. through a MRA between attribute-performance scores and CS), one should additionally assess and compare the attribute's impacts in cases of low-level performance and high-level performance. Attributes showing significant impact-asymmetries should consequently be treated with particular care when setting improvement priorities. As a general guideline, when several attributes have similar levels of average impact, attributes with a negative impact-asymmetry should be assigned higher priority than attributes with a positive impact-asymmetry when performance is low, whereas the latter group of attributes should be assigned higher priority when performance is high, following the rule to decrease dissatisfaction first, before increasing satisfaction. However, it is important to compare only those elements with similar impact-levels, because attributes with a positive impact-asymmetry might have a larger absolute impact on creating dissatisfaction than attributes with a negative impact-asymmetry (Mikulić and Prebežac, 2008).

MULTICOLLINEARITY IN CUSTOMER SATISFACTION RESEARCH

When reviewing studies that employ MRA to prioritize product/service attributes, it becomes evident that multicollinearity is another major issue plaguing researchers. In general, multicollinearity occurs when two or more predictors in a MRA are highly correlated with each other, which frequently occurs when using large numbers of predictors or inadequate measurement models. Possible consequences are inversed signs of regression coefficients, or some predictors appear statistically insignificant though they actually are not, and reversely. The implication for service researchers is that one cannot use any desired number of service-attributes to be tested for their impact on the criterion variable (e.g. CS, CL). In order to circumvent this problem, it appears as if some authors intentionally use smaller sets of predictors, but this is not a satisfactory solution for managers, because such an approach does not allow for capturing facets of a service. Another group of authors approaches the multicollinearity-problem by using bivariate analyses (only two variables are analyzed at a time), like bivariate regression analysis (Ting and Chen, 2002), or correlation analysis (CA). However, one should be cautious when using correlation coefficients (CC) for several reasons. First, a CC does not represent the impact of a predictor on a criterion variable, but the *strength of linear relationship* between two variables. However, social science researchers frequently imply causation when employing CA, and interpret CCs as indicators of an attribute's impact. Second, a CC measures the strength of a *linear* relationship, meaning that the value of CCs is questionable when analyzed relationships in fact are nonlinear and asymmetric. Third, and maybe most important, since in a CA only two variables are analyzed at a time, the *impact and significance* of all the other variables remains unconsidered, resulting thus in a lesser ability of the analysis to discriminate between the *weights* of attributes. However, given the research objective to prioritize attributes, relative attribute-weights are what researchers are actually interested most in.

A more reasonable way to deal with multicollinearity is to use a research design that minimizes the risk of the problem's occurrence. In order to avoid (undesired) inter-predictor correlations, one should adopt the following guidelines in the scale development stage. First, there should not be any overlapping between the conceptual domains of predictors (i.e. service attributes), and predictors should always be on the same level of abstraction. This reduces the risk of one predictor tapping into the domain of other predictors, or of a predictor being a formative sub-attribute of one or more other predictors, which is likely to result in intercorrelations. Second, the number of predictors should generally be kept low because 'lesser predictors' simply means 'lesser danger of intercorrelations'. In this regard, it is suggested using hierarchical designs in the fashion of multilevel measurement models as proposed by e.g. Dagger *et al.* (2007). Adapted to the task of service attribute prioritization, one analysis should be conducted at the first level, comprising the main components of the service, and several analyses should be conducted at the second level, comprising attributes (i.e. facets) of the first-level components. A deeper structure might as well be taken into consideration, depending on the desired level of detail.

METHODOLOGY

The aim of this study is to demonstrate an analytical framework for developing service improvement strategies using a case study of a complex service—i.e. airline passenger services. The framework consists of analyses at two levels. At the first level, improvement priorities of the main service components are derived using a four-dimensional importance-performance analysis (4D-IPA) and a determinance-asymmetry analysis (DAA). At the second level, several DAAs are used to derive improvement priorities of the service attributes forming the service components at the first level. The 4D-IPA and the DAA will be described in detail at the end of the methodology section.

For this study, data were collected for two full-service carriers operating at a major Croatian international airport with similar flight schedules regarding destinations and flight frequencies. The first airline is the focal airline of this study (FAL), whereas the second one is regarded its main competitor (CAL). In total, 718 airline passengers formed the sample for this study (FAL=383; CAL=335). The research instrument was a structured questionnaire which comprised measures for: (i) service attribute-performance; (ii) service component-satisfaction and (iii) intentional loyalty (IL). Service attribute-performance and service component-satisfaction were measured with single item seven point Likert scales, whereas IL was measured with four items derived from the scales of Zeithaml *et al.* (1996) and Taylor and Baker (1994). To generate the initial item pool of airline passenger service attributes, a qualitative study involving several open-ended questions with 30 airline passengers was conducted. The results were paired with items identified in previous research in the relevant literature. A panel of four expert judges then independently grouped the attributes into a smaller number of main components of airline passenger services. The categorizations were then compared by the expert judges, and refined in a three-stage iterative Delphi process.

Based on the results from the qualitative research process, a pre-test questionnaire was constructed which comprised five service components (flight offer; ticket purchase experience; airport experience; flight experience; and relationship experience) with 34 items. The questionnaire was tested on a sample of 100 international airline passengers at a major Croatian airport. In order to explore significant intercorrelations among attributes, correlational matrices were computed. Attributes with high intercorrelations within the proposed service components were reassessed by the judges, who either grouped or excluded such attributes from the final attribute list. By the end of this process, the initial item pool was subsequently reduced to 25 service attributes.

Four-dimensional importance-performance analysis (4D-IPA)

The 4D-IPA is an extension of traditional importance-performance analysis (IPA, Martilla and James, 1978). In the proposed approach, a third dimension is included by using two measures of attribute-importance (AI) commonly used in IPA—i.e. direct AI ratings and weights obtained through MRA. In this study, the MRA was conducted between component-satisfaction scores and IL scores, following the findings of Mittal *et al.* (1998) who proved a direct impact of attribute-performance on IL. Two AI measures were included into the analysis because their combination offers surplus information for managers. What do the two measures actually measure? According to a meta-review of the validity of AI measurement by Van Ittersum *et al.* (2007), direct AI ratings measure the relevance of service components, whereas regression weights measure their determinance (i.e. impact on CS). On the one hand, relevance represents the customer-perceived importance of an element in a service-configuration based on existing industry norms and standards. In this regard, relevance is similar to an attitude. On the other hand, determinance quantifies an attribute's significance in judgment and choice (Myers and Alpert, 1977), and is calculated "…based on the difference in (valuation of) different attribute levels" (Van Ittersum *et al.*, 2007, p.1180). Determinance thus is a dynamic concept, and it is obvious that the two AI measures do not assess identical concepts. Consequently, when using only one AI measure, managers might obtain misleading recommendations, because the importance of relevant components, which appear not to be determinant, might be underestimated when using regression weights, or the importance of determinant components, which appear to be irrelevant, might be underestimated when using direct AI ratings. Using both measures thus decreases the risk of suboptimal prioritizations. Furthermore, a comparison of the two measures facilitates the identification of (i) primary loyalty drivers (high relevance and high determinance), (ii) secondary loyalty-drivers (low relevance, but high determinance), (iii) spurious loyalty-drivers (high relevance, but low determinance), and (iv) low importance components (low relevance and low determinance). Additionally, main competitor performance was included as a fourth dimension to IPA. However, since a 4D representation would be confusing, a 2D-grid was constructed using scores of the relevance and determinance of service components, whereas components with satisfaction scores below average (i.e. below the grand mean of component-satisfaction scores) were marked with a minus (-), and components with satisfaction scores above average were marked with a plus (+). Moreover, components with a performance-level below CAL performance were presented in italics. In order to keep the questionnaire length at a reasonable level, relevance scores (i.e. direct importance scores) were collected only for the five main service components. Thus, the 4D-IPA was conducted only at the service component level.

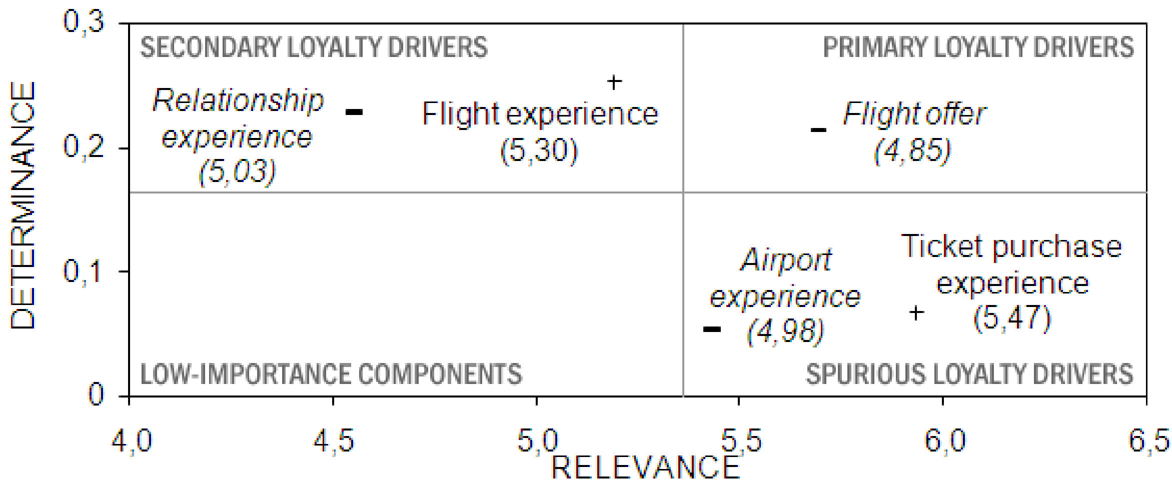
Determinance-asymmetry analysis (DAA)

The DAA was introduced by Mikulić and Prebežac (2008) as a research tool for categorizing service attributes according to their range of impact on OCS, and the degree of asymmetry of their impact on OCS. To remain consistent with the terminology used in the previous section, the range of impact on OCS will be referred to as determinance, and the asymmetry of impact will be referred to as determinance-asymmetry (DA). Determinance scores are obtained through a MRA with scores of component-satisfaction (attribute-performance) as predictors, and IL (component satisfaction) as the criterion variable. DA is calculated in two steps. First, a MRA is conducted using two sets of dummy variables for each component (attribute) as predictors, and scores of IL (component satisfaction) as the criterion variable. The first dummy is created by coding highest scores to 1, whereas all other scores are coded as 0. This set is used to quantify the impact on the criterion in case of very high perceptions (reward coefficient). The second set is created by coding lowest scores to 1, whereas all other scores are coded as 0. This set is used to quantify the impact on the criterion in case of very low perceptions (penalty coefficient). In the second step, reward coefficients and penalty coefficients for each component (attribute) are divided by their sum, and the resulting ratios are subtracted to obtain DA scores ranging from -1 to +1. A DA score of -1 means the component (attribute) has only dissatisfaction-generating potential (DGP), whereas a score of +1 means it has only satisfaction-generating potential (SGP). By depicting scores of determinance and DA along the axes of a two-dimensional grid, the analysis facilitates the identification of low-, medium- and high impact components (attributes), as well as a categorization of components (attributes) based on the degree of their DA.

ANALYSIS AND RESULTS

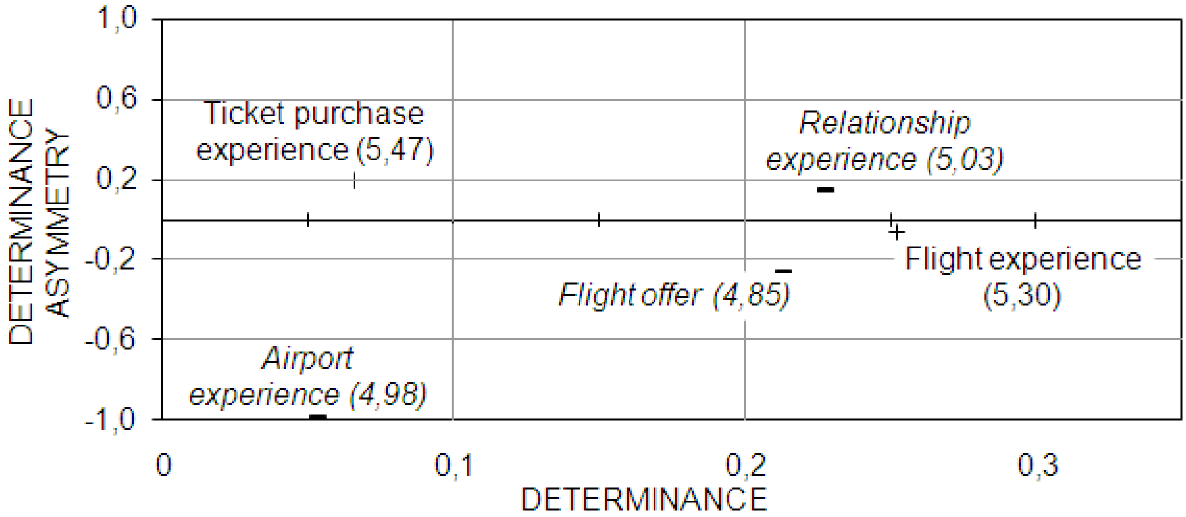
The 4D-IPA (Figure 1) revealed that one service component is a primary loyalty driver (flight offer). This component is perceived important by customers when choosing an airline, and it indeed strongly impacts IL. As the satisfaction-level of this component is quite low (4.85), and below the CAL level (5.21), the airline should assign this component highest priority in improvement strategies. Moreover, two components are categorized as secondary loyalty drivers (relationship experience; flight experience). These components are considered less important in airline choice, but they nevertheless strongly influence IL. As the satisfaction-level of relationship experience is below average (5.03) and below the CAL level (5.11), it should be improved right after *flight offer*. The remaining two components (airport experience; ticket purchase experience) are categorized as spurious loyalty drivers. Passengers consider them very important when choosing an airline, but, in fact, they do not strongly influence IL.

Figure 1 4D-IPA for airline service components



In the next step a DAA was performed to get a detailed insight into asymmetries in the relationship between component-satisfaction and IL (Figure 2). The analysis revealed that four of five components approximately linearly impact IL depending on the level of satisfaction (ticket purchase experience; flight offer; relationship experience; flight experience). Only one component showed an extreme negative asymmetry in the satisfaction-IL relationship (airport experience), meaning the component has no positive influence on IL, even in case of very high satisfaction levels.

Figure 2 DAA for airline service components



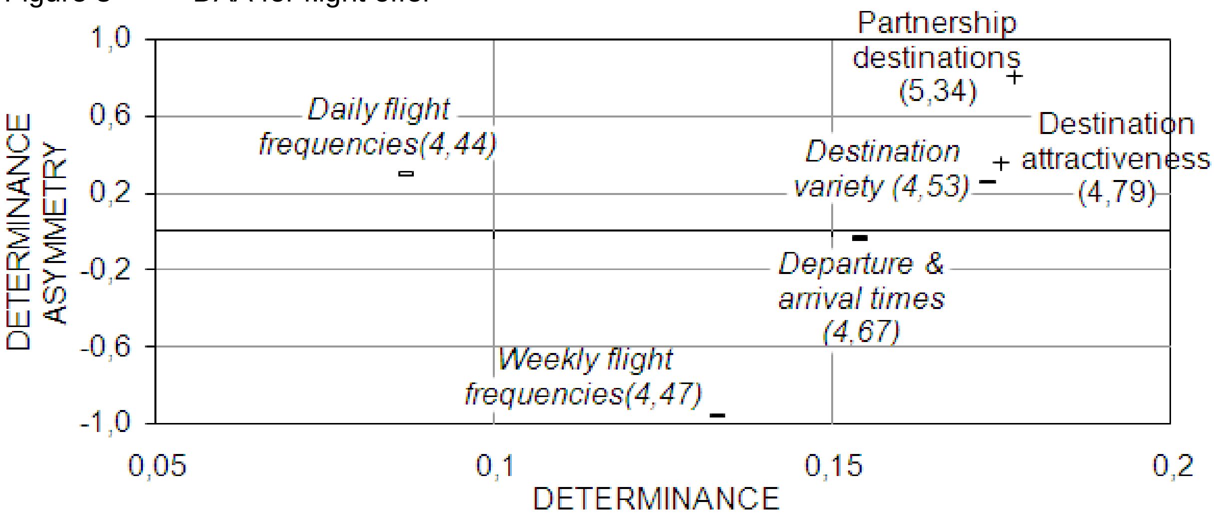
To explore the key-drivers of service component satisfaction, in the following step DAAs were conducted at the attribute level. Table 1 provides an overview of the results.

Table 1 Results of the attribute-level analysis

Service component	Component attributes	FAL perf.	CAL perf.	Deter-minance	DA
Flight offer[1]	Destination variety	4.53	5.19	0.173***	0.256
	Destination attractiveness	4.79	4.78	0.175***	0.350
	Weekly flight frequencies	4.47	4.76	0.133***	-0.958
	Daily flight frequencies	4.44	5.04	0.087**	0.296
	Departure and arrival times	4.67	5.49	0.154***	-0.037
	Partnership destinations	5.34	5.19	0.177***	0.802
Ticket purchase experience[2]	Ease of reservation	5.25	5.43	0.206***	0.939
	Reservation flexibility	5.37	4.79	0.129***	-0.626
	Reservation personnel	4.49	5.30	0.162***	0.589
	Ease of payment	5.28	5.24	0.392***	0.379
Airport experience[3]	Check-in efficiency	5.11	5.30	0.127***	-0.003
	Check-in personnel	5.40	5.38	0.084**	0.576
	Information availability	5.17	5.42	0.099***	0.824
	Airport lounge attractiveness	4.15	4.77	0.192***	-0.642
	Boarding efficiency	4.80	5.20	0.144***	0.429
	On-time performance	4.68	4.80	0.278***	-0.046
Flight experience[4]	On-board catering	4.50	4.58	0.034*	-0.982
	On-board entertainment	4.08	4.55	0.020*	0.333
	Cabin/flight staff	5.38	5.35	0.204***	0.675
	Comfort level of aircraft	4.81	4.59	0.221***	-0.574
	Cleanliness of aircraft	5.58	5.65	0.209***	-0.126
Relationship experience[5]	FFP quality	4.71	5.01	0.135***	0.052
	Treatment in case of failures	4.51	4.90	0.227***	-0.252
	Care for customer wishes	4.78	5.02	0.342***	0.702
	Trustworthiness of airline	5.14	5.31	0.211***	0.473

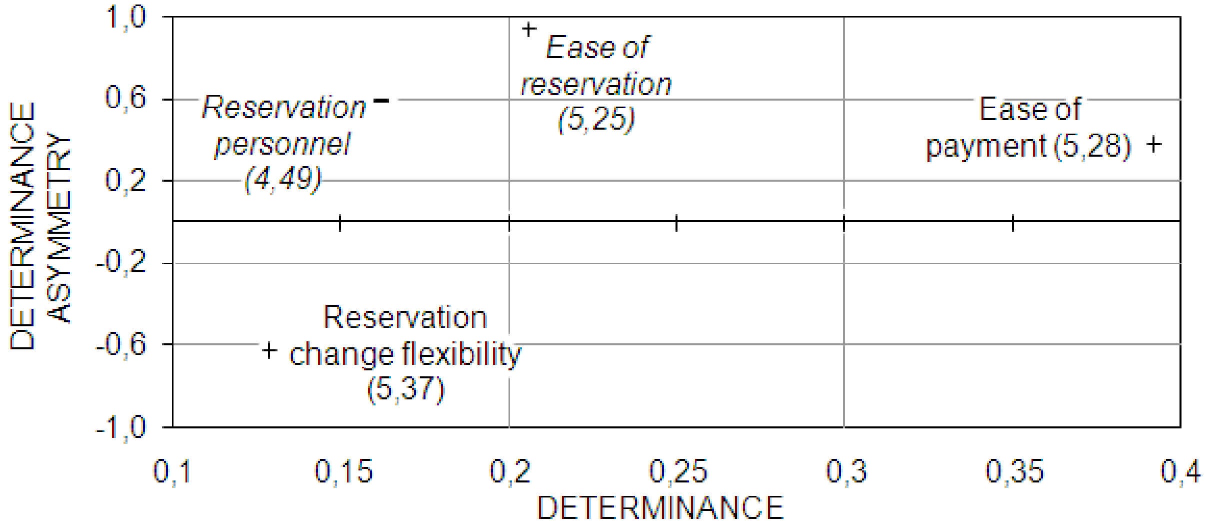
Notes: ***p<0.001; **p<0.01; *p<0.10; [1]R^2=0.489; [2]R^2=0.606; [3]R^2=0.613; [4]R^2=0.577; [5]R^2=0.635; determinance scores are unstandardized regression coefficients

Figure 3 DAA for flight offer



The DAA for the component *flight offer* (Figure 3), which is a primary loyalty driver, revealed three highly determinant attributes which have a significantly larger SGP than DGP (partnership destinations, destination attractiveness; destination variety). Two of them have a performance-level above average, and above CAL performance (destination attractiveness; partnership destinations). However, one attribute performs below both component-average and CAL level (destination variety), why it should be assigned highest priority in this component. High priority should as well be assigned to *attractiveness of departure and arrival times*.

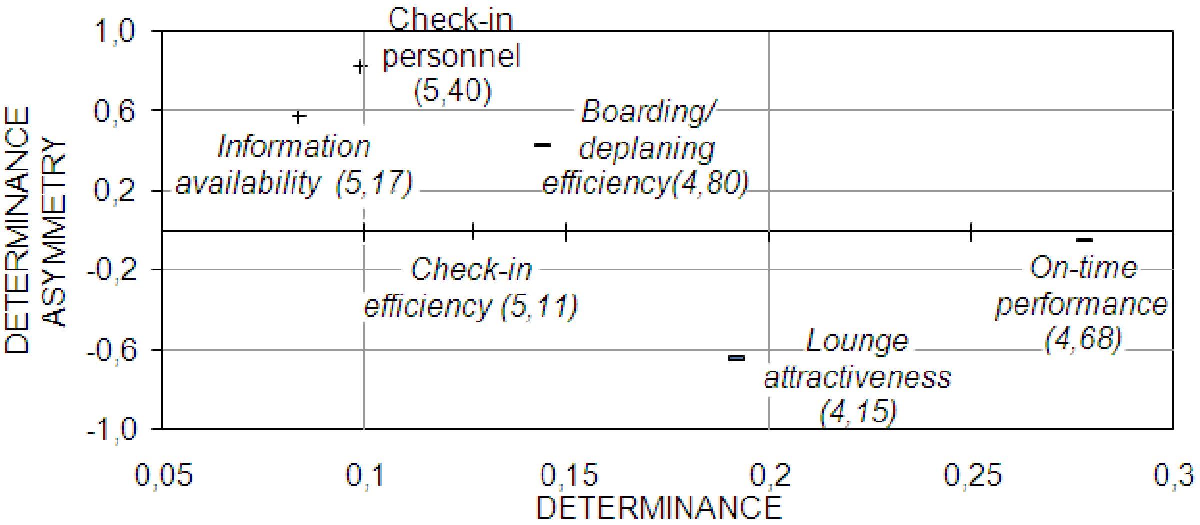
Figure 4 DAA for ticket purchase experience



Key-drivers of satisfaction with the *ticket purchase experience*, which is a less important component in explaining IL, are shown in Figure 4. The DAA revealed that one attribute is dominant in determining component satisfaction (ease of payment). This attribute has a significantly larger SGP than DGP, and performs above both component-average and CAL level, why it does not necessitate managerial action.

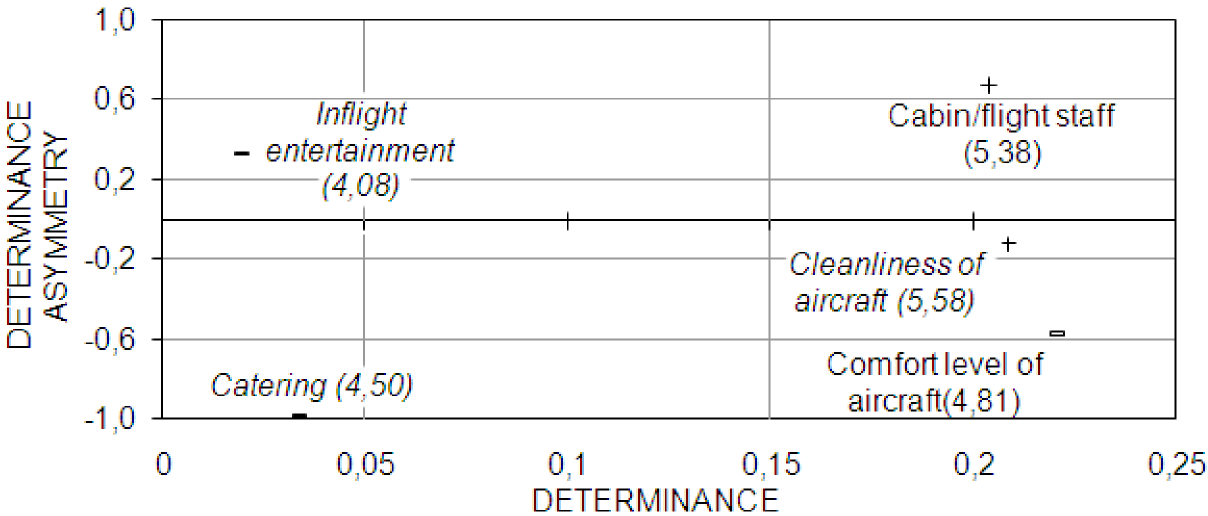
The DAA for the *airport experience* (Figure 5), which is a spurious key-driver of IL, revealed that the airline should mainly focus on two attributes in order to increase component-satisfaction (on-time performance; airport lounge attractiveness). Both attributes are highly determinant and have a performance-level below component-average and CAL level.

Figure 5 DAA for airport experience



As *lounge attractiveness* has a significantly larger DGP than SGP, and a much lower performance-level than *on-time performance*, the management should consider assigning it highest improvement priority within this component.

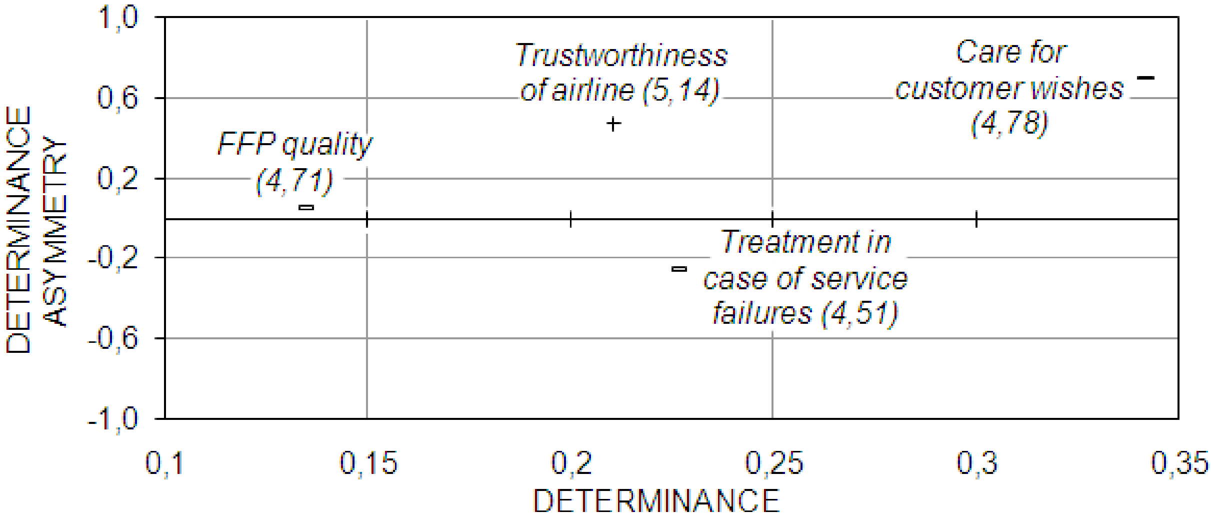
Figure 6 DAA for flight experience



The DAA for the *flight experience* (Figure 6), which is a secondary loyalty driver, revealed that three attributes largely determine the level of component-satisfaction (cabin/flight staff; cleanliness of aircraft; comfort level of aircraft). The management should assign highest improvement priority to *comfort level of aircraft*, as this attribute has a significantly larger DGP than SGP, and is performing quite low (though above the CAL level). After having resolved the performance problems of this attribute, the management should consider improving *cleanliness of aircraft*, as its performance-level is below the CAL level. The attribute *cabin/flight staff* does not necessitate any action, as its performance-level is above both component-average and CAL level. The remaining two attributes are less important in determining component-satisfaction (in-flight entertainment; catering). However, both have very low performance-levels (below component-average and CAL level), why they should be considered for improvement after having resolved the previously mentioned performance shortfalls.

The key-drivers of the *relationship experience*, which is a secondary loyalty driver, are presented in Figure 7.

Figure 7 DAA for relationship experience



All four attributes forming this category perform below the CAL level, why this category should generally be assigned high priority in improvement strategies. The attribute with the largest influence on component-satisfaction is *care for customer needs and wishes*. This attribute has a significantly larger SGP than DGP, and its performance-level is very low (4.78). The airline should therefore assign this attribute highest improvement-priority within this component, since this attribute bears a large potential to increase component-satisfaction, and consequently IL.

CONCLUSION

This study introduced a new analytical framework for developing service improvement strategies, which was demonstrated in a case study using a complex service—i.e. airline passenger services. In a first step, a four-dimensional importance-performance analysis and a determinance-asymmetry analysis (DAA) were used to derive improvement-priorities of the main components of airline passenger services, whereas several DAAs were used, in a second step, to prioritize the service attributes forming the main service components. The key advantages of the proposed framework are that it considers (i) asymmetric effects in customer satisfaction and loyalty; (ii) multicollinearity in customer satisfaction data; as well as (iii) the existence of competitors in a market. The fact that a lack of awareness about these issues might result in misleading recommendations regarding service attribute prioritization, underpins the high managerial value of the framework.

REFERENCES

Brandt, R.D. (1987), “A procedure for identifying value-enhancing service components using customer satisfaction survey data“, in Suprenant, C.F. (Ed.), Add value to your service: The key to success, AMA, Chicago, 61-64

Dagger, T.S., Sweeney, J.C., and Lester W., (2007), "A hierarchical model of health service quality: scale development and investigation of an integrated model, *Journal of Service Research* 10 (2) 123-142

Füller, J., Matzler, K. and Faullant, R., (2006), “Asymmetric effects in customer satisfaction”, Annals of Tourism Research 33 (4) 1159-1163

Herzberg, F., Mausner, B., and Snyderman, B., (1958), “The motivation to work”, Wiley, New York

Heskett,J.L., Jones, T.O., Loveman,G.W., Sasser,W.E.J., and Schlesinger, L.A., (1994), "Putting the service-profit chain to work", *Harvard Business Review*, 72 (2) 164-174

Kahneman, D. and Tversky, A., (1979), "Prospect theory: An analysis of decisions under risk," *Econometrica*, 47 (March) 263-291

Kano, N., Seraku, N., Takahashi, F., and Tsuji, S., (1984), "Attractive quality and must-be quality", Hinshitsu, The Journal of the Japanese Society for Quality Control, April 39-48

Martilla, J.A., and James, J.C., (1977), "Importance-performance analysis", *Journal of Marketing* 41 (1) 77-99

Matzler, K., Sauerwein, E., and Heischmidt, K.A., (2003), "Importance-performance analysis revisited: the role of the factor structure of customer satisfaction", *The Service Industries Journal*, 23 (2) 112-129

Mikulić, J., and Prebežac, D. (2008), "Prioritizing improvement of service attributes using impact range-performance analysis and impact-asymmetry analysis", *Managing Service Quality* 18 (6) 559-576

Mittal, V., and Baldasare, P.M., (1996), "Impact analysis and the asymmetric influence of attribute performance on patient satisfaction", *Journal of Health Care Marketing*, 16 (3) 24-31

Mittal, V., Ross, W.T., and Baldasare, P.M., (1998), "The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions", *Journal of Marketing*, 62 (January) 33-47

Myers, J.H., and Alpert M.I., (1977), "Semantic confusion in attitude research: salience vs. importance vs. determinance", *Advances in Consumer Research* 4 106–110

Taylor, S.A., and Baker, T.L., (1994), "An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions," *Journal of Retailing*, 70 (2) 163-178

Ting, S.C., and Chen, C.N. (2002), "The asymmetrical and nonlinear effects of store quality attributes on customer satisfaction", *Total Quality Management*, 13 (4) 547-569

Van Ittersum, K., Pennings, J.M.E., Wansink, B., and van Trijp, H.C.M., (2007), "The validity of attribute-importance measurement: A review", *Journal of Business Research*, 60, 1177-1190

Zeithaml, V.A., Berry, L.L., and Parasuraman, A., (1996), "The behavioral consequences of service Quality," *Journal of Marketing*, 60 (April) 31-46