GOIZUETA BUSINESS SCHOOL
EMORY

Zyman Institute of
Brand Science

ZIBS Technical Report

Modeling Consumer Choice via Aggregate Generalized Nested Logit: An Application to the Lodging Industry

Sriram Venkataraman, Vrinda Kadiyali

“Brand-Driven Business Performance”

Technical Report Series

The Zyman Institute of Brand Science (ZIBS) is an innovation research group based at Emory University's Goizueta Business School. ZIBS fosters rich collaborations between business managers and leading scholars worldwide. The institute supports advanced research in brand management with the ultimate goal of developing positively useful knowledge that drives superior business performance.

Relevancy in research is core to ZIBS. Research topics are elected by a steering committee consisting of executive managers and leading scholars. ZIBS supports research in next practices (vs. best practices) with resources and funding. ZIBS insures the application of this research through knowledge sharing activities like executive briefings, consultative interventions, roundtables, and conferences.

Stay Informed. Visit zibs.com to signup for our mailing list.

Report Access
Corporate Membership: ZIBS member companies receive all ZIBS reports free of charge. To learn more about corporate membership visit zibs.com.

Purchasing Single Reports: ZIBS makes some reports available for individual purchase. To purchase these reports please visit zibs.com

Report Archive: ZIBS reports for all years are accessible on zibs.com for member firms.

Member Services
Change of Address: Send both old and new address to services@zibs.com.

For Academics
Classroom Use: Upon written request working papers may be copied for one-time classroom use free of charge. Contact services@zibs.com to obtain permission.

Special Memberships: ZIBS can arrange special memberships for Academic Members and Market Research Members. Please visit

Working Paper Submissions: Faculty researchers in branding are encouraged to submit working papers to ZIBS, regardless of where the research funding originated. The research must represent an advancement over existing literature and not be widely disseminated elsewhere. Funding is available to interpret the working papers into a format that best conveys managerial insights.

Research Proposals: ZIBS supports researchers worldwide with research resources (funding, data, analysis, etc). We encourage researchers to submit proposals for consideration. Visit zibs.com/research for more information and to download the research proposal template.

Working Paper Status
Articles that appear in ZIBS Technical Research Reports have not undergone a formal academic peer review. They are released as working papers, which means they are a document released for discussion and represent cutting edge knowledge but are not considered an authoritative text until they pass peer review.

Intellectual Property

Modeling Consumer Choice via Aggregate Generalized Nested Logit: An Application to the Lodging Industry

Sriram Venkataraman*

Vrinda Kadiyali

---

* Sriram Venkataraman is at Emory University's Goizueta Business School and svenka2@emory.edu. Vrinda Kadiyali is at Cornell University's S.C. Johnson Graduate School of Management and kadiyali@cornell.edu. We thank Sachin Gupta, Vithala R. Rao, Doug Bowman and J. P. Dube for their extensive comments and suggestions on this paper. We also thank participants of INFORMS Pricing Conference at Cornell University, Marketing Science 2004, the Zyman Institute of Brand Science (ZIBS) at Emory University and seminar participants at Northwestern-Kellogg, NYU-Stern, UCLA-Anderson and UC-Irvine. The usual disclaimers apply.

Abstract

In this paper, using aggregate data, we demonstrate the ability of the generalized nested logit (Wen and Koppelman, 2000; GNL henceforth) to better capture consumer choice under conditions where consumer tradeoffs among choice items is not ex-ante obvious to the researcher, and data on attributes of consumer choice are incomplete. We extend existing GNL models by using more readily available aggregate data (rather than individual data) while accounting for consumer heterogeneity and endogeneity of firm-choice variables. The empirical application is to the lodging (hotel) industry. The industry classifies properties on the basis of price tiers; it also recognizes that consumers appear to have two idea points (of downtown and airport) for location. However, it appears possible that , a consumer might see a property in the same location but a different price tier as a closer substitute than a property of the same price tier at a different location. Hence for this industry, an ex-ante nesting structure based on price tiers or location alone might not capture the complexities of consumer choices. We find that GNL provides a better fit to these data than aggregate logit or aggregate probit. Our results provide managerially useful insights into who might comprise competition for any hotel, i.e. is it a nearby property of the same/different quality tier or is it a distant property of the same/different quality tier. We also briefly discuss the implications of consumer choices for firm profitability by estimating a supply-side model.

Key words: Consumer choice models, aggregate data, generalized nested logit.

1. Introduction

Understanding consumer choice is the cornerstone of marketing. A key step to quantifying consumer choice is choosing a demand model appropriate to the choice situation. In this paper, using aggregate data, we demonstrate the ability of the generalized nested logit (Wen and Koppelman, 2000; GNL henceforth) to better capture consumer choice under conditions where consumer tradeoffs among choice items is not ex-ante obvious to the researcher. We extend existing GNL models by using more readily available aggregate data (rather than individual data) while accounting for consumer heterogeneity and endogeneity of firm-choice variables. The empirical application is to the lodging (hotel) industry. We demonstrate that for this choice situation, the usual aggregate demand models used in marketing are not as informative.

Recent developments in economics and marketing facilitate the estimation of consumer preference distribution from aggregate data (Berry et al. 1995; Sudhir, 2001; Chintagunta, 2001). The demand specification in these studies is either aggregate logit and/or aggregate probit, i.e. models where individual-level consumer heterogeneity is modeled using aggregate data. In our model -- aggregate/mixed generalized nested logit (AGNL) -- we model consumer demand by simulating individual consumer-level choices of hotel properties. Next, the simulated choices are aggregated to generate the property-level market shares. A non-linear numerical optimization algorithm is used to minimize the error between the predicted market shares and observed market shares. Using this approach and accounting for consumer heterogeneity and econometric endogeneity of price and location, we compare our results with aggregate logit and aggregate probit. Since the estimated elasticities have implications for decisions like pricing, margins and location choice etc., understanding the extent of the elasticity differences is of interest to marketing scholars and practitioners alike.

Our choice of the GNL-based demand model is motivated by features of consumer choice in the lodging industry. The industry classifies hotels or properties based on price (quality) tiers. The industry also acknowledges two ideal points for hotel locations in any metropolis/city - airport and downtown. Consider some common choice structures and how they might estimate choices in this specific choice context. A logit demand model would classify all properties, regardless of price and location, as being part of the same choice set. A nested logit (Ben-Akiva, 1973) specification might propose that consumers choose on the basis of price, and then on location. Given that the nested logit requires properties to be present only once anywhere in the choice/decision tree, hotels might be classified in exclusive price-based or location-based nests.

But what if the tradeoffs in price and location are more complex rather than simple price-based alone or simple location-based alone or a simple combination of the two (as seen via a logit lens) or even simple sequential nested logit choices (first price, then location or vice-versa, as seen by a nested logit)? For example, a property located at a favorable location might be preferred to a property in the same price tier but at an unfavorable location. Similarly, a consumer might see a property in the same location but a different price tier as a closer substitute than a property of the same price tier at a different location. Hence for this industry, an ex-ante nesting structure based on price tiers or location alone might not capture the complexities of consumer choices.

Specifically, such complex tradeoffs could be because a property may have some observed and unobserved attributes similar to elements in multiple nests. These could include free shuttle service to the airport, onsite fitness center, shopping arcade, free continental breakfast, etc. Therefore it would be useful to have a demand model in which the unobserved utility for one alternative be correlated with that of other alternatives within the same nest and also correlated, though to a different degree, with the unobserved utility of alternatives in other nests. The GNL demand model is one such model where any choice alternative can be present in multiple nests. The overlapping alternatives are fractionally allocated across multiple nests, (with all fractions adding up to 1) and these allocations are estimated from the data.

In the GNL, a (logsum) parameter measures how similar items are within a nest. Unlike the nested logit, each nest has its own logsum parameter. The combination of different logsum parameters for each nest, and allocation parameters of hotels across nests, provide for different levels of correlations between alternatives within and across nests. This again can be achieved if, and only if, choice alternatives can be available in multiple nests. Therefore in modeling consumer choice of lodging properties, the GNL might provide both intuitive and data-driven advantages. This model is equally applicable to industries where nested choice alternatives overlap and the degree of overlap is ex-ante unknown to the researcher.

For completeness and based on industry classifications, we fit multiple nesting structures, namely a) price alone, b) location followed by price, c) hotel segment (quality tier) followed by location and d) location alone. We estimate our model on monthly data for the city of Austin, Texas, for the time period 1991-2003. Our empirical findings support a combination of location, and price-based choice with overlapping nesting structure; the best-fitting nest structure is one where properties are classified as belonging to concentric circles of five and ten miles around the airport and downtown locations.[1] We find support for the industry notion that there is a captive audience for both downtown and airport; where consumers are willing to pay \$1.435 to be one mile closer to the airport and \$3.783 to be one mile closer to downtown. GNL provides a better fit to these data than aggregate logit or aggregate probit. Our results provide managerially useful insights into who might comprise competition for any hotel i.e. whether a rival is a nearby property of the same/different quality tier or is it a distant property of the same/different quality tier. As we show via an example, competition comprises not necessarily other hotels in the same price class or the same location, but a more complex combination of the two and these effects vary by property.

To understand the implications of this consumer choice model for firm pricing decisions, we also estimate a supply-side model of firms choosing prices to maximize profits per period. We do not model the choice of locations but instead control for the econometric endogeneity of location choices. Drawing on the economic geography literature in economics, we model a property's marginal costs as a function of exogenous cost shifters and geographic proximity from the airport and downtown. We find that firm marginal costs decrease moving away from downtown, and towards the airport. Therefore, at downtown location, there is greater revenue because of consumer ideal point location but higher marginal costs; near the airport there is both a revenue cluster and lower marginal costs. Therefore, firms have to make tradeoffs on demand and marginal costs while making pricing decisions.

In the following sections, we describe our proposed GNL demand model in more detail (section 2), contrast it to existing studies and discuss the data (section 3) and the results (section 4). In section 5, we present the supply model and its results. We conclude by summarizing and offering suggestions for future research (section 6).

2. Consumer Choice Model

2.1. Consumer Utility Specification

1 Some properties belong to only one nest, i.e. if they are within five miles of either the airport or downtown. The others can belong partially to various nests. See section 4.2 for more details.

There is a long history of utility-based structural demand modeling in marketing (McFadden, 1986; Baltas and Doyle, 2001). While previous literature in marketing estimated these demand models using disaggregate data, recent advances in econometric analysis have facilitated the use of random effects in multinomial logit and nested logit models, to account for consumer heterogeneity using aggregate data (Sudhir, 2001). Unlike finite mixture models which provide discrete representations of consumer preferences (Kamakura and Russell, 1989), continuous mixture based aggregate logit/aggregate probit models recover the underlying distribution of consumer preferences from aggregate data (Chintagunta, 2001). Accounting for consumer heterogeneity in this fashion relaxes the IIA restriction (Sudhir, 2001) at the aggregate demand level[2].

The aggregate probit model (Chintagunta, 2001) provides the major advantage of complete flexibility in the variance-covariance structure of the error terms, and therefore can serve as a good alternative model to our GNL specification. However, given the unique features of hotel choices as discussed in the previous section, it is plausible that the aggregate logit and aggregate probit (Chintagunta, 2001) might not offer the required level of flexibility for our consumer choice context.

We now turn to our application of GNL to hotel demand. In our data we observe monthly sales of room nights at the property level. Our model begins with a consumer choosing a property from the available set of properties. The unit of demand is assumed to be a room-night. According to D.K. Shifflet & Associates, Ltd., of all business travelers staying in a hotel, 40 percent spend one night per year, 24 percent spend two nights per year, and 36 percent spend three or more nights per year. Of leisure travelers staying in a hotel, annually 47 percent spend one night per year, 26 percent spend two nights per year, and 27 percent spend three or more nights per year. The average number of business trips per year is 2.3, while the average number of vacation trips per year is 2.7 (Source: http://espn.go.com/mediakit/research/industry_landscapes/travel.html). Therefore an approximate number of overnight stays/trip is .852 [(.40+.24*2+.36*3)/2.3] for business travelers and .666 [(.47+.26*2+.27*3)/2.7] for leisure travelers, i.e. under one night in either case. This means that the number of multiple night trips is not very large. Hence our monthly unit level demand model might not be a very restrictive.[3]

When the individual demand is aggregated and averaged across all consumers within a month, we obtain the monthly shares of room nights for each property. Thus, at the aggregate level the unit of observation is property level monthly market shares, where any property in a city can compete with any other property in the same city, and not just those in the same geographic location (e.g. airport or downtown) or quality tier. In contrast, recent papers in economics that examine the role of geography either limit the consumer choice set of retailers ex-ante based on the retailers' proximity from highway exits (Davis, 2002 study on highway

---

2 Note, in the case of aggregate logit, the demand model still suffers from IIA at the individual level. This is because at the individual level the aggregate logit is the traditional MNL (Chintagunta, 2001). The aggregate nested logit or aggregate probit, however, do not suffer from IIA even without accounting for consumer heterogeneity.

3 A very small set of consumers could still potentially buy multiple room nights on a travel occasion. In our aggregate data we are unable to account for this behavior. Integrating both quantity and choice models may be a promising direction for future research in this area (see Hendel 1999). Since the main objective of this study is to contrast the fits and estimated elasticities across aggregate demand model specifications, the unit demand misspecification would be a common bias across all demand specifications and appears unlikely to favor GNL over other demand models. This limitation does not, therefore, detract us from our stated goal of comparing GNL fit to other demand model fits.

motels) or divide up the market into grids and assume that the transportation cost for a consumer is the same for all retail outlets within the same cell within the grid (Seim, 2001 study on the video rental market). As we mentioned earlier, if geographic markets are specified ex-ante, e.g. all hotels within two miles of downtown/airport, then patterns of substitution are in effect very restrictive and may result in inaccurate representations of consumer choices.

In our study, on each travel occasion to a given market, a consumer selects a property from the available set of properties based on brand name, price per room night, and its geographic location; all properties in a city are included in the choice set. Literature on the hospitality industry (e.g. Wall et al., 1985) indicates that travelers, who are the primary consumers of lodging services, prefer hotels that are located at one of two locations - central business district (CBD) and airport (AIR)[4]. We treat these two locations as ideal points of consumers. It is possible that consumers have other ideal points. For example, another ideal point can be a tourist attraction, or a convention center etc. Our qualitative findings do not change when we added the convention center, for our focal metropolitan market. This is most likely to be because the convention center is very close to CBD in our market. Hence we do not present results with this additional ideal point. Given likely differences across cities, for any additional ideal point specification (if these were to exist), one needs to generate additional distance measures from the respective ideal points and include these measures in the demand model. That is, our two ideal point operationalization is not a limitation imposed by our demand model, but a model chosen for the city of choice.

Drawing on the Wall et al. (1985) study, we assume that all lodging properties are part of the consumer choice set and use the distance of each property from the two ideal points as its location attributes. Consumers of hotel lodging predominantly do not belong to the local area and therefore local area demographics cannot be used to capture consumer heterogeneity. Therefore, differences in consumer transportation costs to the same property are captured via heterogeneous location preference parameters.

Specifically, the utility for consumer *i* from choosing property *h,* at time *t* where $h \in \{1, ...H+1\}$ is given by

$$U_{iht} = \gamma_{ih} - \alpha_i p_{ht} + g_i(d_{hCBDt}, d_{hAIRt}) + \xi_{ht} + \varepsilon_{iht} \tag{1}$$

where γ_{ih} is consumer *i*'s intrinsic preference for property *h*, α_i is i's marginal utility of price, $g_i(d_{hCBDt}, d_{hAIRt})$ is i's transportation cost and ξ_{ht} reflects factors that affect the consumer's utility from *h* at time *t* but are unobserved by the researcher, and ε_{iht} is a mean zero stochastic term. ξ's include numerous hotel characteristics like external appearance; residence services such as gym, laundry, shuttle service etc.; temporary construction or blockade of highway exits that lead to the property, resulting in reduced access, and hence utility to the consumer; and so forth. Note that these property-specific factors in each market are unobservable to us in the data but observable to both the consumer and the firm. Therefore price *p* might be correlated with the unobserved factors that affect demand, causing econometric endogeneity of price. Choice of property location might also be affected by ξ; for example, proximity to highway exits and recreational facilities. Thus like price, correlation between unobserved factors and geographic location causes endogeneity in location choice. In this respect, we depart from

4 In several studies (Hauser and Shugan 1983; Choi, DeSarbo and Harker, 1990), the consumers' ideal point/points are endogenously determined in their proposed model. In our study they are fixed and exogenous based on industry surveys mentioned previously.

papers that assume location to be econometrically exogenous (e.g., Kamita, 2001; Davis, 2002 and Thomadsen, 2001).

The transportation cost function for any consumer i is $g_i(d_{hCBDt}, d_{hAIRt})$ and depends on two variables - distance of property h from airport and from CBD, denoted as d_{hAIRt} and d_{hCBDt} respectively. We do not impose a linear city structure; i.e. specifying the distance of any property from any one ideal point does not uniquely describe the location of the property. Following Huff (1966) and Anderson and De Palma (1992), we specify a linear quadratic transportation cost function as below

$$g_i(d_{hCBDt}, d_{hAIRt}) = -\phi_{1i} * d_{h,AIRt} + \phi_{2i} * d_{h,AIRt}^2 - \phi_{3i} * d_{h,CBDt} + \phi_{4i} * d_{h,CBDt}^2 \quad (2)$$

We postulate that the disutility from choosing a property increases at a diminishing rate in distance from the consumer's ideal point. This concavity in consumer disutility is modeled as a quadratic transportation cost function, and ensures an internal solution (Anderson and dePalma, 1992). Combining equations (1) and (2) we get

$$U_{iht} = \gamma_{ih} - \alpha_i p_{ht} + -\phi_{1i} * d_{h,AIRt} + \phi_{2i} * d_{h,AIRt}^2 - \phi_{3i} * d_{h,CBDt} + \phi_{4i} * d_{h,CBDt}^2 + \xi_{ht} + \varepsilon_{iht} \quad (3)$$

This utility specification is an extension of single ideal point parameterizations, to accommodate multiple ideal points. In our case, airport and CBD are the ideal points.

To complete the demand specification, we formulate the utility from the outside good as follows

$$U_{iot} = \varepsilon_{iot}$$

(4)

Equation (3) implicitly assumes that a consumer i will select property h over all other properties if

$$U_{iht} > U_{ikt} \;\; \forall k \in \{1,...H\};\; k \neq h \quad (5)$$

If θ_i is the vector of individual response parameters $\theta_i = [\gamma_{ih}, \alpha_i, \phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i}]$, and if A_{ht} represents the set of θ_i's that result in choice of h at time t, then assuming ties occur with zero probability, the market share of property h in market t can be expressed as

$$s_{ht} = \int_{A_{ht}} dP(\varepsilon) \quad (6)$$

where $P(\varepsilon)$ denotes population distribution function of the stochastic demand shock ε.

Next we discuss the homogenous GNL model followed by the heterogeneous GNL model. Note, GNL is consistent with a utility maximizing decision theoretic approach under some conditions as described in the next section.

2.2. GNL (Wen and Koppelman, 2001)

The GNL is a GEV model derived from the p.d.f. for

$$F(\varepsilon_1, \varepsilon_2, \varepsilon_3, ...\varepsilon_J) = \sum_n \left(\sum_{j \in N_n} (a_{jn} e^{V_j})^{\frac{1}{\mu_n}} \right)^{\mu_n} \quad (7)$$

where n denotes nest and N_n is a set of alternatives in nest n. For the purpose of illustration, let us assume consumer homogeneity. If ε in equation 6 were distributed with GNL p.d.f., the expressions for aggregate shares is given by

$$s_{ht} = \sum_n s_{ht|n} * s_{nt} = \sum_n \left[\frac{(a_{hn} e^{V_{ht}})^{1/\mu_n}}{\sum_{k \in H_n} (a_{kn} e^{V_{kt}})^{1/\mu_n}} \right] \left[\frac{\left(\sum_{k \in H_n} (a_{kn} e^{V_{kt}})^{1/\mu_n} \right)^{\mu_n}}{1 + \sum_n \left(\sum_{k \in H_n} (a_{kn} e^{V_{kt}})^{1/\mu_n} \right)^{\mu_n}} \right] \tag{8}$$

where H_n is the set of all alternatives included in nest *n*, a_{hn} is the allocation parameter which characterizes the portion of alternative *h* assigned to nest *n.* If allocation parameter is zero or econometrically insignificant for any alternative in any nest, it implies that the alternative does not belong in that nest. A value of one indicates that this alternative is not fractionally allocated but rather only belongs in a single nest.

The share in nest *n* of alternative *h* in market and share of nest *n* at time *t* are given by the equations 9 and 10 respectively.

$$s_{ht|n} = \frac{(a_{hn} e^{V_{ht}})^{1/\mu_n}}{\sum_{k \in H_n} (a_{kn} e^{V_{kt}})^{1/\mu_n}} \tag{9}$$

$$s_{nt} = \frac{\left(\sum_{k \in H_n} (a_{kn} e^{V_{kt}})^{1/\mu_n} \right)^{\mu_n}}{1 + \sum_n \left(\sum_{k \in H_n} (a_{kn} e^{V_{kt}})^{1/\mu_n} \right)^{\mu_n}} \tag{10}$$

To simplify interpretation of the allocation parameters, we restrict these such that a_{hn} satisfy the condition that $\sum_n a_{hn} = 1 \; \forall \mathrm{h}$. The allocation parameters do not have any explicit structural interpretation, but with the above restriction, can be interpreted as the population-level likelihood or probability of an alternative in a nest. μ_n is the logsum parameter that captures the dissimilarity between alternatives in nest *n*. The GNL model is consistent with random utility maximization if the condition $0 \le \mu_n \le 1$ (for all n) is satisfied. A high value for this parameter means alternatives in a nest are dissimilar, and a low value means alternatives in the nest are similar.

The own and cross price effect of share of alternative *h* with respect to price of alternative *k* are given by

$$-\frac{\sum_n s_{nt} s_{ht|n} \left[(1 - s_{nt}) + \left(\frac{1}{\mu_n} - 1 \right) \left(1 - s_{ht|n} \right) \right]}{s_{nt}} \alpha s_{ht}$$

$$\left(s_{ht} + \frac{\sum_n \left(\frac{1}{\mu_n} - 1 \right) s_{nt} s_{ht|n} s_{kt|n}}{s_{kt}} \right) \alpha s_{ht} \qquad \text{where } k \in \{1....H\}; \;\; k \notin h \tag{11}$$

Note that the elasticity expressions above are functions of nest logsum parameters and allocation parameters (via shares). When the logsums are all set to be equal to 1 and all alternatives belong to one common nest, the above equation simplifies to the cross-elasticity expression for multinomial logit. Under the restriction that alternatives belong to only one-- but not one common-- nest and the allocation parameters sum to 1, the above expression simplifies to the cross-elasticity expression for nested logit. Hence the restricted GNL nests within it both the multinomial logit and nested logit demand models.

Like the nested logit, GNL requires a priori specification of number of nests and nest members, but unlike the nested logit allows for overlapping nest membership. The optimization (estimation) procedure will recover the allocation parameters (if any) of the overlapping members in the respective nests. One can test several nesting structures, each of which are informed by the researcher's conjectured consumer choice structure. This flexibility of the demand model comes at a cost. As a result of the non-linear nature of the demand model, the Hessian of the log-likelihood function need not be negative semi-definite in all regions of the gradient search optimization procedure. Therefore one needs to start the gradient search using different starting values to ensure that the search hasn't stopped at a local optimum.

2.3. Aggregate GNL (With Heterogeneity)

In practice consumers might differ in their preference for price, location and latent preference for each choice alternative. The extant literature in this area employs both continuous parametric distribution (Sudhir, 2001) and discrete support for the random coefficients distribution (Besanko et al., 2004).We use a continuous support random coefficients demand model, and draw the consumer-specific taste parameters from an empirical distribution (Nevo, 2001). Recently, there have been concerns about identifying segments using aggregate data (Bodapati and Gupta, 2004), and therefore we choose the continuous parameterization.

We integrate the individual-level choice probabilities (in the homogenous case this is equivalent to market share expressions) in equation 8, over its continuous support. However, unlike Nevo (2001), consumers in our industry are transient to the local market, and therefore we cannot use the local demographics' distribution to model consumer heterogeneity. Instead, like Besanko et al. (2004), we rely on draws from a parametric distribution. Since we do not observe individual-level choices, we simulate individual choice probabilities by drawing the individual taste heterogeneity parameters as follows

$$\begin{bmatrix} \gamma_{ih} \\ \alpha_{ih} \\ \phi_{i1} \\ \phi_{i2} \\ \phi_{i3} \\ \phi_{i4} \end{bmatrix} = \begin{bmatrix} \gamma_{h} \\ \alpha \\ \phi_{1} \\ \phi_{2} \\ \phi_{3} \\ \phi_{4} \end{bmatrix} + \begin{bmatrix} \varsigma_{\gamma_{ih}} \\ \varsigma_{\alpha_i} \\ \varsigma_{\phi_{i1}} \\ \varsigma_{\phi_{i2}} \\ \varsigma_{\phi_{i3}} \\ \varsigma_{\phi_{i4}} \end{bmatrix} \tag{12}$$

and log-sum heterogeneity parameters

$$\mu_{in} \sim N(\overline{\mu_n}, \sigma_{\mu_n}) \tag{13}$$

where $\gamma, \alpha, \phi_1 - \phi_4$ are the population-level mean preference parameters and $\bar{\mu}_n$ is the vector of population-level mean log-sum parameters.

Respecifying the utility function as a function of household varying and invariant terms, we have

$$U_{iht} = V_{ht} + V_{iht} + \varepsilon_{iht} \tag{14}$$

where the household invariant term is given by

$$\overline{V}_{ht} = \gamma_h - \alpha\, p_{ht} - \phi_1 d_{h,CBD,t} + \phi_2 d^2_{h,CBD,t} - \phi_3 d_{h,AIR,t} + \phi_4 d^2_{h,AIR,t} + \xi_{ht} \tag{15}$$

and the household varying term is given by

$$V_{iht} = \varsigma_{\gamma_{ih}} - \varsigma_{\alpha_i} p_{ht} - \varsigma_{\phi_{i1}} d_{h,CBD,t} + \varsigma_{\phi_{i2}} d^2_{h,CBD,t} - \varsigma_{\phi_{i3}} d_{h,AIR,t} + \varsigma_{\phi_{i4}} d^2_{h,AIR,t} \tag{16}$$

Thus the individual choice probability expressions are given below:

$$P_{iht} = \sum_n P_{iht|nt} P_{\text{int}} = \sum_n \left[\frac{\left(a_{hn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}}{\sum_{j\in n}\left(a_{jn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}} * \frac{\left(\sum_{j\in n}\left(a_{jn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}\right)^{\overline{\mu_n}+\varsigma_{\mu_{in}}}}{\sum_n\left(\sum_{j\in n}\left(a_{jn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}\right)^{\overline{\mu_n}+\varsigma_{\mu_{in}}}} \right] \tag{17}$$

Summing over and taking the average choice probabilities across individuals, we obtain the simulated market share expression for alternative *h* in market *t* as given below

$$MS_{ht} = \frac{1}{NS_i}\sum_i\left(\sum_n P_{iht|nt}P_{\text{int}}\right) = \frac{1}{NS_i}\sum_i\left(\sum_n \left[\frac{\left(a_{hn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}}{\sum_{j\in n}\left(a_{jn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}} * \frac{\left(\sum_{j\in n}\left(a_{jn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}\right)^{\overline{\mu_n}+\varsigma_{\mu_{in}}}}{\sum_n\left(\sum_{j\in n}\left(a_{jn} e^{\overline{V}_{ht}+V_{iht}}\right)^{\frac{1}{\overline{\mu_n}+\varsigma_{\mu_{in}}}}\right)^{\overline{\mu_n}+\varsigma_{\mu_{in}}}} \right]\right) \tag{18}$$

where NS_i is the number of simulated individuals for market *t*.

To summarize, equation 18 represents the system of equations for consumer choice estimation. That is, we will estimate AGNL-based market share model (equation 18) using aggregate data. We simulate individual-level choices using the GNL demand specification, and these individual-level choices are aggregated to generate predictions for market share. We account for individual heterogeneity in two separate ways in equations (12 and 13). First, we estimate: preference heterogeneity- via the parameters $\gamma, \alpha, \phi_1 - \phi_4$. This captures heterogeneity in consumer preference for attributes. Second, we account for covariance heterogeneity among consumers in their evaluation of the similarity/dissimilarity among alternatives in a nest. That is, we allow for error heterogeneity (Swait and Bernardino, 2000; Kamakura et al., 1996) in each of the *n* logsum parameters μ_n. Error heterogeneity enables more flexible patterns in competition between alternatives. Our specification allows for two sources of differential correlation between alternatives belonging to different nests – allocation parameters and consumer varying log-sum parameter.[5]

---

5 Consumer choice/decision rule heterogeneity is another possible source of consumer heterogeneity (see Swait and Bernardino (2000)). Another form of heterogeneity is structural heterogeneity, which captures consumer heterogeneity in nesting structures. (Thanks to a reviewer for bringing this point to our attention). A model that accounts for this form of heterogeneity would be analogous to the Kannan and Wright (1991) study, wherein the estimation yields nesting-structure choice probabilities and demand-model parameters for each nesting structure. Data limitations prevent us from estimating a demand model that accommodates this form of heterogeneity as well. Since neither segmentation nor recovery of the underlying consumer choice process is the focus of this study, we believe this limitation of our model is not severe.

3. Data

The model is applied to the Austin lodging industry. The Texas State Comptroller requires every lodging property to report taxable and non-taxable revenues on a quarterly basis. Based on this, our data source (which we cannot identify as part of our contract with them) aggregates and augments this (public) information to obtain data from 1991 through 2003. In the lodging literature, hotel brands are categorized into sectors (Full-Service, Limited-Service or Extended Stay) and segments (Deluxe, Luxury, Upscale, and Midscale with Food and Beverage, Midscale without Food and Beverage, Economy, Budget, Upper-tier Extended Stay and Lower-tier Extended Stay). In order to get a broader understanding of consumer substitution patterns, in our estimation sample we include properties from all three service sectors. We limit our empirical analysis to properties that exist for the entire span of our data[6]. Market-level demand is modeled as the sum of the demand served by these properties and the outside good.

We chose Austin, TX, for the following reasons: a) manageable number of properties to estimate property-level intercepts, b) representativeness of brands and quality segments and c) appropriateness of the two ideal point model specification as explained in the estimation section. The final estimation sample is made up of 23 properties, spanning three service sectors and fifteen brands. We have 166 monthly cross-sections which results in 3818 (23*166) observations. As is seen in Table 2, the focal properties account for 47.9% of the market share, with each focal property serving on average 2% of the market. There is a significant amount of price and location variation both within and across service sectors.

In addition to information on property name, capacity and revenue, our data contain information on each property's average daily rates (ADR), brand affiliation and the number of days the property is open throughout the year. These are collected by our data source and from financial reports, information from appraisers and chain and AAA directories. The ADR is the pre-sales tax price. Like Sudhir (2001) and Berry et al. (1995), our price measure is the average/aggregate price offered to consumers in a particular market. ADR is a very important marketing-decision variable and one that property managers pay a lot of attention to (Kalnins, 2004).

In the lodging industry, yield management practice might lead different consumers to receive different prices. If the focus of this study were targeting, then price aggregation leads to identification problems since the simulated household price-response coefficient could be the combined effect of response and price heterogeneity. However, pricing or segmentation is not the focus of this paper. Instead our objective is to compare and contrast the results of different demand specifications using data with aggregate prices. Since the aggregation bias is common to all the compared demand-model specifications, it appears improbable that superior fit (if any) of AGNL in the data over mixed probit and mixed logit is arising from price aggregation[7].

The data required to estimate the model include market shares, prices, brand characteristics and local firm cost variables. Therefore, we augment the property-level data with city-county level information including the indices for electric utilities and in-house

---

6 All properties (new and old) are included in the analysis however, the focal properties (non-outside good option properties) are limited to properties that entered prior to January 1991 and survived until December 2003. A detailed explanation of this is in the estimation section.

7 If we had access to individual-level transacted prices, conducting the analysis using this micro-level data or a combination of micro and macro data as done in Sudhir et al. (2005) would be a very interesting extension of this study.

services; these last two data series are to be used along with other exogenous variables as instruments to correct for demand-side endogeneity. Endogeneity correction is explained in detail in the next section.

For each property we also collect location information via Mapquest.com. Given the street address of a property, we obtained the driving distance in miles and driving time in minutes from two fixed loci - the airport, and the CBD. The correlations between distance and time measures were 0.87 and 0.96 for CBD and the airport respectively. Given the high degree of correlation between the two measures, we chose distance rather than driving-time measures in our empirical analysis.

Some additional issues arise in the implementation of the model to the data. First, in the context of hotel demand, it is not obvious what no-purchase means. For example, it could be consumers who visit the town but do not stay in a lodging facility, but data for this measure is hard to obtain. Moreover, there are multiple ways of translating the number of visitors to hotel room demand (e.g., does each visitor translate to one room or double occupancy?). According to Mintel Reports on the Lodging Industry, well over 50% of travelers (U.S., not Austin specific) stay in a hotel if they cannot stay with friends or family. Therefore treating the demand served by the non-focal properties alone as the outside good would result in biased estimates by understating total demand. Therefore we multiply the total demand served by all properties in the market during each time period by a random number drawn from a uniform distribution from [1,5] (given twice the served demand by focal and non-focal properties approximately equals total market, four times the non-focal demand approximates the size of the market). We use this number as the market size each period. It is this number that is used to calculate the observed market shares for the focal properties and outside good.

Tables 1 through 3 contain descriptive statistics of the final estimation sample. Note that there is significant variation in location and price in a market. Table 1 reflects variance in the entire sample, which includes both focal and non-focal properties. In order to rule out variance coming from pooling the three industry segments alone, we provide descriptive statistics by segment in Table 2. Table 2 provides evidence of price and location variance within and across the three segments. This, therefore, attenuates any concern of location and/or price clustering across segments, i.e. not all properties of one segment are around the airport and properties of other segments are around downtown. The same logic applies to prices as well. As one would expect, limited service properties are the least expensive ($49.47), followed by Full-Service ($85.24) and Extended-Stay ($93.91). In our data there are brands with multiple properties in the Austin, Texas, market. To ascertain brand-specific variance (not all properties of the same brand have same prices/location), we provide brand-specific descriptive statistics in Table 3. Here, too, the data indicate a significant variation in both location and prices.

---Insert Tables 1-3 here---

4. Choice Model Results

<u>4.1: Estimation Procedure</u>

We estimate the parameters of the aggregate demand function, explicitly accounting for endogeneity in price and location. This is achieved by an estimation procedure similar to Chintagunta (2001). The contraction-mapping procedure employed by Berry et al. (1995) cannot be used in estimating our proposed demand model for reasons explained below. Among other things, the contraction-mapping procedure facilitates use of linear instrumenting techniques to correct for endogeneity. The Berry et al. (1995) contraction-mapping procedure can be used when the structural error term can be inverted directly as a function of model demand parameters. Note that in the proposed demand model the structural error term cannot be represented as a linear function of demand model parameters. In order to still use linear instrumenting techniques to correct for endogeneity, rather than relying on Berry et al.'s (1995) contraction-mapping procedure, we use a multi-stage non-linear constrained optimization routine in conjunction with Generalized Method of Moments (GMM) estimator. We briefly outline the procedure used for estimation below. For a more detailed explanation, please refer to Figure 1, which presents a flow chart of the estimation procedure.

---Insert Figure 1 here---

The estimation involves estimating the system of equations as described by equation 18. Recall that while we specify a demand model at the individual level, the data we have is at the aggregate level. As shown below, we rely on recent advances in econometric analysis in marketing and economics to recover our demand-model parameters (Sudhir, 2001).

Stage 1: Make N (N=200) draws of the vector of individual specific parameters, i.e. $[\varsigma_{\gamma_i}, \varsigma_{\alpha_i}, \varsigma_{\phi_{i1}}, \varsigma_{\phi_{i2}}, \varsigma_{\phi_{i3}}, \varsigma_{\phi_{i4}}, \varsigma_{\mu_i}] = \theta_i$. These are drawn from a multivariate normal distribution $\theta_i \sim MVN(0, \sigma^2)$. Since σ^2 is unknown, we guess these as well. Once the individual-specific deviations are drawn, they are held fixed for now.

Stage 2: Next we draw two random vectors: a) a vector of size (H+1)*166 and set each (H+1)th element to be zero and b) A , a vector of size $\sum_{h=1..H}$ number of allocation parameters per property h .

The first vector serves as the vector of individual invariant variables contribution to market share expression in equation 18, i.e. $\overline{V}_{ht}$. The second vector is the vector that determines the allocation of overlapping properties across pre-specified nests. Using the fixed values of vector draws [θ_i, σ^2, A] from Stage 1 and Stage 2 draws $\overline{V}_{ht}$, we first compute the individual-choice probability's for each N for each time period. We then average the choices probabilities within each time period to generate the predicted property-specific market shares.

Stage 3: Using the GAUSS's constrained optimization (CO) routine, we minimize the distance between the observed market shares and predicted market shares by iterating over the space of [$\overline{V}_{ht}$]. Note, this is where we differ from BLP (1995) because our demand specification does not lend itself to the traditional log-linear transformation that works for the logit demand model. Chintagunta (2001) also relies on a non-linear optimization technique to minimize the error between predicted shares and observed shares. We have to rely on constrained optimization techniques because we have to constrain the property-level allocation parameters so that they sum to 1 for properties belonging to multiple nests.

This step yields estimates for the population-level mean utilities for each property as intercepts $\overline{V}_{hm}$ where as shown in Figure 1

$$\overline{V}_{ht} = \gamma_h - \alpha p_{ht} - \phi_1 d_{hCBDt} + \phi_2 d^2_{hCBDt} - \phi_3 d_{hAIRt} + \phi_4 d^2_{hAIRt} + \sum_n \overline{\mu}_n I_{hn} + \xi_{ht}$$
$$where\ I_{hn} = 1\ if\ \ h \in n\ else\ 0 \tag{19}$$

Stage 4: As motivated in section 2, both price and location measures can be correlated with ξ, making the OLS estimator inconsistent for estimating the population level mean parameters $\gamma, \alpha, \phi_1 - \phi_4$. Note, however, that the recovered $\overline{V}_{hm}$ is a linear function of observed and unobserved product attribute (ξ_{hm}). Therefore we use linear instruments-based 2SLS estimator to obtain unbiased estimates of $\gamma, \alpha, \phi_1 - \phi_4$. Like Nevo (2001), controlling for property specific means, we assume that market specific valuations are independent. The intuition is that while prices and location might be correlated across markets, the demand shocks are uncorrelated across markets. We select three period-lagged variables for each property, and monthly gas-price indices for Austin, TX, as instruments for price and location.

Stage 5: The property specific ξ's are obtained as residuals of the 2SLS regression from Stage 4.

$$\overline{V}_{ht} - [\hat{\gamma}_h - \hat{\alpha} p_{ht} - \hat{\phi}_1 d_{hCBDt} + \hat{\phi}_2 d^2_{hCBDt} - \hat{\phi}_3 d_{hAIRm} + \hat{\phi}_4 d^2_{hAIRt}] = \hat{\xi}_{ht} \qquad (20)$$

Stage 6: The structural error term vector $\hat{\xi} = \{\hat{\xi}_{ht}\}\ h = 1...H,\ t = 1...T$ is then interacted with the instruments used in Stage 3 to provide the GMM objective function, which is then used for estimating the (individual specific) parameters.

$$E[\hat{\xi}] = 0 \qquad (21)$$

We construct moment conditions of the data-generating process based on the proposed model via orthogonality conditions. The model parameters are estimated by minimizing the sample analogue of the orthogonality conditions, i.e. setting these conditions to zero. We can construct orthogonality conditions using any set of covariates $\boldsymbol{Z}$ that are mean-independent on $\hat{\xi}$. These include exogenous property characteristics and cost shifters. The moment conditions rely of the conditional mean-independence assumption $\boldsymbol{E(\hat{\xi} \otimes Z \mid Z) = 0}$ and restrictions a) $\boldsymbol{E(\hat{\xi}_t \hat{\xi}_t' \mid Z)} = \Theta$ and b) $\boldsymbol{E(\hat{\xi}_t \hat{\xi}_{t'}' \mid Z) = 0}\ \forall \boldsymbol{t \neq t'}$.

Stage 7: Since our demand model is highly non-linear, the grid search procedure to compute the optimal values of $\overline{V}_{ht}$ can be quite sensitive to starting values. Therefore we repeated the exercise with different starting values and seeds for the random-sequence generator. Given the high dimensionality of the parameter space, starting values generated by traditional random-sequence generators may have certain limitations (Bhat, 1999; Revelt and Train, 2000). Instead we generate a Halton sequence of draws for starting values.

Unlike traditional robustness-testing methods where the objective is true randomness to ensure that the same convergence point is reached regardless of starting point, in high-dimension search spaces, true randomness might not achieve the objective of an efficient and thorough search of the parameter space. The Halton sequence procedure is an alternative process for efficient coverage of the parameter space in such situations. Numerical analyses have found that a smaller number of Halton draws is as effective as a larger number of pseudo-random draws using a random-number generator (Revelt and Train, 2000). Different starting values are drawn from a Halton sequence[8]. Using the Halton sequence generator, stages 1-6 are repeated until we obtain robust results and convergence criteria are met[9].

8 Like Nevo (2001), we use draws from a multivariate normal distribution to simulate an individual's preference for attributes. One limitation of this approach as pointed out by Kim et al. (1995) and Brownstone and Train (2002) is that it allows for both positive and negative individual coefficients for price and location. They suggest draws from an alternative distribution within a restricted range (truncated normal distribution). Our results do not change when we estimated our models with truncated normal draws instead of draws from a normal distribution.

9 A detailed description of the estimation procedure can be obtained from the authors. A Maximum Likelihood procedure was also implemented. It relied on a normality assumption on the distribution of the ξ, unlike the GMM procedure. Only the GMM results are presented in the paper. MLE results can be obtained from authors. We also tested for the endogeneity of price and location using a Hausman (1978) test comparing SUR and 2SLS estimates.

4.2: Estimates and Model Fit

Several GNL specifications were tested. These include:

a) Location alone (Model 1): exogenously specified three nests each comprising of : i) properties closer to downtown than airport, ii) properties closer to airport than downtown and iii) outside good, respectively.
b) Location alone (Model 2): exogenously specified five nests, each comprised of : i) properties within 5 miles of downtown, ii) within five miles of the airport, iii) within 5-10 miles of downtown, iv) within 5 to 10 miles of the airport and v) the outside option. Properties that met multiple criteria were forced into only one nest (no overlaps) based on whether it was closer to downtown or the airport[10].
c) Price alone (Model 3): Here we specified four nests--one for full-service properties, one for limited-service, one for extended-stay and one for the outside option.
d) Location followed by Price (Model 4): We specified a three-level GNL model where we first nested properties based on location criteria similar to Model 1, and under each of these nests (apart from the outside-good nest) we created sub-nests based on price tiers as in Model 3.
e) Price followed by Location (Model 5): We specified a three-level GNL model where we first nested properties based on price criteria similar to Model 3, and under each of these nests (apart from outside good nest) we created sub-nests based on location tiers as in Model 1.
f) Location Model with Overlaps (Model 6: We specify a nesting structure similar to Model 2. However, we allow a property that meets multiple nesting criteria to be available in each of the condition-meeting nests.

Using the Smith (1992) model-selection test for non-nested GMM based models, Model 6 was chosen over Models 1 through Model 5. This ex-ante nesting structure yields 15 property-nest overlaps. Table 4 shows the distribution of these overlaps across nests.

---Insert Tables 4, 5(a) and 5 (b) here------

While several competing models were estimated, results for homogenous MNL (with and without endogeneity correction), aggregate logit, aggregate probit, homogenous GNL (Model 6) and aggregate GNL (Model 6) are only presented in Table 5 (a). Since one of the main objectives of this study is to demonstrate the usefulness of AGNL as another demand model, we report results for aggregate logit and aggregate probit as a benchmark, given their widespread usage in marketing and economics (Sudhir, 2001; Chintagunta, 2001).

We compare our results across models that account for both heterogeneity and endogeneity. The signs of the price and location demand parameters are maintained across all models. We find the population mean response coefficient of price is smaller for AGNL than either the aggregate logit or aggregate probit specifications. This could be in part because of the flexible nature of the substitution patterns that GNL affords relative to other demand models. Market share difference between properties with similar prices but at different locations would be attributed to location-specific heterogeneity. Similarly, market share

10 The distance between the airport and downtown in Austin is 12 miles. We chose 5-mile intervals to maximize geographic market coverage while maintaining a manageable number of nests, i.e. as the interval reduces, the number of nests increases. More nests result in more nest specific logsum parameters to be estimated. However, we did test our model using intervals of 3- miles and 6-miles. The 5-mile interval model had the best predictive power (lowest RMSE and best out-of-sample predictions) and was chosen over others among competing GNL model specifications using Smith (1992) non-nested GMM model-selection criteria .

difference between two properties in the same location but at different price tiers, is attributed to heterogeneity in price. In GNL, some of the variance that is previously attributed to heterogeneity, is now being explained by the combination of the nest-specific correlations (log-sums) between alternatives and overlapping alternatives. We also find that as we specify more flexible demand specifications at the individual level, we obtain smaller estimates of population-level heterogeneity in location-response parameters, but higher price- response parameters. This result can be attributable to final location based demand clustering that we observe in the data.

---Insert Table 6 here---

Next consider the estimates of allocation parameters shown in Table 6. Recall that properties that do not belong a particular nest have an exogenously specified nest-allocation value set to 0. Similarly, a property that belongs to only one nest has its allocation parameter fixed at 1. Properties that have fixed-nest specific allocation parameters are indicated by a superscript *f*. We constrain the sum of nest-specific allocations per property to be 1, so that each allocation parameter can be interpreted as that property's fraction of total market share that comes from that specific nest. Properties that have presence in two nests require estimation of only one allocation parameter. If a_{hn} is the allocation parameter of hotel h in nest n (h is only two nests), then h's allocation parameter in the second nest is ($1- a_{hn}$), as is indicated in Table 6. Note that the allocation structure of our model is specific to the Austin, TX, market that we study and not a limitation of the GNL model specified. For example, if there were additional ideal points in the market of interest, one would have to test a different set of nesting structures and use a model-selection criteria to chose between alternative demand specifications. It is important to note that the best fitting demand structure for our market might not necessarily be ideal for other geographic markets. That said, the prescribed model can easily accommodate changes in market conditions. Note that in Table 6, all but one of the estimated allocation parameters are significantly different from 0 or 1, thereby negating the existence of a simple nested logit.

---Insert Table 7 here---

Note from Table 7 that there isn't a pure quality tier or pure location clustering in nesting structure that yields best fit. If competition were purely quality based, then we would expect that each nest be made up predominantly or exclusively of properties belonging to the same industry quality segment. In this case the allocation parameters of overlapping properties belonging to different quality tiers will be zero, which from Table 6 we know is not the case. Such a result would also imply that Model 3 (price based nested demand model), be the best fitting model of all the models estimated. If Model 3 were chosen then the predicted nest shares shown in Table 7, will be an algebraic sum of predicted shares of all properties belonging to the quality tier specific to that nest. Note from Table 7, this is not the case. Nest 2 and Nest 3 include share contributions from all three industry quality tiers. Recall, however, that these nest-specific segment shares are not equivalent to summing up the share of sales per property of that segment within a nest, as is the case in nested logit, because of the overlapping feature of GNL.

Following similar logic as above for what results might look like if competition were purely price based, consider what model results would look like if competition were location based alone. Then either Model 1 or 2 should be chosen over other competing models. Also, we would also expect price promotion by one property in a specific nest to draw more share from another property belonging to the same nest than from a property belonging to another nest. As explained later in this section, due to overlapping feature of GNL, we find that the cross-price elasticity between alternatives belonging to similar quality tiers and different nests is sometimes larger than between alternatives belonging to the same nest and different quality tiers. This implies that there is some price based competition at work as well.

Since overlapping alternatives can be present in differing degrees in multiple nests, the degree of similarity between properties within a nest is both a function of membership within the nest, but also allocations of each of these members. The degree of similarity between nest alternatives seen in Tables 7 is also captured in the logsum parameters shown in Table 5(a). Note that the logsum of the nest with outside good is set to be 1 for identification purposes. These structural nest parameters, or logsums, are significantly different from one another, suggesting violation of IIA property in the data (therefore, the logit specification will not be appropriate for this situation). They also lie between 0 and 1 and are hence consistent with random utility maximization. Nest 3 nest has the highest logsum parameter (0.394), or has the most dissimilar properties of the four nests, and Nest 1 (close to airport) has the most homogenous membership (0.221). The logsum results are outcomes of the nest composition as seen in Table 7, and property-specific nest allocation parameters discussed earlier. We find the estimated mean price and mean location parameters for mixed GNL to be statistically different from other aggregate-demand models. We also compute the own and cross-price elasticities and respective standard errors at the property level across four models – aggregate logit, aggregate probit, homogenous GNL and AGNL. We report the average measures of own and cross-price elasticities along with their ranges across models in Table 8.

---Insert Table 8 here------

We find both the mean and range of own and cross-price effects to be qualitatively different from their aggregate probit and aggregate logit counterparts[11]. Consistent with Chintagunta (2001), we find the range of own and cross-price effects to be larger for aggregate probit than aggregate logit. This is attributed to the differences in error variance across the alternatives. The mean own price elasticity for AGNL is larger than aggregate probit, which is larger than aggregate logit (-4.222>-3.671>-3.236). The range of own price elasticity for aggregate GNL is larger than aggregate probit (4.771 vs. 3.978), which is larger than aggregate logit (3.978 vs. 2.845). Similarly, the mean cross-price elasticity for AGNL is larger than aggregate probit, which is greater than aggregate logit (.993>.845>.782). Much like the range of own price elasticities, we find the range of cross-price elasticities to be largest for AGNL and smallest for aggregate logit (.636>.488>.351).

As stated earlier, AGNL captures consumer choice outcomes better than aggregate logit or aggregate probit for these data. Managers can use these results to gain insights into their competitive set. Consider the following example, where cross-price elasticity with own-price tier property is lower than cross-price elasticity with a different-price tier property. Let us look at three properties listed in Table 6. DoubleTree (Full Service) and Motel 6 (Limited Service) have presence in Nest 4 only while Ramada (Limited Service) has presence in Nest 2 and Nest 4. Since Ramada has a smaller presence in Nest 4 (.334) and a larger presence in Nest 3 (.666), its cross-price elasticity with Motel 6, which belongs to the same quality tier as itself (i.e. Limited Service), is .217**[12].On the other hand, its cross-price elasticity with DoubleTree, which belongs to a different quality, is .381**. Hence the cross-price elasticity between alternatives in our

11 Due to space limitations we have not included the elasticity tables (23*23) matrix for aggregate logit, aggregate probit, homogenous GNL and aggregate GNL. Interested readers can obtain these tables by contacting the authors.
12 ** implies significance at the 95% confidence level.

model is a function of their nest-specific allocations, and if these properties meet in multiple nests, the pair-wise nest-specific elasticities will vary by nest. Such information can be very useful for managers in price-setting; these insights are obtainable because of the flexible GNL demand structure.

Table 5b has results for model selection. Based on the sum of squares errors alone, mixed GNL outperforms the other models. Aggregate GNL was chosen over other competing demand models aggregate logit, aggregate probit and aggregate nested logit (this is a restrictive case of the aggregate GNL), using both the Cox criteria and Encompassing Test criteria for non-nested competing models using GMM as proposed in Smith (1992)[13]. These non-nested test criteria evaluate how the alternative models explain the residual unobserved determinants of market share of a chosen null model. The two tests proposed in Smith (1992) are non-nested tests for competing regression models that are estimated using instrumental variables and GMM estimation.

The AGNL model is chosen over aggregate logit and aggregate probit based on: a) Smith's (1992) non-nested GMM based model selection criteria, b) in-sample root mean square errors (RMSE) and c) out-of sample predictions. The out-of-sample predictions were conducted in three steps. First, the recovered demand-model parameters are used to recover the structural error terms for each property-time period. We have to rely on this because the structural error terms serve as residuals of our regression analysis and are not observed in the data. Second, we take the last thirty months of the data and demand-model parameters to predict property-level market shares. Third, the mean percentage error for all models are computed and then used for model comparison. The nest-specific out-of-sample predictions are reported in Table 9[14].

---Insert Table 9 here------

Remember that understanding tradeoffs that consumers make in price and location is an objective of this study. To that extent, our proposed model results confirm the attractiveness of the two ideals points for consumers. While consumers prefer to be closer to the two ideal points, they need to make tradeoffs with higher prices to be afforded this benefit. This is because firms closer to consumer ideal points charge higher prices than those properties located farther away. We find that consumers are willing to pay more money to be closer to downtown than the airport ($3.783 vs. $1.435).

To summarize, a comparison of the results across several demand-models indicates differences in the response parameters of the utility function, own and cross price elasticities and range of elasticities. Benchmarking our demand specification, i.e. AGNL with the commonly used aggregate logit and aggregate probit, is an important contribution of this paper, since the resulting aggregate elasticities affect pricing decisions. Therefore, AGNL appears to be a good model to describe consumer choices in this industry with ex-ante unknown/complex choice heuristics. The allocation parameters indicate that nest membership is not simply a function of price or location alone, but a complex combination of these and other factors.

---

13 If H_0: E_0[Full Model]=0 and H_{alt}: E_1[Non-Nested Variant Model]=0 are the two competing hypotheses, then the form of the Cox statistic $\square_t(H_o \mid H_{alt})$ is that of a GMM specification test (Newey, 1985), and is amenable to local power analysis. Similarly, the form of the Encompassing test statistic $E£_t(H_o \mid H_{alt})$ has a limiting chi-squared distribution.

14 A similar exercise was carried out for aggregate probit and aggregate logit. Interested readers can obtain these results from the authors.

Our model provides useful managerial insights as well. For example, managers can use our model to identify key competitors. If a property has a presence in multiple nests depending on its allocation parameter in each nest, its closest competitors (in terms of price elasticity or ability to steal market share) could be a different set of properties within each nest. In one nest it could be a property of the same quality tier and in another it could be one of a different quality tier. Consider the following two situations. First, the focal property has a large allocation in a nest where other properties of the same quality tier have small allocations and other quality tier properties have high allocations. Second, the property has high allocation in a nest where there is a high allocation of properties belonging to different quality tiers The appropriate competitive choices of pricing and other forms of differentiation (vertical and horizontal) are likely to be different across these two situations (and depend on the cost of vertical and horizontal differentiation, among other things). While directly answering these questions is outside the scope of this paper, our results provide demand-side estimates that should alert managers that competitive sets are complex and need to be thought of more deeply.

5. Impact of Consumer Choices on Firm Profits

To understand the impact of consumer choices on firm profits, we estimate a simple supply-side model. The components of this are standard in the empirical industrial organization literature, so our explanation will be brief.

5.1: Modeling Cost and Competition

We assume that the monthly (given frequency of data) marginal cost per room for property *h* is a function of property-specific cost shifters, distance from ideal points, and a random cost shock as shown below

$$mc_{ht} = \omega_{1h} + \kappa_1 * elect_{ht} + \kappa_2 * serv_{ht} + \rho_1 * d_{hCBDt} + \rho_2 * d^2_{hCBD} + \rho_3 * d_{hAIR} + \rho_4 * d^2_{hAIRt} + \upsilon_{ht} \tag{22}$$

Here ω_{1h} is the property-varying component of the marginal cost and υ_{ht} is a zero-mean time-varying stochastic shock. Electricity costs and service costs (like security, cleaners, etc.) are exogenous cost components. We specify a quadratic distance-based cost function to allow for non-linear effects as firms move away from ideal points. Costs can be related to distance from ideal points for a number of reasons. For example, hotels might need to offer additional services like shuttle service to downtown and the airport, luggage check-in, additional dining alternatives, etc. to attract location-sensitive consumers. These additional service offerings increase marginal costs for the property. It is also possible, and even likely, that consumer ideal points attract a cluster of competitors, and hence create a common labor market or lower delivery costs for suppliers (Hannan and Freeman, 1977; Baum and Haveman, 1997). Another possiblity is that a cluster of competitors can increase the marginal costs of labor due to the risk of employee turnover to competitors. The stochastic marginal cost error term captures cost shocks (e.g. due to travel shuttle service costs, refurnishing, new on-site features) that are unobserved to the econometrician but known to the firm. We assume that υ_{ht} is independent of ε_{iht}.

Given this cost function and location pre-determined prior to the start of our data period, firms maximize profits by choosing monthly prices in a Bertrand-Nash manner. Note that in estimation, we treat location as econometrically endogenous. In the data, firms have multiple properties in each market, and are modeled as maximizing profits across their product line. Let each firm *f* own $\boldsymbol{H_{ft}}$ is a set of properties owned or managed by firm *f* at time t.

The product line profit function for firm *f* over all its properties h at time t is given by

$$\Pi_{ht} = \sum_{h \in H_{ft}} (p_{ht} - mc_{ht}) S_{ht} * M_t \tag{23}$$

where S_{ht} = share of property *h*, M_t is the size of market including the outside good. Differentiating the profit function with respect to price, we obtain price first-order conditions as follows

$$p_{ht}^{*} = mc_{ht} + \frac{\sum_{z \in H_{ft}; z \neq h} (p_{zt} - mc_{zt}) \left[\frac{1}{NS_i} \sum_i \left(\hat{P}_{iht} + \frac{\sum_n \left(\frac{1}{\mu_{in}} - 1 \right) \hat{P}_{int} \hat{P}_{iht|n} \hat{P}_{izt|n}}{P_{izt}} \right) \alpha_i \hat{P}_{iht} \right] + \hat{S}_{ht}}{\frac{1}{NS_i} \sum_i \frac{\sum_n \hat{P}_{int} \hat{P}_{iht|n} \left[(1 - \hat{P}_{int}) + \left(\frac{1}{\mu_{in}} - 1 \right) \left(1 - \hat{P}_{iht|n} \right) \right]}{\hat{P}_{int}} \alpha_i P_{iht}} + \upsilon_{ht} \tag{24}$$

Using standard empirical IO assumptions, marginal costs are not observed. In the literature, there are two broad approaches to jointly estimating demand and supply. Multiple approaches have been proposed in the current literature to account for this. Yang et al. (2004) estimate a hierarchical Bayes-based joint supply demand system. The advantage of their approach is they can recover agent-level (household and firm-specific) response parameters. However, this requires household-level data and making distributional assumptions on the structural and marginal cost errors. Berry et al. (1995) estimate the demand and supply-side model using aggregate data and the generalized method of moments (GMM) estimator. Their approach requires estimating the joint system of demand and supply-side moment conditions.

Newey and McFadden (1994) estimate the model in two stages -- demand model in stage one and supply model in the second stage. They rely on obtaining consistent estimates of the demand system and on the equilibrium assumption to compute the marginal costs implied by equation 23. We employ a two-stage estimation procedure similar to Newey and McFadden (1994) for several reasons. First, because equilibrium is not enforced in the demand estimation procedures, consistency of demand-parameter estimates is independent of the particular equilibrium assumptions made, and is therefore robust to a wide set of possible assumptions. Second, there is no need to solve the equilibrium pricing problem during demand- model estimation, thereby greatly reduceing the computational burden of the estimation procedures involved with estimating the proposed non-linear demand model. The disadvantage of the approach is that greater efficiency could be obtained in the estimates by enforcing equilibrium during estimation. Third, the two-stage estimation procedure employed here does not require us to posit any distributional assumption on ξ in the consumer-utility specification, as done in the Villas-Boas and Zhao (2003) study where the demand and supply side models are jointly estimated.

5.2: Results and Discussion

The estimates of marginal cost are in Table 9.

---Insert Table 9 here------

Marginal cost decreases at a decreasing rate as a property moves farther away from CBD. Marginal costs increase at an increasing rate as a property moves away from the airport. These distance-based costs can capture a variety of variables not already in our cost specification, including different tax rates, costs related to agglomeration, etc. Several of the brand-specific intercepts and cost elements in the marginal cost function are statistically significant (only means reported to conserve space). The overall cost for a luxury brand is higher than for non-luxury brands. That is, the correlation between price and estimated marginal-cost is very high (.92). As an ex-post check of our marginal cost specification, we also calculated the correlation between estimated marginal costs and the number of properties per brand per city. This correlation is very low (.003), ruling out economies of scope. The correlation between capacity and marginal costs is .42, and capacity and prices is .54. This can be interpreted as market power of larger hotel via fixed/sunk costs.

Therefore, it appears that though CBD is an ideal point from the consumer/revenue perspective, it is not from a cost perspective. The airport location is unambiguously good from revenue and cost perspectives for any given firm. (Of course, in equilibrium, this is not true for all firms simultaneously, given there is finite demand at any ideal point). Our demand-side and supply-side results demonstrate the importance of geographic location both as an important attribute of consumer demand and as an element of the cost structure of a firm.

6. Conclusions

In this paper, we study the significance of price and geographic location as important elements of consumer choice in the lodging industry. In this industry and our data, choice heuristics are not obvious, and there is missing information on several attributes likely to matter to consumer choice. Therefore, we show that the GNL is a good consumer choice model. We also briefly examined the implications of this for firm profits.

The contributions of the paper are both methodological and substantive. We demonstrate for our industry and dataset, that -- a GNL based demand model -- that also accounts for taste and error heterogeneity offers distinct benefits and statistical superiority over the popular mixed logit and aggregate probit. Its ability to accommodate flexible patterns of substitution between choices results in more accurate predictions of the price-and-location response parameters, and a more accurate picture of the consumer choices.

The methodology employed in this paper can be used to test for price-location interactions in other retail environments where price and/or location are strategic variables. The flexibility of substitution patterns offered by GNL makes it a good alternative for situations in which nesting structures are overlapping. Examples of such an application are models of store-brand choices for studying retail competition, car choices, financial investment choices, college choices etc., where an alternative can be viewed as being a member of multiple nests and/or alternative nesting structures are possible.

To enrich our understanding of the consumer decision-making in these situations, it would be useful to complement market-level data with individual-level choice data (Berry et al., 2004) and Sudhir et al. (2004), unlike the present paper that relies only on aggregate data. Another extension would be to use the GNL framework to not just characterize the final choice but also the antecedent consideration set formation (Swait, 2000). We look forward to future work in this area.

References

Anderson, S.P. and dePalma, A., 1992. Multiproduct firms: a nested logit approach. *Journal of Industrial Economics*, 40, pp. 261-276.

Baum, Joel A.C., and Heather A. Haveman, 1997. Love thy neighbor? Differentiation and spatial agglomeration in the Manhattan hotel industry. *Administrative Science Quarterly*, 42: 304-338.

Bhat, C.R., 1999. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. Working paper, University of Texas, Austin

Besanko, David; Dube, J-P; Gupta, Sachin, 2003. Competitive Price Discrimination Strategies in a Vertical Channel with Aggregate Retail Data. *Management Science*.

Berry, Steve, James Levinsohn and Ariel Pakes, 1995. Automobile Prices in Market Equilibrium. *Econometrica*, 841-890 Vol. 63 (4) pp. 841-890.

Berry, Steve, James Levinsohn and Ariel Pakes, 2005. Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*, (Forthcoming).

Bodapati, A. and S. Gupta, 2004. The Recoverability of Segmentation Structure from Store-Level Aggregate Data. *Journal of Marketing Research* (Forthcoming).

Chintagunta, Pradeep, 2001. Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data. *Marketing Science*, 20, 4, 442-456

Choi, S. C., W. S. DeSarbo and P. T. Harker, 1990. Product Positioning under Price Competition. *Management Science*, 36, pp. 175-199.

Davis, Peter, 2002. Entry, Cannibalization and Bankruptcy in the U.S. Motion Picture Exhibition Market, Mimeo MIT.

Hannan, Michael T. and Freeman, John, 1977.The Population Ecology of Organizations, A*merican Journal of Sociology*.

Hauser, John R., and Steven M. Shugan, 1983. Defensive Marketing Strategy. *Marketing Science*, Vol. 2, No. 4,(Fall), 319-360.

Hausman, Jerry A, 1978. Specification Tests in Econometrics. *Econometrica*, Vol. 46 (6) pp.1251-71.

Huff, L.D, 1966. Programmed Solution for Approximating an Optimum Retail Location, *Land Economics*, 42 (3): 293-303

Kalnins, Arturs, 2004. An Empirical Analysis of Territorial Encroachment in Franchised and Company-owned Branded Chains, *Marketing Science*, 23(4): 476-489.

Kamakura, Wagner and Gary J. Russell, 1989. A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research*, 26, pp. 379-90.

Kamakura, Wagner, Kim Byung-Do, and Jonathan Lee, 1996. Modeling Preference and Structural Heterogeneity in Consumer Choice. *Marketing Science*, Vol. 15 (2)

Kamita, Rene, 2001. Merger Analysis in Geographically Differentiated Industries with Municipal and Private Competitors: The Case of Solid Waste Disposal. Mimeo UCBerkeley.

Kannan, P. K. and Gordon P. Wright, 1991. On Testing Competitive Market Structures. *Marketing Science*, Vol. 10, No. 4, pp. 338 - 34.

McFadden, D., 1973. Conditional Logit Analysis of Quantitative Choice Behavior, in Zaremmbka P. (ed.), *Frontier of Econometrics*, Academic Press, New York.

McFadden, D., 1989. A Method of Simulated Moments for Estimation of Discrete Choice Response Models without Numerical Integration. *Econometrica*, Vol.57, No.5, pp.995 1026.

Nevo, A., 2001. Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69(2), 307-342.

Newey, W.K., 1985. Generalized Method of Moments Specification Testing. *Journal of Econometrics*, 29, 229-256.

Newey, W.K. and D.L. McFadden, 1994. Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics*, Vol.4, Amsterdam: North-Holland.

Revelt, D., and K. Train, 2000. Customer Specific Taste Parameters and Mixed Logit Models. Working Paper, University of California, Berkeley.

Seim, Katja, 2000.Geographic Differentiation and Market Structure: The Video Retail Industry, mimeo Stanford University.

Smith, Richard J., 1992. Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments. *Econometrica*, Vol. 60, No. 4, 973-980.

Sudhir, K., 2001. Competitive Pricing Behavior in the Auto Market: A Structural Analysis, *Marketing Science*, 20 (1),42-60.

Sudhir, K, P.B. Seetharaman and Hai Che, 2004. Beyond the Tradeoff between Richness and Representativeness: Optimally Combining Individual and Store level Data. Working Paper.

Swait, J., (2000). Choice Set Generation Within the Generalized Extreme Value Family of Discrete Choice Models, *Transportation Research Part B*.

Swait, Joffre, and Adriana Bernardino. 2000 . Distinguishing taste variation from error structure in discrete choice data. *Transportation Research Part B* 34(1) 1 –15.

Thomadsen, R., 1999. Price Competition in Industries with Geographic Differentiation: The Case of Fast Food, Mimeo, Stanford University, Stanford, CA.

Train, E. Kenneth, 2003. *Discrete Choice Models with Simulation*. Cambridge University Press.

Villas-Boas, M, and Zhao, Ying, 2003. The Ketchup Marketplace: Retailer, Manufacturers, and Individual Consumers. Working Paper.

Vovsha, P., 1997. The cross-nested logit model: Application to mode choice in Tel Aviv metropolitan area, Conference presentation, 76th Transportation Research board meetings, Washington, DC.

Wen, C.H. and Koppelman, F.S., 2001. The Generalized Nested Logit Model. *Transportation Research-B*, V 35, N 7, pp 627-641.

Wall, G, Dudycha, D and Hutchinson, J., 1985. Point pattern analyses of accommodation in Toronto, *Annals of Tourism Research*, 12(4) pp. 603-618.

Yang, Sha, Yuxin Chen and Greg M. Allenby (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1, 251-275.

Figure 1: Flow Chart for Demand Model Estimation

Step 1

$$\begin{aligned}
V_{iht} &= (\gamma_h - \alpha p_{ht} - \phi_1 d_{hCBD} + \phi_2 d_{hCBD}^2 - \phi_3 d_{hAIR} + \phi_4 d_{hAIR}^2 + \sum_n I_{hn}.\overline{\mu_n} + \xi_{ht}) \\
&\quad + (\zeta_{\gamma_{ih}} - \zeta_{\alpha_{ih}} p_{ht} - \zeta_{\phi_{1i}} d_{hCBD} + \zeta_{\phi_{2i}} d_{hCBD}^2 - \zeta_{\phi_{3i}} d_{hAIR} + \zeta_{\phi_{4i}} d_{hAIR}^2 + \sum_n I_{hn}.\mu_{in}) \\
&= \overline{V_{ht}} + (\zeta_{\gamma_{ih}} - \zeta_{\alpha_{ih}} p_{ht} - \zeta_{\phi_{1i}} d_{hCBD} + \zeta_{\phi_{2i}} d_{hCBD}^2 - \zeta_{\phi_{3i}} d_{hAIR} + \zeta_{\phi_{4i}} d_{hAIR}^2 + \sum_n I_{hn}.\mu_{in})
\end{aligned}$$

Step 2

Guess - $\sigma_{\gamma_{ih}}, \sigma_{\alpha_{ih}}, \sigma_{\phi_{1i}}, \sigma_{\phi_{2i}}, \sigma_{\phi_{3i}}, \sigma_{\phi_{4i}}, \sigma_{\mu_{in}}$ and Draw - $\zeta_{\gamma_{ih}}, \zeta_{\alpha_{ih}}, \zeta_{\phi_{1i}}, \zeta_{\phi_{2i}}, \zeta_{\phi_{3i}}, \zeta_{\phi_{4i}}, \mu_{in}$

Step 3

Fix - $\zeta_{\gamma_{ih}}, \zeta_{\alpha_{ih}}, \zeta_{\phi_{1i}}, \zeta_{\phi_{2i}}, \zeta_{\phi_{3i}}, \zeta_{\phi_{4i}}, \mu_{in}$ and $\underset{\overline{V}_{ht}}{Min}\left[S_{ht}^{obs} - S_{ht}^{pred} \right]$

Step 4

$$\overline{V_{ht}} = \gamma_h - \alpha p_{ht} - \phi_1 d_{hCBD} + \phi_2 d_{hCBD}^2 - \phi_3 d_{hAIR} + \phi_4 d_{hAIR}^2 + \sum_n I_{hn}.\overline{\mu_n} + \xi_{ht}$$

Regress $\overline{V_{ht}}$ on predictor variables using 2SLS

Step 5

Recover the vector of ξ_{ht}'s from Step 4 and interact these with instrument matrix Z used in Step 4. These moment conditions serve as the GMM objective function for Step 6

Step 6

$$\underset{\zeta_{\gamma_{ih}}, \zeta_{\alpha_{ih}}, \zeta_{\phi_{1i}}, \zeta_{\phi_{2i}}, \zeta_{\phi_{3i}}, \zeta_{\phi_{4i}}, \mu_{in}}{\text{Min}} \quad E(\xi * Z)$$

Recover heterogeneity parameters

Table 1 – Descriptive Statistics

Statistic	Outside Good Share	Distance to Airport (miles)	Distance to Downtown (miles)	Market Share of Property	Daily Price ($)
Min	0.273	0.000	0.000	0.003	21.657
Max	0.702	18.000	8.800	0.075	257.077
Mean	0.521	8.465	3.909	0.021	73.930
Std. Dev.	0.121	3.752	2.240	0.011	33.311

Number of observations = 3818 (166 months * 23 properties)

Table 2 – Descriptive Statistics by Industry Segment

Service Sector	N	Market Share Mean (Std Dev)	Dist. From Airport (miles) Mean (Std Dev)	Dist. From Downtown (miles) Mean (Std Dev)	Property Market Share Mean (Std Dev)	Daily Price ($) Mean (Std Dev)
Extended Stay	166	0.063 (0.017)	8.000 (2.831)	3.077 (1.473)	0.021 (.008)	93.910 (23.40)
Full Service	166	0.319 (0.077)	8.519 (4.271)	3.742 (2.457)	0.027 (.012)	85.240 (35.248)
Limited Service	166	0.097 (.029)	8.559 (3.166)	4.473 (1.985)	0.012 (.005)	49.474 (13.452)

Table 3-Descriptive Statistics by Brand

Brand	N	Market Share Mean (Std Dev))	Daily Price Mean (Std Dev)	Dist. from Airport Mean (Std Dev)	Dist. from Downtown Mean (Std Dev)
Courtyard	166	0.018 (0.005)	71.379 (21.375)	12.000 (0.000)	2.700 (0.000)
Doubletree	166	0.032 (0.009)	79.781 (12.162)	12.000 (0.000)	5.300 (0.000)
Embassy	332	0.025 (0.007)	104.848 (20.694)	9.000 (3.005)	2.615 (1.617)
Four Seasons	166	0.029 (0.008)	166.576 (48.034)	7.000 (0.000)	0.000 (0.000)
Hampton	166	0.012 (0.004)	61.333 (11.051)	12.000 (0.000)	5.440 (0.000)
Hawthorn	166	0.013 (0.004)	72.033 (8.824)	6.000 (0.000)	4.000 (0.000)
Ramada	166	0.008 (0.003)	56.127 (8.517)	8.240 (0.000)	6.000 (0.000)
Super 8	166	0.008 (0.002)	41.090 (6.475)	7.000 (0.000)	2.000 (0.000)
Holiday Inn	498	0.022 (0.008)	68.118 (12.027)	9.667 (5.913)	4.000 (2.947)
Hyatt	166	0.043 (0.010)	99.857 (21.120)	8.000 (0.000)	1.000 (0.000)
LaQuinta	664	0.012 (0.004)	56.148 (9.278)	6.718 (1.684)	3.860 (1.012)
Marriott	166	0.037 (0.010)	103.475 (19.329)	9.000 (0.000)	2.600 (0.000)
Motel 6	332	0.015 (0.005)	34.389 (6.923)	11.300 (3.605)	6.450 (2.354)
Omni	332	0.030 (0.007)	87.908 (22.997)	3.105 (3.110)	5.150 (2.153)
Red Lion	166	0.026 (0.009)	65.512 (10.929)	10.780 (0.000)	5.000 (0.000)

Table 4 – Nesting Rules and Nest Membership Overlaps

Nesting rule

Recall that:
If distance from airport ≤ 5 miles then Nest 1
If distance from downtown ≤ 5 miles then Nest 2
If distance from airport is (5,10] miles then Nest 3
If distance from downtown is (5,10] miles then Nest 4
Nest 5 only contains the outside good

Overlaps

The diagonal elements of the matrix below indicate the number of properties that only belong to that nest. Off diagonal elements correspond to properties that overlap across nests.

Total number of property specific estimated allocation parameters = 2+1+1+9+1+1=15

	Nest 1	Nest 2	Nest 3	Nest 4
Nest 1	0	2	1	1
Nest 2	2	2	9	1
Nest 3	1	9	1	1
Nest 4	1	1	1	5

Table 5 (a) - Estimated Demand Model Parameters

	Homog. Logit	Mixed Logit	Mixed Probit	Homog. GNL	Mixed GNL
Mean Response (Std. Err)					
Price	-.022 (.000)	-.034 (.001)	-.026 (.000)	-.021 (.000)	-.023 (.000)
Dist. From Airport	-.048 (.014)	-.059 (.024)	-.048 (.014)	-.026 (.004)	-.033 (.001)
Dist. From Downtown	-.216 (.019)	-.216 (.020)	-.179 (.019)	-.014 (.011)	-.087 (.009)
Dist. From Airport^2	.001 (.000)	.005 (.000)	.001 (.000)	.001 (.000)	.001 (.000)
Dist. From Downtown^2	.024 (.002)	.033 (.008)	.024 (.002)	.003 (.000)	.003 (.000)
LogSum (Nest 1)	-	-	-	.228 (.011)	.221 (.009)
LogSum (Nest 2)	-	-	-	.251 (.041)	.261 (.029)
LogSum (Nest 3)	-	-	-	.410 (.031)	.394 (.028)
LogSum (Nest 4)	-	-	-	.281 (.017)	.310 (.019)
Heterogeneity (Std. Err)					
Price	-	.017 (.000)	.019 (.000)	-	.024 (.000)
Dist. From Airport	-	.041 (.000)	.052 (.004)	-	.029 (.000)
Dist. From Downtown	-	.128 (.019)	.214 (.001)	-	.053 (.000)
Dist. From Airport^2	-	.004 (.000)	.010 (.001)	-	.002 (.000)
Dist. From Downtown^2	-	.001 (.005)	.002 (.000)	-	.001 (.000)
LogSum (Nest 1)	-	-	-	-	.011 (.000)
LogSum (Nest 2)	-	-	-	-	.017 (.009)
LogSum (Nest 3)	-	-	-	-	.161 (.022)
LogSum (Nest 4)	-	-	-	-	.010 (.000)

Table 5 (b) - Model Selection

Within Sample					
RMSE	.430	.398	.373	.352	.298
GMM Objective Function	-	3.445 e-003	3.236 e-003	2.467 e-003	2.145 e-003
Out of Sample Predictions					
RMSE	.245	.221	.214	.174	.165

Notes

- The Austin market contains 23 properties; therefore [24*(24-1)/2]-1 parameters of the covariance matrix are uniquely identified.
- In case of homogenous GNL and Heterogeneous GNL, the nest allocation parameters are estimated for each property-nest pair.
- The GNL results are only for the best-fitting GNL model (i.e. Model 6). Results for other GNL models are available on request.

Table 6 - Estimated Allocation Parameters

Property	Nest 1 Parameter (Std. Err.)	Nest 2 Parameter (Std. Err.)	Nest 3 Parameter (Std. Err.)	Nest 4 Parameter (Std. Err.)
Courtyard, Full Service	0^f	1^f	0^f	0^f
DoubleTree, Full Service	0^f	0^f	0^f	1^f
Embassy Suites, Extended Service	0^f	0^f	1^f	0^f
Embassy Suites, Extended Service	0^f	0^f	0^f	1^f
Four Seasons, Full Service	0^f	.831 (.011)	(1-.831)	0^f
Hampton Inn, Limited Service	0^f	0^f	0^f	1^f
Hawthorn, Extended Service	0^f	.382(.301) ^{ns}	(1-.382)	0^f
Holiday Inn, Full Service	.712(.084)	(1-.712)	0^f	0^f
Holiday Inn, Full Service	.689 (.017)	0^f	(1-.689)	0^f
Holiday Inn, Full Service	0^f	0^f	0^f	1^f
Hyatt, Full Service	0^f	.739 (.003)	(1-.739)	0^f
La Quinta, Limited Service	.422 (.024)	(1-.422)	0^f	0^f
La Quinta, Limited Service	0^f	.589 (.003)	(1-.589)	0^f
La Quinta, Limited Service	0^f	.641 (.000)	(1-.641)	0^f
La Quinta, Limited Service	0^f	0^f	.408 (.000)	(1-.408)
Marriott, Full Service	0^f	.728 (.000)	(1-.728)	0^f
Motel 6, Limited Service	0^f	.605 (.071)	(1-.605)	0^f
Motel 6, Limited Service)	0^f	0^f	0^f	1^f
Omni, Full Service	.746 (.000)	0^f	0^f	(1-.746)
Omni, Full Service	0^f	.501 (.009)	(1-.501)	0^f
Ramada, Limited Service	0^f	.666 (.045)	0^f	(1-.666)
Red Lion, Full Service	0^f	1^f	0^f	0^f
Super8, Limited Service	0^f	.704 (.000)	(1-.704)	0^f

Note – Superscript 'f' denotes fixed values. Since none of the properties overlapped across more than two nests, for each overlapping property we only had to estimate one nest-specific allocation parameter. The allocation in the other nest is simply 1 minus the estimated allocation parameter in the other nest.

Table 7 – Full Model Nest Shares

	Nest 1		Nest 2		Nest 3		Nest 4	
Service Sector	N	Market Share Average (Std Dev)	N	Market Share Average (Std Dev)	N	Market Share Average (Std Dev)	N	Market Share Average (Std Dev)
Extended Stay	0		166	0.063 (0.017)	166	0.063 (0.017)	0	
Full Service	166	0.049 (0.009)	166	0.229 (0.056)	166	0.252 (0.064)	166	0.090 (0.021)
Limited Service	166	0.011 (0.003)	166	0.058 (0.016)	166	0.086 (0.027)	166	0.039 (0.013)

Table 8- Average Own and Cross Price Elasticities

Model	Average Own Price Elasticities	Range Of Own Price Elasticities	Average Cross Price Elasticities	Range Of Cross Price Elasticities
Aggregate Logit	-3.236	2.888	.782	.351
Aggregate Probit	-3.671	3.978	.845	.488
Aggregate GNL	-4.222	4.771	.993	.636

Note – The elasticities were estimated at the property level. We only report average measures in the table above.

Table 9 – Out-of-Sample Market Share Predictions

	Nest 1 Market Share		Nest 2 Market Share		Nest 3 Market Share		Nest 4 Market Share	
Sector	O* Mean (Std. Dev)	P** Mean (Std Dev)	O* Mean (Std. Dev)	P** Mean (Std Dev)	O* Mean (Std. Dev)	P** Mean (Std Dev)	O* Mean (Std. Dev)	P** Mean (Std Dev)
Extended Stay			0.063 (0.017)	.058 (.014)	0.063 (0.017)	.058 (.023)		
Full Service	0.049 (0.009)	.052 (.013)	0.229 (0.056)	.049 (.051)	0.252 (0.064)	.026 (.048)	0.090 (0.021)	.131 (.026)
Limited Service	0.011 (0.003)	.013 (.001)	0.058 (0.016)	.061 (.018)	0.086 (0.027)	.091 (.031)	0.039 (0.013)	.043 (.009)

Notes:
O* = observed
P** = predicted

Table 10 – Average Supply-Side Estimates

Parameters	Mean Estimates (Std Error)
Dist. to CBD ρ_1	-1.008 (0.001)
Dist. to CBD^2 ρ_2	0.011 (0.014)
Dist. to Airport ρ_3	2.980 (0.114)
Dist. to Airport^2 ρ_4	0.117 (0.023)
Electricity Utilities	0.118 (0.002)
Services	1.614 (.029)

Our Mission

The Zyman Institute of Brand Science pursues the advancement of brand-driven business performance.

Founded in 2004, the Zyman Institute of Brand Science is the definitive source for cutting-edge knowledge and thinking about brands. The Institute is an independently managed organization within the Goizueta Business School at Emory University.

Who we work with

The Institute works with top management teams to solve pressing issues in brand strategy.

What we do

The Institute's lauded scholars lead the development of new knowledge and insights for building, maintaining, enhancing, and revitalizing persistently profitable brands. We help companies link brand equity to business and shareholder value. We help management make decisions in managing brands to accelerate cash flows, enhance cash flows, and reduce vulnerability and volatility of cash flows, and optimize the long-term value of the organization.

How we do it

The Institute works collaboratively with its sponsors, other universities, research organizations, and its members to devise cross-disciplinary solutions for managing real world problems in brand strategy. We play an integrative role in problem solving, and capability building. We focus on viable actions

Web: zibs.com

About Emory University

Emory University is an internationally recognized institution based in Atlanta that was founded over 150 years ago. Emory sustains top ranked programs in medicine, law, business, theology, nursing, public health, arts & sciences. It supports world-class research across its disciplines.

Web: emory.edu

About Goizueta Business School

The Goizueta Business School was established in 1919 and is named for Roberto Goizueta, former president of The Coca-Cola Company. The school is considered a top-tier school and is consistently ranked in the top 20. It offers degrees in business including Bachelor, Master and PhD. Its Executive MBA is ranked 10 globally by both BusinessWeek and U.S. News.

Web: goizueta.emory.edu

Contacts:

For further information, please visit www.zibs.com or contact:

Zyman Institute of Brand Science
Goizueta Business School

Emory University
1300 Clifton Road
Atlanta, Georgia 30322 USA
Main: +1 (404) 727-9172
Fax: +1 (404) 727-1082

Web: www.zibs.com

Raj Srivastava
Executive Director
Professor of Marketing
Phone: +1 (404) 727-4858
Email: raj@zibs.com

Academic Facing Contacts

Sundar Bharadwaj
Director of Academic Programs
Professor of Marketing
Phone: 404-727-2646
Email: sundar@zibs.com

Doug Bowman
Director of Academic Programs
Professor of Marketing
Phone: 404-727-4613
Email: doug@zibs.com

Sriram Venkataraman
Director, Analytics
Assistant Professor of Marketing
Phone: +1 (404) 727-5275
Email: sriram@zibs.com

Ashish Sood
Director, Analytics
Assistant Professor of Marketing
Phone: +1 (404) 727-4226
Email: ashish@zibs.com

Business Facing Contacts

Greg Thomas
Director of Research and Programs
Phone : +1 (404) 727-4613
Email: greg@zibs.com

Dana Page
Program Manager
Phone: +1 (404) 727-1078
Email: dana@zibs.com

Reshma Shah
Director of Business Relations
Assistant Professor of Marketing
Phone: +1 (404) 727-6302
Email: reshma@zibs.com

Susan Hogan
Director of Business Relations
Assistant Professor of Marketing
Phone: +1 (404) 727-5516
Email: reshma@zibs.com

Leanne Fesenmeyer
Director of Alliances
Phone: +1 (770) 817-4141
Email: leanne@zibs.com