

An Analysis of Consumer Search and Buying Behaviour in the US Airline Industry using Big Data

Christopher P. Holland^{ab} and Eleni Georghiades^a

^aManchester Business School
University of Manchester, UK
chris.holland@mbs.ac.uk

^bDepartment of Information Systems
University of Münster, Germany

Abstract

Online search and buying behaviour are analyzed using online panel data from ComScore, and passenger data from the Federal Aviation Authority in the US market. Most online activity is e-service. The online consideration set is measured longitudinally and the 2013 average is 2.46, which is a relatively narrow search process and different to theory predictions. The correlation between sales through all channels and online search is very high at 0.76, which demonstrates the importance of the online channel for search, as well as sales. The variability of the search and sales patterns is very low in the legacy airlines, and this demonstrates that consumers perceive little difference between the competitive offers of legacy airlines. Conversely, search and sales variability in SouthWest Airlines is significantly higher, which indicates a more active search process.

Keywords: consumer search, online panel data, buying behaviour, strategic groups

1 Introduction

The importance of attracting visitors and new customers to a company's website is well documented in the academic and business literature (Yadav and Pavlou 2013). In addition to new customer acquisition, a website is also an important service channel for existing customers (Toufaily et al. 2013). Previous research into the evaluation of online marketing has tended to focus on the use of internal data regarding online spending and online customer behavior (Buckinx et al. 2007). Valuable insights about online users can be gained but web server monitoring tells a company almost nothing about its *relative performance against its competitors*. This research uses online panel data and sales data to evaluate online consumer behaviour in the U.S. airline market.

2 Theory

The consideration set concept is a power concept in consumer marketing and has been used widely in marketing research pre-Internet (Hauser and Wernerfelt 1990). It is equally valid in the online channel and is termed the online consideration set. Online search is generally thought to be highly effective, efficient, and convenient. It requires less time compared to visiting physical shops or reading magazines and brochures. For example, McKinsey reported that a consumer will generally find time to perform ten online searches compared to two offline searches for a purchase (Bughin et al. 2011). Reduced cost of online search is therefore predicted to encourage consumers to

perform more extensive search on the Internet compared to other channels. However, empirical results from research into the size of online consideration sets using panel data show relatively small online consideration sets, e.g. Johnson et al. (2004), reported an average online consideration set of just 1.8 for travel sites, Zhang et al. (2006) reported 3.3 for travel sites, and Holland and Mandry (2013) reported values of between 2.1 and 2.8 in a multi-sector study of US and UK markets.

2.1 Research Questions

Research Question 1. What is the extent of the search process measured by the online consideration set and the distribution of online users across competitor websites?

Research Question 2. What is the relationship between share of unique visitors and market share in the US airline industry?

3 Methods

Two main data sources are used: (1) ComScore’s US online panel data of one million users and passenger data from the Federal Aviation Administration (2014). **Research Question 1.** The pre-Internet consideration sets were taken from the authoritative review conducted by Hauser and Wernerfelt (1990) of consideration set sizes. The statistical ‘t’ test was used to measure the significance of the difference between the pre-Internet and online measurements. **Research Question 2.** A linear regression model was used to measure the strength of the statistical correlation between overall sales and online visitors for all matched data points over the time period 2011 - 2013.

4 Results

4.1 Research Question 1

An online user that visits one website only is categorized as an e-service user, i.e. someone who uses the Internet to manage their relationship with a single airline. Those visiting two or more websites are termed online searchers. This approach is based on the definition of the consideration set by Hauser and Wernerfelt (1990), and Shocker (1991). The results are shown in Table 1.

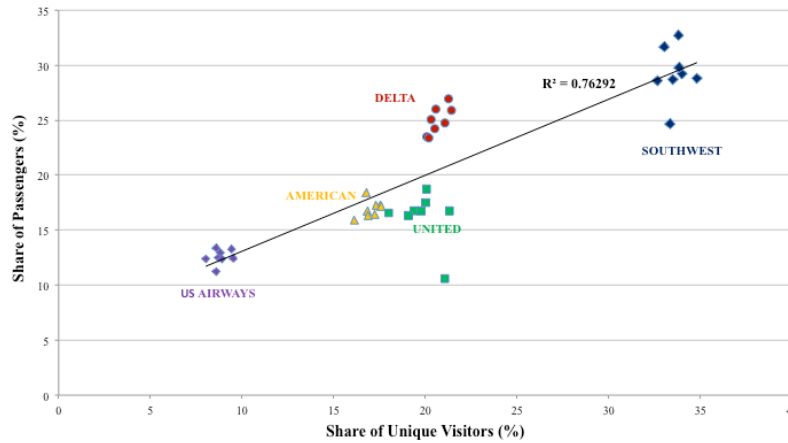
Table 1. The Search Process Measured by the Distribution of Online Searchers and Online Consideration Set

US Airline Online Consideration Set 2013	Number of Websites Visited			
	2	3	4	5
2.46	70%	23%	8%	1%

4.2 Research Question 2: Online Search and Total Sales

The data were normalized to share of passengers (i.e. market share) and share of unique visitors and the results are shown in Figure 1.

Figure 1. Share of passengers in time period (t) plotted against share of unique visitors in period (t-1)



There is a very strong correlation of 0.76 that is compelling evidence of the importance of online search to subsequent sales. The plot in Figure 1 shows a clear clustering pattern for each airline. A method of statistical cluster analysis is shown in Figure 2.

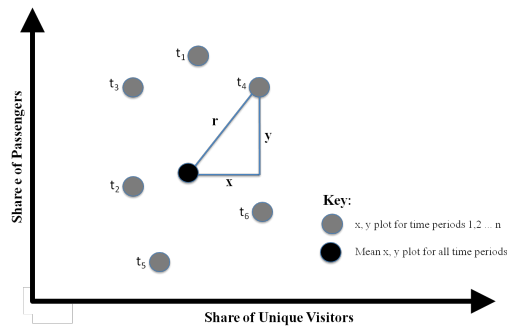


Figure 2. Cluster Analysis of an Online Sales Map

An illustrative x,y plot for one airline is shown in Figure 2 for six time periods, t1 – t6. The mean x,y position is the average for all time periods. The cluster analysis uses the radial distance of from each measurement to the mean and is shown for time period 4 in the diagram. The variance in the share of unique visitors is a function of ability to attract online visitors relative to competitors and is an indication of the level of interest from consumers in actively looking and switching between airline websites. The variance in sales arises from differences in search to buyers. The radius is therefore a composite measure of search and buying variance. For each individual airline, the variance of the set $r_1, r_2 \dots r_n$, where n is the number of time periods, represents the variance in search and buying behaviour. The results are shown in Table 2.

Table 2. Cluster Analysis, Q3 2011 – Q3 2013, United post-merger 2012-2013

	Cluster Analysis	
	Average	Variance
Southwest	1.82	2.35
Delta	1.16	0.37
American	0.78	0.16
United	2.13	2.04
United post-merger	1.68	0.51
US	0.67	0.16

There are two sets of data for United: United including the earlier time periods when it was going through a period of transition with the acquisition of Continental Airlines, and post-merger Q1 2012 – Q3 2013. The high level of variance is because of the activity in 2011. The post-merger cluster analysis shows a much lower level of variance. The variance of SouthWest’s cluster is larger than all of the legacy airlines if we take the United post-merger results. To test whether the differences in variance between the legacy airlines and the discount airline are significant, a matrix of F-test results is shown in Table 3.

Table 3. F test comparison of cluster variances between each pair of airlines.

	SouthWest	Delta	American	United	United post-merger	US
SouthWest	-	0.026	0.002	0.854	0.082	0.002
Delta	0.026	-	0.278	0.039	0.679	0.306
American	0.002	0.278	-	0.003	0.147	0.949
United	0.854	0.039	0.003	-	n/a	0.004
United post-merger	0.082	0.679	0.147	n/a	-	0.164
US	0.002	0.306	0.949	0.004	0.164	-

The United results include noise from the Continental merger, as evident from Table 1 where its passenger numbers jump by 71% from Q4 2011 to Q1 2012, and are excluded from the analysis. The US post-merger results are used instead. The F tests are significantly different for SouthWest. Conversely there are no differences in variance between the legacy airlines at the 10% level of significance. This demonstrates that search and buying behaviour of SouthWest Airline customers are significantly different to those of the legacy airline customers.

6 Conclusions

Only 26% of online consumers are actively searching and of these searchers, their average online consideration set is 2.46. These search patterns indicate a much lower level of search than is predicted from theory in marketing, Information Systems and economics. Sales through all channels are strongly correlated with online search and the results have similarities with the approach taken by Hu et al (2014).

The cluster analysis showed that the variance of the legacy airline clusters were very low, which indicates that consumers perceive little difference between the offers of the legacy airlines. The significantly higher level of variability in the SouthWest Airline cluster indicates that consumers are actively searching out this airline, resulting in variability of its share of unique visitors and in sales conversions.

Online panel data is a specific example of big data that combines a very large sample with very detailed tracking of behaviour through clickstream data. In the classification of research into computer-mediated marketing (Yadav and Pavlou 2014), this research is in the broad category of consumer-firm interactions, which is concerned with consumer behaviour towards suppliers in the online channel. It is part of an emerging field of research into network level phenomena. The advantage of online panel data is that it gives a network perspective of activities, i.e. the airline market can be analysed in a holistic manner as a group of competitors, rather than focusing on search behaviour towards just one airline.

7 References

- Buckinx, W., Verstraeten, G., and Van den Poel, D. 2007. "Predicting Customer Loyalty Using the Internal Transactional Database," *Expert Systems with Applications* (32:1), pp. 125-134.
- Bughin, J., Corb, L., Manyika, J., Nottebohm, O., Chui, M., de Muller Barbat, B., and Said, R. July 2011. "The Impact of Internet Technologies: Search." McKinsey&Company, High Tech Practice.
- Federal Aviation Administration. 2014. "Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports Collection & Use of Data." from:http://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/?sect=collection
- Hauser, J.R., and Wernerfelt, B. 1990. "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research* (16:4), pp. 393-408.
- Holland, C.P., and Mandry, G.D. 2013. "Online Search and Buying Behaviour in Consumer Markets," in: *46th Hawaii International Conference on System Sciences*. Maui, Hawaii: IEEE, pp. 2918-2927.
- Hu, Y., Du, R.Y., and Damangir, S. 2014. "Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data," *Journal of Marketing Research*.
- Johnson, E.J., Moe, W.W., Fader, P.S., Bellman, S., and Lohse, G.L. 2004. "On the Depth and Dynamics of Online Search Behavior," *Management Science* (50:3), pp. 299-308.
- Toufaily, E., Ricard, L., and Perrien, J. 2013. "Customer Loyalty to a Commercial Website: Descriptive Meta-Analysis of the Empirical Literature and Proposal of an Integrative Model," *Journal of Business Research* (66:9), pp. 1436-1447.
- Yadav, M.S., and Pavlou, P.A. 2013. "Marketing in Computer-Mediated Environments: Research Synthesis and New Directions," *Journal of Marketing* (78:1), pp. 20-40.
- Zhang, J.J., Fang, X., and Liu Sheng, O.R. 2006. "Online Consumer Search Depth: Theories and New Findings," *Journal of Management Information Systems* (23:3), pp. 71-95.