

Using Revealed- and Stated-Preference Customer Choice Models for Making Pricing Decisions in Services: An Illustration from the Hospitality Industry

Leo MacDonald, Kennesaw State University
Chris K. Anderson, Cornell University
Rohit Verma, Cornell University

This article presents an overview of discrete choice modeling for making pricing decisions in services. During recent years, discrete choice modeling has emerged as an effective approach for developing analytical models and for estimating relative weights of parameters based on empirical data. For estimation purposes, typically one of two forms of data is used: transactional data captured in databases (revealed-preference data); or primary experimental data (stated-preference data). In this article, we provide detailed illustration of both approaches for pricing decisions for hospitality services. Finally, we discuss the managerial implications of the discrete choice modeling approach described earlier in the article.

Solving mathematical puzzles without worry concerning context can provide, for some, a satisfactory exercise. But the science and art of management calls for more. An application is when the context is understood, the theory is relevant and the decision process is influenced. Theory may become a waste of time for all but the theorists when there is no concern for relevance or application beyond the self-perpetuation of the club. (Shubik, 1987)

The vast proliferation of goods and services, increased emphasis on mass customization and customer experiences, as well as new technologies requires that firms carefully evaluate the drivers of customer choices for their service offerings. For example, in the best-seller book *The Paradox of Choices*, Schwartz (2004) suggests that both mundane and involved decisions such as ordering a cup of coffee, choosing a health-care provider or setting up a retirement plan are becoming increasingly complex because of the abundance of choices available to the consumers in the marketplace. The same scenario can be derived also for many B2B markets and their customers.

The underlying problem in predicting customer choices resides much more in the fact that purchasing decisions are made on the basis of (potentially) many different criteria simultaneously, such as brand, quality, performance, features, channel while also considering price (McFadden, 1986). This problem is further confounded in 'service' applications in which a customer may consider non-tangible features and characteristics of the market offerings (for example, service quality, safety and trust; interactions between service providers and customers and so on). Using the hospitality sector as an example, customers might select a particular hotel based on price as well as its location (close to airport, tourist location and downtown), brand name, various amenities (swimming pool, golf course and fitness centers), and loyalty program, among other things. At the same time, given resource constraints, firms must make trade-offs on the basis of what they do best, their cost structure, what their competitors are offering and what criteria they think matter most to their customers. In summary, pricing decisions for even simple service products can become relatively complex.

It is also well known that current service consumers have available a vast array of data, including both price and inventory availability from multiple service operators, along with a host of product attributes (for example, amenities and other services) available when making their purchase decisions. For example, hotel customers can easily compare competitive offerings at online reservation channels such as Expedia, Orbitz, Kayak and so on. This increased market transparency creates both opportunities and risks to the service companies that must operate within this environment. As various service providers face knowledgeable and sophisticated consumers, there is an increasing urgency to gain an understanding of trade-offs in consumer's choices so that appropriate pricing decisions can be made. The reader is referred to Talluri and Van Ryzin (2004a, b) for a comprehensive review of this literature with respect to pricing and revenue management.

To that end, service providers such as airlines, hotels and so on have expressed a heightened interest in analytic techniques that allow explicit modeling of the drivers of consumer choice, thus providing the understanding that service providers require to effectively design and price their product offerings. The purpose of this article therefore is to illustrate the use of an effective market-utility-based approach known as discrete choice analysis (DCA) in making pricing decisions in services. During the recent years, discrete choice modeling has emerged as an effective approach for developing analytical models and for estimating relative weights of service parameters based on empirical data. For estimation purposes, typically one of the two forms of data is used: transactional data captured in the databases (revealed-preference (RP) data); or primary experimental data (stated-preference (SP) data). In this article, we provide detailed illustration of both approaches for pricing decisions for hospitality

services. Finally, we discuss the managerial implications of the discrete choice modeling approach described earlier in the article.

The remainder of this article is organized in the following manner: first, we briefly review the theory of DCA; second, we provide examples of discrete choice pricing models developed from available archival data in the hospitality industry; third, we provide examples of pricing models derived from experimental DCA; fourth, we discuss the strengths and weaknesses of each approach as well as available methods for combining the pricing models based on two sources of data; finally we present discussions, conclusions and managerial implications of this research.

Discrete Choice Analysis

DCA provides a systematic way to identify the implied relative weights and attribute trade-offs revealed by the choices of decision makers (for example, a customer or a manager). DCA has been used to model choice behavior in many business and social science fields, and introductions to and extensions of DCA can be found in sources such as Ben-Akiva and Lerman (1991), Hensher and Johnson (1980) and Louviere (1988), Gensch and Recker (1979), Green and Krieger (1996), Guadagni and Little (1983), Louviere and Timmermans (1990), and McFadden (1986). Thus, rather than repeat what is already well known, we only summarize the main ideas behind the approach in this article. Naturally, DCA is not the only approach that has been used to understand and model consumer decision making, but it has proved particularly valuable in many hundreds of applications since its introduction by McFadden (1986). The applicability of a particular model depends on the assumptions, theoretical foundations and scientific methods used in modeling, data collection and analysis. Therefore, although the following subsections only provide a general overview of DCA, it should be noted that the approach can be adapted to fit many specific research situations and applications.

Information Integration Theory in psychology (for example, Anderson, 1981, 1982) and random utility theory in econometrics (for example, Hensher and Johnson, 1980; Ben-Akiva and Lerman, 1991) provide the theoretical basis for DCA. In particular, research suggests that after acquiring information and learning about possible alternatives, decision makers define a set of determinant attributes to use to compare and evaluate alternatives. After comparing available alternatives with respect to each attribute, decision makers eliminate some alternatives and form a final choice set containing a few alternatives. They then form impressions of each alternatives' position on the determinant attributes, value these attribute positions vis-à-vis one another (that is, make trade-offs) and combine the attribute information to form overall impressions of each alternative.

Economic choice theory assumes that individuals' choice behavior is generated by maximization of preferences or Utility. Louviere (1988) defines utility as 'judgments, impressions, or evaluations that decision makers form of products or services, taking all the determinant attribute information into account'. The idea of Utility Maximization and its relation to human choice behavior is not new. For example, McFadden (1986) quotes from a 1912 economics text by Taussig: 'An object can have no value unless it has utility. No one will give anything for an article unless it yield him satisfaction. Doubtless people are sometimes foolish, and buy things, as children do, to please a moment's fancy; but at least they think at the moment that there is a wish to be gratified'.

It is now well known that the conditional probability of choosing an alternative from a given choice set can be expressed as a multinomial logit (MNL) model if the random components of utility (the errors in the valuation process) are IID Gumbel-distributed random variates (Gensch and Recker, 1979; Guadagni and Little, 1983; Louviere, 1988; McFadden, 1986; Louviere and Timmermans, 1990; Ben-Akiva and Lerman, 1991). Other forms of choice models (for example, nested logit (NL) models) can be derived by relaxing the IID error assumption, but for purposes of illustration and exposition, this article focuses on the MNL model. The MNL model is expressed as

(1)

$$(P_j|C_n) = \frac{e^{V_j\mu}}{\sum_{k=1}^n e^{V_k\mu}}$$

where V_j represents the systematic component of utility of alternative j in a choice set C_n that includes n alternatives. The parameter μ represents constant scale for underlying Gumbel distribution.

Furthermore, V_j can be written as

(2)

$$V_j = \sum_{a \in A} \beta_a X_{aj}$$

where β_a is the relative utility associated with attribute a of the alternative.

In practice, either actual transactional data captured in databases (known as RP data) or experimental choice analysis (known as SP data) is used to estimate β_a associated with equations 1 and 2 using maximum likelihood estimation techniques. The next section of the article provides examples of each of the two estimation approaches for making pricing decisions in hospitality services.

Revealed-Preference Models

For the RP models, we first describe the data and preliminary preparations involved before estimation of the discrete choice model, and then present the results of the models developed.

Data and Analysis

Hotel booking data were obtained via a third-party online service provider. The data consisted of the browsing history as well as the consumer's final purchase decision that being the choice of a hotel property for a given stay duration and rate. Only sessions in which a purchase occurred are recorded, with the choice set consisting of all available properties viewed as part of the web search. Thus, the RP data used in the subsequent analysis may be characterized as a consideration choice set, as only the properties viewed or considered are included in the database. The raw data consisted of 45 000 + entries or records pertaining to 268 individual consumer shopping sessions, where a session is defined as all properties viewed from the start of the search process until the final purchase decision is made. Thus, within a given session, the same property may be included more than once if it was viewed multiple times. In addition, some variable values for identical properties may change as the search progresses depending on the search criteria employed. The records were combined data from four separate markets.

A unique Session ID tracked the properties viewed during each session, and a Property ID established the final purchase decision for each session. Attributes for each property included the average daily rate (ADR), hotel rating (star rating), distance from search location (if specified), distance from the nearest major airport, duration of stay, days before arrival based on the booking date and the display position for each property as a function of the consumer's search. Table 1 lists the key data entry items and describes each. As is common with RP data, information pertaining to consumer demographics was not available, including the type of trip involved (that is, business versus leisure), household income and so on. Also lacking were property brand or availability of specific amenities at a given property.

An initial data cleansing step was performed to facilitate data analysis and fitting the discrete choice models. The data removed primarily consisted of multiple views of the same property, based upon the Property ID, multiple times during an individual session. These duplicates were removed so that only a single record for each property was contained in a shopping session, as the multiple viewings of a single property does not constitute additional properties available in the choice-set. Also removed were sessions that involved only a single property (the property selected), as these choice sets are

deterministic and contain no information regarding how the consumer made trade-offs among attributes. Finally, sessions in which the ‘booked property id’ did not match properties in the session were excluded, as there was no way of retroactively identifying the correct choice. After data cleansing, the remaining data consisted of 6700 + entries for 242 individual booking sessions.

Preliminary Analysis

To simplify the descriptive statistics, and allow for ease of comparison to the SP data later, for our analysis, we grouped the hotels into three categories, economy (up to 2.5 stars), midrange (3–4 stars) and upscale (4.5–5 stars). The choice set was composed of 27 per cent economy, 69 per cent mid-range and 4 per cent upscale properties. There were 297 unique properties within the economy choice set, 528 in the mid-range and 19 in the upscale. The choice set was proportionally similar to the final choice selections, which were 28 per cent, 66 per cent and 7 per cent, respectively. Nine of the 242 bookings, or less than 4 per cent, involved Saturday night stays. As the majority of the data consists of trips with no weekend bookings, it is similar to the overall statistics. However, trips involving a Saturday stay had a much higher proportion of economy hotel bookings, and, in particular, it is noted that final bookings did not include a single upscale hotel. Further, the length of stay was much higher for Saturday versus no Saturday stay, with Saturday stays averaging 6.4 days in length (range of 5–11 days) compared to the average of 1.9 days (range of 1–5 days) for bookings with no Saturday stay. Table 2 summarizes the choice sets for the different hotel types.

Table 1: Revealed-preference database fields

<i>Entry</i>	<i>Description</i>
Session ID	Identifies the user session activity that resulted in a booking
Property ID	A unique code for the property being displayed in the shopping results
Average Daily Rate (ADR)	The average daily rate in US dollars for the room type and property
Hotel (Star) Rating	Numeric rating for this property
Distance from Search Location	Hotel distance in miles from unique search point (if specified)
Distance from Airport	Hotel distance in miles from nearest major airport
Duration of Stay	Booking length in days
Days before Arrival (DBA)	Number of days property was booked before arrival
Display Position	Position of the property being displayed relative to other properties

Average prices for all available economy hotels were US\$77, \$199 for mid-range and \$451 for upscale hotels. Average prices for booked properties across the three categories were \$77, \$199, and \$357, respectively.

All of the bookings were made within a few days before arrival (DBA), with the largest DBA being 3 days. This short reservation window, as well as the lack of a Saturday night stay would indicate that the data set is composed of primarily business travelers.

Revealed-Preference Choice Models

The simplest and most commonly employed choice model is the MNL model, at least in part owing to its analytical tractability. In general, all of the attributes listed in Table 1 are available to develop a choice model to predict the selection of a particular property. However, as is often encountered in RP data, issues related to the data and collection methods make it difficult or impossible to incorporate in the modeling process. For ‘distance from search location’, though likely to be significant (or at least relevant) to those entering or inquiring of a specific reference point, the values would only apply to a particular search, and therefore of little value for a hotel property in determining the probability of being selected by an individual customer for the data set. Further complicating the use of this variable is that the reference point was not always included for all properties in a given session, possibly reflecting changing search preferences throughout the shopping process. However, for properties located in and around popular attractions, there is value in realizing the weight or importance placed on relative proximity to these attractions.

Table 2: Hotel type choice set

<i>Hotel type</i>	<i>Overall</i>		<i>Sat stay</i>		<i>No Sat stay</i>	
	<i>Choice set (%)</i>	<i>Chosen (%)</i>	<i>Choice set (%)</i>	<i>Chosen (%)</i>	<i>Choice set (%)</i>	<i>Chosen (%)</i>
Economy	27	28	26	26	36	67
Mid-Range	69	66	69	67	62	33
Upscale	4	7	4	7	1	0

Further, as there is little variation in data associated with Duration of Stay or DBA, we are unable to obtain any measure of the significance of either parameter estimates. Finally, the lack of detailed information on the search process before selection of the property makes it difficult to utilize/interpret the display position variable. In addition, for a given session, the same property may occur multiple times with different display position as the search is re-sorted or otherwise refined. Thus, although this may be an interesting variable to investigate, with no information on the search process, we cannot logically incorporate the display position into the model.

Excluding the above variables, the 'base' MNL model includes 'Average Daily Rate', 'Hotel Rating' and 'Distance from Airport'. Here, we interpret the Hotel Rating data as categorical with a five-star hotel as the reference variable, resulting in the parameter estimates summarized in Table 3.

For all the attributes entered in the MNL model, the parameter estimates are significant at the 0.05 level, with McFadden's $r^2 = 0.040$. Further, the signs of these ADR and Hotel Rating parameters are as expected, with utility increasing in the rating of the hotel while decreasing in price. For Distance from Airport, it is more difficult to interpret the sign, as it could be highly dependent on the purpose of the trip. We could logically expect that proximity to the airport may be more important, and thus decrease utility with increasing distance for business travelers who wish to make quick trips into and out of their destinations. For leisure travelers, it may be plausible to argue that proximity to the airport is a negative, and utility would increase with increasing distance. As the evidence indicates that the data are primarily that of business travelers, the negative parameter value would appear to be consistent with these expectations.

To further facilitate our discussion with respect to the stated-choice model to be presented in the subsequent section, we evaluate a slightly different approach to modeling the data set given. Here, we use the broad categories (economy, mid-range and upscale) defined previously to define the hotel types as opposed to the direct Hotel Ratings above. As hotel types are strictly categorical, we use dummy variables in our model, with two dummy variables, economy and mid-range, and upscale representing the reference category. The results are presented in Table 4.

Similar to the above base model, we see that ADR and the hotel rating, given as categories, are still significant at the 0.05 level. However, the P-value for the Distance from Airport variable has increased slightly, but is still significant at the 0.10 level. In addition, signs are again consistent with expectations. Further, the relative magnitudes of the categorical variables (economy, mid-range) are reasonable as we expect them to be negative, relative to the reference category upscale, and that the economy category would be more negative than that of the mid-range property.

A final model run was to generate parameter estimates, using only the hotel categories and ADR variables, for later comparison to results of the SP data. The results are presented in Table 5. As above, parameter estimates are highly significant and all correctly signed, with only slight changes in the estimated values. Elimination of the Distance from Airport variable, which was not included in the SP choice experiments, allows for easier comparison of the price parameter across the two data types.

Table 3: Parameter estimates – base MNL model

<i>Attributes</i>	<i>Estimate</i>	<i>SE</i>	<i>t-value</i>	<i>P> t </i>
Average Daily Rate (USD)	-0.00705	0.001201	-5.87	<0.0001
Distance from Airport (miles)	-0.0148	0.007172	-2.07	0.0384
1 star	-1.7256	0.653	-2.64	0.0082
2 star	-1.2526	0.4112	-3.05	0.0023
3 star	-1.0859	0.3609	-3.01	0.0026
4 star	-0.7574	0.3298	-2.3	0.0216

N=242 observations; $\rho^2=0.040$; Akaike Information Criterion (AIC)=1401.

Table 4: Parameter estimates – hotel categories model

<i>Attributes</i>	<i>Estimate</i>	<i>SE</i>	<i>t-value</i>	<i>P> t </i>
Average Daily Rate (USD)	-0.00616	0.001025	-6.01	<0.0001
Economy	-0.9343	0.3793	-2.46	0.0138
Mid-Range	-0.7351	0.3236	-2.27	0.0231
Distance from Airport (miles)	-0.0136	0.007078	-1.93	0.0541

N=242 observations; $\rho^2=0.037$; AIC=1402.

The above models, with or without the Distance from Airport variable and including either the direct Hotel Ratings or the Hotel Categories can be incorporated into a simple predictive model of market share for a given price and set of available properties. For example, take a market containing a single unit in each of the economy, mid-range and upscale categories, at the average prices given previously of \$77, \$199 and \$357, respectively. Using the parameters given in Table 5 in equations (1) and (2), we would estimate the probability of sale of 0.44, 0.29 and 0.27, respectively. We can also compare the overall market share based on the choices made to that predicted for all economy, mid-range and upscale properties. Taking the overall choice set, we evaluate the market share for each category as the sum of probabilities for the given ADR and Distance from Airport. Using the model parameters in Table 4, we see that the predicted results, given in Table 6, are of similar magnitude to the actual choices.

Stated-Preference Models

During the recent years, a number of studies have used experimental DCA within the context of services. For example, based on discrete choice data collected at a large international airport, Pullman et al (2000) developed a framework matching the needs of multiple market segments with service offerings. Eastman and Pullman (2001) developed a mathematical modeling formulation of the sellers' utility problem within the context of new service design using discrete choice data. Verma et al (2001) presented a non-linear optimization model linking customer preferences obtained from discrete choice

analysis, production cost and operating difficulty. Verma et al (2004) described the similarities and differences for choice of online financial services for different market segments. In another paper by Garrow et al (2007), the authors design a SP survey to examine airline customers' willingness to pay for products offered through online channel for business and leisure travelers. In addition to price, the survey evaluated departure and total travel time as well as qualitative issues such as legroom. They then estimate price sensitivity to the varying combinations via MNL and NL models and how the different attributes affect final preferences.

Table 5: Parameter estimates – price category-only model

<i>Attributes</i>	<i>Estimate</i>	<i>SE</i>	<i>t-value</i>	<i>P> t </i>
Average Daily Rate (USD)	-0.00521	0.000871	-5.98	<0.0001
Economy	-0.9796	0.3779	-2.59	0.0095
Mid-Range	-0.77	0.3239	-2.38	0.0174

N=242 observations; $\rho^2=0.034$; AIC=1404.

Table 6: Actual versus predicted

<i>Hotel type</i>	<i>Choice set</i>	
	<i>Chosen (%)</i>	<i>Predicted (%)</i>
Economy	28	34
Mid-Range	66	63
Upscale	7	3

Stated-Preference Data and Analysis

Experimental DCA involves careful design of service profiles; in this case, a specific hotel and choice sets (a number of service alternatives) in which two or more service alternatives are offered to decision makers and they are asked to evaluate the options and choose one (or none). Each subject in a DCA experiment typically receives several choice sets to evaluate (for example, 8–32 sets), with two or more hypothetical services to choose from in each set. The design of the experiment is under the control of the researcher, and, consequently, the decision makers' choices (dependent variable) are a function of the attributes of each alternative, personal characteristics of the respondents and unobserved effects captured by the random component (for example, unobserved heterogeneity or omitted factors).

DCA applications based on discrete choice experiments typically involve the following steps: (1) identification of attributes, (2) specification of attribute levels, (3) experimental design, (4) presentation of alternatives to respondents and (5) estimation of the choice model.

Attributes and Experimental Design

Before finalizing the experimental attributes and levels with the discrete choice customer survey, we conducted extensive qualitative research (Verma et al, 1999). We interviewed managers from economy, mid-range and upscale hotels and several business and leisure hotel customers. On the basis of qualitative data and a review of academic and practitioner's literature on the topic, we identified five broad constructs of hotel attributes to be varied in discrete choice experiments. They are – hotel type; price; loyalty/frequent user programs; eating options; office facilities and technology options; customization options; and hotel amenities. Each of these constructs was further expanded into attributes (each with two or more levels).

The 'hotel type' construct consisted of three attributes: economy, mid-range and upscale. Each attribute was represented by four of the six experimental levels: motel, bed & breakfast inn, independent boutique hotel, standardized hotel affiliated/operated by recognized chain, boutique hotel operated by a recognized chain and convention style hotel. The 'technology' construct was described by three attributes: internet access in room (none, free, \$5 or \$10/day), business center (not available; full-service & centrally located; multiple business kiosks, in-room printer, fax and so on), and availability of internet reservations (yes, no). The last construct, customization, includes different alternatives that match a person's lifestyle. Customization was described by five attributes: ability to bring small pets to room (yes, no), availability of flexible check-in/ check-out times (yes, no), ability to personalize in-room décor (yes, no), childcare (not available, fee-based nanny and/or kids club for infants and up to 12-year-old kids) and in-room kitchen facilities (none, coffeemaker, microwave, refrigerator and full kitchenette). Finally, pricing construct was nested within the hotel type. The economy hotel prices varied between \$40 and \$ 70; mid-priced hotel prices varied between \$85 and \$130; and upscale hotel prices varied from \$140 and \$200.

After finalizing the list of attributes and their experimental levels, we designed 64 orthogonal profiles that allowed us to reliably estimate the main effects of all the hotel attributes described above (Verma et al, 1999). To enhance the realism of the task, a full-profile approach was used in presenting the choice sets (Green and Srinivasan, 1990), that is, each profile shown to the respondents simultaneously described some combination of all the attributes. Within the actual survey, three hotel

profiles (one economy, one mid-range and one upscale) were shown to respondents at the same time and they were asked to choose one hotel (or neither) which varied from each other on numerous attributes simultaneously. Each respondent evaluated eight hotel choice-sets. In addition to the hotel choice task, the survey instrument included several questions about respondents' past hotel visits as well as demographics.

We pre-tested the survey with 25 randomly selected hotel customers to ensure ease and comprehension of the task, as well as to ensure reliable data collection methods. Average time for completing the entire survey was approximately 20 min and respondents did not indicate difficulty in comprehension.

Sampling Frame and Data Collection

The population of interest consisted of business and leisure travelers who stayed in economy, mid-range or upscale hotels. To obtain a representative sample (or as close to it as possible), we acquired from a third-party vendor a reliable electronic mailing list of 4000 potential respondents with residences scattered across the United States from a well-reputed marketing research company. The mailing list contained a sample of respondents balanced according to the US census data validated by various demographics criteria. Each of the potential respondents received an email invitation to participate in the survey from the lead researcher. By participating in the survey, a respondent had the ability to participate in a raffle to win one of the 10 gift certificates for \$100. From the initial list of potential respondents, approximately 2500 chose to participate in the survey. Approximately 40 per cent of the respondents answered negatively to the screening question (have you taken a business or leisure trip during the last 1 year which required a hotel stay?) and were not allowed to continue with the survey. At the end of a 3-week data collection period, a total of 930 respondents completed and returned the survey (each received a second email reminder). As there was no indication of any response bias, the analysis presented in this article is based on the survey data collected from all the respondents.

Stated-Preference Choice Models

We estimated hotel choices MNL models for two groups of respondents: business travelers and leisure travelers. Both models were statistically significant. Furthermore, as a direct comparison of utilities or β parameters for two MNL models is inappropriate, because the models contain an embedded Gumbel scale parameter (μ), we used an appropriate χ^2 -test developed by Swait and

Louviere (1993) to test the similarities or differences across the two estimated MNL models. This procedure first identifies the optimal Gumbel scale for the second model relative to the first, and then compares the two models using a χ^2 statistic. The two models were shown to be statistically different from each other. A similar example is presented in Iqbal et al (2003).

While the discrete choice experiment described earlier was quite elaborate in terms of number of variables used, for the sake of clarity and brevity we only discuss the results related to hotel pricing parameters. Furthermore, the results are presented in a graphical format below for the ease of understanding. Figure 1 shows the relative β parameters for business travelers for economy, mid-priced and upscale hotels. A higher number implies that customers assign higher utility to the specific price. As expected, utilities reduce as the prices increase for mid-priced and upscale hotels. For economy hotels, we observe an almost flat utility curve. This implies that within the economy hotel category within the \$40–\$70 price ranges, the customers were relatively price-insensitive compared to the other attributes varied within the choice experiment.

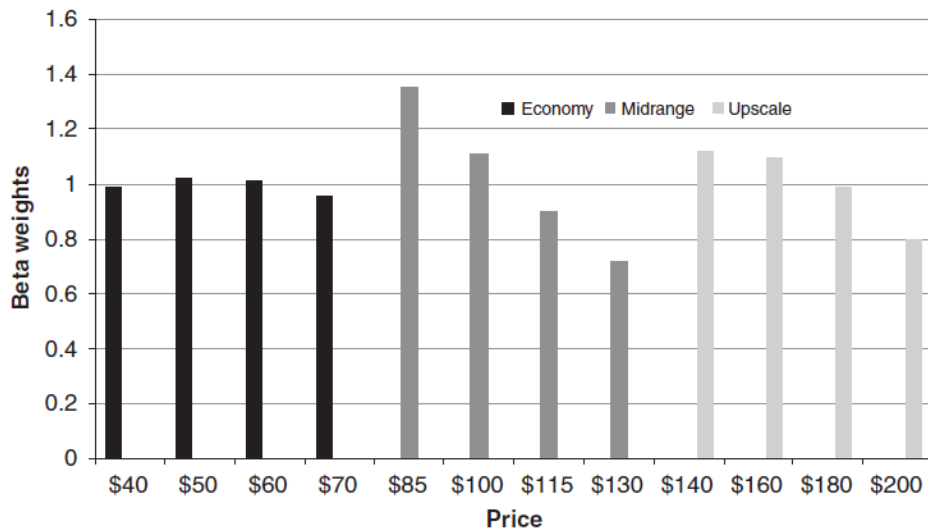


Figure 1 Price Coefficients for Business Travelers

Figure 2 shows the relative b parameters for leisure travelers for economy, mid-priced and upscale hotels. In this case, all three price curves show a downward slope, and the shapes of the curves are slightly different from the business travelers models.

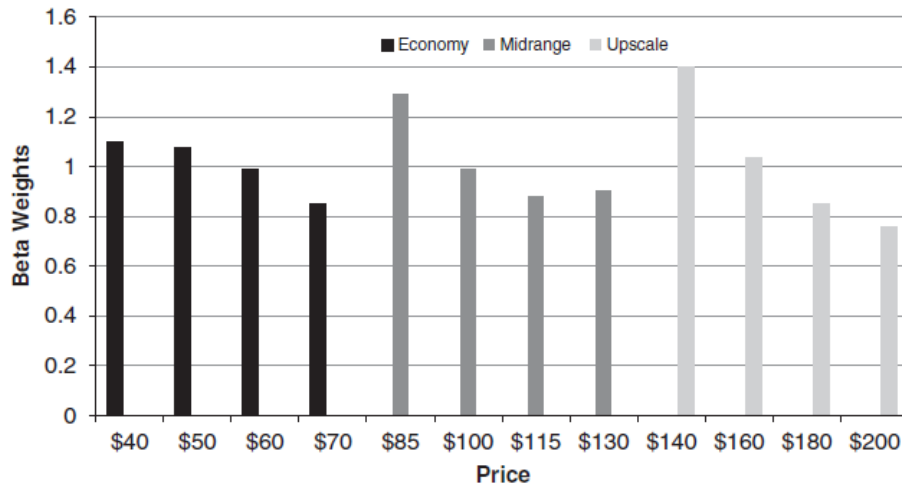


Figure 2 Price Coefficients for Leisure Travelers

Discussion

In the preceding sections, we discussed and developed two separate types of discrete choice models, based on two different types of data, RP and SP data. For the RP data-based MNL model, primary parameters (ADR, Hotel Ratings) were statistically significant and as importantly correctly signed. Miles from the airport was also significant at a minimum of the 0.10 level. The importance of this particular parameter may be reflective of the type of data, or more to the point, the type of consumer represented, in that there is reasonable evidence that the data set consists primarily of business travelers. However, the extent of the correlation among the parameters, in particular the ADR with both Hotel Rating and Miles from the Airport, is cause for concern. RP data has the advantage of ‘face-visibility’ as it is consumer’s actual choices faced with real constraints on their own resources and the options available. The data set used for the RP model, similar to that of Anderson and Xie (2009), is from complete shopping sessions at an online travel agent containing data on both completed transactions as well as sessions where no sale was made. This is unique RP data that allow traditional maximum likelihood estimation of parameters unlike other RP data, similar to Vulcano et al (forthcoming), where the data consist of only sales data and require use of methods similar to the EM algorithm for parameter estimation.

However, the weaknesses tend to mirror the strengths of SP data. The data often lack variation in attributes and are restricted to ‘what is’ available, which in turn leads to difficulties in establishing significance of the parameter estimates, although this was not the case in the above model. Further, the availability of ‘good’ data is somewhat limited, and it generally lacks customer-specific attributes. Often,

a significant amount of data cleansing is required before usage. The primary limitation of the RP data and the subsequent model above is the lack of detailed knowledge of the properties that, along with price, drive the consumer's final choice, along with specific customer attributes. Also missing are individual consumer data that would allow us to more accurately segment our customer base, and make better decisions depending on the type of demand we face. Finally, the level of correlation is a common problem with actual (revealed) data sets. However, obtaining information of this type is one of the particular strengths of the SP method. Using a properly designed choice experiment, we are able to collect and assess the impacts of various combinations of attributes on the choices of various market segments. Further, as the levels of the attributes are controlled, the researcher is able to eliminate the problems of correlations observed in RP data.

One of the advantages of using consumers' stated preferences, as opposed to revealed-preferences, based on market data, is that it allows for wider variation in the available attributes, including price, than that which may occur in the market. This allows estimation of attribute elasticity and cross-elasticities, which may not be possible otherwise, and to gain the desired understanding of the consumer's choice behavior. Further, we can control for desired consumer-specific socioeconomic factors, increasing our ability to segment or differentiate our customers and account for customer heterogeneity in the purchase decision. Finally, although not important in our current context, it allows for investigation of possible or planned offerings not currently available. Of course, the main drawback of the SP method is that consumers do not necessarily behave as they say they will. Thus, there is often a response bias, with decisions or choices made during choice experiments not consistent with actual behavior. Further, the flexibility in defining attributes and product combinations can result in choices that are not reasonable or practical in actual decision-making settings. For example, in the above models, to control the size of the choice experiment presented, hotels were categorized as Economy, Mid-Range and UpScale, whereas customers choices are actually based upon hotel ratings, typically ranging from 1 to 5 'stars', similar to that available in the RP data set.

Recent Advances in Discrete Choice Modeling and Analysis

Similar to any analysis tool, the science of discrete choice modeling continues to evolve as researchers in various academic disciplines pursue research projects with varied focus and emphases. At the same time, the art of choice modeling is also evolving rapidly as information technology makes it possible to develop more realistic choice experiments. Some trends relevant to service sector applications are described below.

Emergence of Multimedia-Driven Choice Experiments

Even a few years ago, a typical implementation of experimental choice modeling involved developing paper-pencil surveys in which respondents were subjected to a series of preconfigured, table-like formatted choice scenarios. Choice sets were presented as static tables with little room for customization to zero-in on the respondent's most interesting purchase drivers. However, the most recent advances in information technology – including broadband internet connections, digital imaging and streaming video technologies, and almost unlimited computing resources and sophisticated programming languages – allow researchers to develop very realistic and highly customizable choice experiments specific to each respondent, resulting in visually appealing and easy-to-use formats triggering a high level of respondents' involvement. As mentioned earlier, these advances are highly relevant to service sector applications because they allow the researchers to construct realistic experimental scenarios.

In our recent studies, we have extensively used web-based technologies (with hyperlinked pictures or written illustrations, brand logos, audio and video files) to realistically illustrate choice scenarios to respondents in service applications. For example, in an ongoing study, several descriptions of 'service scripts' in face-to-face customer interactions in a hospitality setting are being first acted by professional actors (Victorino and Verma, 2007). The video clips of the service scripts along with other features of the service interactions are being presented to the customers in the form of a discrete choice experiment. In another study in a retail setting, we first used a series of screens (each with several pictures and detailed descriptions) to describe the customer service, shopping experience, parking convenience at a futuristic shopping center (Verma et al, 2008). Later when the respondents were presented with the discrete choice exercise, the earlier descriptions were also available as hyperlinks for ready reference.

When choice experiments require transferring huge amount of data, we either mail high-capacity portable storage devices (for example, USB storage keys that can contain dozens of MBs for data) to respondents or conduct interview at any site on a wired or wireless laptop computer. Although such options have been available for a while, only recently they have become relatively cost-effective and easy to implement. In fact, we are anxiously anticipating the day when three-dimensional virtual reality technologies will become cheaply available to truly create 'information accelerated' choice experiments. We already see early indications of use of such technologies in limited fashion (for example, launch of a prototype W Hotel in the virtual reality world: SecondLife.Com).

Advances in Experimental Design and Estimation Processes

Although information technology's role in designing realistic experiments is impressive, even more impressive is the 'behind the scene' hard work of statisticians, mathematicians and management science researchers who have been developing advanced procedures for estimating and fine-tuning econometric models to assess the plethora of customer choice situations. For example, most recent advances in Bayesian statistics allow us to estimate choice models for each individual respondent and therefore enable us to fine-tune market segment memberships on a case-by-case basis. Innovative optimization procedures such as chaos theory, neural networks, simulated annealing, genetic algorithm and simulation modeling are being used in various applications to identify optimal product-service design configurations and to link choice modeling results with other managerial decisions (for example, labor scheduling, capital-based resource constraints).

During the early days of choice modeling, researchers often debated about 'how many' market drivers in a choice exercise should be considered as 'too much' information for the respondents. Researchers also debated on how many choice scenarios should be shown to each respondent for developing robust choice models. Although there is still no agreement on many such theoretical and methodological issues, advanced experimental design procedures and relative ease of data collection from larger number of respondents will relieve some of these academic tensions in the future. For example, we used semi-to-completely randomized experimental designs in combination with statistical blocking, and partial experimental profiles to allow respondents to assess a highly complex choice situation in a consumer-oriented service environment. Other advances in the choice experiment design include developing sophisticated hierarchical choice experiments combined with nested and partial profile designs. Although the use of such procedures increase complexity in designing choice studies, data analysis and econometric models estimation, it allows researchers to significantly reduce the choice-task complexity and time requirement for respondents by only showing a few market drivers within each choice set at one time.

Integration with the SP and RP Data Sources

During the last few years, firms have invested heavily in customer relationship management (CRM) systems and information technology in general. Such implementation generates huge amount of customer transaction data (for example, hotel check-in records; use of various facilities; reservation and credit card usage patterns; frequent user/loyalty card records), which can be used to monitor customer preferences over a long period of time. Effective use of CRM data can allow organizations to customize

product-service offerings to usage patterns of individual customers, thereby increasing satisfaction, retention and loyalty. We believe that organizations can gain valuable insights on the impact of new market drivers by combining existing databases with customer responses to carefully constructed choice experiments. As a matter of fact, within the domain of choice experiments, new market drivers can be varied and their relative utilities estimated; thus, choice-modeling results combined with econometric models developed from existing CRM databases can realistically estimate the impact of any new product-service offering within a chosen business context. We believe that the end result of such triangulation will lead to development of highly robust predictive models. The reader, however, should note that extreme caution is needed for such data-merging techniques to isolate any statistical differences owing to the use of multi-methods, otherwise the resulting models might be confounded with random errors. For example, it is possible that mean and/or variance estimates (and therefore the scale parameter) for CRM and choice experiments- based models differ from each other simply because of differences in data collection and estimation techniques. Therefore, the researcher needs to make appropriate corrections within the model estimation procedures to isolate the impact of such errors.

Summary and Concluding Remarks

The purpose of this article was to introduce discrete choice modeling as an approach for assessing customer choices in the service industry. For the ongoing SSME momentum to be successful, we believe that it is necessary that sophisticated customer choice approach, such as discrete choice modeling, become an essential component of the framework. In this article, we have provided several examples of discrete choice studies conducted for a variety of service sector applications. We have also discussed how the science and art of discrete choice modeling continues to evolve rapidly. We hope that researchers interested in service science management and engineering will find discrete choice modeling useful in their future research and applied projects.

References

- Anderson, C.K. and Xie, X. (2009) A Choice Based Dynamic Programming Approach for Setting Opaque Prices. Cornell University. Working Paper.
- Anderson, N.H. (1981) *Foundations of Information Integration Theory*. New York: Academic Press.
- Anderson, N.H. (1982) *Methods of Information Integration Theory*. New York: Academic Press.
- Ben-Akiva, M. and Lerman, S.R. (1991) *Discrete Choice Analysis*. Massachusetts: The MIT Press.

- Easton, F. and Pullman, M.E. (2001) Optimizing service attributes: The seller's utility problem. *Decision Sciences* 32(2): 251–275.
- Garrow, L.A., Jones, S.P. and Parker, R.A. (2007) How much airline customers are willing to pay: An analysis of price sensitivity in online distribution channels. *Journal of Revenue and Pricing Management* 5(4): 271–290.
- Gensch, D.H. and Recker, W.W. (1979) The multinomial, multiattribute logit choice model. *Journal of Marketing Research* 16: 124–132.
- Green, P.E. and Krieger, A.M. (1996) Individualized hybrid models for conjoint analysis. *Management Science* 42: 850–867.
- Green, P.E. and Srinivasan, V. (1990) Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing* 54: 3–19.
- Guadagni, P.M. and Little, J.D.C. (1983) A logit model of brand choice based on scanner data. *Marketing Science* 2: 203–238.
- Hensher, D.A. and Johnson, L. (1980) *Applied Discrete-Choice Modelling*. London: Croom-Helm.
- Iqbal, Z., Verma, R. and Baran, R. (2003) Understanding consumer choices and preferences in transaction-based e-services. *Journal of Service Research* 6(1): 51–66.
- Louviere, J.J. (1988) *Analyzing Decision Making: Metric Conjoint Analysis*. Newbury Park, CA: SAGE Publications.
- Louviere, J. and Timmermans, H. (1990) Stated preference and choice models applied to recreation research: A review. *Leisure Sciences* 12(1): 9–32.
- McFadden, D. (1986) The choice theory approach to market research. *Marketing Science* 5(4): 275–297.
- Pullman, M., Verma, R. and Goodale, J. (2000) Service design and operations strategy formulation in multicultural markets. *Journal of Operations Management* 19(2): 239–254.
- Schwartz, B. (2004) *The Paradox of Choices: Why More is Less*. HarperCollins Publishers.
- Shubik, M. (1987) What is an application and when is theory a waste of time? *Management Science* 33(12): 1511–1522.
- Swait, J. and Louviere, J.J. (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* 30: 305–314.
- Talluri, K.T. and van Ryzin, G.J. (2004a) *The Theory and Practice of Revenue Management*. Boston, MA: Kluwer Academic Publishers, New York.
- Talluri, K. and van Ryzin, G. (2004b) Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50(1): 15–33.

- Verma, R., Iqbal, Z. and Plaschka, G. (2004) Understanding customer choices in E-financial services. *California Management Review* 46(4): 43–67.
- Verma, R., Plaschka, G., Hanlon, B., Livingston, A. and Kalcher, K. (2008) Predicting customer choices in services using discrete choice analysis. *IBM Systems Journal* 47(1): 179–191.
- Verma, R., Thompson, G.M. and Louviere, J.J. (1999) Configuring service operations in accordance with customers needs and preferences. *Journal of Service Research* 1(3): 262–274.
- Verma, R., Thompson, G.M., Moore, W.L. and Louviere, J.J. (2001) Effective design of products/services: An approach based on integration of marketing and operations management decisions. *Decision Sciences* 32(1): 165–193.
- Victorino, L. and Verma, R. (2007) Scripting the Service Encounter: An Empirical Analysis. Working Paper.
- Vulcano, G., van Ryzin, G.J. and Char, W. (forthcoming) Choice-based revenue management: An empirical study of estimation and optimization. MSOM, in press.