Travel Demand Modeling with Behavioral Data

Juan L. Nicolau

1 Introduction

Today's travelers demand personalized and comprehensive experiences, and guided by their personal motivations, they try to back their decisions on recommendations expressed on the Internet. Besides, they write on official and unofficial websites their personal preferences, and tell other travelers about their intentions on their next destinations, plan the itinerary of the visit, compare, make reservations and pay with a few clicks from home just seating at their computer. Also, with their cell phones they build a unique story with pictures and comments on what they see and feel while at the destination. Fuchs, Abadzhiev, Svensson, Höpken, and Lexhagen (2013) indicate that this customer-generated data can be divided into explicitly-provided information through the use of surveys and e-reviews or implicitly-given information via information traces such as Internet-navigation data, online requests, booking, payment data, or tourists' spatial movements through sat navs; distinguishing between structured data (e.g. surveys) and unstructured data (e.g. e-reviews with free text) (Höpken, Fuchs, Keil, & Lexhagen, 2011). Only if firms and analysts were able to manage this amount of information—structured and unstructured—they could identify consumers' preferences and, more importantly, anticipate their decisions to adapt companies' services in real time and in a personalized way (Invat·tur Report, 2015). Certainly, in tourism more than in any other industry, the 3Vs of big data reflect purely the essence of the intricacies that entail managing such plethora of information: in line with Dolnicar and Ring (2014), "Big Data implies the availability of significantly larger, often gigantic, amounts of data (volume) on a continuous basis and often in real time (velocity) from a range of diverse data sources (variety)".

J.L. Nicolau (✉)
University of Alicante, Alicante, Spain
e-mail: jl.nicolau@ua.es

Z. Xiang, D.R. Fesenmaier (eds.), *Analytics in Smart Tourism Design*, Tourism on the Verge, DOI 10.1007/978-3-319-44263-1_3

Observe that one of the great opportunities that Big Data offers the tourism industry resides in the Smart Cities and more specifically in smart destinations, which facilitate the experience and interaction between the destination and the tourist by ensuring sustainable development. As the production and consumption take place simultaneously in tourism, it means that information is being massively produced through all the stages the tourist is going through. Also, one of the great appeals of massive data is its potential to predict phenomena, anticipate behavior, expectations and future needs of tourists, so that smarter and safer business decisions are made (Invat·tur Report, 2015). For example, adjusting prices quickly and competitively in response to an analytically predictable change in travel demand represent an edge over rivals.

It is evident that these advantages come with challenges. While most companies have myriads of data (surveys, internal customer transaction data, quality data (complaints), secondary research reports about trends and markets, and online data), there is a strong need to coordinate all the information sources and put together the data in a manageable way. When it comes to a destination level, this challenge is even more acute as huge volumes of data on customer transactions, needs and behaviour are stored by different stakeholders of the destination (Vriens & Kidd, 2014). In their knowledge destination framework architecture, Fuchs et al. (2013) emphasize the fact that different data sources require different techniques for the data extraction, so that heterogeneous data from distinct data sources should be mapped into a homogeneous data format.

Another challenge is analysis. Be it through methods of data mining (e.g. techniques of machine learning and artificial intelligence) or traditional methods (e.g. regression-like procedures), detection of patterns and relationships in the data is not fully guaranteed unless some empirical-related issues are considered when the modeling of the travel demand is carried out. In the increasing use of Big Data in tourism research (Mellinas, Martínez María-Dolores, & Bernal García, 2015), the review of the literature identifies three relevant aspects of demand analysis (Radojevic, Stanisic, & Stanic, 2015): (1) tourist heterogeneity; (2) the ability to identify all the alternatives available to the tourists when they make their choices; and (3) the inherently hierarchical character of the data at the destination level (e.g. hotels are nested within destinations, destinations within countries).

2 Empirical Results

Vriens and Kidd (2014) outline the key areas where advanced analytics derived from Big Data can provide solutions with special added value. These are market forecasting (especially if a firm operates in multiple markets), quantifying customer needs and motivations (with an emphasis on quantitatively determined emotional states which leads to an improved ability to understand customer needs), analyzing drivers of brand share (e.g. the predictive power of brand perceptions), product and pricing optimization (to find the best mix of attributes to optimize volume, share or profitability), marketing efficiency modelling (to detect how well marketing efforts are working), and customer dynamics (e.g. which customers are most likely to defect and when, or how to determine lifetime value of customers). However, in all cases, when modeling individual behavior with big data three issues are to be considered: heterogeneity, choice set and information hierarchy.

2.1 Heterogeneity in Tourists

The existence of strong heterogeneous demand looking for product and service provision adapted to its specific needs, along with the intensification of competition in the market, has led to heterogeneity identification becoming fundamental to the marketing strategies of organizations and tourism destinations. As the heterogeneity of the market reflects the existence of a diversity of needs and desires and, therefore, of differentiated consumer behaviour among individuals, understanding heterogeneity in tourist preferences is of paramount importance in many tourism marketing actions. Strategically, knowing the distribution of people's responses to destination attributes would guide the design decisions of the tourism products (this insight would not be detected if the preference is observed only at the mean). Operationally, modeling individual-level responses to marketing actions allows tourist firms to adjust allocation of resources across regions, establishments, and tourists.

Despite the fact that segmentation allows the definition of different market segments that group consumers with shared behaviour and needs, nowadays there is more and more importance attached to personalised service for each client. More pro-active consumers and an intense competition increase the demand for better service, better adapted to their individual needs and, therefore, personalized. Tourists expect to be treated as individual clients. This situation leads to the appearance of one-by-one marketing, which entails individual consideration of consumers and a one-by-one service. This approach is the basic pillar of relationship marketing -and, therefore, the application of CRM (*Customer Relationship Management*)-, which is designed to create, strengthen and maintain relationships between companies. Mass marketing has been transformed into fragmented or micro-segmented marketing to satisfy the demands of smaller and smaller segments, even down to the level of the individual customer. So, the key question in the context of Big Data is how to analyze and detect individual preferences of tourists by introducing heterogeneity.

Tourists process and integrate information to choose an alternative (e.g. destination, type of accommodation or method of transport) that maximises their utility. The objective or subjective character with which the researcher examines the result of this choice process determines the different approximations of choice analysis. The study of tourist behaviour and, therefore, of the way in which they process, evaluate and integrate the information used to make a decision, is traditionally made in two ways. The first approximation is centred on the analysis of the *real choices* made by individuals. This approach is based on the Neoclassical Economic Theory and the Theory of Discrete Choice, and assumes the existence of *preferences* that are unobservable to the analyst but that tourists implicitly consider when ranking alternatives, and which are only *revealed* through the real purchase choice. Therefore, this approximation is known as the *Revealed Preferences* approach.

The second approach examines the ranking or scoring according to *preferences*, given by individuals to hypothetical choice alternatives. This approximation is based on the Information Integration Theory and the Social Judgement Theory, and assumes that the decision maker is capable of ranking alternatives according to his/her preferences. In contrast to the previous case, the analyst does not observe the real purchase choice, given that the individual only makes a *declaration of intent* based on their preferences (i.e. which alternative would be chosen if they had to choose from the given possibilities). This approximation, therefore, is known as the *Stated Preferences* approach.

To give an example, an individual declares that Hawaii is the destination he/she would like to go to on his/her next holiday. In other words, the individual selects Hawaii from a series of destinations and, through this *declaration*, preferences are analysed. However, this aspect has been widely criticized, due the fact that this approach does not reflect reality in the sense that the declaration of the preferred alternative of an individual does not necessarily coincide with his/her real behaviour, i.e. with the alternative that is really chosen. The fact that an individual *declares* that he/she would like to go to Hawaii on his/her next summer holiday does not necessarily mean that he/she will go there in the end.

Conversely, the *Revealed Preferences Approach* analyses the real choices made by tourists in order to obtain their preferences. In the example above, the individual *reveals* his/her preferences when, from a group of destination choices, he/she chooses and goes to Hawai. However, one of the weak points of the *Revealed Preferences Approach* derives from the fact that the estimation of preferences is made at a global sample level, which does not allow representation of individual level preferences. If U_{in} is the utility of alternative i for tourist n, explained through the personal characteristic x_n of individual n and through attribute z_i of the same alternative i, then the utility function is expressed as

$$U_{in} = \alpha_i + x_n\beta_i + z_i\gamma_i + \varepsilon_{in}$$

where α_i is the utility constant, β_i and γ_i are the parameters that measure (respectively) the effects of characteristic x_n of the individual and attribute z_i on the utility of alternative i and ε_{in} is the error term.

Specifically, β_i and γ_i represent the marginal utilities of individuals of alternative i; and these parameters allow us to answer questions such as "If a destination improves one of its attributes (for example, the quality and cleanliness of its water), to what extent would preferences for this destination increase?" The value of this tool for the decision making of tourism organisations is unquestionable, as it allows them to know the responses of a series of people to this improvement.

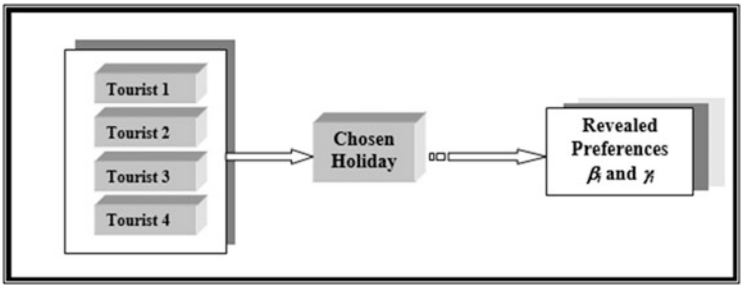

Fig. 1 Linking *sample revealed preferences* through choices made

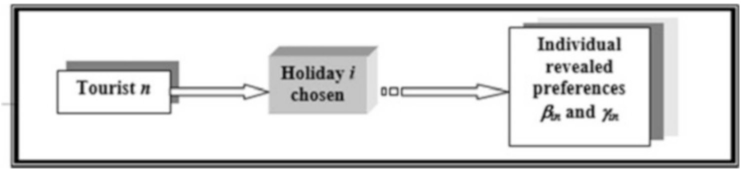

Fig. 2 Individual revealed preferences through choices made

However, note that the estimations of parameters β_i and γ_i are made at the global sample level (see Fig. 1).

What if the estimation of these parameters could be made tourist by tourist? This way, the resulting equation would be

$$U_{in} = \alpha_i + x_n\beta_{in} + z_i\gamma_{in} + \varepsilon_{in}$$

where, in this case, β_{in} and γ_{in} represent the preferences of tourist n around alternative i. Note that now we obtain a parameter for each tourist (and not for the whole sample) (see Fig. 2).

The main implication of knowing the tourist by tourist preference structure is that it allows the adaptation of each product to each individual, as well as the formation of groups of individuals with similar preferences.

In the context of Big Data, most user-generated data is observed data, so revealed preferences can be obtained through this modeling; thus, the introduction of the heterogeneity of tourist preferences into the analysis of the choice process is a major issue.

One of the procedures proposed in the literature to incorporate heterogeneity of preferences assumes the existence of differentiated response parameters for each individual. The most used models in this approach are the random effects models, which model heterogeneity with the assumption that the coefficients of the utility functions of each individual vary according to the probability distribution, either continuous -which gives rise to the Random Coefficients Logit Model- or discrete -which leads to the Latent Class Logit Model-. Initially, the Latent Class Logit Model has been widely accepted in the literature due to the fact that the estimation of the *mass probabilities* -or points where the distribution reaches the greatest *probability masses* allows identification of *latent segments* in the market, which are represented by groups of individuals with similar response profiles. Moreover, in order to segment the market, discrete distribution has an advantage over continuous distribution in that there is no need to assume a concrete probability distribution, as the segments are obtained through empirical data. However, the discrete approach has two important limitations (Allenby & Rossi, 1999): (1) the estimation becomes complex with six or more *mass probabilities*, which hinders the capture of the complete sample heterogeneity; and (2) the impossibility of identifying the preferences of individuals situated beyond a certain threshold of the distribution function (e.g. in the distribution tails).

Because of this, some authors consider that the optimum method of capturing market heterogeneity is to estimate the parameters of each individual, as this allows the capture of any individual preference structure (Allenby & Rossi, 1999). In fact, this model has enough flexibility to provide a tremendous range within which to specify individual unobserved heterogeneity. This flexibility can even offset the specificity of the distributional assumptions.

2.2 Choice Set

One major issue in demand analysis is the definition of the choice set, that is, the alternatives from which the tourist selects the preferred option. The analyst is always uncertain about the set of alternative that the individual considered when making the decision. Obviously, the more data exists, the more alternatives, and the more the potential error of omitting alternatives considered by the tourist but not regarded by the analyst will increase. Therefore, when using Big Data, not only is important to collect information on the selected alternative, but also on the whole choice set. So, in the analysis of Big Data of hotels and airlines, all alternatives on the "screen" presented to the customer should be stored in a database. The first challenge here is to store the data. The size of the data increases 50-fold if the choice set has 50 alternatives. After storing the data in json files, the next challenge is how to analyze it. Bookings with choice sets can be used in discrete choice models. The fact that there are individuals who have been presented with different sets of alternatives can be easily managed with these models (Train, 2009).

2.3 Information Hierarchy

In the context of Big Data, researchers examine data from multiple entities (e.g. hotels and destinations). Certainly, this type of data is inherently hierarchical, as hotels are nested within destinations, and destinations within countries, and ignoring this effect might reduce the validity of results and conclusions (Radojevic et al., 2015). In an attempt to mimic and reflect the way people process information, hierarchical decision processes should be considered when analyzing travel demand. This statement is based on the idea that, when confronted with many alternatives, people tend to follow strategies of the "satisficing" type (satisfice = *satis*fy + suf*fice*), as defended by Simon (1955), where alternatives are considered sequentially. This proposal is further backed by: (1) The Associative Network Theory (Collins & Loftus, 1975) which, through "cognitive networks", explains the way the information on alternatives is represented, processed and activated in consumers' memory through nested links. Specifically, this theory proposes that information is held in the memory through an interrelated structure of "cognitive networks", in which each cognitive network has various "nodes" and "links" between different nodes. (2) The Cybernetic model of decision making (Steinbruner, 2002), which explains how the consumer can follow a hierarchical choice process to reduce uncertainty and complexity in the decision task. Destination choice has numerous factors for consideration and problems related with available information, so they are inclined to use a hierarchical strategy for their choice to reduce uncertainty to a certain manageable level.

Radojevic et al. (2015) use a four-level mixed linear model with random intercepts for country of origin, destination, and hotel, so that the implicit hierarchy is considered in the analysis; and Park, Nicolau, and Fesenmaier (2013) propose the Destination Advertising Response (DAR) model in which they examine the advertising information effects on a sequential travel decision process, including different travel products advertised. Specifically, the choice in the first stage is between visiting and not visiting a destination. Once individuals decide to visit a tourism destination in the first stage, those travelers go on to a second stage where they make a decision whether or not to purchase advertised items. People who select advertised items in the second stage go on to a third stage choice among six different advertised items.

Let us imagine a group of people has to decide the hotel where they are staying. Accordingly, the previous sections show the issues that must be considered when modelling this decision. First, as not all people behave the same way, it means that their preferences are dissimilar or, more formally, there is heterogeneity. Second, if each of them has searched the hotel availability in different periods of time (say, different days), the set of alternatives (e.g. types of hotels) that each individual has been confronted with might be different. Third, before selecting the hotel, they have had to choose the destination; and even before, they have had to decide on whether they take a vacation or not; thus, a hierarchical structure is implied.

With regard to heterogeneity of preferences, a choice model that allows the coefficients of the preferences to vary over tourists is required. Therefore, the utility of an alternative i for tourist t is defined as $U_{it} = X_{it}\beta_t + \varepsilon_{it}$ where X_{it} is a vector that represents the attributes of the alternative and the characteristics of tourists; β_t is the vector of coefficients of these attributes and characteristics for each individual t which represent personal tastes; and ε_{it} is a random term that is iid extreme value. This utility specification leads to a Random Coefficient Logit Model (RCL) in which its coefficients β_t vary over tourists with density $g(\beta)$. Thus, the non-conditional probability is the integral of $P_t(i/\beta_t)$ over all the possible values of β_t:

$$P_i = \int_{\beta_t} \frac{\exp\left\{\sum_{h=1}^{H} x_{ih}\beta_{th}\right\}}{\sum_{j=1}^{J_t} \exp\left\{\sum_{h=1}^{H} x_{jh}\beta_{th}\right\}} g(\beta_t|\theta) d\beta_t \tag{1}$$

where J_t is the number of alternatives the tourist t has been presented to, g is the density function of β_t, and θ are the parameters of this distribution (mean and variance). So far, this model considers both heterogeneity and the existence of different choice sets for each individual. As for the hierarchical structure, the RCL model is flexible enough to represent different correlation patterns among non-independent alternatives; in fact, it does not have the restrictive substitution patterns of traditional Logit models, allowing representation of any random utility model (McFadden & Train, 2000). In particular, an RCL model can approximate a Nested Logit (NL), which is appropriate for non-independent and nested choice alternatives. Following Browstone and Train (1999), the RCL model is analogous to an NL model in that it groups the alternatives into nests by including a dummy variable in the utility function which indicates which nest an alternative belongs to. Technically, the presence of a common random parameter for alternatives in the same nest allows us to obtain a co-variance matrix with elements distinct from zero outside the diagonal, obtaining a similar correlation pattern to that of an LN model. Regarding the previous example, the analyst should consider that all the hotels in destination A belong to the same nest. So, this fact has to be included in the model. Let us assume that the utility function of alternative i is $U_{it} = \beta x_t + \mu_t z_i + \varepsilon_{it}$, where μ is a vector of random terms with zero mean and variance σ^2_μ, and ε_{it} is independently and identically distributed extreme value with variance σ^2_ε. The non-observed random part of the utility is $\eta_i = \mu_t z_i + \varepsilon_{it}$, which can be correlated with other alternatives depending on the specification of z_i. For example, assume that four alternatives "Hotel 1 in Destination A" (H1A), "Hotel 2 in Destination A" (H2A), "Hotel 1 in Destination B" (H1B) and "Hotel 2 in Destination B" (H2B) have the following utility functions:

$$\begin{aligned}
U_{H1A,t} &= \beta x_t + \mu_t + \varepsilon_{H1A,t} \\
U_{H2A,t} &= \beta x_t + \mu_t + \varepsilon_{H2A,t} \\
U_{H1B,t} &= \beta x_t + \varepsilon_{H1B,t} \\
U_{H2B,t} &= \beta x_t + \varepsilon_{H2B,t}
\end{aligned}$$

If two alternatives H1A and H2A are truly correlated, their covariance is Cov $(\eta_A, \eta_B) = E(\mu_t + \varepsilon_{At})(\mu_t + \varepsilon_{Bt}) = \sigma^2_\mu$, which permits identification of correlated non-independent alternatives. Therefore, if the parameter of the variance σ^2_μ, is significantly different from zero, it implies that the alternatives are correlated and must be "closer to each other" and even at the same level of decision. In the context of this example, it means that the two hotels belong to the same "nest", i.e. the same destination (Fig. 3) The advantage of this procedure is that you can test as many nest combinations as "paths to the final decision" the tourist might have in mind. If one were to hypothesize that a tourist, for some reason, chooses the "type of hotel" first (say, the number of stars a hotel has: for example, Hotel 1 means five stars, Hotel 2 means four stars, and so on) and then selects the destination (Fig. 4), the model can accommodate this situation just by defining the non-observed random part of the utility function. Accordingly, assuming that H1A and H1B are hotels with the same number of stars (and the same happens with H2A and H2B), the specification of the utility function would be like this:

$$\begin{aligned} U_{H1A,t} &= \beta x_t + \mu_t + \varepsilon_{H1A,t} \\ U_{H2A,t} &= \beta x_t + \varepsilon_{H2A,t} \\ U_{H1B,t} &= \beta x_t + \mu_t + \varepsilon_{H1B,t} \\ U_{H2B,t} &= \beta x_t + \varepsilon_{H2B,t} \end{aligned}$$

This way, the model tests whether the tourists follow the hierarchical decision "first the hotel type and second the destination", rather than "first the destination and then the hotel type". An illustration of testing different hierarchical structures in tourist decisions can be found in Nicolau and Mas (2008).

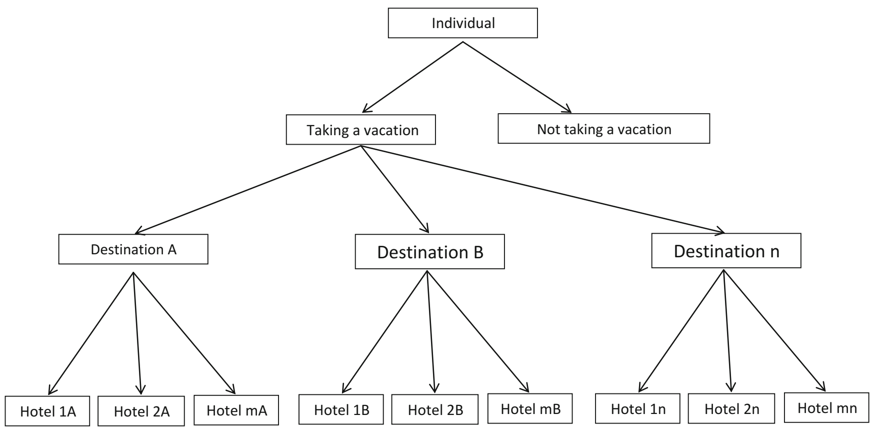

Fig. 3 Hierarchical hotel decision with n destinations and m hotels

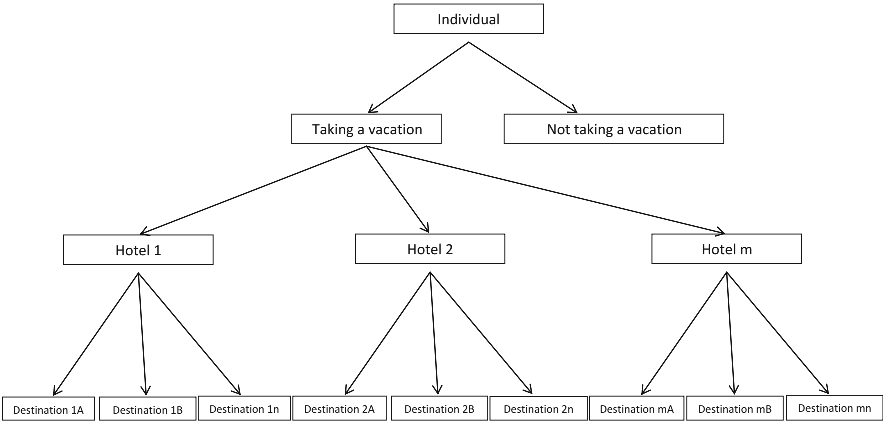

Fig. 4 Hierarchical hotel decision with m hotels and n destinations

3 Research Avenues

This third section explores new avenues for research so that several potential applications are described.

First, knowing the individual utility of a specific tourist gives us information about him or her; information that, he or she himself/herself is not aware that they employed to make the decision. In fact, the estimation of the individual parameters of the utility function of each individual reveals his/her preference structure and allows us to operate with precise information on each individual. At a time when tourists are increasingly demanding and insist on service provision adapted to their specific needs, knowledge of the profile of each tourist allows tourism organizations to offer the most suitable products. Also note that the analysis is based on *real purchase choices* made by individuals (and not on *declarations of intent*), which allows a more accurate representation of the behavior of each tourist.

Second, turning the "market model" into the "click model". The market model is a finance model used to measure the returns of a firm trading on the stock exchange market (for an application in tourism, see Nicolau, 2002). In particular, the rate of returns on the share price of firm i on day t is expressed as:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it}$$

where R_{it} is the rate of returns on the share price of firm i on day t, R_{mt} is the rate of returns on a market portfolio of stocks on day t. The parameters α_i and β_i are the constant and the systematic risk of stock i, respectively, and ε_{it} is the error term. The analogy would consist of estimating the demand of a product by looking at the number of "clicks" (purchasing clicks, liking clicks, acknowledgment clicks, etc.) where the "clicks portfolio of the market" would be the average number of clicks of the top companies in a industry. Actually, this model would permit the estimation of the expected demand (of clicks) on a specific day. Plus, in the same way that we can estimate the difference between the actual and expected returns by calculating the so-called abnormal returns through the formula:

$$AR_{it} = R_{it} - \left(\hat{\alpha}_i + \hat{\beta}_i R_{mt}\right)$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the estimates obtained from the regression of R_{it} on R_{mt} over an estimation period, we could estimate the difference of the actual number of clicks and the expected amount of clicks. This analysis would give information on the success of a new tourism product or the success of an advertising campaign. For example, if "twitter" were treated as a market where information is exchanged, and the number of "tweets" were considered as a measure of repercussion (or hype), it could be interesting to observe the expectations generated by, say, an innovation announcement on a specific day. Paralleling the market model, it would imply observing whether the amount of "exchanged information" (tweets) derived from a firm's release of news on a given day is abnormally superior to the quantity of "exchanged information" in a normal day, and *whether* and *how many good things* are said.

Third, measuring success of anticipation. The WTTC Report (2014) tells the case of "a match made in heaven": *A passenger boards a transatlantic flight, expecting to plug in the earphones for ten hours straight. But much to her surprise, the passengers on either side of her are also journalists heading to the same conference. Big Data has allowed the airline to engineer the seating arrangement; passengers remember the flight with much more fondness.*

The magazine Hosteltur, in a 2013 article, tells the story of the American writer Janine Driver went to a conference in Nashville and told his audience that the Loews Vanderbilt Hotel where he was staying had visited his profile on Facebook, had downloaded a photo of his newborn son and his older brother, had printed it and left on her bedside table. Driver praised the experience. Both cases are the result of the application of Big Data. However, the following step is to determine how satisfied the people involved are. Not just the specific individuals involved in these two previous examples, but in general terms. Would this strategy be generally favored or would it be considered as interference on one's personal life?

Fourth, in line with Nicolau and Mas (2015), detection of the positioning of both collective and individual brands in people's mind can be done without asking the individuals themselves, just by looking at their decisions and actions. Base on the idea that the meaning of a brand is first individually determined according to people's perceptions; it means that these perceptions will have an influence on the way they will socialize and place their ideas about the brand into social discourse. This social discourse can be examined to discover, not only where they went (so that the analyst can build choice models) but also what they think (so that the analyst can uncover destination positioning strategies).

Fifth, the literature shows that the size of the effect of online reviews depends on whether they are positive or negative, giving rise to asymmetric effects, that is, people perceive extreme ratings (positive or negative) as more useful and enjoyable than moderate ratings (Park & Nicolau, 2015). On account of the importance of online reviews for travel demand, more dimensions can be analyzed and even the ratings of specific attributes of a hotel, airline or destination can be examined.

Sixth, blind booking is a strategy in which an airline offers you different known prices for several unknown destinations. The individual's choice is the "price", not the "destination". This context opens up new research lines as the core element for the tourist's decision, i.e. the destination, is no longer essential. So, people choose prices and their preferences are not based on destination attributes other than price.

Seventh, upselling through auctions. When upselling is the result of auctions, large amounts of data can be obtained that can delve into people's psychology as to the effects of prices.

4 Conclusions

This chapter discusses developments and potential analytic approaches to travel demand modeling with *behavioral Big Data*, with the ultimate goal of generating customer-based knowledge through tourists' feedback and information traces. The advantages linked to the use of Big Data are accompanied with challenges. Accordingly, coordination of the different levels of information is a requisite to properly use this flood of information. This is even more relevant when dealing with destinations as the distinct information is stored by different stakeholders of the destination; thus, heterogeneous data from distinct data sources should be mapped into a homogeneous data format.

Regarding the analysis of Big Data, three empirical problems are to be considered: (1) tourist heterogeneity; (2) the ability to identify all the alternatives available to the tourists when they make their choices; and (3) the inherently hierarchical character of the data at the destination level (e.g. hotels are nested within destinations, destinations within countries). Finally, several new avenues for research are presented. The basic idea is that with the use of Big Data and correctly choosing the analytical tool, we can have a profound understanding of today's travelers' preferences; preferences that they might not even be fully aware that they have.

References

Allenby, G. M., & Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics, 89*, 57–78.

Browstone, D., & Train, K. (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics, 89*, 109–129.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychology Review, 82*(6), 407–428.

Dolnicar, S., & Ring, A. (2014). Tourism marketing research: Past, present and future. *Annals of Tourism Research, 47*, 31–47.

Fuchs, M., Abadzhiev, A., Svensson, B., Höpken, W., & Lexhagen, M. (2013). A knowledge destination framework intelligence application from Sweden. *Tourism, 61*(2), 121–148.

Höpken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2011). The knowledge destination—a customer information-based destination management information system. In R. Law, M. Fuchs, & F. Ricci (Eds.), *Information and communication technologies in tourism 2011* (pp. 417–429). New York: Springer.

Hosteltur. (2013). Las redes sociales, arma para personalizar la experiencia del cliente en hoteles de lujo. http://www.hosteltur.com/118410_redes-sociales-arma-personalizar-experiencia-cliente-hoteles-lujo.html

Invat·tur Report. (2015). Bid Data: retos y oportunidaddes para el turismo. http://invattur.gva.es/estudio/big-data-retos-y-oportunidades-para-el-turismo/

McFadden, D., & Train, K. (2000). Mixed MNL models of discrete response. *Journal of Applied Econometrics, 15*, 447–270.

Mellinas, J. P., Martínez María-Dolores, S.-M., & Bernal García, J. J. (2015). Booking.com: The unexpected scoring system. *Tourism Management, 49*, 72–74.

Nicolau, J. L. (2002). Assessing new hotel openings through an event study. *Tourism Management, 23*(1), 47–54.

Nicolau, J. L., & Mas, F. (2008). Sequential choice behaviour: Going on vacation and type of destination. *Tourism Management, 29*, 1023–1034.

Nicolau, J. L., & Mas, F. J. (2015). Detecting free riders in collective destination brands through a hierarchical choice process. *Journal of Travel Research, 54*, 288–301.

Park, S., & Nicolau, J. L. (2015). Asymmetric effect of online consumer reviews. *Annals of Tourism Research, 50*, 67–83.

Park, S., Nicolau, J. L., & Fesenmaier, D. (2013). Assessing advertising in a hierarchical decision model. *Annals of Tourism Research, 40*, 260–282.

Radojevic, T., Stanisic, N., & Stanic, N. (2015). Solo travellers assign higher ratings than families: Examining customer satisfaction by demographic group. *Tourism Management Perspectives, 16*, 247–258.

Simon, H. (1955). A behavioural model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

Steinbruner, J. (2002). *The cybernetic theory of decision*. Princeton, NJ: Princeton University Press.

Train, K. E. (2009). *Discrete choice methods with simulation*. Nueva York: Cambridge University Press.

Vriens, M., & Kidd, P. (2014). What every marketer needs to know about advanced analytics. *Marketing Insights*, 22–29.

WTTC Report. (2014). *WTTC Big Data Report*.http://inspirationhub.imexexhibitions.com/p/4020527602/2014/04/30/wttc-big-data-report