

Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content

Anindya Ghose, Panagiotis G. Ipeirotis, Beibei Li

Stern School of Business, New York University, New York, New York 10012
{aghose@stern.nyu.edu, panos@stern.nyu.edu, bli@stern.nyu.edu}

User-generated content on social media platforms and product search engines is changing the way consumers shop for goods online. However, current product search engines fail to effectively leverage information created across diverse social media platforms. Moreover, current ranking algorithms in these product search engines tend to induce consumers to focus on one single product characteristic dimension (e.g., price, star rating). This approach largely ignores consumers' multidimensional preferences for products. In this paper, we propose to generate a ranking system that recommends products that provide, on average, the best value for the consumer's money. The key idea is that products that provide a higher surplus should be ranked higher on the screen in response to consumer queries. We use a unique data set of U.S. hotel reservations made over a three-month period through Travelocity, which we supplement with data from various social media sources using techniques from text mining, image classification, social geotagging, human annotations, and geomapping. We propose a random coefficient hybrid structural model, taking into consideration the two sources of consumer heterogeneity the different travel occasions and different hotel characteristics introduce. Based on the estimates from the model, we infer the economic impact of various location and service characteristics of hotels. We then propose a new hotel ranking system based on the average utility gain a consumer receives from staying in a particular hotel. By doing so, we can provide customers with the "best-value" hotels early on. Our user studies, using ranking comparisons from several thousand users, validate the superiority of our ranking system relative to existing systems on several travel search engines. On a broader note, this paper illustrates how social media can be mined and incorporated into a demand estimation model in order to generate a new ranking system in product search engines. We thus highlight the tight linkages between user behavior on social media and search engines. Our interdisciplinary approach provides several insights for using machine learning techniques in economics and marketing research.

Key words: user-generated content; social media; search engines; hotels; ranking system; structural models; text mining; crowdsourcing

History: Received: January 15, 2010; accepted: November 13, 2011; Peter Fader served as the special issue editor and Jean-Pierre Dubé served as associate editor for this article. Published online in *Articles in Advance* April 10, 2012.

1. Introduction

As social media and user-generated content (UGC) become increasingly ubiquitous, consumers rely on a large variety of Internet-based sources prior to making a purchase. However, although product search engines have access to lots of UGC (not only on their own site but also across other social media channels), they typically fail to effectively leverage and present such product information going beyond simple numerical ratings. Moreover, existing ranking algorithms typically induce consumers to focus on only one single product characteristic dimension (e.g., price, star rating, number of reviews). This rudimentary approach largely ignores consumer heterogeneity and their multidimensional product preferences.

In this paper, we propose a new ranking system for product search engines that aims to maximize the expected utility gain for consumers from a purchase. We instantiate our study by looking into the hotel industry. According to a study by comScore (Lipsman 2007), more than 87% of customers rely on the online UGC to make purchase decisions for hotels, higher than any other product category. This situation motivates the need for a robust ranking mechanism on travel search engines that can more efficiently incorporate the publicly available knowledge within and across a large variety of social media platforms. Toward this goal, we propose a system that ranks each hotel according to the expected utility gain across the consumer population. The advantage of this system is that it uses consumer utility theory and

characteristics theory (e.g., Fader and Hardie 1996) to design a scalar utility score with which to rank hotels while incorporating all the dimensions of hotel quality observed from diverse information sources. No established measures currently exist that quantify the economic impact of various internal (service) and external (location) characteristics on hotel demand. By analyzing UGC from social media, we are able to estimate consumer preferences toward different hotel characteristics and recommend products that provide the best value for money on an average.

We use a unique data set of hotel reservations from Travelocity.com. The data set contains complete information on transactions conducted over a three-month period, from November 2008 to January 2009, for 1,497 hotels in the United States. We have data on UGC from three sources: (i) user-generated hotel reviews from two well-known travel search engines, Travelocity.com and TripAdvisor.com; (ii) social geotags generated by users identifying different geographic attributes of hotels from GeoNames.org; and (iii) user-contributed opinions on important hotel characteristics based on user surveys from Amazon Mechanical Turk (AMT). Moreover, because some location-based characteristics are not directly measurable based on UGC, we use image classification techniques to infer such features from the satellite images of the area. We then merge these different data sources to create one comprehensive data set that summarizes the location and service characteristics of all the hotels.

In the first step of our analysis, we determine the particular hotel characteristics customers value most and thus influence the aggregate demand of the hotels. Beyond the directly observable characteristics (e.g., the “number of stars”) most third-party travel websites provide, many users also tend to value specific location characteristics, such as proximity to the beach or downtown. We incorporate satellite image classification techniques and use both human and computer intelligence (in the form of social geotagging and text mining of reviews) to infer these location features. In the second step, we use demand estimation techniques (Berry et al. 1995, Berry and Pakes 2007, Song 2011) to quantify the economic influence and relative importance of location and service characteristics. Our empirical modeling and analyses enable us to compute the “expected utility gain” from a particular hotel based on the estimation of price elasticities and average utilities. In the third step, we use this measure of expected utility gain to propose a new ranking system in which a hotel that provides a comparably higher average utility gain would appear at the top of the list displayed by a travel search engine. By doing so, we can provide customers with the “best-value” hotels early on, thereby improving

the quality of online hotel search. In the final step, we validate the superiority of our proposed ranking system by conducting online experiments across 7,800 users on AMT, across six different cities based on a number of benchmark systems.

Our key results are as follows:

1. Five location-based characteristics have a positive impact on hotel demand: the number of external amenities, presence near a beach, presence near public transportation, presence near a highway, and presence near a downtown area. The textual content and style of reviews also demonstrate a statistically significant association with demand. Reviews that are less complex, have shorter words, and have fewer spelling errors influence demand positively, as do reviews with more characters and those written in simple language. Consumers prefer hotels with reviews that contain objective information (such as factual descriptions of hotels) rather than subjective information, indicating they trust third-party information over descriptions provided by the hotels. Consumers also prefer to stay in hotels with reviews written in a “consistent objective style” rather than a mix of objective and subjective sentences.

2. We examine interaction effects between travel purpose, price, and hotel characteristics. Business travelers are the least price sensitive, whereas tourists are the most price sensitive. In addition, business travelers have the highest marginal valuation for hotels located closer to a highway and with easy access to public transportation. In contrast, romance travelers have the highest marginal valuation for hotels located closer to a beach and those with a high service rating.

3. A comparison between the model that conditions on the UGC variables and a model that does not shows that the former outperforms the latter in both in-sample and out-of-sample analyses. Additional model fit comparisons suggest that the model’s predictive power is weakest when excluding all the location variables, followed by the service variables, and then the UGC variables. Moreover, within the set of UGC variables, we find textual information (e.g., text features, review subjectivity, readability) has a significantly higher impact than numerical information on the model’s predictive power.

There are three key contributions of this paper. First, we illustrate how researchers can mine UGC from multiple and diverse sources on the Internet to examine the economic value of different product attributes, using a structural model of demand estimation. Customers today make their decisions in an environment with a plethora of available data. Some consumers might research a hotel using tour guides and mapping applications, or they may consult online review sites to determine a hotel’s quality and amenities. To replicate this decision-making

environment, we construct an exhaustive data set, collecting information from a variety of sources and using a variety of methodologies, such as text mining, on-demand annotations, and image classification. We demonstrate the marginal contribution from different information sources by conducting model fit comparisons between models that condition for one set of variables versus another.

Second, our empirical estimates enable us to propose a new ranking system for hotel search based on the computation of each hotel's expected utility gain, which measures the "net value" a consumer gets from the transaction. The key notion is that in response to a search query, the system would take into account consumers' multidimensional preferences in order to recommend and rank higher those hotels that provide a higher "value for money." Thus, our paper shows how businesses can leverage UGC in social media to generate a ranking system in product search engines that improves the quality of choices available to consumers. The methodological approach used in this paper can be applied toward ranking any product or service that has multiple attributes, and hence the applicability of this paper is very wide. This is also the first study that demonstrates the linkages between user behavior on social media platforms and search engines.

Third, to evaluate the quality of our ranking technique, we conducted several user studies based on online surveys on AMT across six different markets in the United States. Using 7,800 unique user responses for comparing different rankings, we unequivocally demonstrate that our proposed ranking system performs significantly better than several baseline-ranking systems currently used by travel search engines. A follow-up survey reveals that users strongly preferred the diversity of the retrieved results, given that the list produced by our method consisted of a mix of hotels cutting across several price and quality ranges. This finding indicates that customers prefer a list of hotels with each specializing in a variety of characteristics, rather than a variety of hotels with each specializing in only one characteristic. Besides providing consumers with direct economic gains, such a ranking system can lead to nontrivial reductions in consumer search costs. Furthermore, directing customers to hotels that are better matches for their interests can lead to increased usage of travel search engines.

The rest of this paper is organized as follows. Section 2 discusses related work and places our work in the context of prior literature. Section 3 discusses the work related to the data preparation, including the methods used to identify important hotel characteristics, the steps undertaken to conduct the surveys on AMT to elicit user opinions, and the text-mining

techniques used to parse user-generated reviews. Sections 4 and 5 provide an overview of our econometric approach and discuss empirical results, respectively. Section 6 discusses how one can apply our approach to design a real-world application, such as a ranking system for hotel search. Section 7 concludes.

2. Prior Literature

Our paper draws from multiple streams of work. A key challenge is to bridge the gap between the textual and qualitative nature of review content and the quantitative nature of discrete choice models. With the rapid growth and popularity of the UGC on the Web, a new area of research has emerged that applies text-mining techniques to product reviews. The first stream of this research focused on the sentiment analysis of product reviews (Hu and Liu 2004, Pang and Lee 2004, Das and Chen 2007). This focus stimulated additional research on identifying product features in which consumers expressed their opinions (Hu and Liu 2004, Scaffidi et al. 2007). The automated extraction of product attributes has also received attention in the recent marketing literature (Lee and Bradlow 2011).

Meanwhile, the hypothesis that product reviews affect product sales has received strong support in prior empirical studies (e.g., Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Dellarocas et al. 2007, Duan et al. 2008, Forman et al. 2008, Moe 2009). However, these studies focus only on numeric review ratings (e.g., the valence and volume of reviews) in their empirical analysis. Researchers using only numeric ratings have to deal with issues such as self-selection bias (Li and Hitt 2008) and bimodal distribution of reviews (Hu et al. 2006). More importantly, the matching of consumers to hotels in numerical rating systems is not random. A consumer only rates the hotel she frequents (i.e., the one that maximizes her utility). Consequently, the average star rating for each hotel need not reflect the population's average utility. Because of the above drawbacks, the average numerical star rating a product receives may not convey a lot of information to a prospective buyer.

To the best of our knowledge, only a handful of empirical studies have formally tested whether the textual information embedded in online UGC can have an economic impact. Ghose et al. (2009) estimate the impact of buyer textual feedback on price premiums sellers charge in online secondhand markets. Eliashberg et al. (2007) combine natural-language processing techniques and statistical learning methods to forecast the return on investment for a movie, using shallow textual features from movie scripts. Netzer et al. (2012) combine text mining and semantic network analysis to understand the brand associative

network and the implied market structure. Decker and Trusov (2010) use text mining to estimate the relative effect of product attributes and brand names on the overall evaluation of the products.

None of these studies focuses on estimating the impact of user-generated product reviews on influencing product sales beyond the effect of numeric review ratings, which is one of the key research objectives of this paper. Controlling for the volume of blog posts, Dhar and Chang (2009) show that changes in the number of friends of an artist are correlated with sales of music albums. The papers closest to ours are Ghose and Ipeirotis (2011) and Archak et al. (2011), who explore multiple aspects of review text to identify important text-based features and to study their impact on review helpfulness (Ghose and Ipeirotis 2011) and product sales (Ghose and Ipeirotis 2011, Archak et al. 2011). However, these studies do not have data on actual product demand and do not use structural models, nor do they examine the use of UGC in developing a ranking system for product search in online markets.

Our work is related to models of demand estimation such as BLP (Berry et al. 1995). Because of the limitations of the product-level “taste shock” in logit models, Berry and Pakes (2007) proposed a new model based on pure product characteristics. The pure characteristics model (hereafter, PCM) differs from the BLP model in the sense that it does not contain the product-level taste shock. It describes the consumer heterogeneity, based purely on consumers’ different tastes toward individual product characteristics, without consideration of the tastes of certain products as a whole (i.e., brand preferences). However, as Song (2011) points out, whether one should introduce the product-level taste shock depends on the context of the market. Keeping in mind the two levels of consumer heterogeneity introduced by different travel categories (i.e., family trip, romantic trip, or business trip) and different hotel characteristics, we propose a random coefficients-based hybrid structural model to identify the latent weight distribution consumers assign to each hotel characteristic. The outcome of our analysis enables us to compute each hotel’s expected utility gain and rank the hotels accordingly on a travel search engine.

Finally, our paper is related to the work in online recommender systems. By generating a novel ranking approach for hotels, we aim to improve the recommendation strategy for travel search engines and provide customers with the “best-value” hotels early on in the search process. The marketing literature has proposed several model-based recommendation systems to predict preferences for recommended items (Ansari et al. 2000, Ying et al. 2006, Bodapati 2008). A more recent trend along this line is adaptive

personalization systems (Ansari and Mela 2003, Rust and Chung 2006, Chung et al. 2009).

3. Data Description

Our data set, compiled from various sources, consists of observations from 1,497 hotels in the United States. We have three months of hotel transaction data from Travelocity.com (November 1, 2008 to January 31, 2009) that contains the average transaction price per room per night and the total number of rooms sold per transaction.

Our work leverages three types of UGC data:

- on-demand user-contributed opinions through Amazon Mechanical Turk (AMT),
- location description based on user-generated geotagging and image classification, and
- service description based on user-generated product reviews.

We first discuss how we leverage AMT to collect information on user preferences for different hotel characteristics. Their responses suggest we can lump these characteristics into two groups: location and service. Once we identify the set of consumer preferences, we use other UGC to infer the external location characteristics, the internal service characteristics, and the textual characteristics of hotel reviews that can influence consumer purchases. We present the data sources, definitions, and summary statistics of all variables in Tables 1 and 2.

3.1. Identification of Hotel Characteristics

Our analysis first requires knowledge of which aspects of a hotel are most important to consumers and therefore determine the aggregate prices of the hotels. We use a survey of potential hotel customers to identify these aspects.

To reach a wide demographic, we relied on AMT, an online marketplace used to automate the execution of micro tasks that require human intervention (i.e., cannot be fully automated using data-mining tools). Task requesters post simple micro tasks, known as hits (human intelligence tasks), in the marketplace. The marketplace provides proper control over the task execution, such as validation of the submitted answers or the ability to assign the same task to several different workers. It also ensures the proper randomization of the assignments of tasks to workers within a single task type. Each user receives a small monetary compensation for completing the task.

Our main goal was to obtain a diversity of consumer opinions. Therefore, we wanted to first ensure the participants were representative of the overall Internet population. Therefore, we constructed a survey in which we asked AMT workers to provide

Table 1 Summary of Different Methods for Extracting Hotel Characteristics

Category	Hotel characteristics	Methods
Transaction data	Transaction price (per room per night) Number of rooms sold (per night)	Travelocity
Service based	Hotel class Hotel amenities	TripAdvisor
Review based	Number of customer reviews Overall reviewer rating Disclosure of reviewer identity information Subjectivity Mean probability Std. dev. of probability Readability Number of characters Number of syllables Number of spelling errors Average length of sentence SMOG index (Additional) Breakfast Hotel staff Bathroom Bedroom Parking	Travelocity and TripAdvisor
Location based	Near the beach Near downtown External amenities (number of restaurants/shopping destinations) Near public transportation Near the interstate highway Near the lake/river City annual crime rate	Image classification; tags from GeoNames.org and social annotations from Amazon Mechanical Turk Microsoft Virtual Earth geomapping search SDK Tags from GeoNames.org; social annotations from Amazon Mechanical Turk Social annotations from Amazon Mechanical Turk FBI online statistics

Table 2 Definitions and Summary Statistics of Variables

Variable	Definition	Mean	Std. dev.	Min	Max
<i>PRICE</i>	Transaction price per room per night	126.59	79.47	12	978
<i>CHARACTERS</i>	Average number of characters	766.54	167.13	121	2,187
<i>COMPLEXITY</i>	Average sentence length	16.41	3.95	2	44
<i>SYLLABLES</i>	Average number of syllables	245.48	53.77	37	700
<i>SMOG</i>	SMOG index	9.91	0.63	3	19.80
<i>SPELLERR</i>	Average number of spelling errors	1.10	0.37	0	3.33
<i>SUB</i>	Subjectivity—mean	0.99	0.03	0.05	1
<i>SUBDEV</i>	Subjectivity—standard deviation	0.02	0.02	0	0.25
<i>ID</i>	Disclosure of reviewer identity	0.77	0.14	0	1
<i>CLASS</i>	Hotel class	3.02	0.93	1	5
<i>CRIME</i>	City annual crime rate	195.09	123.11	3	1,310
<i>AMENITYCNT</i>	Total number of hotel amenities	16.38	3.21	2	23
<i>EXTAMENITY</i>	Number of external amenities within one mile, i.e., restaurants or shops	4.95	7.37	0	27
<i>BEACH</i>	Beachfront within 0.6 miles	0.24	0.43	0	1
<i>LAKE</i>	Lake or river within 0.6 miles	0.23	0.42	0	1
<i>TRANS</i>	Public transportation within 0.6 miles	0.11	0.31	0	1
<i>HIGHWAY</i>	Highway exits within 0.6 miles	0.68	0.47	0	1
<i>DOWNTOWN</i>	Downtown area within 0.6 miles	0.69	0.46	0	1
<i>TA_REVIEWCNT</i>	Total number of reviews (TripAdvisor)	127.81	164.22	0	999
<i>TA_REVIEWCNT²</i>	Square of <i>TA_REVIEWCNT</i>	28,573.16	70,943.83	0	998,001
<i>TA_RATING</i>	Overall reviewer rating (TripAdvisor)	3.49	0.59	1	5
<i>TL_REVIEWCNT</i>	Total number of reviews (Travelocity)	25.26	29.77	0	202
<i>TL_REVIEWCNT²</i>	Square of <i>TL_REVIEWCNT</i>	731.40	1,794.81	0	40,804
<i>TL_RATING</i>	Overall reviewer rating (Travelocity)	3.87	0.74	1	5

Notes. Number of observations = 8,099. Time period is January 11, 2008 to January 31, 2009.

information about their place of origin and residence, gender, age, education attainment, income, marital status, household size, and number of children. We also asked them how much time they spend every week on AMT, how much work they complete, how much payment they receive, and their reasons for participating on AMT. We conducted the survey monthly for six months and found the results were robust.

The results of the survey indicate that most of the workers—70% to 80%—are based in the United States. More than 60% of the workers had a university education, and more than 15% of them had graduate degrees. The age of the workers varied widely but with an overrepresentation of young ages (21–30). Because the participants are therefore marginally younger than the overall Internet population, their income levels are lower and they have smaller families. Overall, despite some differences, we see that the AMT population is generally representative of the overall U.S. Internet population and more representative than surveys conducted using only locally available participants.¹

We also asked survey participants about their previous experiences with Travelocity.com. Of the 92.5% that had visited the website before, 55% had made hotel reservations.

Having determined that AMT workers are representative of the general Internet user population, we used them as the sample for our following survey. As part of this survey, we asked 100 random AMT users which hotel characteristics they considered important. We grouped and coded the answers (see Table 1), identifying two broad categories of hotel characteristics:

1. location-based characteristics (e.g., near a beach, near a waterfront (lake/river), near public transportation, near downtown) and
2. service-based characteristics (e.g., hotel class, number of internal amenities).

Next, we describe how we use UGC to collect information about the variables that are either too difficult to collect otherwise (e.g., density of shops around the hotel) or are likely to be subjective (e.g., quality of service).

3.2. Extraction of Location and Service Characteristics

For the location-based characteristics, we combine UGC with automatic techniques to scale our data collection and generate data sets that are comprehensive at the national and international levels (i.e., tens or hundreds of thousands of hotels). A first automatic approach is to use a service such as the Microsoft Virtual Earth Interactive SDK (software development kit) that enables us to compute location characteristics such as “near restaurants and shops” for a given hotel location on a map. Using the application programming interface from Microsoft, we can automatically perform such local search queries.

However, the presence of a characteristic such as “near a beach” or “near downtown” cannot be retrieved using existing mapping services. To measure such characteristics, we use a combination of user-generated geotagging and automatic classification of satellite images of areas near each hotel in our data set. The concept of geotagging has been popularized lately by photo-sharing websites on which users annotate their photos with the exact longitude and latitude of the location. The concept has been extended and is now used in “wiki”-style websites on which users annotate maps with tags such as “bridge,” “lake,” or “park.” In our study, we extracted the location characteristics “near public transportation,” “near a beach,” and “near the downtown” via GeoNames.org. For the characteristics “near public transportation,” “near a lake/river,” and “near the interstate highway,” we extracted the features using on-demand annotations from a set of workers from AMT. Such geotagging and on-demand annotations enable us to generate a richer description of each hotel’s location, using features that are not directly available through existing mapping services.

Regardless of how comprehensive the tagging is, users may not have yet tagged *all* locations. Therefore, we need to leverage the tag database and allow for the automatic tagging of nontagged areas. We therefore use automatic image classification techniques of satellite images to tag location features that can influence hotel demand. See Online Appendix §F-1 for details. Finally, we collect the “crime rate” at the city level from the FBI website. It contains the total number of violent crimes (e.g., murder, aggravated assault) and property crimes (e.g., burglary, larceny theft, arson).

We use two broad service-based characteristics: *hotel class* is an internationally accepted standard ranging from one to five stars, representing low to high hotel grades; and the *number of internal amenities* includes, for example, indoor swimming pools, high-speed Internet, free breakfasts, hair dryers, and parking facilities. We extracted this information from TripAdvisor.com using fully automated

¹ In Online Appendix G (at <http://mktsci.journal.informs.org/>), we provide the exact analysis of the survey and a comparison of the demographics, with the demographics of Internet users in the United States according to the data provided by comScore. To compensate for the differences in the population, we also stratified the responses from the sample based on demographics and placed appropriate weights on the responses so that the results would match the composition of the Internet user population in the United States.

parsing. Because the website does not explicitly list hotel amenities, we retrieved them by following the link provided on the hotel Web page directing users to one of the hotel's partner websites (i.e., Travelocity.com, Orbitz.com, Expedia.com, Priceline.com, or Hotels.com).

3.3. Extraction of Linguistic Style of Customer Reviews

We collected customer reviews from Travelocity.com. We also collected reviews from a neutral, third-party site and the world's largest online travel community, TripAdvisor.com, to account for the indirect influence of word of mouth. We collected all available online reviews and reviewers' information up to January 31, 2009 (the last date of transactions in our database).

Consistent with prior work, we use the total number of reviews and the numeric reviewer rating to control for word-of-mouth effects. In addition, given that the actual quality of reviews affects product sales, we looked into two text style features—subjectivity and readability, both of which can influence consumers' purchase decisions (Ghose and Ipeirotis 2011). For more details, see Online Appendix SF-2.

Five broad types of characteristics are present in this category: (i) total number of reviews, (ii) overall review rating, (iii) review subjectivity (mean and variance), (iv) review readability (the number of characters, syllables, and spelling errors; complexity; and SMOG index), and (v) disclosure of the reviewer's identity.

4. Model

In this section, we will discuss our random coefficients-based structural model and how we apply it to estimate the distribution of consumer preferences toward different hotel characteristics. The estimates from these analyses are then used toward computing the expected utility gain from each hotel, which, in turn, is used in designing the ranking system described in §6.

4.1. Model Setup

Our model is motivated directly by Song (2011), who proposes a hybrid discrete choice model of differentiated product demand. Whereas Song (2011) has one random coefficient on price, we have multiple random coefficients on price as well as hotel characteristics. Note that his hybrid model is a combination of the BLP and PCM approaches. It is called a hybrid model because it resembles the random coefficient logit demand model in describing a brand choice (BLP) and the pure characteristics demand model in describing a within-brand product choice (PCM). This hybrid model is a discrete choice model of differentiated product demand in which product groups are

horizontally differentiated, whereas products within a given group are vertically differentiated conditional on product characteristics. These two types of differentiation are distinguished by a group-level “taste shock,” which is assumed to be independently and identically distributed with a Type I extreme value distribution. This taste shock represents each consumer's specific preference toward a product group that product characteristics—both observed and unobserved—do not capture. Song (2011) refers to a product group that contains vertically differentiated products as a “brand.” This hybrid model identifies the preference for product characteristics in a similar way as the PCM, the main difference being that the hybrid model compares products of each brand on the quality ladder separately, whereas the PCM compares all products on it at the same time. Hence, the quality space is much less crowded in the hybrid model.²

In our context, a hotel “travel category” represents a “brand,” and the hotels within each travel category represent “products.” In particular, the market share function of hotel j^k within travel category k can be written as the product of the probability that travel category k is chosen and the probability that hotel j^k is chosen given that travel category k is chosen. The former probability is similar to the choice probability in BLP and the latter to that of the PCM.

We define a consumer's decision-making behavior as follows. A consumer needs to locate the hotel whose location and service characteristics best match her travel purpose. For instance, if a consumer wants to go on a romantic trip with a partner, she will be interested in the set of hotels located close to a beach as well as to a downtown area with such amenities as nightclubs and restaurants. She is also aware that hotels specializing in the romance category are more likely to satisfy such location and service needs. Each hotel can belong to one of the following eight types of travel categories: family trip, business trip, romantic trip, tourist trip, trip with kids, trip with seniors, pet-friendly trip, and disability-friendly trip. Each travel category is defined and chosen according to the information gleaned from TripAdvisor.com, which allows

² This hybrid model provides more efficient substitution patterns according to its basic assumptions and model foundations. As Song (2011) describes, it distinguishes between two types of cross substitutions: within-travel and between-travel category substitution. The former is confined to hotels within the same travel category and has the same substitution pattern as in the PCM. The latter determines the substitution pattern for hotels in different travel categories and has a similar pattern as in BLP but with a distinct difference: the impact of a change (in price or availability) on other travel categories is confined to hotels of similar quality. As a result, a hotel will have fewer substitutes in our model than in the BLP (Berry et al. 1995) model.

reviewers to specify their main trip purpose (travel category) when posting a review.

We have data on all the hotel reviews posted by users for a given hotel right from the time the first review was posted until the last date of our transaction data set (January 31, 2009). We classify a hotel into a specific travel category based on reviewers' most frequently mentioned travel purpose for that hotel. Hence, each hotel belongs to a single travel category. To capture the heterogeneity in consumers' travel purpose, we introduce an idiosyncratic taste shock at the travel category level. This shock is similar to the product-level taste shock in the BLP model.

Each travel category has a hotel that maximizes a consumer's utility in that category. We refer to this as the "best" hotel in that category. To find the best hotel within each travel category, we use the PCM, which enables us to capture the vertical differentiation among hotels within the same travel category. A rational consumer chooses a travel category if and only if her utility from the best hotel in that category exceeds her utility from the best hotel in any other travel category. Thus, in our model, the utility for consumer i from choosing hotel j with category type k in market t can be represented as follows:

$$u_{ijkt} = X_{jkt}\beta_i + \alpha_i P_{jkt} + \xi_{jkt} + \varepsilon_{ikt}, \quad (1)$$

where i represents a consumer, j^k represents hotel j with travel category type k ($k \in \{1, 2, \dots, 7, 8\}$), and t represents a hotel market (city-week combination). In this model, β_i and α_i are individual-specific random coefficients that capture consumers' heterogeneous tastes toward different observed hotel characteristics, $X = [X^1, X^2, \dots, X^Z]$, and toward the average price per night, P , respectively. Note that α_i is a scalar, whereas β_i is a Z -dimensional vector corresponding to Z hotel characteristics. ξ_{jkt} represents hotel characteristics unobservable to the econometrician. ε_{ikt} with a subscript k represents a travel category-level taste shock. Note that in our model, the travel category-level shock is independently and identically distributed across consumers and travel categories, consistent with Song (2011).³

We define a "market" as the combination of "city-week." Correspondingly, we calculate the market share for each hotel based on the number of rooms

sold for that hotel in that market (i.e., city-week) divided by the total size of that market. With regard to market size, in our main estimation, we applied the same idea as in the demand estimation literatures (e.g., Berry et al. 1995, Nevo 2001, Song 2011), computing the market size by estimating the potential consumption in a market. That is, we estimate the total potential market consumption to be proportional to the total number of rooms available in the existing hotels in a certain market (including the hotels whose transactions appear in our current choice set and those whose transactions we do not observe).⁴ We acquired the total number of existing hotel rooms in each market via TripAdvisor.com. Under this measure, the outside good is defined as "no purchase from the current choice set."⁵

Alternatively, in our robustness checks, we define the market size as the total number of rooms all hotels in that city sold during that week, based on the transaction data from Travelocity.com. Recall that our main data set comes from two sources: Travelocity.com-generated transaction data and TripAdvisor.com hotel listing data. The data set we use is the set of hotels at the intersection of the two sources, which means the hotel choice set for each market includes those hotels that not only have a transaction generated via Travelocity.com but also have information available from user-generated reviews on TripAdvisor.com. Because not every hotel that has a Travelocity.com-generated transaction is listed on TripAdvisor.com, we define our "outside good" as the set of hotels listed in the original Travelocity.com transaction data but not TripAdvisor.com.

We follow Berry et al. (1995) and model the distribution of consumers' taste parameters as multivariate normal, conditional on demographics:

$$\begin{aligned} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} &= \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix} + \Pi D_i + \Sigma v_i, & D_i &\sim P_D^*(D), \\ v_i &\sim N(0, I_{Z+1}), \end{aligned}$$

where D_i is a $d \times 1$ vector of consumer demographic variables, $P_D^*(D)$ is a nonparametric distribution observed from other data sources, and v_i is a $(Z+1) \times 1$ vector capturing the additional unobserved consumer-specific preference toward hotel characteristics. It follows a multivariate normal distribution. Π is a $(Z+1) \times d$ matrix of coefficients

³ Besides our model, which incorporates a travel category-level taste shock, at least three other plausible modeling approaches exist in this context: (i) a model with only a hotel-level taste shock approach; (ii) a model with both travel category-level and hotel-level taste shocks, with the travel category at the top hierarchy, resembling the nested logit model; and (iii) a model with no taste shocks at either the travel category level or the hotel level, resembling the PCM (Berry and Pakes 2007) approach. We have estimated all these models and found that our hybrid model provides the best performance in both precision and deviation. Details are provided in §5.3.

⁴ For this estimation of market size to be valid, we assume that the total number of room nights rented in a certain market is proportional to the number of individuals in that market.

⁵ Because our transaction data set is a random sample from Travelocity, it is an unbalanced panel data set. Based on Yamamoto (2011), the BLP-type of model (i.e., mixed logit model) is a more general form of the "varying choice set logit model." Therefore, our estimation is able to account for the varying choice set bias that standard logit models suffer.

that measures how consumers' taste parameters vary with observed demographics. Σ is a $(Z + 1) \times (Z + 1)$ scaling matrix. This specification allows the observed demographics D_i and the unobserved factor v_i to determine the consumer-specific taste.

As we know, a consumer's income, y_i , normally affects his taste. In particular, we notice that y_i affects the consumer's taste primarily through price. Thus we propose our basic model with two assumptions: (i) D_i contains only the consumer income y_i , which follows the empirical income distribution $F_y(\cdot)$ and can be derived from the U.S. Census data; and (ii) Π is zero in all but one row. The nonzero row corresponds to the price coefficient. We relax both assumptions in our extended model in §4.4.

Given these assumptions, we basically model α_i and β_i as a function of consumer income and the unobserved characteristic: $\alpha_i = \bar{\alpha} + \alpha_y y_i + \alpha_v v_i$ and $\beta_i = \bar{\beta} + \beta_v v_i$. Then we rewrite our model as follows:

$$u_{ijk} = \delta_{jk} + X_{jkt} \beta_v v_i + \alpha_y y_i P_{jkt} + \alpha_v v_i P_{jkt} + \varepsilon_{ikt}, \quad (2)$$

where $\delta_{jk} = X_{jkt} \bar{\beta} + \bar{\alpha} P_{jkt} + \xi_{jkt}$ represents the mean utility of hotel j with category type k in market t ; $\bar{\alpha}$, $\bar{\beta}$, α_y , α_v , and β_v are the parameters to be estimated.

4.2. Estimation

As mentioned in the previous subsection, our goal here is to estimate the mean and deviation of β_i and α_i . We apply methods similar to those used in Berry and Pakes (2007) and Song (2011). In general, with a given starting value of $\theta_0 = (\alpha_y^0, \alpha_v^0, \beta_v^0)$, we look for the mean utility δ , such that the model-predicted market share is equal to the observed market share. We then form a generalized method of moments (GMM) objective function using the moment condition that the mean of unobserved characteristics is uncorrelated with the instrumental variables. We then update the parameter value of $\theta_1 = (\alpha_y^1, \alpha_v^1, \beta_v^1)$ and use it as the starting point for the next-round iteration. This procedure is repeated until the algorithm finds the optimal value of θ that minimizes the GMM objective function. The algorithm searches only over a subset of parameters because we concentrate on the mean utility parameters that enter linearly, out of the search. We conduct the estimation in three stages.

To calculate the market share for a particular hotel, we need to know (1) the size of a certain consumer segment and (2) the probability of that segment choosing this hotel. Multiplying the two gives us the overall market share. See Online Appendix D for the mathematical details for the derivation.

We next identify the mean utility δ by equating the estimated market share with the observed market share conditioning on a given $\theta = (\alpha_y, \alpha_v, \beta_v)$. The solution to this problem satisfies a system of nonlinear equations. In our case, $\sum_{k=1}^K J^k$ nonlinear

equations (where J^k is the total number of hotels within travel category type k) and $\sum_{k=1}^K J^k$ unknown variables exist (δ being a $\sum_{k=1}^K J^k$ dimension vector). To solve, we apply the Newton–Raphson method per Song (2011), which works well when the number of products per market is up to 20. To guarantee the robustness of the results when the number of products is larger than 20, we tried different initial values in the iteration and found the final solution was consistent. In practice, this approach locates the closest solution for our settings, whereas the iteration procedure provides a closed form to locate the roots rapidly.

To account for the endogeneity of price, we use a GMM estimator and form an objective function by interacting the unobservable parameter, ξ , with a set of instrumental variables. We use the Nelder–Mead simplex algorithm to update the parameter values for α_y , α_v , and β_v , and we use them as the starting points to recalculate the market share and solve for the new mean utility. This process allows us to extract the new structural error ξ and form the GMM objective function. This entire procedure iterates until the algorithm finds the optimal combination of α_y , α_v , β_v , and δ that minimizes the GMM function.

4.2.1. Identification. A critical issue in the estimation process is price endogeneity. To separate the exogenous variation in prices (i.e., as a result of differences in marginal costs) and endogenous variation (i.e., as a result of differences in unobserved valuation), we use the average price of the same-class hotels in the other markets as an instrument for price. This is similar to Hausman's (1996) approach. The identification assumption is that, controlling for class-specific means and demographics, market-specific valuations are independent across markets (but are allowed to be correlated within a market). Hence, prices of the same-class hotels in two markets will be correlated as a result of the common marginal cost but because of the independence assumption will be uncorrelated with market-specific valuation.

In addition, we used three other sets of instruments. First, we followed Villas-Boas and Winer (1999) and Archak et al. (2011) and used lagged prices as instruments in conjunction with Google Trends data. The lagged price may not be an ideal instrument because common demand shocks may be correlated over time. Nevertheless, common demand shocks that are correlated through time are essentially trends. Controlling for trends through our use of search volume data for different major hotel brands should alleviate most, if not all, such concerns.

Second, cost-side variables that are correlated with prices but uncorrelated with factors that are reflected in the unobserved characteristics term are widely used as instruments for price (see, e.g., Chintagunta et al. 2005). We used region dummies as proxies for the marginal costs, as suggested by Nevo (2001).

The marginal costs include costs related to production (e.g., facilities, labor, energy) and distribution (e.g., transportation, labor, storage space). Because production costs exhibit little variation over time, they constitute too small a percentage of marginal costs to be correlated with prices (Nevo 2001). Thus, instead of finding instruments for production costs, our model uses brand dummies and hotel class variables to control for such costs. Because distribution costs contribute to most of the marginal costs that are correlated with prices, we used region dummy variables as proxies for the distribution costs.

Third, we used BLP-style instruments. Specifically, we used the average characteristics of the hotels with the same class in the other markets. All these alternative estimations yielded similar results. See Table 3, columns 5–7, for the corresponding estimation results using alternative instruments.

We performed an F -test in the first stage for each of the instruments. In each case, the F -test value was well over 10, suggesting that our instruments are valid (i.e., the instruments are not weak). In addition, the Hansen's J -test could not reject the null hypothesis of valid overidentifying restrictions. See Online Appendix E for the detailed estimation algorithm.⁶

4.3. Model Extension 1: Additional Text Features

So far, we have not fully exploited the information about hotel service characteristics from the data, which is embedded in the natural-language text of the consumer reviews. For example, the helpfulness of the hotel staff is a service feature one can assess by reading the consumer opinions. Toward extracting such information, we build on the work of Hu and Liu (2004), Popescu and Etzioni (2005), and Archak et al. (2011).

First, we extract the important hotel features. Following the automated approach introduced previously (Archak et al. 2011), we use a part-of-speech tagger to identify the frequently mentioned nouns and noun phrases, which we consider candidate hotel features. We then use WordNet (Fellbaum 1998) and a context-sensitive hierarchical agglomerative clustering algorithm (Manning and Schütze 1999) to further cluster the identified nouns and noun phrases into clusters of similar nouns and noun phrases. The resulting set of clusters corresponds to the set of identified product features mentioned in the reviews. For our analysis, we kept the top five most frequently mentioned features, which were hotel staff, food quality, bathroom, parking facilities, and bedroom quality.

⁶ Further information on the proof of existence and uniqueness of the mean product quality (delta parameter that matches the model predicted market shares with observed market share) is available in Song (2011). This is in addition to Berry et al. (2004), who provide support for their arguments regarding the asymptotic properties for the multidimensional pure characteristics model with Monte Carlo simulations.

For sentiment analysis, we extracted all the evaluation phrases (adjectives and adverbs) that were used to evaluate the individual service features (for example, for the feature “hotel staff,” we extracted phrases such as “helpful,” “smiling,” “rude,” “responsive”). The process of extracting user evaluation phrases can be automated. To measure the meaning of these evaluation phrases, we used AMT to exogenously assign explicit polarity semantics to each word. To compute the scores, we used AMT to create our ontology, with the scores for each evaluation phrase. Our process for creating these “external” scores was done using the methodology of Archak et al. (2011). Finally, to handle the negation (e.g., “I didn’t think the staff was helpful”), we built a dictionary database to store all the negation words (e.g., “not,” “hardly”) using an approach similar to NegEx (<http://code.google.com/p/negex>; accessed February 1, 2012). For more details on how we extracted the text features together with the corresponding sentiment analysis, see Online Appendix SF-3.

4.4. Model Extension 2: Interactions with the Travel Category

As discussed in §4.1, we simplify our basic model framework by making two assumptions: (i) D_i contains only the consumer income, $INCOME_i$; and (ii) Π is zero in all but one row, which corresponds with the price coefficient. However, other consumer demographic characteristics are also likely to affect consumers’ tastes. Moreover, other interaction effects might also exist beyond the one between income and price. Based on the basic model, we now relax these assumptions by considering interaction effects with the demographic variables. This extension is done similar to that of Nevo (2001) by enabling interactions between consumer travel purposes and hotel characteristics. More specifically, we extend our basic model by allowing D_i to contain both consumer travel purposes and income. We also allow Π to be nonzero in all its elements. We define T_i as an indicator vector with identity components representing consumer travel purpose:⁷

$$T_i' = [Family_i, Business_i, Romance_i, Tourist_i, Kids_i, Seniors_i, Pets_i, Disability_i].$$

⁷ The empirical distribution of T_i can be acquired from online consumer reviews and reviewers’ profiles. After writing an online review for a hotel, a reviewer is asked to provide additional demographic and trip information—for example, “What was the main purpose of this trip? (Select one from the eight choices.)” The distribution of T_i is derived based on reviewers’ responses to this question. Our robustness test showed that consumers’ demographics derived from different online resources stay consistent (Jensen-Shannon divergence = 0.03). Note that because there are eight travel purpose dummies, we use seven of them in estimating the interaction effects.

Table 3 Main Estimation Results

Variable	Coefficients (Std. error)						
	I	II	III	A1	A2	A3	T
				Means			
PRICE (log)	−0.149*** (0.002)	−0.146** (0.003)	−0.142** (0.002)	−0.149*** (0.007)	−0.158** (0.027)	−0.143*** (0.005)	−0.149*** (0.001)
CHARACTERS (log)	0.010*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	0.015*** (0.003)	0.016*** (0.004)	0.010** (0.004)	0.009*** (0.002)
COMPLEXITY	−0.011*** (0.002)	−0.012*** (0.002)	−0.011*** (0.003)	−0.013*** (0.003)	−0.007** (0.003)	−0.011*** (0.002)	−0.011*** (0.001)
SYLLABLES (log)	−0.044*** (0.007)	−0.045*** (0.008)	−0.044*** (0.007)	−0.038*** (0.006)	−0.032*** (0.007)	−0.042*** (0.006)	−0.043*** (0.006)
SMOG	0.079*** (0.020)	0.077** (0.024)	0.080** (0.028)	0.065** (0.022)	0.093** (0.033)	0.072** (0.026)	0.077** (0.027)
SPELLERR (log)	−0.125*** (0.003)	−0.126*** (0.004)	−0.129*** (0.004)	−0.120*** (0.006)	−0.131*** (0.008)	−0.129*** (0.004)	−0.125*** (0.006)
SUB	−0.132*** (0.006)	−0.141*** (0.005)	−0.141*** (0.004)	−0.149*** (0.009)	−0.124*** (0.025)	−0.133*** (0.014)	−0.135*** (0.007)
SUBDEV	−0.403*** (0.011)	−0.412*** (0.009)	−0.420*** (0.016)	−0.437*** (0.021)	−0.396*** (0.033)	−0.414*** (0.013)	−0.400*** (0.010)
ID	0.058** (0.020)	0.056* (0.025)	0.066** (0.023)	0.046 (0.034)	0.031 (0.034)	0.057* (0.029)	0.051* (0.026)
CLASS	0.035*** (0.009)	0.034*** (0.008)	0.041*** (0.009)	0.040*** (0.010)	0.043*** (0.009)	0.036*** (0.008)	0.034*** (0.009)
CRIME (log)	−0.024* (0.017)	−0.025* (0.017)	−0.020* (0.011)	−0.019* (0.010)	−0.015 (0.014)	−0.022* (0.014)	−0.022* (0.016)
AMENITYCNT (log)	0.005* (0.002)	0.006* (0.003)	0.006*** (0.001)	0.007** (0.002)	0.010** (0.004)	0.007*** (0.002)	0.006* (0.003)
EXTAMENITY (log)	0.007*** (0.002)	0.008*** (0.001)	0.011*** (0.002)	0.012*** (0.001)	0.015*** (0.002)	0.009*** (0.001)	0.007*** (0.002)
BEACH	0.155*** (0.003)	0.156*** (0.004)	0.167*** (0.004)	0.160*** (0.017)	0.165*** (0.021)	0.153*** (0.010)	−0.025*** (0.003)
LAKE	−0.109*** (0.031)	−0.106*** (0.029)	−0.108*** (0.031)	−0.122*** (0.036)	−0.117* (0.059)	−0.111** (0.041)	−0.127 (0.092)
TRANS	0.158*** (0.003)	0.163*** (0.007)	0.175*** (0.007)	0.165*** (0.006)	0.162*** (0.008)	0.162*** (0.019)	0.158*** (0.004)
HIGHWAY	0.067* (0.025)	0.070** (0.025)	0.075** (0.026)	0.077*** (0.022)	0.088** (0.030)	0.063** (0.022)	0.066* (0.026)
DOWNTOWN	0.045*** (0.002)	0.049*** (0.004)	0.044*** (0.003)	0.039*** (0.005)	0.033*** (0.004)	0.046*** (0.009)	0.042*** (0.001)
TA_RATING	0.039** (0.018)	0.044** (0.020)	0.038** (0.018)	0.045** (0.019)	0.046** (0.022)	0.043** (0.019)	0.042** (0.014)
TL_RATING	0.034*** (0.008)	0.035*** (0.008)	0.035*** (0.007)	0.039*** (0.010)	0.048*** (0.012)	0.035*** (0.006)	0.034*** (0.006)
TA_REVIEWCNT (log)	0.186*** (0.043)	0.182*** (0.041)	0.188*** (0.045)	0.190*** (0.041)	0.175*** (0.035)	0.169*** (0.037)	0.187*** (0.042)
TA_REVIEWCNT ² (log)	−0.053*** (0.005)	−0.051*** (0.006)	−0.052*** (0.006)	−0.068*** (0.008)	−0.076** (0.031)	−0.055*** (0.011)	−0.052*** (0.004)
TL_REVIEWCNT (log)	0.014*** (0.002)	0.014*** (0.002)	0.015*** (0.002)	0.016*** (0.001)	0.019*** (0.004)	0.016*** (0.002)	0.013*** (0.002)
TL_REVIEWCNT ² (log)	−0.021*** (0.005)	−0.023*** (0.005)	−0.025*** (0.005)	−0.027*** (0.004)	−0.031*** (0.007)	−0.024*** (0.004)	−0.020*** (0.002)
Constant	0.039*** (0.002)	0.033*** (0.005)	0.036** (0.006)	0.044*** (0.010)	0.057** (0.021)	0.034*** (0.009)	0.041*** (0.003)
Brand control ^a	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instruments	Comp price in other markets ^b	Comp price in other markets	Comp price in other markets	Lag price with Google Trends	Cost–region dummies	BLP-style instruments	Comp price in other markets
Distribution of idiosyncratic error term	Type I extreme value	Type I extreme value	Type I extreme value	Type I extreme value	Type I extreme value	Type I extreme value	Type I extreme value
HIGH TEMP	—	—	—	—	—	—	0.078 (0.066)
HIGH TEMP × LAKE	—	—	—	—	—	—	0.020*** (0.005)
HIGH TEMP × BEACH	—	—	—	—	—	—	0.179*** (0.031)
							Normal distribution
							—
							—
							—

Table 3 (Cont'd.)

Variable	I	II	III	A1	A2	A3	T	N
	Interaction effect (α_i) and standard deviations (α_i)							
$PRICE(\log) \times INCOME(\log)$	0.023*** (0.002)	0.026*** (0.002)	0.021*** (0.002)	0.017*** (0.003)	0.014*** (0.004)	0.025*** (0.004)	0.021*** (0.005)	0.032*** (0.008)
$PRICE(\log)$	0.011 (0.101)	0.014 (0.098)	0.018 (0.103)	0.022 (0.082)	0.029 (0.116)	0.009 (0.088)	0.016 (0.108)	0.027 (0.071)
	Standard deviations (β_i)							
CLASS	0.025*** (0.005)	0.026*** (0.006)	0.033** (0.014)	0.028** (0.011)	0.042* (0.024)	0.033*** (0.004)	0.024*** (0.004)	0.037** (0.014)
CRIME(log)	0.006 (0.011)	0.012 (0.021)	0.012 (0.018)	0.016 (0.018)	0.019 (0.016)	0.010 (0.007)	0.007 (0.011)	0.016 (0.018)
AMENITYCNT(log)	0.016 (0.029)	0.023 (0.037)	0.025 (0.044)	0.020 (0.027)	0.025 (0.038)	0.015 (0.026)	0.015 (0.022)	0.024 (0.035)
EXTAMENITY(log)	0.003 (0.014)	0.005 (0.019)	0.004 (0.020)	0.006* (0.003)	0.009 (0.015)	0.009 (0.016)	0.003 (0.014)	0.005 (0.004)
BEACH	0.061*** (0.012)	0.056*** (0.015)	0.055*** (0.014)	0.066*** (0.017)	0.072* (0.039)	0.064*** (0.013)	0.063*** (0.011)	0.068*** (0.020)
LAKE	0.112* (0.078)	0.104* (0.058)	0.097* (0.056)	0.114 (0.092)	0.107* (0.060)	0.101† (0.067)	0.114*** (0.021)	0.117* (0.069)
TRANS	0.129** (0.054)	0.118* (0.064)	0.123** (0.061)	0.122*** (0.028)	0.124*** (0.024)	0.117** (0.040)	0.118** (0.049)	0.133** (0.051)
HIGHWAY	0.065* (0.035)	0.052 (0.043)	0.068 (0.047)	0.053 (0.047)	0.077 (0.092)	0.069** (0.024)	0.067* (0.036)	0.052 (0.076)
DOWNTOWN	0.031** (0.011)	0.034*** (0.009)	0.045*** (0.007)	0.025*** (0.002)	0.039*** (0.010)	0.036** (0.014)	0.032* (0.017)	0.038*** (0.007)
GMM obj. value	8.689e-4	8.115e-4	7.345e-4	5.972e-4	6.001e-4	7.145e-4	8.016e-4	6.638e-4

Notes. I is based on the main data set (at least one review from either TripAdvisor or Travelocity). II is based on the main data set with review count ≥ 5 . III is based on the main data set with review count ≥ 10 . A1 denotes alternative instruments 1—lag price with Google trends. A2 denotes alternative instruments 2—region dummy variables (northeast, south, midwest, southwest, west). A3 denotes alternative instruments 3—BLP-style instruments (average characteristics of the same-star hotels in other markets). T is based on data set I, considering interactions of temperatures with “lake/river” and with “beach.” N denotes Normal distribution of the idiosyncratic error term.

^aWe use dummy variables to control for nine major hotel brands: Accor, Best Western, Cendant, Choice, Hilton, Hyatt, InterContinental, Marriott, and Starwood.

^bIn the main estimation, we used the average price of the “same-star rating” hotels in the other markets as instruments.

***Significant at a 0.1% level; **significant at a 1% level; *significant at a 5% level; †significant at a 10% level.

For example, if consumer i is on a business trip, the corresponding travel purpose vector is

$$T'_i = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0].$$

Thus, the extended model can be rewritten as

$$u_{ijk_t} = \delta_{jk_t} + X_{jk_t}\beta_y y_i + X_{jk_t}\beta_T T_i + X_{jk_t}\beta_v v_i + \alpha_y y_i P_{jk_t} + \alpha_T T_i P_{jk_t} + \alpha_v v_i P_{jk_t} + \varepsilon_{ijk_t}. \quad (3)$$

In the next section, we discuss the empirical results from our basic and extended models.

5. Empirical Analysis and Results

In §5.1, we discuss the main results obtained from the main data set. In §5.2, we discuss our robustness tests using (1) the same model based on different samples using alternative levels of online review data and (2) a different model based on the same data sets. Then, in §5.3, we further discuss the results on model validation by comparing our model with the current competitive ones. In §5.4, we provide some managerial implications by conducting counterfactual policy experiments. Finally, in §5.5, we briefly discuss the results from our extended model.

5.1. Results from the Basic Model

Five location-based characteristics have a positive impact on hotel demand: external amenities, proximity to the beach, public transportation, a highway, and downtown. Hotels providing easy access to public transportation (e.g., subways or bus stations), highway exits, restaurants, shops, or a downtown area can have a much higher demand. *BEACH* also has a positive impact on demand. Most beach-based hotels in our data set were located in the South, where the weather typically stays warm year round. Therefore, the desirability of a “walkable” beachfront did not lessen even in the winter (the time of our data).

Two location-based characteristics have a negative impact on hotel demand: the annual crime rate and proximity to a lake. The higher the average reported crime rate in a local area, the lower the desirability of that area’s hotels. This result indicates that neighborhood safety plays an important role in the hotel industry. The second of these characteristics is interesting because one would expect people to choose—rather than avoid—a hotel near a lake. However, most waterfront-based hotels in our data set were located in places where the weather becomes extremely cold from November to January. A waterfront location is therefore going to be less desirable to travelers in winter.

To further examine the impact of lakefront locations, we collected weather data from the National

Oceanic and Atmospheric Administration (NOAA) on the average temperature from November 2008 to January 2009 for all cities in our data set. Then we defined two dummy variables: *HIGH TEMP*, which equals 1 if the average temperature is higher than 50°, and *LOW TEMP*, which equals 1 if the average temperature is lower than 40°. We interacted *HIGH TEMP* and *LOW TEMP* separately with *LAKE* in our model. The results show that the interaction of *LOW TEMP* with *LAKE* has a significantly negative effect. This finding supports our earlier argument. Meanwhile, the interaction of *HIGH TEMP* with *LAKE* showed a significantly positive effect, suggesting that warmer weather may help the lake area to attract more visitors. As a robustness check, we conducted a similar analysis for *BEACH* conditional on high and low temperatures. The results show a similar trend. Column 8 of Table 3 shows the corresponding estimation results considering the interactions with the temperature.

Class and amenity count both have a positive impact on hotel demand. Hotels with a higher number of amenities and higher star levels have higher demand, controlling for price. Reviewer rating is also positively associated with hotel demand. With regard to the *TL_REVIEWCNT* and *TA_REVIEWCNT* variables, we find a positive sign for their linear form and a negative sign for their quadratic form. This finding indicates that the economic impact from the customer reviews is increasing in the volume of reviews but at a decreasing rate, as one would expect.

The textual quality and style of reviews demonstrated a statistically significant association with demand. All the readability and subjectivity characteristics had a statistically significant association with hotel demand. Among the readability subfeatures, complexity, syllables, and spelling errors had a negative sign and therefore are negatively associated with hotel demand. This finding implies that reviews with higher readability characteristics (shorter sentences and less complex words) and reviews with fewer spelling errors are positively associated with demand. On the other hand, the sign of the coefficients on *CHARACTERS* and *SMOG* is positive, implying that longer reviews that are easier to read are positively associated with demand.⁹ These results indicate that consumers can form a judgment about the quality of

⁸ We tried other combinations to classify high versus low temperatures ($\geq 70^\circ$ as high and $\leq 30^\circ$ as low (or) $\geq 60^\circ$ as high and $\leq 20^\circ$ as low), but they all yielded qualitatively similar results.

⁹ To alleviate any possible concerns with multicollinearity between *SMOG* and *SYLLABLES*, we reestimate our model after excluding the *SMOG* index variable. We found no change in the qualitative nature of the results across the different data sets.

a hotel by judging the quality of the (user-generated) reviews.

Both “mean subjectivity” and “subjectivity standard deviation” are negatively associated with demand. This finding implies that consumers tend to believe reviews that contain objective information (e.g., factual description of a room) over reviews that contain subjective information (e.g., comfort of a room). With respect to the subjectivity standard deviation, our findings suggest that people prefer a “consistent objective style” from online customer reviews compared to a mix of objective and subjective sentences. The last review-based characteristic was disclosure of reviewer identity (*ID*). This variable demonstrated a positive association with hotel demand. This result is consistent with previous work (Forman et al. 2008), suggesting that identity information about reviewers in the online travel community can positively shape community members’ judgment of hotels. *PRICE* has a negative sign, which is as expected.¹⁰

Besides the above qualitative implications, we also quantitatively assess the economic value of different hotel characteristics. More specifically, we examined the magnitude of marginal effects on hotel demand for the location-, service-, and review-based hotel characteristics. The presence of a nearby beach increases hotel demand by 18.23%, on average. In contrast, a nearby lake or river decreases demand by 12.83%. Meanwhile, easy access to transportation and to highway exits increases demand by 18.32% and 7.87%, respectively. Presence near a downtown area increases demand by 5.29%. With regard to service-based characteristics, a one-star increase in hotel class leads to an increase in demand by 4.13%, on average. Moreover, the presence of one more internal or external amenity increases demand by 0.06% or 0.08%, respectively. Demand decreases by 0.28% if the local crime rate increases by one unit.

With regard to the review-based characteristics, the SMOG index (which represents the readability of the review text) was associated with the highest marginal influence on demand on average. A one-level increase in the SMOG index is associated with an increase in hotel demand by 9.3%, on average. A one-unit increase in the number of characters is associated with an increase in hotel demand by 0.12%, whereas a one-unit increase in the number of spelling errors, syllables, or complexity is associated with a decrease in hotel demand by 1.41%, 0.50%, and 1.18%, respectively. In terms of review subjectivity, a 10% increase in the average subjectivity level

is associated with a decrease in hotel demand by 1.55%, and a 10% increase in the standard deviation of subjectivity reduces demand by 4.74%. Finally, a 10% increase in the reviewer identity-disclosure levels is associated with an increase in hotel demand by 0.68%.

Note that during the period of our data collection, Travelocity displayed 5 reviews per page, whereas TripAdvisor displayed 10 per page. To minimize the bias Web page design might cause (as some customers may only read the reviews on the first page of each site), we decided to consider two more alternatives besides our main data set: data set (II), with hotels that have at least 5 reviews, and data set (III), with hotels that have at least 10 reviews. Controlling for brand effect, the estimation results from these three data sets are illustrated in columns 2–4 of Table 3. For normalization purposes, we used the logarithms of price, characteristics, syllables, spelling errors, crime rate, internal amenities, external amenities, and review count (both TripAdvisor and Travelocity) in all the analyses in this paper. See Table 3, columns 5–7, for the corresponding results. The estimation results from the three data sets are highly consistent. In general, all the coefficients illustrate a statistical significance, with a *p*-value equal to or below the 5% level across all three data sets. Moreover, a large majority of variables present a high significance, with a *p*-value below the 0.1% level.

5.2. Robustness Checks

To assess the robustness of our estimation model and results, we report some additional checks. First, we estimated the same model on alternative sample splits. We considered three alternative data sets: data set (IV), containing hotels with at least one review from TripAdvisor.com; data set (V), containing hotels with at least one review from Travelocity.com; and data set (VI), containing hotels with at least one review from both sites. Table A.1 in Appendix A presents the results. We found that the coefficients from the estimations are qualitatively similar to our main results. Moreover, similar to those in the main results, most variables in the robustness tests illustrate statistical significance at or below the 5% level or stronger. Thus our estimation results, based on the hybrid random coefficient model, are consistent across different data sets.

Second, we conducted another group of tests using an alternative model commonly used in the industrial organization and marketing literature, the random coefficient logit model (BLP; Berry et al. 1995). As mentioned in §4, the key difference between the BLP approach and our model is that BLP introduces a demand taste shock at each product level (i.e., hotel)

¹⁰ We also considered the covariates “airport,” “convention centers,” and “number of rooms.” The estimation results are consistent with our current results, but the coefficients for the three characteristics are statistically insignificant.

rather than at a group level (i.e., the travel category), as in our model. Consequently, the substitution space for BLP is different in the sense that BLP does not distinguish between the two types of cross substitutions—the within-travel and between-travel categories. Rather, it would treat all hotels as possible substitutes for each other. We added two sets of dummy variables, one for brand and the other for travel category. We conducted the same set of estimations based on data sets (I)–(VI). Table A.2, columns 2–7, in Appendix A contain the results. In addition to an alternative specification with homogeneous coefficients on the travel category dummies, we further considered consumers' heterogeneous preferences by assigning random coefficients to these dummies. The last column in Table A.2 in Appendix A shows the corresponding results.

The results from the BLP model are consistent with our main estimation results using the hybrid model. Specifically, the coefficients from the BLP estimation demonstrate three trends: (i) they have the same signs as our main results from the hybrid model, which means the economic effects are consistent in direction; (ii) they exhibit lower levels of statistical significance compared with our main results; and (iii) the magnitude of these coefficients is generally higher compared with our main results. These three trends are also consistent with the findings in Song (2011). In the next subsection, our model validation results further confirm this finding.

Third, to alleviate concerns regarding whether UGC has a causal effect on demand, we conducted an additional robustness test using a regression discontinuity (RD) design as suggested by Luca (2011). More specifically, our test builds on the special “rounding mechanism” used by both Travelocity.com and TripAdvisor.com. These two websites generate their overall rating for each hotel by rounding the average review ratings to the closest half star. For example, if the average rating across all reviewers is 3.24, the site rounds that number to 3; if the average rating is 3.25, the site rounds that number to 3.5. Thus, we looked at those hotels with an unrounded average rating *just below* and *just above* each rounding threshold. Then we looked for any discontinuous jumps in sales patterns following discontinuous jumps in the rounded overall rating while controlling for the continuous unrounded rating and other hotel characteristics. Similar to Luca (2011), we based this design on the assumption that all sales-affecting predetermined characteristics of hotels become increasingly similar when approaching both sides of a rounding threshold. We found a significant positive treatment effect, suggesting that keeping all else the same, the discontinuous pattern in the sales is caused by the discontinuous pattern in the rating. This finding strongly

suggests user-generated reviews have a causal impact on hotel demand.

As a further robustness check, we tried different bandwidths (bin size) of the neighborhood near the rounding threshold. Our results are consistent and insensitive to the bin size. Moreover, to eliminate the possibility of self-selection bias (e.g., as addressed by Hartmann et al. 2011) that could potentially invalidate the RD design (e.g., hotels may submit the reviews themselves to pass the threshold), we performed an additional McCrary density test (McCrary 2008) as suggested by Luca (2011). In particular, we divided the range of rating into small bins with a range of 0.05. Then we checked for whether the density for the number of submitted reviews is disproportionately large in the bins just above the rounding threshold (e.g., 3.25–3.3). We performed this check because if hotels are gaming the system and submitting reviews themselves, we would expect to see such a pattern. We did not find any significant difference in the density from our data, suggesting no evidence for hotel “gaming” behavior. In summary, the additional robustness tests using RD design together with the McCrary density test give us more confidence in the causal impact of online reviews on product demand.

5.3. Model Comparison

For comparison purposes, we estimated three baseline models: the BLP model, the PCM, and the nested logit model with the travel category at the top hierarchy. Based on a study by Steckel and Vanhonacker (1993), we randomly partitioned our main sample data set (I) into two parts: a subset with 70% of the total observations as the estimation sample and a subset with 30% of the total observations as the hold-out sample. To minimize any potential bias from the partition procedure, we performed a 10-fold cross validation. We conducted this validation process for our random coefficient model and the three baseline models. Furthermore, to examine the model's ability to capture a deeper level of consumer heterogeneity, we compared an extended version of our model with an extended version of the BLP model when incorporating additional interaction effects (i.e., travel purpose interacted with price and hotel characteristics).

To examine the significance of the UGC-based, location-based, and service-based hotel characteristics, we compared the original hybrid model with the same model but excluding the UGC, location, and service variables, respectively. Finally, to evaluate the usefulness of different aspects of UGC in modeling the demand, we further conducted model comparison using the hybrid model but excluding the numerical ratings and the textual review features,

respectively. We also evaluated models without each of the textual features, such as readability, subjectivity, and reviewer-identity variables, respectively. We have done the above work for both in-sample and out-of-sample comparisons. Tables B.1–B.8 in Appendix B contain the results.¹¹

The results show that conditioning on UGC variables significantly improves a model's predictive power. With respect to out-of-sample root mean square error (RMSE), the model fit improves by 36.16% when adding the UGC variables. Similar trends in improvement in our model fit occur with respect to the other two metrics, mean square error (MSE) and mean absolute deviation (MAD), in both in-sample and out-of-sample analyses.

Our out-of-sample results in Table B.5 illustrate that our model improves by 12.86% in RMSE compared to the BLP model with no random coefficients on travel category dummies. This number becomes 53.85%, 63.28%, and 9.64% for the PCM, the nested logit model, and the BLP model with random coefficients on travel category dummies, respectively. Thus, our model provides the best overall performance in both precision (i.e., RMSE, MSE) and deviation (i.e., MAD) of the predicted market share. The nested logit model demonstrates the weakest predictive power. Moreover, as illustrated in Table B.6, when incorporating interaction effects, although both models show improvement in predictive power, the extended hybrid model performs much better than the extended BLP model.

Table B.7 shows that by including the UGC, location-based, and service-based variables, our model fit improves by 36.16%, 55.77%, and 53.56%, respectively, in RMSE. Similar trends in improvement in our model fit occur with respect to MSE and MAD. Therefore, our results suggest that the model's predictive power would decrease the most if we were to exclude the location-based variables from our model, followed by the service-based variables, and finally by the UGC variables. This finding strongly indicates location- and service-based characteristics are indeed the two most influential factors for hotel demand.

Moreover, Table B.8 shows that of all the UGC-related features, textual information improves the model's predictive power significantly more than the

numerical features by 35.17% and 21.06%, respectively, in RMSE. In addition, within the set of textual features, the review readability and subjectivity show a higher impact than the reviewer-identity information.

5.4. Counterfactual Experiments

A key advantage of structural modeling is its potential for normative policy evaluation. To explicitly measure the economic impact of strategic policies, we conducted a counterfactual experiment. Specifically, we looked into how a price change in one type of hotel affects the demand for other types of hotels to examine the extent of competition (and consequent substitution patterns) between hotels. We focused on hotels with different star ratings for this analysis.

We assumed a price cut by 20% for all four-star hotels in an effort to determine the demand changes for the five-, three-, two-, and one-star hotels. We find the demand for four-star hotels increases 2.9%, whereas the demand for all other hotel classes decreases. Demand for five-star hotels drops the most (5.34%), followed by three-star hotels (3.88%), one-star hotels (2.87%), and finally two-star hotels (2.6%). We also conducted similar analyses for hotels from other classes. For example, by assuming a 20% price cut for the three-star hotels, we find the demand for three-star hotels increases by 2.79%, and the demand for four- and two-star hotels drops the most—5.14% and 5.01%, respectively.

The basic findings from the above set of experiments are as follows: (i) a price cut for a particular class of hotels tends to cause a demand drop for all hotels in the lower-level classes, and (ii) the closest substitutes for four-star hotels are five-star hotels, the closest substitutes for three-star hotels are four- and two-star hotels, and the closest substitutes for two-star hotels are one-star hotels.

5.5. Results from Model Extensions

As discussed in §§4.3 and 4.4, we also estimated two extended models. Table 4 shows the estimation results for the extended model with additional text features. We see the qualitative nature of our main results remains the same. The three features that have a positive and statistically significant impact on demand are food quality, hotel staff, and parking facilities. Among these three features, food quality presents the highest positive impact, followed by hotel staff and parking. In contrast, bedroom quality has a negative impact on demand. This negative sign may seem surprising. One possible explanation is that consumers often use bedroom quality as a cue for price, especially given that quality in our data is a proxy for the number of beds and size of the room (full, queen, king, etc.). This situation may occur when prices are obfuscated

¹¹ With regard to the unobserved characteristics required for out-of-sample prediction using the hybrid model, BLP model, and PCM, we applied the same method as suggested in Athey and Imbens (2007). We drew the unobserved characteristics for the holdout sample randomly from the marginal distribution of unobserved characteristics from the estimation sample. This method has also been used in the marketing literature. See, for example, Nair et al. (2005), who infer the structural error for the holdout sample from the marginal distribution of the structural error across different markets derived from the estimation sample.

Table 4 Extended Model (I) with Additional Text Features

Variable	Coefficients (Std. error)		
	I	II	III
Means			
PRICE (log)	−0.144*** (0.015)	−0.150*** (0.014)	−0.157*** (0.014)
CHARACTERS (log)	0.008*** (0.001)	0.009*** (0.002)	0.009*** (0.002)
COMPLEXITY	−0.015*** (0.003)	−0.014*** (0.002)	−0.012*** (0.002)
SYLLABLES (log)	−0.043*** (0.012)	−0.044*** (0.012)	−0.045*** (0.012)
SMOG	0.081** (0.029)	0.078** (0.027)	0.076** (0.029)
SPELLERR (log)	−0.132*** (0.031)	−0.132*** (0.026)	−0.139*** (0.023)
SUB	−0.149*** (0.032)	−0.151*** (0.036)	−0.162*** (0.039)
SUBDEV	−0.408*** (0.100)	−0.412*** (0.095)	−0.417*** (0.102)
ID	0.055* (0.031)	0.063* (0.034)	0.066* (0.034)
CLASS	0.039*** (0.009)	0.040*** (0.009)	0.045*** (0.009)
CRIME (log)	−0.033** (0.012)	−0.032* (0.017)	−0.028* (0.015)
EXTAMENITY (log)	0.008*** (0.002)	0.007*** (0.001)	0.007*** (0.002)
BEACH	0.157*** (0.004)	0.165*** (0.004)	0.163*** (0.004)
LAKE	−0.118*** (0.030)	−0.111*** (0.031)	−0.112*** (0.033)
TRANS	0.163*** (0.003)	0.167*** (0.006)	0.173*** (0.009)
HIGHWAY	0.065* (0.028)	0.070*** (0.021)	0.073** (0.024)
DOWNTOWN	0.044*** (0.004)	0.047*** (0.004)	0.048*** (0.005)
TA_RATING	0.034* (0.018)	0.041** (0.018)	0.044** (0.021)
TL_RATING	0.036*** (0.005)	0.037*** (0.005)	0.038*** (0.006)
TA_REVIEWCNT (log)	0.177*** (0.038)	0.180*** (0.042)	0.183*** (0.043)
TA_REVIEWCNT ² (log)	−0.059*** (0.006)	−0.063*** (0.010)	−0.062*** (0.009)
TL_REVIEWCNT (log)	0.017*** (0.002)	0.016*** (0.002)	0.018*** (0.002)
TL_REVIEWCNT ² (log)	−0.025*** (0.006)	−0.031*** (0.008)	−0.032*** (0.008)
FOOD	0.115** (0.045)	0.122*** (0.034)	0.124** (0.042)
STAFF	0.059** (0.024)	0.059** (0.020)	0.064** (0.024)
BATHROOM	0.046 (0.103)	0.047 (0.105)	0.045 (0.110)
BEDROOM	−0.015* (0.007)	−0.016 (0.009)	−0.016 (0.011)
PARKING	0.036*** (0.007)	0.037*** (0.007)	0.040*** (0.009)
Constant	0.031 (0.019)	0.032 (0.022)	0.035 (0.026)
Brand control	Yes	Yes	Yes
Interaction effect (α_y) and standard deviations (α_y)			
PRICE (log) ×	0.020*** (0.004)	0.026*** (0.005)	0.022*** (0.007)
INCOME (log)	0.016 (0.087)	0.012 (0.092)	0.013 (0.106)
PRICE (log)			
Standard deviations (β_y)			
CLASS	0.025*** (0.006)	0.031** (0.011)	0.033** (0.012)
CRIME (log)	0.013 (0.022)	0.015 (0.026)	0.016 (0.022)
AMENITYCNT (log)	0.024 (0.037)	0.023 (0.035)	0.029 (0.043)
EXTAMENITY (log)	0.007 (0.023)	0.012 (0.033)	0.012 (0.029)
BEACH	0.065*** (0.015)	0.063*** (0.017)	0.056** (0.021)
LAKE	0.114** (0.044)	0.103** (0.041)	0.099** (0.038)
TRANS	0.132* (0.078)	0.133* (0.083)	0.134* (0.081)
HIGHWAY	0.077* (0.043)	0.065 (0.049)	0.067 (0.048)
DOWNTOWN	0.036*** (0.009)	0.039*** (0.011)	0.044*** (0.014)
GMM obj. value	8.412e−4	8.066e−4	8.137e−4

Notes. I is based on the main data set (at least one review from either TA or TL). II is based on main data set with reviews ≥ 5 . III is based on main data set with reviews ≥ 10 .

*** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$; † $p \leq 0.1$.

on the main results page and are only available just before checkout.

Next, we estimate the model with additional interaction effects to better understand how the distribution of consumers' demographic information influences the distribution of consumers' heterogeneous preferences. More specifically, the extended model is able to capture six interaction effects:

1. Between *travel category* and *hotel characteristics* (e.g., location, service).
2. Between *travel category* and *price*.
3. Between *income* and *hotel characteristics*.
4. Between *income* and *price*.
5. Between *unobserved consumer characteristics* and *hotel characteristics*.
6. Between *unobserved consumer characteristics* and *price*.

Note that the basic model already captures interaction effects 4–6, whereas the extended model further relaxes the model assumptions by considering three additional interaction effects. The corresponding results for the extended model are provided in Tables 5(a) and 5(b).

First, we notice that consumers' travel purposes can explain their heterogeneous tastes toward price. For example, from Table 5(a), we see the mean price coefficient is -0.145 . From Table 5(b), we can infer that if a consumer is on a business trip, her price

Table 5(a) Extended Model (II) Mean Coefficients

	Mean coefficients (Std. error)
PRICE (log)	-0.145^{***} (0.003)
CHARACTERS (log)	0.009^{***} (0.002)
COMPLEXITY	-0.012^{***} (0.003)
SYLLABLES (log)	-0.045^{***} (0.008)
SMOG	0.083^{**} (0.029)
SPELLERR (log)	-0.129^{***} (0.003)
SUB	-0.138^{***} (0.007)
SUBDEV	-0.403^{***} (0.016)
ID	0.055^{*} (0.030)
CLASS	0.037^{***} (0.008)
CRIME (log)	-0.025^{*} (0.016)
AMENITYCNT (log)	0.005^{**} (0.002)
EXTAMENITY (log)	0.007^{***} (0.001)
BEACH	0.158^{***} (0.005)
LAKE	-0.111^{***} (0.021)
TRANS	0.159^{***} (0.003)
HIGHWAY	0.064^{*} (0.030)
DOWNTOWN	0.045^{***} (0.002)
TA_RATING	0.033^{**} (0.012)
TL_RATING	0.031^{**} (0.011)
TA_REVIEWCNT (log)	0.180^{***} (0.046)
TA_REVIEWCNT ² (log)	-0.055^{***} (0.007)
TL_REVIEWCNT (log)	0.014^{***} (0.003)
TL_REVIEWCNT ² (log)	-0.021^{**} (0.008)
Constant	0.037^{**} (0.017)

*** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$; † $p \leq 0.1$.

Table 5(b) Extended Model (II) Additional Interaction Effects and Standard Deviations of Coefficients

Variable	v_i	INCOME (log)	Family	Business	Romance	Tourist	Kids	Seniors	Pets
PRICE (log)	0.012 (0.076)	0.021^{***} (0.003)	-0.011^{***} (0.000)	0.038^{***} (0.005)	-0.003^{***} (0.000)	-0.015^{***} (0.004)	0.032^{***} (0.004)	-0.007 (0.071)	-0.018 (0.043)
CLASS	0.019 (0.028)	0.017^{***} (0.002)	0.046^{***} (0.012)	0.023^{**} (0.011)	0.108^{***} (0.028)	-0.014^{***} (0.003)	0.021^{*} (0.011)	0.033 (0.027)	-0.052 (0.081)
CRIME (log)	0.012 (0.023)	0.002 (0.007)	-0.033^{*} (0.017)	0.087 (0.065)	-0.037 (0.034)	-0.023 (0.075)	-0.046^{***} (0.003)	0.012 (0.011)	0.026 (0.044)
AMENITYCNT (log)	0.021 (0.043)	0.015 (0.012)	0.027 (0.029)	0.035 (0.040)	0.008^{***} (0.001)	-0.004 (0.006)	0.001 (0.024)	-0.005 (0.009)	-0.011 (0.027)
EXTAMENITY (log)	0.014 (0.022)	0.035 (0.050)	0.063 (0.097)	0.011^{***} (0.004)	0.017^{***} (0.004)	0.003 (0.004)	0.009^{***} (0.001)	-0.014 (0.011)	0.049 (0.042)
BEACH	0.029 (0.026)	0.037^{**} (0.014)	-0.059 (0.045)	-0.040^{*} (0.023)	0.165^{***} (0.010)	-0.062 (0.049)	0.153^{***} (0.025)	0.037 (0.031)	0.015 (0.017)
LAKE	0.053 (0.083)	0.011 (0.024)	-0.206 (0.183)	0.033 (0.064)	0.022 (0.018)	-0.125 (0.201)	-0.027 (0.025)	-0.050 (0.121)	0.028 (0.040)
TRANS	0.077 (0.081)	-0.125 (0.231)	-0.025^{**} (0.011)	0.157^{***} (0.014)	-0.011^{*} (0.006)	-0.033 (0.056)	0.041 (0.038)	0.166 (0.202)	-0.039 (0.031)
HIGHWAY	0.042 (0.039)	-0.009^{*} (0.005)	-0.021^{***} (0.005)	0.120^{***} (0.031)	-0.022^{***} (0.002)	0.039^{***} (0.009)	0.063^{***} (0.013)	0.062 (0.100)	0.177 (0.158)
DOWNTOWN	0.026 (0.017)	-0.018 (0.021)	0.188^{***} (0.010)	0.041 (0.037)	-0.002 (0.017)	-0.029^{*} (0.015)	0.121^{***} (0.023)	-0.040 (0.036)	0.136 (0.101)

*** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$; † $p \leq 0.1$.

coefficient will increase by 0.038 relative to the average traveler. In contrast, if a consumer is on a family trip or a romantic trip, her price coefficient is 0.011 or 0.003, respectively, below that of the average traveler. Among all types of travelers, all else being equal, tourists (i.e., travelers with a large group tour) tend to be the most price sensitive, with a price coefficient 0.015 below that of the average traveler, whereas business travelers are the least price sensitive. These findings are consistent with those from the marginal effects of price change on demand. For instance, we find that a 20% increase in hotel price leads to a 1.8% demand drop from business travelers, a 2.6% drop from family travelers, a 2.48% demand drop from romantic travelers, and a 2.67% demand drop from tourists.

Furthermore, we find travel purpose also influences consumer heterogeneity toward different hotel location and service characteristics. For instance, business travelers have the highest marginal valuation for close proximity to a highway and public transportation. From Table 5(a), we see the mean coefficients for *HIGHWAY* and *TRANS* are 0.064 and 0.159, respectively. According to the estimated interaction effects in Table 5(b), the coefficients from business travelers for *HIGHWAY* and *TRANS* are higher than the mean coefficients, with an increase to 0.120 and 0.157, respectively. Likewise, the presence of an interstate highway near a hotel increases demand from business travelers by 19.20%, compared to a 6.68% increase from the average traveler and a 5.34% increase from romantic travelers. Similarly, the presence of public transportation near a hotel increases hotel demand from business travelers by 35.74%, compared to a 17.98% increase from the average traveler and a 15.16% increase from family travelers.

In contrast, romantic travelers are more sensitive to *CLASS* and *BEACH* compared with other types of travelers. For example, the presence of a beach near a hotel increases demand from romantic travelers by 37.11%, compared to an 18.15% increase from the average traveler and a 13.54% increase from business travelers. Similarly, a one-star improvement in hotel class leads to an increase in hotel demand from romantic travelers by 21.95%, compared to a 5.61% increase from the average traveler and a 3.47% increase from tourists.

Finally, Table 5(b) shows that after we account for the interaction effects, all the estimates on the unobserved consumer characteristics (i.e., v_i) become insignificant at conventional significance levels. This finding is consistent with Nevo (2001), who shows that the observable—rather than the unobservable—demographics explain most of the consumer heterogeneity.

6. Utility Gain-Based Hotel Ranking

Having estimated the parameters from the model in §4, we next derive the utility gain a consumer with a particular travel purpose receives from staying in a given hotel. This helps us propose a new ranking system for hotels based on the average utility gain from transactions in each hotel. As discussed in §4, to capture consumer heterogeneity, we model the utility from each hotel for each consumer as consisting of two parts: the mean and the deviation. The mean utility provides us with a good estimation of how much consumers can benefit from choosing a particular hotel, whereas the deviation of utility describes the variance of this benefit. We are interested in knowing consumers' utility gain on an aggregate level from choosing a certain hotel. Therefore, we define the *average utility gain* from hotel j with travel category type k as the sum of its mean utility across all weeks:

$$\text{Utility Gain}_{jk} = \sum_t \bar{\mu}_{jkt}. \quad (4)$$

6.1. Ranking Hotels

In the ranking approach we propose, a hotel that provides a comparably higher average utility gain than others would appear at the top of our list. Using the coefficients for hotel characteristics estimated from our model, we are able to compute the average utility gain based on Equation (4). Notice that in this equation, $\bar{\mu}_{jkt}$ is the average value of the estimated utility gain over the consumer population in market t :

$$\bar{\mu}_{jkt} = \frac{1}{N} \sum_i (\delta_{jkt} + X_{jkt} \beta_v v_i + \alpha_y y_i P_{jkt} + \alpha_v v_i P_{jkt} + \varepsilon_{ikt}), \quad (5)$$

where N represents the total number of consumers involved in the estimation. This definition takes into account all the sources of uncertainty, such as the random coefficients and model errors. Since by assumption $X_{jkt} \beta_v v_i + \alpha_y y_i P_{jkt} + \alpha_v v_i P_{jkt} + \varepsilon_{ikt}$ is mean-zero (e.g., Nevo 2001), in our computation, Equation (5) can be simplified as follows:

$$\bar{\mu}_{jkt} = \frac{1}{N} \sum_i (\delta_{jkt}). \quad (6)$$

Using the estimates from the previous analysis, we compute $\bar{\mu}_{jkt}$. We determine the final average utility gain, *Utility Gain*_{jk}, by summing over $\bar{\mu}_{jkt}$ across all markets. As the final ranking criterion, the average utility gain provides us with a new metric to rank hotels in response to a user query on the travel search engine.

6.2. User Study Based on Online Survey

To evaluate the quality of our ranking technique, we conducted an extensive user study in which we

designed and executed several online surveys using Amazon Mechanical Turk. We computed the expected utility for each hotel from our parameter estimates and ranked the hotels in each city according to their average utility gain. Then we generated different rankings for the top 10 hotels in accordance with several existing baseline criteria deployed by travel search engines: most booked, price low to high, price high to low, hotel class, hotel size (number of rooms), and number of internal amenities. We also considered four other benchmark criteria based on UGC: customer rating from TripAdvisor.com, customer rating from Travelocity.com, mixed rating from TripAdvisor.com and Travelocity.com, and maximum online review count. Moreover, to examine the significance of the UGC and of the comprehensive model to the overall performance of the ranking scheme, we generated two more baselines using the BLP model (as described in §5.2) and the same hybrid model, but excluding all UGC variables. Finally, we also generated a combined ranking using combined criteria of price and hotel class to examine whether a ranking that attempts to introduce diversity artificially can compete with our utility-based ranking. We did this by interlacing the top five hotels with “the lowest price” and the top five hotels with “the highest number of stars.”¹²

In our study, we presented our model-generated ranking together with one of the above-mentioned alternative rankings and asked users to compare each pair of rankings—that is, our ranking paired with one of the existing benchmarks. To avoid any potential bias, we did not release any information to the users about the criteria for generating those rankings, and we randomized the presentation order of the rankings. The studies in our ranking evaluations were blind pairwise tests in which we presented the two rankings side, by side and the user had to pick one of them without having any information beyond the list of the hotels. This setting resulted in 13 different surveys for each of the six cities (Los Angeles, New Orleans, New York, Orlando, San Francisco, and Salt Lake City), giving us 78 surveys with 100 participants each, equaling 7,800 user comparisons of different ranking lists. To further control for any biases, we conducted this user study with stringent controls in the design and execution of the survey. Appendix C provides more details on the user study.

¹² We also tried interlacing rankings with different criteria, such as “the highest price” and the “lowest number of stars,” or “the lowest price” and the “lowest number of stars,” or “the highest price” and the “highest number of stars.” The results are similar, which suggests that customers prefer a list of hotels that specialize in a variety of characteristics, rather than a variety of hotels, each specializing in only one characteristic.

Our results (see Table 6) show that for each of the 13 comparisons in each of the six cities, the majority of customers preferred our ranking ($p = 0.05$, sign test). Notice that in all 78 surveys, we observe a statistically significant difference for our ranking ($p = 0.05$, sign test). The overall set of results (i.e., in *none* of the 78 surveys was our ranking deemed worse) indicates that our ranking strategy is preferable to the existing baselines. (Figure C.1 in Appendix C provides a screenshot of sample tasks from the online survey.) Moreover, users preferred our ranking based on the hybrid model with UGC variables over the one without UGC variables and over the one generated based on the BLP model. This finding further demonstrates the importance of incorporating UGC variables in any demand estimation model that generates a ranking system for products in shopping search engines. Furthermore, we checked the click-throughs of the top-ranked hotel in each ranking list. On average (over a total of 78 comparison tasks), the top-ranked hotel in our utility-based ranking list received approximately two times the clicks compared to its counterpart in the competitor ranking list. This provides additional support that our ranking can help users locate the best-value hotel early on.

To better understand how users interpret the utility-based ranking, we also asked consumers why they chose a particular ranking. The majority of users indicated our utility-based ranking promoted the idea that price was not the main factor in rating a hotel’s quality. Instead, a good ranking recommendation satisfies customers’ multidimensional preferences for hotels. Moreover, users strongly preferred the diversity of the retrieved results, given that the list consisted of a mix of hotels cutting across several price and quality ranges. In contrast, the other ranking approaches tended to list hotels of only one type (e.g., very expensive for “star ratings” or mainly three-star hotels for “most booked”). Notice that even the ranking baseline with the combined criteria showed a similar trend. This finding further indicates customers prefer a list of hotels where each specializes in a variety of characteristics rather than a variety of hotels where each specializes in only a few characteristics.

Of course, diversity of results is a well-known factor of user satisfaction in Web search (Agichtein et al. 2006). Although we could potentially try to imitate solutions from Web search and introduce diversity in the results in an exogenous manner, we observe that the approach based on consumer-utility theory introduces diversity naturally in the results. This result seems intuitive: if a specific segment of the market systematically appears to be underpriced—hence, introducing a nondiverse set of results—market forces would modify the prices for the whole segment

Table 6 Ranking User Study Results

	Rating (TA) (%)	Rating (TL) (%)	Rating (mixed ^a) (%)	Most booked (%)	Price low to high (%)	Price high to low (%)	Hotel class (%)	No. of reviews (%)	No. of rooms (%)	No. of amenities (%)	Combined Price with Rating (%)	No UGC (%)	BLP ^b (%)
New York	64	68	64	66	62	74	70	68	66	62	66	70	68
Los Angeles	62	64	66	66	64	71	74	64	66	65	62	74	70
San Francisco	64	68	75	78	62	72	73	66	62	62	66	70	68
Orlando	66	66	69	71	68	76	70	63	69	62	64	76	72
New Orleans	64	69	68	62	66	70	70	72	72	74	63	72	70
Salt Lake City	66	68	66	64	64	63	64	64	66	66	64	62	64
Significance level				$P = 0.05$ ≥ 59	$P = 0.01$ ≥ 62	$P = 0.001$ ≥ 66					(Sign test, $N = 100$)		

Notes. The percentages in the table indicate how often users preferred our ranking scheme when presented side by side with an alternative. For example, in one of the surveys for New York City, 64% users chose our ranking over the alternative ranking based on the TripAdvisor rating. This rate is statistically significant at $P < 0.01$ level according to the sign test. TA, TripAdvisor.com; TL, Travelocity.com.

^aMixed rating strategy: (i) average of TripAdvisor and Travelocity ratings when both are available, (ii) equal to one of the two ratings if the other one is missing, and (iii) zero when both ratings are missing.

^bBLP with homogeneous coefficients on travel category dummies.

accordingly. Thus, these results dovetail well with our empirical estimation, which suggests that our utility-based ranking model can capture consumers' true purchase motivations.

Moreover, our user study indicates a star rating system would not come close to achieving the same goal. Apparently, one could interpret a subject's star rating as a discrete approximation of her utility for a hotel; thus, a ranking based on star rating should perform as well as a ranking based on utility, as the latter is just a money-metric transformation of the former. However, this argument is not true, because the matching of consumers to hotels in star rating systems is not random. A consumer only rates a hotel she has previously chosen (i.e., the one that maximizes her perceived utility gain). Consequently, the average star rating for each hotel need not reflect the population average utility but rather consumers' satisfaction with their own choices. Thus, a ranking based on average star ratings need not reflect a ranking based on average utility.

7. Conclusions and Implications

In this paper, we estimate the economic value of different location- and service-based characteristics of hotels, given the associated local infrastructure. We build a discrete choice structural model, taking into consideration the two sources of consumer heterogeneity introduced by the different travel occasions and different hotel characteristics. Using techniques from text mining, image classification, social geotagging, human annotations, and geomapping tools, we estimate this model based on a unique data set consisting of actual transactions for hotels located in the United States, their external and internal attributes, and multiple aspects of review text.

Based on the estimates, we propose a new hotel ranking system in which a hotel that provides a comparably higher average utility gain would appear at the top of our list. By doing so, we can provide customers with the best-value hotels early on, thereby improving the quality of local searches for such hotels. The estimation models are privacy friendly, as they do not require individual consumer data but rather rely on aggregate data.

On a broader note, the objective of this paper is to illustrate how user-generated and crowdsourced content on the Internet can be mined and incorporated into a demand estimation model. Our interdisciplinary approach can provide insights for using text-mining and image classification techniques in economics and marketing research. Simultaneously, such research can also highlight the value of using an economic context to computer scientists to estimate both the intensity and the polarity of the UGC. Toward this end, we empirically estimate the economic value of different hotel characteristics, including both service- and location-based characteristics.

Our research enables us to not only quantify the economic impact of hotel characteristics but also, by reversing the logic of this analysis, to identify the characteristics that most influence demand for a particular hotel. After inferring the economic significance of each characteristic, we incorporate the characteristics in a model of expected utility gain estimation. The end goal is to generate a ranking system that recommends hotels that provide the best value for the money, on average. The key idea is that hotels (or products, in general) that provide consumers with a higher surplus should be ranked higher in response to consumer queries on search engines. We conduct blind tests using real users recruited through AMT to examine how well our ranking

system performs relative to existing alternatives. We find our ranking performs significantly better than several baseline ranking systems that are currently being used.

We should also note that our ranking scheme is *causal*, in the sense that the model can predict what *should* happen when we observe changes in the market. For example, when we see a new product in the marketplace, we can rank it by simply observing its characteristics, without waiting to see consumer demand for the product. Furthermore, we can dynamically change the rankings in response to changes in the products. For example, if we observe a price change or if we observe a hotel closing its pool for renovations, we can immediately adjust the surplus values and reestimate the rankings.

Such research can provide us with critical insights into how people make choices when exposed to multiple ranked lists of options online. Furthermore, by examining product search through the “economic lens” of utilities, we leverage and integrate theories of relevance from information retrieval and microeconomic theory. Our interdisciplinary approach has the potential to improve the quality of results any product search engine displays and to improve the quality of choices available online to consumers.

To better understand the antecedents of consumers’ decisions, future work can look not only at transaction data but also into consumers’ browsing history and learning behavior. For example, our current model assumes consumers engage in optimal utility-maximizing behavior. However, they do not always, as some consumers are more thorough than others in their searches. By leveraging browsing histories, we can build models that explicitly take into consideration the fact that some users are utility optimizers and others simply engage in satisficing behavior. The difference in the conversion rate of users when presented with surplus-based rankings would also be an interesting avenue for study.

Our work has several limitations, some of which can serve as fruitful areas for future research. To better understand the antecedents of consumers’ decisions, future work could look not only at transaction data but also into consumers’ browsing history and learning behavior. Furthermore, by incorporating more individual-level demographics and context information from the time of purchase, one could extend our techniques to infer expected utility gains at a more personalized level. This step would potentially improve the evaluation process by comparing our recommendations with the results from the traditional collaborative filtering or content-based algorithms. Our model has a limited structure with regard to competition, preventing us from studying the impact of entry/exit decisions of hotels in different regions.

Future work can relax this constraint. In our model, the travel category-level shock is independently and identically distributed across consumers and travel categories. However, correlations could be present in the travel category shocks, wherein a consumer combines multiple purposes in one trip occasion. Although our model does not capture this possibility, it is a promising area for future work. Our analysis assumes that each hotel is exogenously endowed with a capacity of rooms, and this could bias results in favor of larger hotels. However, solving the revenue management problem fully is beyond the scope of this work, and hence we leave it for future researchers to address it, including issues such as hotel room availability (Bruno and Vilcassim 2008). Our AMT-based user studies presented a series of default rankings to users without the ability to sort the offers. Future work can enhance the nature and scope of such user studies by letting users choose their own sorting algorithm in a more interactive platform and execute randomized experiments. Future work can also examine the associations between product reviews on UGC sites and travel search engine ranking decisions using keyword-level analyses, as described in Dhar and Ghose (2010). Notwithstanding all these limitations, we believe our paper can pave the way for more research in this increasingly important domain.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mktsci.journal.informs.org/>.

Acknowledgments

Author names are in alphabetical order. The authors thank Susan Athey, Peter Fader, Francois Moreau, Aviv Nevo, Minjae Song, Daniel Spulber, Catherine Tucker, and Hal Varian for helpful comments that have significantly improved the paper. They also thank participants at NBER IT Economics and Productivity Workshop, 2010 Workshop on Digital Business Models, 2010 Marketing Science Conference, 2010 Searle Research Symposium at Northwestern University, Customer Insights Conference at Yale University, 2010 SCECR Conference, 2009 Workshop on Information Technology and Systems (WITS), 2009 Workshop on Economics and Information Systems, and seminar participants at the Columbia Business School, Harvard, George Mason, Georgia Tech, Harvard Business School, University of Maryland MIT Sloan School, and University of Minnesota for helpful comments. A. Ghose and P. Ipeirotis acknowledge the financial support from National Science Foundation [CAREER Awards IIS-0643847 and IIS-0643846], respectively. Support was also provided through a MSI-Wharton Interactive Media Grant (WIMI) and a Microsoft Virtual Earth Award. The authors thank Travelocity for providing the data and Uthaman Palaniappan for research assistance.

Appendix A.

Table A.1 Robustness Test (I) Using Alternative Sample Splits

Variable	Coefficients (Std. error)		
	IV	V	VI
Means			
<i>PRICE</i> (log)	−0.141*** (0.004)	−0.138*** (0.004)	−0.145*** (0.004)
<i>CHARACTERS</i> (log)	0.010*** (0.001)	0.011*** (0.003)	0.011*** (0.002)
<i>COMPLEXITY</i>	−0.011*** (0.002)	−0.012*** (0.002)	−0.014*** (0.002)
<i>SYLLABLES</i> (log)	−0.043*** (0.004)	−0.044*** (0.006)	−0.045*** (0.006)
<i>SMOG</i>	0.081** (0.027)	0.077** (0.028)	0.080** (0.027)
<i>SPELLERR</i> (log)	−0.127*** (0.004)	−0.126*** (0.007)	−0.132*** (0.006)
<i>SUB</i>	−0.142*** (0.012)	−0.152*** (0.009)	−0.152*** (0.008)
<i>SUBDEV</i>	−0.424*** (0.015)	−0.428*** (0.017)	−0.425*** (0.016)
<i>ID</i>	0.053* (0.028)	0.051 (0.037)	0.058† (0.034)
<i>CLASS</i>	0.034*** (0.010)	0.041*** (0.011)	0.042*** (0.012)
<i>CRIME</i> (log)	−0.024* (0.013)	−0.026* (0.014)	−0.021* (0.011)
<i>AMENITYCNT</i> (log)	0.007*** (0.001)	0.006** (0.002)	0.006* (0.003)
<i>EXTAMENITY</i> (log)	0.008*** (0.001)	0.009*** (0.002)	0.007*** (0.002)
<i>BEACH</i>	0.160*** (0.007)	0.159*** (0.007)	0.166*** (0.005)
<i>LAKE</i>	−0.112*** (0.035)	−0.114*** (0.032)	−0.108*** (0.028)
<i>TRANS</i>	0.159*** (0.009)	0.162*** (0.009)	0.169*** (0.005)
<i>HIGHWAY</i>	0.070** (0.031)	0.075** (0.030)	0.078** (0.030)
<i>DOWNTOWN</i>	0.044*** (0.004)	0.049*** (0.005)	0.043*** (0.005)
<i>TA_RATING</i>	0.042* (0.025)	0.044 (0.033)	0.042* (0.024)
<i>TL_RATING</i>	0.035** (0.014)	0.034* (0.019)	0.036** (0.017)
<i>TA_REVIEWCNT</i> (log)	0.187*** (0.043)	0.188*** (0.043)	0.189*** (0.045)
<i>TA_REVIEWCNT</i> ² (log)	−0.056*** (0.007)	−0.056*** (0.008)	−0.057*** (0.007)
<i>TL_REVIEWCNT</i> (log)	0.014*** (0.002)	0.014*** (0.003)	0.014*** (0.003)
<i>TL_REVIEWCNT</i> ² (log)	−0.021*** (0.005)	−0.024*** (0.005)	−0.027*** (0.006)
Constant	0.037*** (0.006)	0.040*** (0.008)	0.046** (0.012)
Brand control	Yes	Yes	Yes
Interaction effect (α_v) and standard deviations (α_v)			
<i>PRICE</i> (log) \times <i>INCOME</i> (log)	0.021*** (0.002)	0.022*** (0.002)	0.018*** (0.001)
<i>PRICE</i> (log)	0.010 (0.088)	0.008 (0.096)	0.011 (0.104)
Standard deviations (β_v)			
<i>CLASS</i>	0.023*** (0.003)	0.029*** (0.005)	0.030** (0.009)
<i>CRIME</i> (log)	0.011 (0.012)	0.013 (0.016)	0.015 (0.014)
<i>AMENITYCNT</i> (log)	0.017 (0.022)	0.021 (0.025)	0.024 (0.025)
<i>EXTAMENITY</i> (log)	0.004 (0.021)	0.006 (0.018)	0.006 (0.015)
<i>BEACH</i>	0.050*** (0.012)	0.056** (0.021)	0.066*** (0.020)
<i>LAKE</i>	0.104* (0.058)	0.113* (0.065)	0.105* (0.055)
<i>TRANS</i>	0.126* (0.067)	0.134** (0.057)	0.132** (0.060)
<i>HIGHWAY</i>	0.051* (0.027)	0.064 (0.049)	0.067 (0.045)
<i>DOWNTOWN</i>	0.031*** (0.003)	0.026** (0.011)	0.028* (0.015)
GMM obj. value	8.561e−4	8.890e−4	9.006e−4

Notes. IV is based on a filtered data set (≥ 1 review from TripAdvisor). V is based on a filtered data set (≥ 1 review from Travelocity). VI is based on a filtered data set (at least one review from both TripAdvisor and Travelocity).

*** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$; † $p \leq 0.1$.

Table A.2 Robustness Test (II) Using the BLP Model

Variable	Coefficients (Std. error)						
	I	II	III	IV	V	VI	R
<i>PRICE</i> (log)	-0.257*** (0.033)	-0.254** (0.034)	-0.251*** (0.031)	-0.255*** (0.040)	-0.263*** (0.045)	-0.266*** (0.046)	-0.256*** (0.034)
<i>CHARACTERS</i> (log)	0.036** (0.012)	0.038*** (0.011)	0.040*** (0.012)	0.031** (0.014)	0.044** (0.015)	0.041** (0.016)	0.037** (0.014)
<i>COMPLEXITY</i>	-0.009 (0.011)	-0.009 (0.011)	-0.009 (0.011)	-0.010 (0.010)	-0.014 (0.011)	-0.014 (0.011)	-0.014 (0.012)
<i>SYLLABLES</i> (log)	-0.143** (0.051)	-0.142** (0.051)	-0.142** (0.053)	-0.145** (0.062)	-0.147** (0.061)	-0.149** (0.064)	-0.143** (0.059)
<i>SMOG</i>	0.159* (0.080)	0.156* (0.079)	0.157* (0.080)	0.159* (0.082)	0.150* (0.085)	0.142† (0.101)	0.157* (0.083)
<i>SPELLERR</i> (log)	-0.088** (0.035)	-0.091** (0.036)	-0.091** (0.037)	-0.104** (0.040)	-0.106** (0.037)	-0.105** (0.040)	-0.098** (0.037)
<i>SUB</i>	-0.251* (0.177)	-0.255* (0.178)	-0.257* (0.178)	-0.253* (0.161)	-0.262* (0.163)	-0.266* (0.171)	-0.249* (0.118)
<i>SUBDEV</i>	-0.577** (0.218)	-0.576** (0.216)	-0.580** (0.218)	-0.589** (0.215)	-0.572** (0.210)	-0.574** (0.211)	-0.576** (0.214)
<i>ID</i>	0.143*** (0.033)	0.145*** (0.034)	0.146*** (0.039)	0.151*** (0.041)	0.155*** (0.043)	0.153*** (0.043)	0.152*** (0.041)
<i>CLASS</i>	0.057*** (0.002)	0.058*** (0.002)	0.056*** (0.002)	0.059*** (0.005)	0.066*** (0.008)	0.071*** (0.011)	0.059*** (0.011)
<i>CRIME</i> (log)	-0.086* (0.049)	-0.088* (0.050)	-0.086* (0.052)	-0.092* (0.054)	-0.099* (0.055)	-0.096* (0.055)	-0.091* (0.053)
<i>AMENITYCNT</i> (log)	0.029* (0.016)	0.032* (0.014)	0.034** (0.014)	0.036** (0.012)	0.038** (0.015)	0.032** (0.015)	0.035** (0.013)
<i>EXTREMITY</i> (log)	0.035** (0.015)	0.034** (0.012)	0.035** (0.014)	0.041*** (0.012)	0.052** (0.014)	0.049*** (0.011)	0.047*** (0.014)
<i>BEACH</i>	0.366*** (0.103)	0.368*** (0.099)	0.373*** (0.098)	0.378*** (0.100)	0.396*** (0.105)	0.391*** (0.101)	0.377*** (0.101)
<i>LAKE</i>	-0.174** (0.071)	-0.166** (0.072)	-0.172** (0.071)	-0.169** (0.076)	-0.171** (0.077)	-0.180** (0.079)	-0.171** (0.080)
<i>TRANS</i>	0.385*** (0.111)	0.382*** (0.111)	0.389*** (0.119)	0.400*** (0.121)	0.408*** (0.120)	0.410*** (0.121)	0.395*** (0.112)
<i>HIGHWAY</i>	0.404** (0.184)	0.404** (0.183)	0.404** (0.192)	0.407** (0.187)	0.416** (0.189)	0.420** (0.191)	0.401** (0.182)
<i>DOWNTOWN</i>	0.190* (0.106)	0.192* (0.102)	0.193* (0.102)	0.201* (0.100)	0.202* (0.099)	0.201* (0.098)	0.194* (0.099)
<i>TA_RATING</i>	0.103** (0.037)	0.105** (0.037)	0.110** (0.035)	0.112** (0.036)	0.115** (0.041)	0.117** (0.039)	0.109** (0.031)
<i>TL_RATING</i>	0.110* (0.059)	0.109* (0.058)	0.108* (0.058)	0.106* (0.060)	0.107* (0.058)	0.108* (0.059)	0.108* (0.054)
<i>TA_REVIEWCNT</i> (log)	0.325* (0.171)	0.329* (0.173)	0.326* (0.170)	0.330* (0.172)	0.351* (0.184)	0.352* (0.182)	0.342* (0.177)
<i>TA_REVIEWCNT²</i> (log)	-0.063*** (0.011)	-0.061*** (0.012)	-0.062*** (0.012)	-0.069*** (0.011)	-0.073*** (0.012)	-0.070*** (0.012)	-0.065*** (0.012)
<i>TL_REVIEWCNT</i> (log)	0.045** (0.021)	0.048** (0.021)	0.045** (0.021)	0.054** (0.023)	0.062** (0.027)	0.062** (0.026)	0.050** (0.021)
<i>TL_REVIEWCNT²</i> (log)	-0.037* (0.019)	-0.037* (0.020)	-0.036* (0.019)	-0.041** (0.018)	-0.043** (0.019)	-0.041** (0.017)	-0.039** (0.015)
Constant	0.058 (0.197)	0.052 (0.194)	0.060 (0.189)	0.067 (0.201)	0.071 (0.206)	0.075 (0.209)	0.063 (0.200)
Brand control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Travel category with homogeneous coef. ^a	Yes	Yes	Yes	Yes	Yes	Yes	No
Travel category with random coef. ^b	No	No	No	No	No	No	Yes
Interaction effect (α_v) and standard deviations (α_v)							
<i>PRICE</i> (log) × <i>INCOME</i> (log)	0.045*** (0.003)	0.047*** (0.003)	0.048*** (0.003)	0.041*** (0.003)	0.051*** (0.003)	0.050*** (0.003)	0.043*** (0.001)
<i>PRICE</i> (log)	0.016 (0.067)	0.012 (0.063)	0.015 (0.065)	0.011 (0.068)	0.020 (0.076)	0.019 (0.071)	0.014 (0.066)

Table A.2 (Cont'd.)

Variable	I	II	III	IV	V	VI	R
			Deviations (β_i)				
CLASS	0.033**	0.034**	0.034**	0.032**	0.038**	0.039**	0.035***
CRIME (log)	0.014	0.015	0.013	0.010	0.009	0.011	0.013
AMENITYCNT (log)	0.015	0.013	0.016	0.016	0.015	0.017	0.014
EXTREMITY (log)	0.006	0.006	0.005	0.005	0.005	0.006	0.006
BEACH	0.094***	0.093***	0.095***	0.097***	0.101***	0.106***	0.095***
LAKE	0.142**	0.143**	0.149**	0.150**	0.165**	0.166**	0.146**
TRANS	0.168**	0.170**	0.167**	0.162**	0.173**	0.177**	0.173**
HIGHWAY	0.077*	0.081*	0.075†	0.074†	0.068	0.063	0.077*
DOWNTOWN	0.041*	0.042*	0.045*	0.049*	0.051**	0.058**	0.049**
GMM obj. value	7.011e-4	7.617e-4	7.201e-4	6.978e-4	7.961e-4	8.262e-4	7.223e-4

Notes. I is based on the main data set (at least one review from either TripAdvisor or Travelocity). II is based on the main data set with review count ≥ 5 . III is based on the main data set with review count ≥ 10 . IV is based on the filtered data set (at least one review from TripAdvisor). V is based on the filtered data set (at least one review from Travelocity). VI is based on the filtered data set (at least one review from both TripAdvisor and Travelocity). R is based on data set I, with random coefficients on travel category dummies.

^aWe consider dummy variables with homogeneous coefficients to control for the eight corresponding travel categories.

^bWe consider dummy variables with random coefficients to control for the eight corresponding travel categories.

***Significant at a 0.1% level; **significant at a 1% level; *significant at a 5% level; †significant at a 10% level.

Appendix B. In-Sample Basic Model Comparison Results

In-sample and out-of-sample results are estimated based on a 10-fold cross validation. The size of the estimation sample for both the in-sample and out-of-sample estimations is 5,669. The size of the holdout sample for out-of-sample estimation is 2,430.

Table B.1 In-Sample Basic Model Validation Results

	Hybrid model	BLP without random coef. on travel categories	BLP with random coef. on travel categories	PCM	Nested logit (random utility maximization)
RMSE	0.0407	0.0518	0.0485	0.0976	0.1158
MSE	0.0016	0.0027	0.0024	0.0095	0.0134
MAD	0.0133	0.0185	0.0167	0.0318	0.0379

Table B.2 In-Sample Extended Model Validation Results

	Hybrid model with interaction effects	BLP with interaction effects
RMSE	0.0347	0.0426
MSE	0.0012	0.0018
MAD	0.0100	0.0161

Table B.3 In-Sample Model Validation Results by Excluding Certain Features

(Hybrid model)	Without UGC variables	Without location variables	Without service variables
RMSE	0.0743	0.1159	0.1112
MSE	0.0055	0.0134	0.0124
MAD	0.0328	0.0360	0.0353

Table B.4 In-Sample Model Validation Results by Excluding Certain UGC Features

(Hybrid model)	Without all text features	Without readability	Without subjectivity	Without numeric rating	Without reviewer identity
RMSE	0.0678	0.0642	0.0539	0.0513	0.0435
MSE	0.0046	0.0041	0.0029	0.0026	0.0019
MAD	0.0309	0.0289	0.0201	0.0217	0.0156

Table B.5 Out-of-Sample Basic Model Validation Results

	Hybrid model	BLP without random coef. on travel categories	BLP with random coef. on travel categories	PCM	Nested logit (random utility maximization)
RMSE	0.0881	0.1011	0.0975	0.1909	0.2399
MSE	0.0078	0.0102	0.0095	0.0364	0.0576
MAD	0.0276	0.0362	0.0387	0.0524	0.1311

Table B.6 Out-of-Sample Extended Model Validation Results

	Hybrid model with interaction effects	BLP with interaction effects
RMSE	0.0865	0.0922
MSE	0.0075	0.0085
MAD	0.0253	0.0287

Table B.7 Out-of-Sample Model Validation Results by Excluding Certain Features

(Hybrid model)	Without UGC variables	Without location variables	Without service variables
RMSE	0.1380	0.1992	0.1897
MSE	0.0190	0.0397	0.0360
MAD	0.0965	0.1276	0.1155

Table B.8 Out-of-Sample Model Validation Results by Excluding Certain UGC Features

(Hybrid model)	Without all text features	Without readability	Without subjectivity	Without numeric rating	Without reviewer identity
RMSE	0.1359	0.1252	0.1176	0.1116	0.0964
MSE	0.0185	0.0157	0.0138	0.0125	0.0093
MAD	0.0812	0.0618	0.0607	0.0583	0.0303

More Details on the Design of the User Study

To further control for biases, we conducted the user study with stringent controls in the design and execution of the online survey. This led to a set of 7,800 user comparisons. Our additional design takes into account the following issues:

- First, to prevent each AMT worker from seeing the same ranking multiple times, we restricted each worker to only participate in at most one ranking comparison for each city.

- Second, to make sure each AMT worker is exposed to the full decision-making environment as a “real” visitor, in addition to the hotel name, address, price, and class information, we provided the URLs for each hotel’s main Web pages from five different major (travel) search engine websites: TripAdvisor.com, Expedia.com, Hotels.com, Travelocity.com, and Google.com. Moreover, we were able to track whether or not a particular AMT worker clicked on any of these URLs for a particular hotel in a particular ranking comparison task.

- Third, we were able to track the exact time each AMT worker spent on a task (i.e., from the moment an AMT worker accepted a task until the moment that worker submitted the result).

Appendix C.**Figure C.1** Screenshot for a Sample Task from the User Study

Below is a pair of hotel ranking lists. Assume you are traveling to the city below and looking for a hotel, please **compare the two ranking lists and choose the better one** (left or right) that you think is more helpful. For detailed information about each hotel, the corresponding URLs are provided from the major travel websites and search engines (e.g., TripAdvisor, Google, Expedia, Hotels.com, Travelocity, etc). You can **click the URLs** to learn more about the hotel information, such as amenities, location, photos, customer reviews, etc.

NOTE: Before you make your choice, please examine carefully the hotels in each ranking list. We appreciate your collaboration.

1) San Francisco, CA									
Left					Right				
	Hotel	Address	Class	Price		Hotel	Address	Class	Price
1	The Maxwell Hotel	386 Geary St. San Francisco, CA 94102	3	\$109	1	Prescott Hotel	545 Post St. San Francisco, CA 94102	4	\$154
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
2	JW Marriott San Francisco	500 Post St. San Francisco, CA 94102	5	\$138	2	Hotel Griffon	155 Stewart St. San Francisco, CA 94105	3	\$240
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
3	Hotel Carlton San Francisco	1075 Sutter St. San Francisco, CA 94109	3	\$83	3	Courtyard San Francisco Downtown	299 2nd Street, San Francisco, CA 94105	3	\$129
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
4	Le Meridien San Francisco	333 Battery St. San Francisco, CA 94111	4	\$124	4	York Hotel	940 Sutter Street, San Francisco, CA 94109	3	\$74
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
5	Chancellor Hotel on Union Square	433 Powell Street, San Francisco, CA 94102	3	\$114	5	Club Quarters San Francisco	424 Clay Street, San Francisco, CA 94111	0	\$141
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
6	Executive Hotel Vintage Court	650 Bush Street, San Francisco, CA 94108	3	\$77	6	Hotel Carlton San Francisco	1075 Sutter St. San Francisco, CA 94109	3	\$83
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
7	Palomar	12 Fourth Street, San Francisco, CA 94103	5	\$196	7	Chancellor Hotel on Union Square	433 Powell Street, San Francisco, CA 94102	3	\$114
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
8	Broadway Manor Inn	2201 Van Ness Ave. San Francisco, CA 94109	2	\$70	8	Harbor Court Hotel	165 Stewart Street, San Francisco, CA 94105	3	\$422
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
9	The Ritz-Carlton	600 Stockton St. San Francisco, CA 94108	5	\$245	9	The Inn at Union Square	440 Post Street, San Francisco, CA 94102	3	\$182
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			
10	The Orchard Hotel	665 Bush Street, San Francisco, CA 94108	4	\$115	10	The Fairmont San Francisco	950 Mason Street, San Francisco, CA 94108	3	\$143
	(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com					(1) TripAdvisor.com (2) Google.com (3) Expedia.com (4) Hotels.com (5) Travelocity.com			

Your Choice is: ☐ Left ☐ Right

• Finally, to control for the quality of the responses, we allowed only those AMT workers with a prior approval rate higher than 95% to participate in the survey. AMT provides an approval rate for each worker based on the frequency with which tasks have been approved by the buyer. This approval rate can provide information on the quality of the workers.

Our finding suggested that, on average, each AMT worker spent 116.8 seconds (~2 minutes) per task. Besides, more than 50% of AMT workers clicked on the provided URLs to facilitate their decisions based on the detailed information of the hotels.

References

- Agichtein, E., E. Brill, S. Dumais, R. Ragno. 2006. Learning user interaction models for predicting Web search result preferences. *Proc. 29th Annual Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval*, ACM, New York, 3–10.
- Ansari, A., C. Mela. 2003. E-customization. *J. Marketing Res.* **40**(2) 131–145.
- Ansari, A., S. Essegai, R. Kohli. 2000. Internet recommendation systems. *J. Marketing Res.* **37**(3) 363–375.
- Archak, N., A. Ghose, P. G. Ipeirotis. 2011. Deriving the pricing power of product features by mining consumer reviews. *Management Sci.* **57**(8) 1485–1509.
- Athey, S., G. Imbens. 2007. Discrete choice models with multiple unobserved choice characteristics. *Internat. Econom. Rev.* **48**(4) 1159–1192.
- Berry, S., A. Pakes. 2007. The pure characteristics demand model. *Internat. Econom. Rev.* **48**(4) 1193–1225.
- Berry, S., J. Levinsohn, A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–890.
- Berry, S., O. B. Linton, A. Pakes. 2004. Limit theorems for estimating the parameters of differentiated product demand systems. *Rev. Econom. Stud.* **71**(3) 613–654.
- Bodapati, A. 2008. Recommender systems with purchase data. *J. Marketing Res.* **45**(1) 77–93.
- Bruno, H. A., N. J. Vilcassim. 2008. Structural demand estimation with varying product availability. *Marketing Sci.* **27**(6) 1126–1131.
- Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.
- Chintagunta, P., J.-P. Dubé, K. Y. Goh. 2005. Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Sci.* **51**(5) 832–849.
- Chung, T. S., R. T. Rust, M. Wedel. 2009. My mobile music: An adaptive personalization system for digital audio players. *Marketing Sci.* **28**(1) 52–68.
- Das, S., M. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the Web. *Management Sci.* **53**(9) 1375–1388.
- Decker, R., M. Trusov. 2010. Estimating aggregate consumer preferences from online product reviews. *Internat. J. Res. Marketing* **27**(4) 293–307.
- Dellarocas, C., N. Awad, M. Zhang. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing* **21**(4) 23–45.
- Dhar, V., E. A. Chang. 2009. Does chatter matter? The impact of user-generated content on music sales. *J. Interactive Marketing* **23**(4) 300–307.
- Dhar, V., A. Ghose. 2010. Sponsored search and market efficiency. *Inform. Systems Res.* **21**(4) 760–772.
- Duan, W., B. Gu, A. B. Whinston. 2008. Do online reviews matter? An empirical investigation of panel data. *Decision Support Systems* **45**(4) 1007–1016.
- Eliashberg, J., S. K. Hui, Z. J. Zhang. 2007. From story line to box office: A new approach for green-lighting movie scripts. *Management Sci.* **53**(6) 881–893.
- Fader, P. S., B. G. S. Hardie. 1996. Modeling consumer choice among SKUs. *J. Marketing Res.* **33**(4) 442–452.
- Fellbaum, C. 1998. *Wordnet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.
- Forman, C., A. Ghose, B. Wiesenfeld. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Systems Res.* **19**(3) 291–313.
- Ghose, A., P. G. Ipeirotis. 2011. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowledge Data Engrg.* **23**(10) 1498–1512.
- Ghose, A., P. G. Ipeirotis, A. Sundararajan. 2009. The dimensions of reputation in electronic markets. Working paper, New York University, New York.
- Godes, D., D. Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Marketing Sci.* **23**(4) 545–560.
- Hartmann, W., H. S. Nair, S. Narayanan. 2011. Identifying causal marketing mix effects using a regression discontinuity design. *Marketing Sci.* **30**(6) 1079–1097.
- Hausman, J. 1996. Valuation of new goods under perfect and imperfect competition. T. Bresnahan, R. Gordon, eds. *The Economics of New Goods, Studies in Income and Wealth*, Vol. 58. National Bureau of Economic Research, Chicago, 207–248.
- Hu, M., B. Liu. 2004. Mining and summarizing customer reviews. *Proc. 10th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining (KDD04)*, ACM, New York, 168–177.
- Hu, N., P. A. Pavlou, J. Zhang. 2006. Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. *Proc. Seventh ACM Conf. Electronic Commerce*, ACM, New York, 324–330.
- Lee, T., E. Bradlow. 2011. Automatic construction of conjoint attributes and levels from online customer reviews. *J. Marketing Res.* **48**(5) 881–894.
- Li, X., L. Hitt. 2008. Self-selection and information role of online product reviews. *Inform. Systems Res.* **19**(4) 456–474.
- Lipsman, A. 2007. Online consumer-generated reviews have big impact on offline purchases. Press release, comScore, Reston, VA. <http://www.marketingcharts.com/direct/online-consumer-generated-reviews-have-big-impact-on-offline-purchases-2577/kelsey-group-comscore-online-review-influence-on-purchase-by-service-type.jpg>.
- Liu, Y. 2006. Word-of-mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* **70**(3) 74–89.
- Luca, M. 2011. Reviews, reputation, and revenue: The case of Yelp.com. Working Paper 12-016, Harvard Business School, Boston.
- Manning, C., H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.
- McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *J. Econometrics* **142**(2) 698–714.
- Moe, W. 2009. How much does a good product rating help a bad product? Modeling the role of product quality in the relationship between online consumer ratings and sales. Working paper, University of Maryland, College Park.

- Nair, H., J.-P. Dubé, P. Chintagunta. 2005. Accounting for primary and secondary demand effects with aggregate data. *Marketing Sci.* **24**(3) 444–460.
- Netzer, O., R. Feldman, J. Goldenberg, M. Fresko. 2012. Mine your own business: Market-structure surveillance through text mining. *Marketing Sci.* **31**(3) 521–543.
- Nevo, A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2) 307–342.
- Pang, B., L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proc. 42nd Annual Meeting Assoc. Computat. Linguistics, Barcelona, Spain*, 271–278.
- Popescu, A., O. Etzioni. 2005. Extracting product features and opinions from reviews. *Proc. Human Language Tech. Conf. Empirical Methods Natural Language Processing (HLT/EMNLP 2005), Vancouver*, 339–346.
- Rust, R. T., T. S. Chung. 2006. Marketing models of service and relationships. *Marketing Sci.* **25**(6) 560–580.
- Scaffidi, C., K. Bierhoff, E. Chang, M. Felker, H. Ng, C. Jin. 2007. Red opal: Product-feature scoring from reviews. *Proc. 8th ACM Conf. Electronic Commerce (EC'07)*, ACM, New York, 182–191.
- Song, M. 2011. A hybrid discrete choice model of differentiated product demand with an application to personal computers. Simon School Working Paper FR 08-09, University of Rochester, Rochester, NY.
- Steckel, J., W. Vanhonor. 1993. Cross-validating regression models in marketing research. *Marketing Sci.* **12**(4) 415–427.
- Villas-Boas, J. M., R. S. Winer. 1999. Endogeneity in brand choice models. *Management Sci.* **45**(10) 1324–1338.
- Yamamoto, T. 2011. A multinomial response model for varying choice sets, with application to partially contested multiparty elections. Working paper, Princeton University, Princeton, NJ.
- Ying, Y., F. Feinberg, M. Wedel. 2006. Leveraging missing ratings to improve online recommendation systems. *J. Marketing Res.* **43**(3) 355–365.

Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowd-Sourced Content

(Online Appendices)

Anindya Ghose, Panagiotis G. Ipeirotis, Beibei Li

Stern School of Business, New York University
(aghose, panos, bli)@stern.nyu.edu

(Online) Appendix D Market Share Calculation

Our model is motivated by the model in Song (2011). A rational consumer with a marginal utility of income α_i chooses travel category k over other travel categories if and only if the best hotel (the one that provides the highest utility) within this travel category exceeds the best hotel within any other travel category:

$$\max_{j^k \in H_k} (\delta_{j^k_t} + X_{j^k_t} \beta_v v_i + \alpha_y y_i P_{j^k_t} + \alpha_v v_i P_{j^k_t}) + \varepsilon_{ikt} > \max_{j^r \in H_r} (\delta_{j^r_t} + X_{j^r_t} \beta_v v_i + \alpha_y y_i P_{j^r_t} + \alpha_v v_i P_{j^r_t}) + \varepsilon_{irt}, \quad \forall r \neq k.$$

Thus, similar to Song (2011), by assuming ε has a type I extreme value distribution, we can calculate the market share for a travel category type k as the probability of this category being chosen:

$$s_k = \iint \frac{\exp(\max_{j^k \in H_k} (\delta_{j^k_t} + X_{j^k_t} \beta_v v_i + \alpha_y y_i P_{j^k_t} + \alpha_v v_i P_{j^k_t}))}{1 + \sum_{r=1}^K \exp(\max_{j^r \in H_r} (\delta_{j^r_t} + X_{j^r_t} \beta_v v_i + \alpha_y y_i P_{j^r_t} + \alpha_v v_i P_{j^r_t}))} f(y) g(v) dy dv. \quad (D1)$$

As a robustness check, we tested different assumptions for ε (e.g., using a normal distribution), consistent with Chintagunta (2001). Our results showed high consistency with the previous estimates (i.e., based on the Type I extreme value distribution), similar to findings of Chintagunta (2001). The results are given in the last column of Table 3.

Furthermore, within travel category k consumer i chooses hotel j^k if and only if its utility exceeds the utility from any of the other hotels within the same travel category:

$$\delta_{j^k_t} + X_{j^k_t} \beta_v v_i + \alpha_y y_i P_{j^k_t} + \alpha_v v_i P_{j^k_t} > \delta_{h^k_t} + X_{h^k_t} \beta_v v_i + \alpha_y y_i P_{h^k_t} + \alpha_v v_i P_{h^k_t}, \quad \forall h^k \in H_k \text{ and } h^k \neq j^k,$$

where H_k represents the subset of hotels with travel category type k . This can be transformed to

$$(\delta_{j^k_t} - \delta_{h^k_t}) + (X_{j^k_t} - X_{h^k_t}) \beta_v v_i + (P_{j^k_t} - P_{h^k_t}) \alpha_v v_i > \alpha_y y_i (P_{h^k_t} - P_{j^k_t}).$$

Similar to Berry and Pakes (2007), we rank the hotels within each travel category in the order of ascending price. Therefore, conditioning on v_i , a consumer with income type y_i will choose hotel j^k if and only if

$$\begin{aligned} y_i &< \min_{j^k > h^k} \frac{(\delta_{j^k_t} - \delta_{h^k_t}) + (X_{j^k_t} - X_{h^k_t}) \beta_v v_i + (P_{j^k_t} - P_{h^k_t}) \alpha_v v_i}{\alpha_y (P_{h^k_t} - P_{j^k_t})} \equiv \bar{\Delta}(\bullet | \theta, v), \\ \text{and } y_i &> \max_{j^k < h^k} \frac{(\delta_{j^k_t} - \delta_{h^k_t}) + (X_{j^k_t} - X_{h^k_t}) \beta_v v_i + (P_{j^k_t} - P_{h^k_t}) \alpha_v v_i}{\alpha_y (P_{h^k_t} - P_{j^k_t})} \equiv \underline{\Delta}(\bullet | \theta, v). \end{aligned} \quad (D2)$$

Let $F_y(\bullet)$ denote the cdf of y_i , and $G(\bullet)$ denote the cdf of v_i . Similar to Song (2011), the market share of hotel j within travel category type k can be calculated as

$$s_{(j|category=k)} = \int [F_y(\bar{\Delta}_j^k(\bullet | \theta, v)) - F_y(\underline{\Delta}_j^k(\bullet | \theta, v))] 1[\bar{\Delta}_j^k(\bullet | \theta, v) > \underline{\Delta}_j^k(\bullet | \theta, v)] dG(v), \quad (D3)$$

Where $1[\bullet]$ is an indicator for the condition, and θ is a vector containing α_y, α_v and β_v . Note here, in order to compute the income upper bound $\bar{\Delta}(\bullet | \theta, v)$ and lower bound $\underline{\Delta}(\bullet | \theta, v)$, we need the value of θ . Given the set of values for θ , this integration is typically not analytically solvable. For this reason, we

use a Monte Carlo simulation to approximate it. Since v_i follows the multivariate normal distribution $v_i \sim N(0, I_{Z+1})$, we can obtain an unbiased estimator of this integral by taking ns_v random draws of v_i :

$$s_{(j|category=k)}(\delta, p, X; \theta, F_y, G_{ns}) \equiv \frac{1}{ns_v} \sum_i^{ns} [F_y(\bar{\Delta}_{j^k}(\bullet, \theta, v_i)) - F_y(\underline{\Delta}_{j^k}(\bullet, \theta, v_i))] 1[\bar{\Delta}_{j^k}(\bullet, \theta, v_i) > \underline{\Delta}_{j^k}(\bullet, \theta, v_i)] \quad (D4)$$

We can further derive the market share, which is the probability that a hotel j within category type k is chosen by consumer type (y_i, v_i) , to be the following

$$s_{j^k} = \iint_{y_i, v_i \in C_{j^k}} \frac{\exp(\delta_{j^k_t} + X_{j^k_t} \beta_v v_i + \alpha_y y_i P_{j^k_t} + \alpha_v v_i P_{j^k_t})}{1 + \sum_{r=1}^K \exp(\max_{j^r \in H_r} (\delta_{j^r_t} + X_{j^r_t} \beta_v v_i + \alpha_y y_i P_{j^r_t} + \alpha_v v_i P_{j^r_t}))} f(y) g(v) dy dv, \quad (D5)$$

where $y_i, v_i \in C_{j^k}$ indicates consumers who choose hotel j in travel category k . Note that there is no max function in the numerator. As shown in Song (2011), this market share function can be rewritten as the product of the equation (D3) and the probability that travel category k is chosen by those consumers who choose hotel j of travel category k . That is

$$s_{j^k} = \int_{v_i \in C_{j^k}} \left([F_y(\bar{\Delta}_{j^k}(\bullet, \theta, v_i)) - F_y(\underline{\Delta}_{j^k}(\bullet, \theta, v_i))] \times \int_{y_i \in C_{j^k}} \frac{\exp(\delta_{j^k_t} + X_{j^k_t} \beta_v v_i + \alpha_y y_i P_{j^k_t} + \alpha_v v_i P_{j^k_t})}{1 + \sum_{r=1}^K \exp(\max_{j^r \in H_r} (\delta_{j^r_t} + X_{j^r_t} \beta_v v_i + \alpha_y y_i P_{j^r_t} + \alpha_v v_i P_{j^r_t}))} h(y) dy \right) g(v) dv, \quad (D6)$$

where

$$h(y) = \frac{f(y)}{F_y(\bar{\Delta}_{j^k}(\bullet, \theta, v_i)) - F_y(\underline{\Delta}_{j^k}(\bullet, \theta, v_i))}.$$

Again, these integrals are not analytically solvable. Hence, we use a Monte Carlo simulation-based approach to approximate their values based on the distributions $G(v)$ and $H(y)$:

$$s_{j^k} = \frac{1}{ns_v} \sum_{v_i \in C_{j^k}} \left([F_y(\bar{\Delta}_{j^k}(\bullet, \theta, v_i)) - F_y(\underline{\Delta}_{j^k}(\bullet, \theta, v_i))] \frac{1}{ns_y} \sum_{y_i \in C_{j^k}} \frac{\exp(\delta_{j^k_t} + X_{j^k_t} \beta_v v_i + \alpha_y y_i P_{j^k_t} + \alpha_v v_i P_{j^k_t})}{1 + \sum_{r=1}^K \exp(\max_{j^r \in H_r} (\delta_{j^r_t} + X_{j^r_t} \beta_v v_i + \alpha_y y_i P_{j^r_t} + \alpha_v v_i P_{j^r_t}))} \right), \quad (D7)$$

where ns is the number of simulated consumers whose $\alpha \in [F_y(\underline{\Delta}_{j^k}(\bullet | \theta, v)), F_y(\bar{\Delta}_{j^k}(\bullet | \theta, v))]$. By restricting the taste shock at a travel category level, this hybrid model combines the choice probabilities of the PCM and the BLP as described in (Song 2011).

(Online) Appendix E Estimation Algorithm for the Random Coefficient Demand Model

The estimation involves two nested loops. In the outer loop, the parameters corresponding to the individual heterogeneity distribution are heuristically learned, whereas the inner loop involves computing the unknown parameters embedded in the mean utility. More specifically, we ran the estimation algorithm in the following seven steps.

1. Generate 500 random draws per market for v_i and y_i , from standard normal and income distribution respectively.
2. Initialize starting values $\theta_0 = (\alpha_y^0, \alpha_v^0, \beta_v^0)$ and δ^0 .
3. Compute market share within a travel category. This corresponds to the conditional probability calculated by equation (D3), which numerical approximation is shown in equation (D4).

3.1 Sort hotels within each travel category in the order of ascending price;

3.2 For each hotel j^k within travel category k in market t , calculate the corresponding Δ value

$$\Delta = \frac{(\delta_{j^k t} - \delta_{h^k t}) + (X_{j^k t} - X_{h^k t})\beta_v v_i + (P_{j^k t} - P_{h^k t})\alpha_v v_i}{\alpha_y (P_{h^k t} - P_{j^k t})},$$

where h^k represents all other hotels in the same market t .

3.3 Now, for each travel category k in market t , consider all hotels h^k ranked before hotel j^k (which means those hotels with lower prices than j^k), compute the upper bound

$$\bar{\Delta}(\bullet | \theta, v) \equiv \min_{j^k > h^k} \frac{(\delta_{j^k t} - \delta_{h^k t}) + (X_{j^k t} - X_{h^k t})\beta_v v_i + (P_{j^k t} - P_{h^k t})\alpha_v v_i}{\alpha_y (P_{h^k t} - P_{j^k t})}.$$

3.4 Similarly, consider all hotels h^k ranked after hotel j^k (which means those hotels with higher prices than j^k), compute the lower bound

$$\underline{\Delta}(\bullet | \theta, v) \equiv \max_{j^k < h^k} \frac{(\delta_{j^k t} - \delta_{h^k t}) + (X_{j^k t} - X_{h^k t})\beta_v v_i + (P_{j^k t} - P_{h^k t})\alpha_v v_i}{\alpha_y (P_{h^k t} - P_{j^k t})}.$$

3.5 If the upper bound is strictly higher than the lower bound, then the market share within travel category k is positive

$$S_{(j|category=k)} = \int [F_y(\bar{\Delta}_{j^k}(\bullet | \theta, v)) - F_y(\underline{\Delta}_{j^k}(\bullet | \theta, v))] dG(v).$$

Compute by Monte Carlo simulation, with v_i from the previous random draws, $ns_v = 500$,

$$S_{(j|category=k)} \equiv \frac{1}{ns_v} \sum_i^{ns} [F_y(\bar{\Delta}_{j^k}(\bullet, \theta, v_i)) - F_y(\underline{\Delta}_{j^k}(\bullet, \theta, v_i))].$$

3.6 Otherwise, the market share within travel category k is zero.

4. Compute the overall market share function s_{j^k} based on equation (D6), which numerical approximation is shown in equation (D7). We achieve this by using the Monte Carlo simulation.

$$s_{j^k} = \frac{1}{ns_v} \sum_{v_i \in C_{j^k}} \left([F_y(\bar{\Delta}_{j^k}(\bullet, \theta, v_i)) - F_y(\underline{\Delta}_{j^k}(\bullet, \theta, v_i))] \frac{1}{ns_y} \sum_{y_i \in C_{j^k}} \frac{\exp(\delta_{j^k_t} + X_{j^k_t} \beta_v v_i + \alpha_y y_i P_{j^k_t} + \alpha_v v_i P_{j^k_t})}{1 + \sum_{r=1}^K \exp(\max_{j^r \in H_r} (\delta_{j^r_t} + X_{j^r_t} \beta_v v_i + \alpha_y y_i P_{j^r_t} + \alpha_v v_i P_{j^r_t}))} \right),$$

where v_i and y_i are from the previous random draws, with $ns_v = ns_y = 500$.

5. The inner loop computation takes place. Keeping the nonlinear parameters fixed at the initial guesses, iterate over the values of the mean utility δ to minimize the distance between the predicted market share and the observed market share. This requires to solve the system of nonlinear equations, $s(\delta)$, where δ is a n -dimension vector of unknown variables ($n = \sum_{k=1}^K J^k$). This can be done by using Newton-Raphson Method.

5.1 Compute the Jacobian matrix $J(\delta)$ for $s(\delta)$:

$$J(\delta) = \begin{bmatrix} \frac{\partial s_1}{\partial \delta_1} & \dots & \frac{\partial s_1}{\partial \delta_n} \\ \dots & \dots & \dots \\ \frac{\partial s_n}{\partial \delta_1} & \dots & \frac{\partial s_n}{\partial \delta_n} \end{bmatrix}.$$

5.2 Given a starting value of δ^0 , solve the nonlinear system by iteration:

$$J(\delta^m)(\delta^{m+1} - \delta^m) = -s(\delta^m), \quad m = m+1, \text{ until } \|\delta^{m+1} - \delta^m\| < \varepsilon.$$

5.3 Given the solved δ , extract the unobserved characteristic ξ

$$\xi = \delta - X\bar{\beta} - \bar{\alpha}P.$$

6. Form a GMM objective function by interacting the unobserved characteristic, ξ , with the instrumental variable IV :

$$GMMobj = E[\xi' \cdot IV].$$

7. The outer loop computation takes place. Use Nelder-Mead Simplex algorithm to update the parameter values for $\theta_1 = (\alpha_y^1, \alpha_v^1, \beta_v^1)$. Assign $\theta_1 = (\alpha_y^1, \alpha_v^1, \beta_v^1)$ and δ (which was computed in step 5) as the new starting value and iterate from step 3, until the algorithm finds the optimal combination of α_y , α_v , β_v and δ , which minimizes the GMM objective function.¹

¹ Dube, Fox and Su (2009) note that a theoretical advantage of Newton-type methods, is that they are quadratically convergent when the iterates are close to a local solution. To make sure our estimates are reliable, we employed 50 starting points in each run of the estimation. We routinely found that our algorithm were able to identify the same local minimum each time. Moreover, as suggested by Knittel and Metaxoglou (2008), we also tried several alternative optimization algorithms, including (i) direct-search algorithms: e.g., the Nelder-Mead simplex method; (ii) derivative-based algorithms: e.g., the Fletcher-Reeves conjugate gradient method and the vector Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (which is a quasi-Newton method). We found that different algorithms were able to recover consistent structural parameters in our data.

(Online) Appendix F More Details on UGC Mining

(F-1) Extraction of Location Characteristics using Social Geotagging and Image Classification

As mentioned in Subsection 3.2, to allow for the automated tagging of areas that lack tags from the human tagging process, we use automatic image classification techniques of satellite images to tag location features that can influence hotel demand. Consider, for example, the case where we are trying to automatically identify whether a hotel is located “Near a beach,” or “Near downtown.” Towards this, we extracted hybrid satellite images (sized 256×256 pixels) using the Bing Maps Tile System (<http://msdn.microsoft.com/en-us/library/bb259689.aspx>), for each of the (thousands) of hotel venues located in the US, with four different zoom levels for each. These $4 \times 1,497$ images were used to extract information about the surroundings of the hotel, through image classification and human inspection using AMT. For better understanding, below are two examples of the images used in our analysis.



Beach



Downtown

To automatically tag satellite images, we first needed to train our classification model. To build a “training set,” we used information from two sources: (i) locations tagged by users on a social tagging site such as Geonames.org or (ii) locations annotated by users on AMT. We built the image classifiers as follows: First, we randomly selected a set of 121 hotels and requested five AMT users to label each example according to its corresponding satellite images from four different zoom levels. The labelers answered whether there is a beach in the image, or whether the image is that of a downtown area. We applied a simple majority voting method to make the final decision from the multi-labels of the example. Second, we trained a Support Vector Machines (SVM) classifier on this dataset and used the trained SVM classifier to classify the images that corresponded to the remaining hotels. Prior work has shown that non-parametric classifiers, such as Neural Networks, Decision Trees, and Support Vector Machines (SVM) provide better results than parametric classifiers in complex landscapes (Lu and Weng 2007). Therefore, we tested various non-parametric classification techniques. These include (i) Decision Trees, which are widely used for training and classification of remotely sensed image data (due to their ability to generate human interpretable decision rules and its speed in training and classification), and (ii) Support Vector Machines (SVM), that are highly accurate and perform well for a wide variety of classification tasks (Fukuda and Hirosawa 2001).

We conducted a small study to examine the performance of the classifier out-of-sample data. We classified the out of sample images using AMT; our results illustrated that our SVM classifier had an accuracy of 91.2% for the “beach” image classification and 80.7% for the “downtown” image classification. We also used the C4.5 algorithm for the classification, and noticed an accuracy increase for “Near a beach” and a decrease for “Near downtown.” The main reason for this is that “beach” images often contain a “sand strip,” together with an “ocean margin” well distributed in density. This typically provides more stable and distinct textural information for the “beach” images, thus making them easier to distinguish.

Moreover, as a robustness check, we extracted the two location characteristics “Near a beach” and “Near downtown” using geo-tagging approach via the site Geonames.org. In particular, the geo-tagging process and the image classification process were conducted independently. We found that the overlapping rate between the two sets of results (i.e., acquired independently from the geo-tagging approach and from the image classification approach) is very high: For “Near a beach,” this overlapping rate is 92.3%, and for “Near the downtown,” this overlapping rate is 85.1%. This additional test provides us with high confidence on our image tagging results.

(F-2) Extraction of Textual Quality of Customer Reviews

In regards to the extraction of textual quality of customer reviews, as discussed in Subsection 3.3, we looked into two text style features “subjectivity” and “readability” as the evaluating criteria. To capture the review textual style comprehensively, we used a multiple-item method for subjectivity and readability. We included two sub-features for subjectivity and five sub-features for readability, each of which measures the review text style.

We observed that there are two types of reviews, from the stylistic point of view. There are reviews that list “objective” information, listing the characteristics of the hotel, and giving an alternate description that confirms (or rejects) the description given by the hotel. The other types of reviews are those with “subjective,” sentimental information, in which the reviewers give a very personal description of the hotel, and give information that, typically, does not appear in the official description of the hotel.

We distinguished the extent of “subjective assessments” in the reviews by deriving a review-level numerical score for the degree of subjectivity. More specifically, we used the methods from Ghose and Ipeirotis (2011) who build on the methods in Pang and Lee (2004). In particular, objective information is considered the information that also appears in the hotel-provided description, and subjective is everything else. To infer the probability of review subjectivity, we trained a classifier by using as “objective” documents the hotel-generated descriptions from the websites of Travelocity and TripAdvisor. We then randomly retrieved 1000 reviews to construct the “subjective” examples of the training set. We conducted the training process by using a 4-gram Dynamic Language Model classifier provided by the lingpipe toolkit. “Lingpipe” is a tool kit provided online for processing text using computational linguistics (More information can be found at <http://alias-i.com/lingpipe/>). After constructing the classifiers, we used the resulting classification models in the remaining, unseen reviews. Instead of classifying each review as subjective or objective, we classified each sentence in each review as either “objective” or “subjective,” keeping the probability of being subjective for each sentence. By doing so, we were able to acquire a subjectivity confidence score for each sentence in a review, hence deriving the mean and standard deviation of this score as the subjectivity measurements for that review. These numerical scores are able to distinguish how likely a review contains subjective assessments as opposed to objective descriptions.

We also look into the impact of “Readability,” which is a proxy for the difficulty faced by people when reading online reviews. Past research has shown that easy-reading text improves comprehension, retention, and reading speed, and that the average reading level of the US adult population is at the eighth grade level (White 2003). Specifically, for each hotel, we collected all existing reviews to examine the average number of characters per review, average number of syllables per review, average number of spelling errors per review, and the average length of the sentence as a “Complexity” measurement (total number of characters divided by the total number of sentences). To avoid idiosyncratic errors peculiar to a specific readability metric, we computed a set of metrics for each review. Specifically, we computed the following: Automated Readability Index, Coleman-Liau Index, Flesch Reading Ease, Flesch-Kincaid Grade

Level, Gunning and SMOG. For brevity, we only show results with SMOG Index in the paper although all the other readability measures yield similar results.

Furthermore, previous studies have shown that the social identity information of reviewers in an online community shapes community members' judgment of the products. In other words, the prevalence of reviewer disclosure of identity information is associated with changes in product sales (Forman et al. 2008). Therefore, consistent with prior work, we include the characteristic that captures the level of reviewers' disclosure of their identity information – “real name or location.” More specifically, this binary characteristic describes whether or not a reviewer had revealed her real name or location information on the reviewer profile page of Travelocity and TripAdvisor.

In sum, our analysis identifies 5 broad types of characteristics in this category: (i) total number of reviews, (ii) overall review rating, (iii) review subjectivity (mean and variance), (iv) review readability (the number of characters, syllables, and spelling errors, complexity and SMOG Index), and (v) the disclosure identity information by the reviewer.

(F-3) Text Feature Extraction and Sentimental Analysis

Towards extracting the additional text features discussed in Subsection 4.3, we build on the work of Hu and Liu (2004), Popescu and Etzioni (2005), Archak et al. (2011). More specifically, we conduct the text mining process in the following three steps:

(1) Text Feature Extraction.

First, we extract the important hotel features. Following the automated approach introduced previously (e.g., Archak et al. 2011), we use a POS (part-of-speech) tagger to identify the frequently mentioned nouns and noun phrases, which we consider candidate hotel features. We then use WordNet (Fellbaum 1998) and a context-sensitive hierarchical agglomerative clustering algorithm (Manning and Schutze 1999) to further cluster the identified nouns and noun phrases into clusters of similar nouns and noun phrases. The resulting set of clusters corresponds to the set of identified product features mentioned in the reviews.

For our analysis, we kept the top-5 most frequently mentioned features, which were hotel staff, food quality, bathroom, parking facilities, and bedroom quality. To select the top 5 features, we first processed all the reviews for each hotel, and extracted text features (i.e., terms) that appeared frequently in the reviews for each hotel. For example, for Hotel A the features extracted based on the reviews for Hotel A can be “bed”, “bathroom” and “pool”; for Hotel B the features can be “bed”, “bathroom” and “restaurant”. Then, we selected the top 5 most frequently extracted features across all hotels. In our example, the features will be “bed” and “bathroom”. The top 5 features that we selected in our study covered 80% of the hotels, which means that for 80% of the hotels the extracted text features contain these 5 features. While technically possible, we did not consider more textual features because the frequency in which the additional features are mentions drops significantly, and therefore we would not be able to have a robust measurement for these textually-inferred features for a very significant fraction of the hotels in our dataset.

Besides, as suggested in Archak et al. (2011), in addition to the fully automated tool we can also use a semi-automated crowd sourcing approach via Amazon Mechanical Turk, by asking AMT workers to manually process each review and extract evaluation phrases for any given product feature.

(2) Sentimental Analysis.

For sentimental analysis, we extracted all the evaluation phrases (adjectives and adverbs) that are being used to evaluate the individual service features (for example, for the feature “hotel staff” we extracted phrases like “helpful,” “smiling,” “rude,” “responsive,” etc). The process of extracting user evaluation phrases can be automated. To measure the meaning of these evaluation phrases, we used AMT to exogenously assign explicit polarity semantics to each word. To compute the scores, we used AMT to create our ontology, with the scores for each evaluation phrase. Our process for creating these “external” scores was done using the methodology of Archak et al. (2011).

We asked nine AMT workers to look at the pair of the evaluation phrase together with the product feature, and assign a grade from -3 (strongly negative) to +3 (strongly positive) to the evaluation. This

resulted in a set of nine, independently submitted evaluation scores; we dropped the highest and lowest evaluation score, and used the average of the remaining seven evaluations as the externally imposed score for the corresponding evaluation-product phrase pair. As an example, when evaluating “hotel staff”, the AMT process resulted in “helpful” having value of 0.9, “rude” to be -0.5, “responsive” to be 0.5, and so on. We should stress that the scoring of the evaluation phrases is only necessary to be done once as the set of hotel features, and the corresponding semantic evaluation phrases are highly unlikely to change over time.

(3) Negation Handling.

Finally, to handle the negation (e.g., “I didn’t think the staff was helpful”), we built a dictionary database to store all the negation words (e.g., not, hardly) using approach similar to NegEx (<http://code.google.com/p/negex/>). In the sentiment analysis process, if the algorithm finds a negation word in the reviews based on the dictionary, it will reverse the sign of the sentiment score of that sentence (e.g., from 3 to -3), indicating an opposite sentiment.

For better understanding, we provide below an example of the final results from our sentimental analysis for the text feature “*food quality*.” Note that there are totally 339 evaluation phrases extracted for this hotel. Due to space limitation, we only show the first 16.

Table F An Example of the Final Text Mining Results for Hotel X

Text Feature	Synonyms Extracted from the Reviews	Evaluation Phrases Extracted from the Reviews	Sentimental Score for Each Evaluation Phrase	Overall Score
Food Quality	breakfast, food, buffet,	good	1.0	0.1833
	complimentary, restaurant,	great	1.6	
	cook, burger, donut, cereal,	free	0.8	
	egg, bagel, fresh, fruit,	dry	-0.2	
	gravy, pancake, pastry,	horrible	-1.8	
	sausage, toast, menu, fish,	special	1.2	
	salmon, chicken, ham, cafe,	various	1.2	
	continental, dinner, lunch,	beautiful	1.4	
	meat, bacon, beverage, tea,	mean	-1.0	
	coffee, snack, appetizer,	minimal	-0.4	
	dessert, avocado, taste, tasty,	humble	-0.2	
	grill, salad, ice cream	erroneous	-1.0	
		feral	-1.4	
		garlic	0.0	
		colder	-0.4	
		favorable	1.2	
		

(Online) Appendix G

Comparison of Mechanical Turk Users with overall US Internet Population

	June 2008	October 2008	December 2008
	US Internet Users comscore Data	Mechanical Turk Users	Mechanical Turk Users
Total Audience	100	100	100
Persons - Age			
Persons: 15+	85.9	100	100
Persons: 18+	80.1	99.6	99.5
Persons: 21+	74.3	92.9	91.1
Persons: 35+	52.4	39.3	37.1
Persons: 50+	24.3	11.2	10.7
Persons: 55+	16.2	5.2	5.4
Persons: 2-11	9.5	0	0
Persons: 2-17	19.9	0.2	0.4
Persons: 6-11	7.4	0	0
Persons: 6-14	12	0	0
Persons: 9-14	8.9	0	0
Persons: 12-17	10.4	0.2	0.4
Persons: 12-24	22.9	19	21.5
Persons: 12-34	38	57.8	60
Persons: 12-49	66.2	87.4	88.2
Persons: 18-24	12.5	18.7	21.1
Persons: 18-34	27.6	57.5	59.7
Persons: 18-49	55.8	87.2	87.8
Persons: 21-34	21.9	53.3	53.9
Persons: 21-49	50	82.9	82
Persons: 25-34	15.1	38.8	38.6
Persons: 25-49	43.2	68.4	66.7
Persons: 25-54	51.3	75.2	72.3
Persons: 35-44	18.7	22.4	21.5
Persons: 35-49	28.2	29.7	28.1
Persons: 35-54	36.2	36.4	33.7
Persons: 35-64	46.8	41.4	38.8
Persons: 45-54	17.6	14	12.2
Persons: 45-64	28.1	19	17.4
Persons: 55-64	10.5	5	5.2
Persons: 65+	5.7	0.7	1.1
Males - Age			
All Males	49.5	28	36.6
Male: 15+	42.1	28	36.6
Male: 18+	39.1	27.8	36.3
Male: 21+	36.1	24.7	32.4
Male: 35+	25.7	9.5	11.3
Male: 50+	12	2.8	2.6
Male: 55+	8.1	1.4	1.1
Male: 2-11	4.9	0	0

Male: 2-17	10.4	0.1	0.2
Male: 6-11	3.9	0	0
Male: 6-14	6.3	0	0
Male: 9-14	4.5	0	0
Male: 12-17	5.5	0.1	0.2
Male: 12-24	11.6	7.5	9.1
Male: 12-34	18.9	17.3	24.2
Male: 12-49	32.5	25	33.9
Male: 18-24	6.1	7.4	8.9
Male: 18-34	13.4	17.2	23.9
Male: 18-49	27.1	24.9	33.7
Male: 21-34	10.4	15.2	21.1
Male: 21-49	24.1	22.9	30.8
Males: 25-34	7.3	9.8	15
Male: 25-49	20.9	17.6	24.8
Male: 25-54	24.8	19	26.3
Males: 35-44	9.1	6	8
Male: 35-49	13.7	7.7	9.7
Male: 35-54	17.5	9.1	11.2
Male: 35-64	22.6	10.6	12.3
Male: 45-54	8.4	3.1	3.2
Male: 45-64	13.5	4.5	4.3
Males: 55-64	5.1	1.4	1.1
Males: 65+	3	0	0.1
Females - Age			
All Females	50.5	72	63.4
Female: 15+	43.8	72	63.4
Female: 18+	41	71.9	63.3
Female: 21+	38.2	68.2	58.7
Female: 35+	26.8	29.8	25.8
Female: 50+	12.3	8.3	8.1
Female: 55+	8.1	3.8	4.3
Female: 2-11	4.6	0	0
Female: 2-17	9.5	0.1	0.1
Female: 6-11	3.6	0	0
Female: 6-14	5.7	0	0
Female: 9-14	4.5	0	0
Female: 12-17	4.9	0.1	0.1
Female: 12-24	11.3	11.5	12.3
Female: 12-34	19.1	40.5	35.9
Female: 12-49	33.6	62.4	54.3
Female: 18-24	6.4	11.5	12.2
Female: 18-34	14.2	40.5	35.8
Female: 18-49	28.7	62.4	54.1
Female: 21-34	11.5	38.1	32.8
Female: 21-49	25.9	60	51.2
Females: 25-34	7.8	28.9	23.6
Female: 25-49	22.3	50.9	41.9
Female: 25-54	26.5	56.2	46
Females: 35-44	9.5	16.4	13.4

Female: 35-49	14.5	21.9	18.4
Female: 35-54	18.7	27.3	22.4
Female: 35-64	24.1	30.8	26.5
Female: 45-54	9.2	10.9	9
Female: 45-64	14.6	14.5	13.1
Females: 55-64	5.4	3.6	4.1
Females: 65+	2.6	0.7	1
HH Income (US)			
HHI USD: Less than 15,000	6	11.4	12.9
HHI US: Under \$25K	9.3	22.8	23.1
HHI US: Under \$60K	44.5	64.8	60.5
HHI US: \$60K+	55.5	34.8	39.1
HHI US: \$75K+	43	22.7	27.5
HHI USD: 15,000 - 24,999	3.4	11.4	10.1
HHI USD: 25,000 - 39,999	9.9	21.8	18.9
HHI USD: 40,000 - 59,999	25.3	20.2	18.6
HHI USD: 60,000 - 74,999	12.6	12.1	11.6
HHI USD: 75,000 - 99,999	17.7	10.2	11.5
HHI USD: 100,000 or more	25.3	12.5	16
<u>Region (US)</u>			
Central Region US:West North	7.6	5.8	7.5
Region US:Mountain	6.9	6.4	7.4
Region US:Pacific	15.4	13.3	15.7
Region US:New England	5.5	6.4	4.7
Region US:Mid Atlantic	14.2	13.9	15.8
Region US:South Atlantic	18.7	19.2	19.9
Central Region US:East South	5.1	8.3	5.2
Central Region US:West South	10.5	10.7	9
Central Region US:East North	16.1	15.7	14.8
Children			
Children:No	39.3	52.7	57.6
Children:Yes	60.7	47.3	42.3
HH Size			
HH Size: 1	4.4	17.7	17.3
HH Size: 2	24.2	28.9	30.6
HH Size: 3	21.4	19.7	19.2
HH Size: 4	25.3	20.5	21.9
HH Size: 5+	24.8	12.9	10.7
HH Size: 1-2	28.5	46.6	47.8
HH Size: 3+	71.5	33.5	32.7
Race			
Race:White	87.3	82.7	82
Race:Black	8	6.5	5.3
Race:Asian	1.6	5.7	6.8
Race:Other	3.1	4.9	5.8

Additional References

- Berry, S. T. 1994. Estimating discrete-choice models of product differentiation. *RAND J. Econom.* 25(2) 242–262.
- Chintagunta, P. K. 2001. Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Sci.* 20(4) 442–456.
- Dube, J.P., J. Fox, L. Su 2009. Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation, Working Paper, University of Chicago.
- Knittel, C. R., K. Metaxoglou. 2008. Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings. Working Paper, U.C. Davis.
- White, S. 2003. The 2003 national assessment of adult literacy (NAAL). Technical Report NCES 2003495rev, Center for Education Statistics (NCES), Institute of Education Sciences, U.S. Department of Education, Washington, DC.