ONLINE APPENDIX- SUPPLEMENTARY FIGURES/TABLES

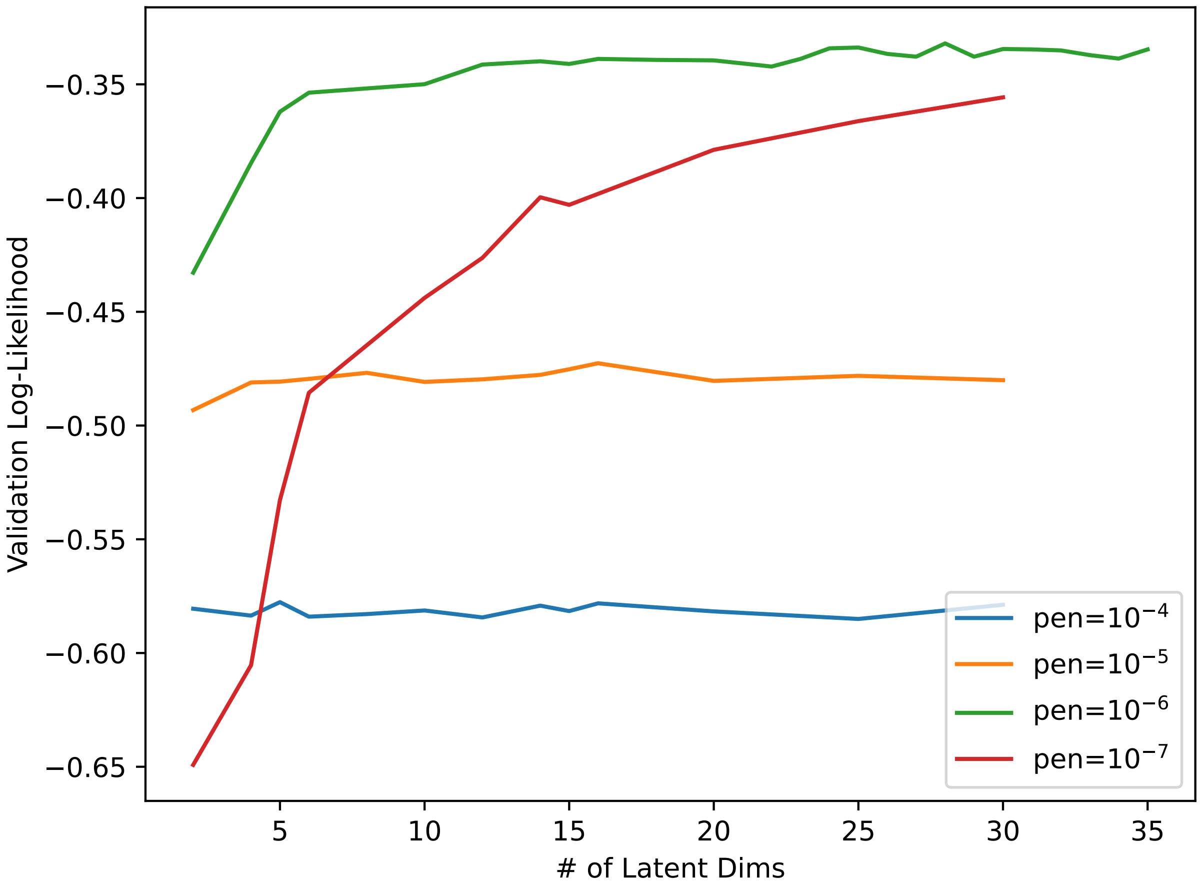


Figure A1: Validation Log-likelihood for Tuning # of Latent Dimensions

	Latent Characteristics			Observable Characteristics		
	# Iterations	Nodes/Layer	Regularization	# Iterations	Nodes/Layer	Regularization
Latitude-Longitude	8325	10	0.0001	4856	5	0.0046
Independent Hotel	430	5	0.0001	735	40	0.0000
Near Airport	180	40	0.0001	456	20	0.0000
Hotel Size Category	160	30	0.0046	873	40	0.2154

Table A1: Hyperparameters Chosen for Predicting Observable Characteristics

Characteristics Preferences	Observables			Price & Latent		
		Normal	Search		Normal	Search
	Price	0.0129	0.564	Price	0.00991	0.978
		(0.00116)	(0.798)		(0.00415)	(0.264)
	Latitude	3.49e-11	0.00434	Latent Dimension 1	-3.02e-09	-1.12
		(4.35e-09)	(0.926)		(1.41e-09)	(1.3)
	Longitude	-2.5e-11	-0.000721	Latent Dimension 2	1.29e-08	-0.808
		(6.14e-09)	(0.539)		(3.25e-08)	(2.34)
	Log(Meeting Space+1)	7.31e-11	0.0284	Latent Dimension 3	-1.05e-09	-0.359
		(9.27e-09)	(21.8)		(2.06e-09)	(2.45)
	Price Segment	-3.13e-11	0.0438	Latent Dimension 4	0.108	-0.0928
		(3.73e-10)	(81.1)		(1.37)	(3.18)
	Hotel Size Category	-1.19e-10	0.0852	Latent Dimension 5	8.1e-10	0.00602
		(2.72e-08)	(73.8)		(1.28e-08)	(4.13)
	Independent Hotel	3.91e-10	0.372	Latent Dimension 6	-3.73e-09	0.355
		(1.15e-08)	(49.3)		(1.54e-08)	(2.08)
	Location = Suburban	-1.22e-10	-1.58	Latent Dimension 7	-4.06e-10	-0.00261
		(2.96e-09)	(221)		(1.16e-09)	(5.76)
	Location = Small Metro/Town	2.49e-10	-0.868	Latent Dimension 8	1.11e-09	-0.386
		(9.36e-09)	(304)		(7.46e-09)	(3.43)
	Location = Urban	-0.23	-33.3	Latent Dimension 9	1.34e-10	0.323
		(25.4)	(48.3)		(1.12e-09)	(0.755)
	Location = Resort	-3.65e-10	-2.38	Latent Dimension 10	-1.26e-10	-0.0616
		(2.53e-08)	(173)		(1.91e-10)	(2.35)
	Location = Airport	-5.62e-10	-17.1	Latent Dimension 11	1.33e-09	-1.11
		(1.63e-08)	(92.8)		(1.41e-09)	(1.19)
	Location = Interstate	-2.18e-10	-31	Latent Dimension 12	-5.56e-10	-0.0416
		(7.28e-09)	(153)		(7.57e-10)	(3.05)

Table A2: Estimated Consumer Heterogeneity Demand Parameters

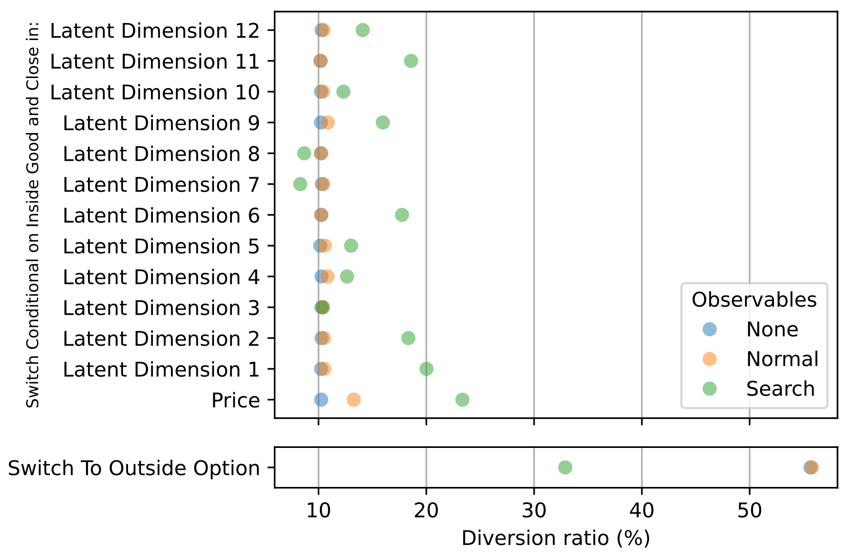
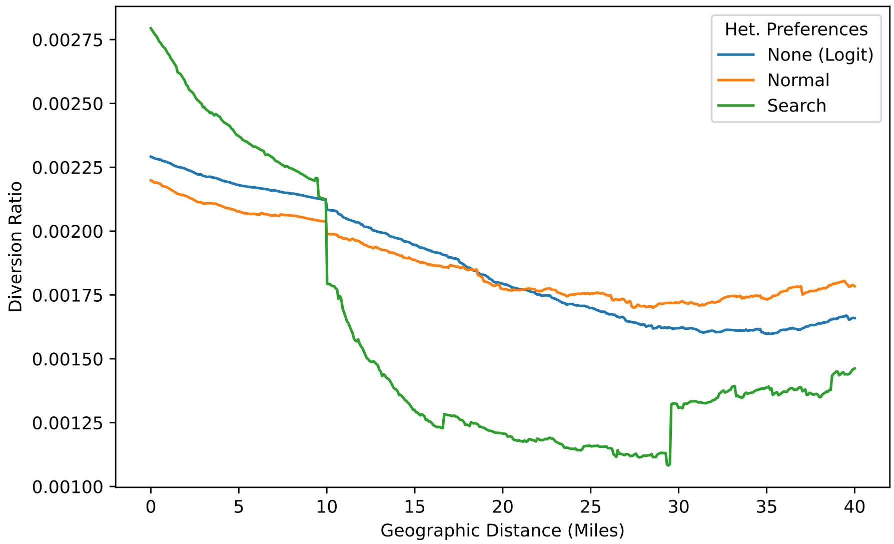
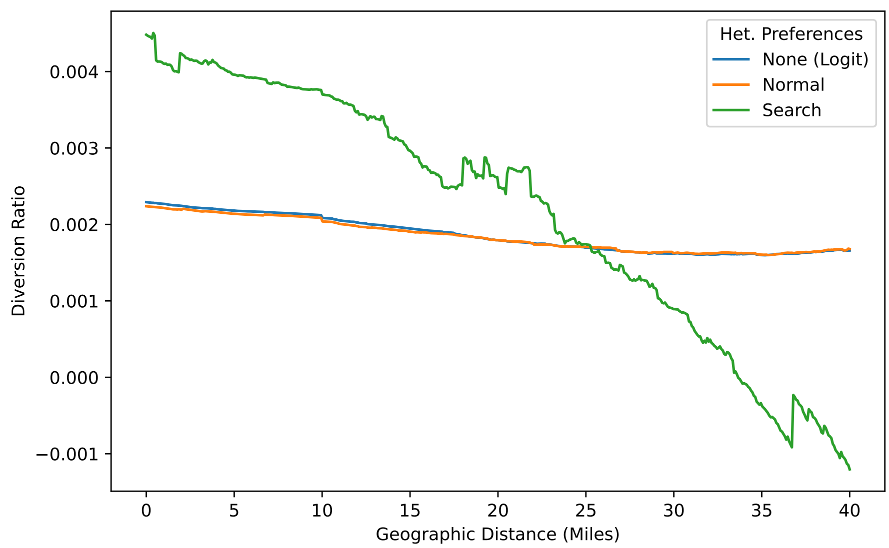


Figure A2: Diversion Ratios Across Demand Models with Latent Characteristics



(a) Observable Characteristics



(b) Latent Search Characteristics

Figure A3: Diversion Ratio As a Function of Distance, by Preference Specification

Figure shows the estimated mean diversion ratio from a kernel regression (uniform kernel, bandwidth = 10 miles) between hotels from our estimated demand model, as a function of the geographical distance between hotels. Het. Preferences denote whether there is no consumer heterogeneity (None), normally distributed heterogeneity (Normal), or heterogeneity based in preferences learned from search data (Search). Panel (a) plots comparisons across models using observables characteristics, while Panel (b) plots comparisons across models using only the price and the learned latent characteristics learned from search data.

A Understanding the Characteristics and Preference Space: Supplementary Evidence

In this section, we provide supplementary evidence to Section 4 in the main text that our recovered latent preferences and product characteristics from search data yield a sensible characterization of the market for hotels.

We can assess whether our estimated preferences are sensible by examining the correlation within-consumer of their preferences over various characteristics of hotels. In Figures A4 and A5, we plot the correlation matrix of preferences over characteristics for each of the two estimated models. In Figure A4, where we estimate preference heterogeneity only over observable characteristics, we find that preferences for location types are strongly negatively correlated, while tastes for meeting space and size category (number of rooms) are positively correlated. The former correlation represents a preference for the overall spatial location of hotels, while the latter represents an overall taste for large hotels both in guest capacity and ability to host large business conferences.

Second, we use an off-the-shelf method for classifying products, a top-k classifier, to see if hotels that are close in the latent attribute space also share observable characteristics. Specifically, given a candidate hotel j, we examine the characteristics of hotels that are the top-10 closest in euclidean distance to j in the latent characteristic space:

$$d(j,k) = ||\gamma_j - \gamma_k||_2 = \sqrt{\sum_{l=1}^{L} (\gamma_{j,l} - \gamma_{k,l})^2} \tag{A1}$$

where $\vec{\gamma}_j$ is the 12-d vector of latent characteristics learned in the latent characteristics only model. We do this for all hotels k within the same STR-defined geographical market as j. We then see what share of hotels classified as close according to the top-10 classifier share the same brand, management company, ownership company, Parent company, and how close they are in miles to the candidate hotel j. We perform this exercise for all hotels in our dataset, then average across the classifiers for all J hotels in our dataset to see if, on average, the hotels close in unobserved characteristics share observable attributes more often than those that are classified as far away in the latent characteristic space.

Table A3 plots the result of the top-10 classifier for our targeted characteristics. Observations differ across target variables because not all hotels have a management, owner, parent company, or brand (e.g. independent hotels have no parent company). We also use as a comparison the results from a top-10 classifier based on the observable characteristics of hotels, to benchmark our results.[1] We find that in general, hotels closer in latent characteristic are more likely to share supply-side characteristics such as brand/chain affiliation, compared to a random hotel in the same market. At the same time, the classifier based on observable characteristics performs better for all characteristics, except distance. To an extent, this is not surprising since some characteristics (such as class/price segment) are defined at the brand level, and many hotel chains implement uniform characteristics across their locations.

[1] We use the Mahalanobis distance on observable characteristics to standardize the scale of each characteristic. Because latitude and longitude is included in our input observable characteristics, we exclude it when classifying for the distance metric.

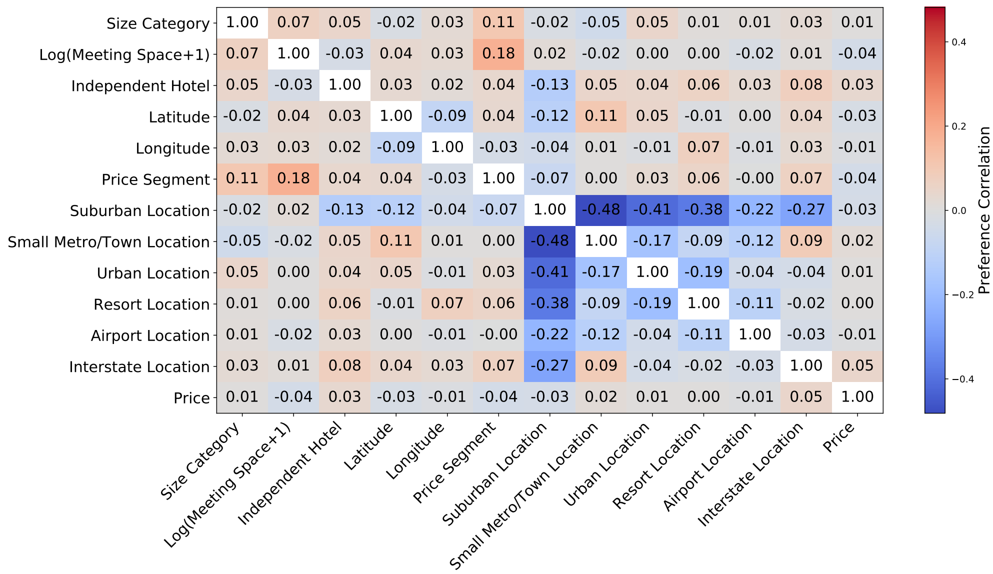


Figure A4: Correlation Matrix of User Search Preferences: Observable Characteristics Model

	Baseline All Hotels In Market	Top-10 Classifier Based on Characteristics Latent	Top-10 Classifier Based on Characteristics Observations	# Observations
Pr(Same Brand)	0.029	0.039	0.093	3438
Pr(Same Mgmt Company)	0.027	0.051	0.098	1690
Pr(Same Owner)	0.037	0.065	0.117	1202
Pr(Same Parent Company)	0.133	0.164	0.251	3438
Distance (Miles)	51.413	30.265	37.261	4218

Table A3: Predictive Performance of Top-10 Classifier based on Distance in Latent and Observable Characteristic Space

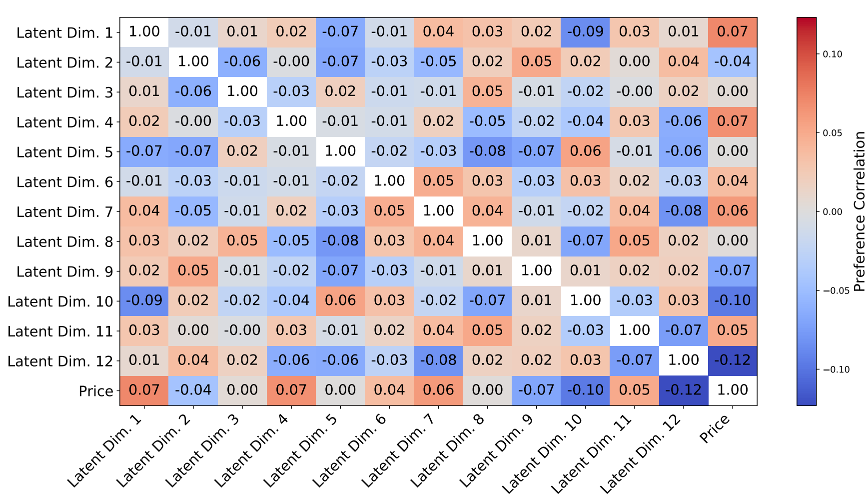


Figure A5: Correlation Matrix of User Search Preferences: Latent Characteristics Model

B Merger Impact on Hotel Conduct

In order to evaluate how effective our proposed demand model is in capturing actual substitution patterns, we exploit a large merger that occurred during our sample period in the hotel industry that induces plausibly exogenous price changes. On November 16, 2015, Marriott International Inc announced that it would be acquiring the Starwood Hotels company. The merger was completed on September 23, 2016 (Dogru, Erdogan, and Kizildag 2018). After the merger completed, Marriott International became the largest hotel chain in the world. After this transaction occurred, prices in markets with high concentration of Starwood / Marriott hotels changed noticeably. Recall that the brand affiliations of each hotel in our STR dataset is an anonymized ID, so we cannot observe which hotels in our transaction data were Marriott or Starwood affiliates before the merger. Through further coordination with STR, we were able to obtain counts of each hotel brand within each geographical market and class segment, as of December 2015. We use this data to construct an "exposure index" to the merger, $Pr_j(\text{Starwood or Marriott}|\text{Market}_j, \text{Class}_j)$, and measure the effect of exposure to the merger on post-merger prices. The rationale behind this exposure index is that if Marriott/Starwood hotels changed conduct in price-setting after a merger, then hotels belonging to the same geographical market/ class segment as Marriott/Starwood hotels will also respond and change their price-setting behavior. The exposure index then captures both "direct effects" of Marriott-Starwood hotels changing their price behavior due to backend changes in costs and increased market power, as well as "indirect effects" of competing hotels responding to the new price-setting behavior of Marriott/Starwood hotels.

We estimate the effect of exposure to the merger via the following event study specification:

$$\log(p_{j,t}) = \alpha_{j,\text{month}(t)} + \delta_t + \sum_{q=2013Q1}^{2019Q1} \beta_q Pr_j(\text{Starwood or Marriott}|\text{Market}_j, \text{Class}_j) + \epsilon_{j,t} \quad \text{(A2)}$$

where $\alpha_{j,\text{month}(t)}$ denotes hotel×month-of-year fixed effects , to capture time-invariant differences in hotel prices as well as seasonalities in pricing structure, δ_t is a geographical market $\times$ month $\times$ year fixed effect, to capture common demand shocks occurring in each market-month-year, and β_s is the effect of the exposure index in quarter q on hotel j. We aggregate the event-study specifications to the quarterly-level due to power concerns given the large number of fixed effects. The control group in this event study are hotels in the same market as Marriott-Starwood hotels but a different class segment. Because consumers have differential demand, hotels belonging to a different class are not as exposed to the changed pricing behavior of Marriott-Starwood hotels after the merger.

Figure A6 plots the estimated event study. The dashed red line represents the time of the merger announcement, while the solid red line represents the date the merger was completed. The blue dots represent the estimated coefficients of the above specification, while the orange dots replace the market-month-year fixed effects with market-month-year-location type (e.g. urban vs suburban hotels in Los Angeles in May 2018) fixed effects.

In general, there do not appear to be strong pre-trends before the merger announcement. We see large price *decreases* following the completion of the merger, which is consistent with discussions surrounding the merger of cost reductions via centralization of sales and customer service operations between the acquiring and target firms (Dogru, Erdogan, and Kizildag 2018).

We use this event study as plausible evidence that the Marriott-Starwood merger led to exogenous price changes by hotels in markets with high exposure to Marriott and Starwood hotels, independent of demand fluctuations. Therefore, measuring the ability of our demand models to accurately predict demand post-merger may serve as a test to whether a demand model augmented with search data may perform “better” in predicting substitution patterns after exogenous price changes.

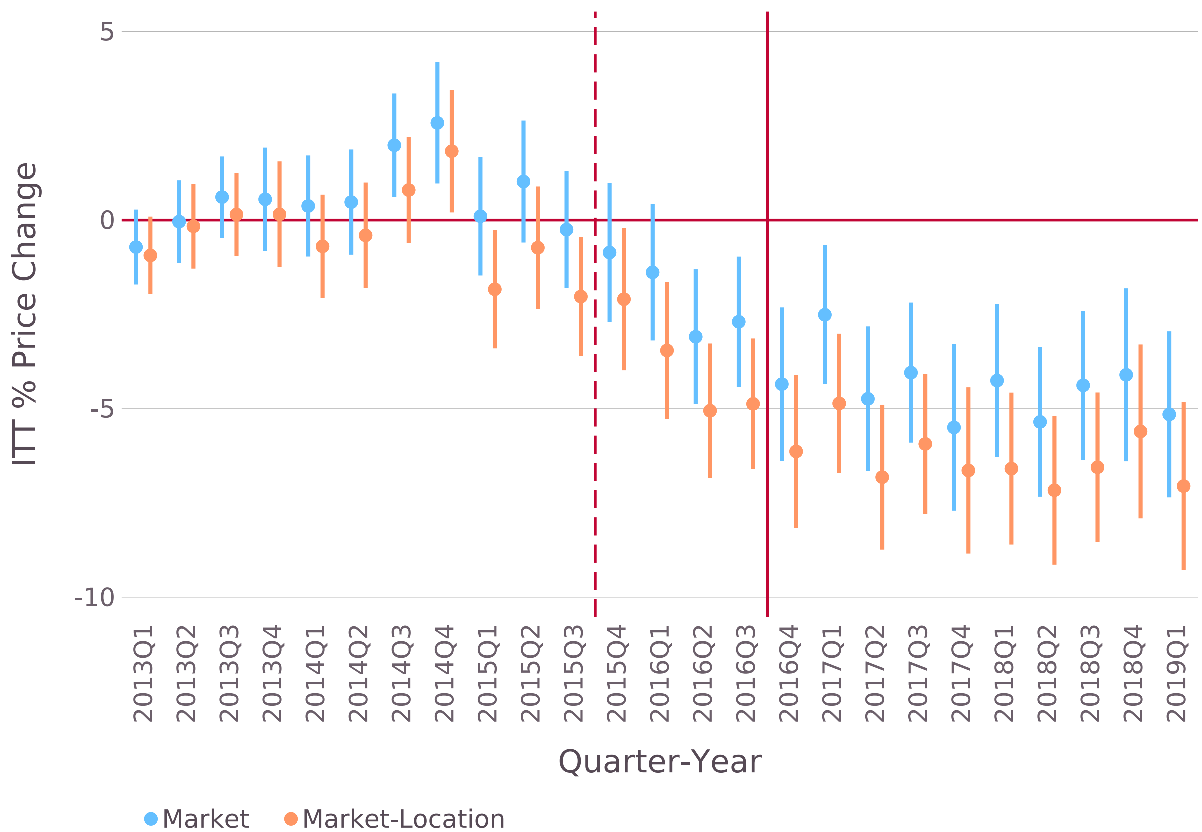


Figure A6: Effect of Marriott-Starwood Merger on Prices in Regions with High Marriott-Starwood Presence

C Derivation of Choice Market Shares

In this section, we derive the choice probabilities implied by the search and choice model presented in Section 3.2 in the main text. Recall that expected utility is defined as

$$E[u_{i,j,t}] = \bar{\delta}_{i,j,t} = \delta_{j,t} - \alpha_i p_{j,t} + \vec{\beta}_i \vec{X}_j + \zeta_{i,j,t} \tag{A3}$$

Since realized utility is $u_{i,j,t} = E[u_{i,j,t}] + (1 - \sigma_\zeta)\tilde{\epsilon}_{i,j,t}$, we can divide by $(1 - \sigma_\zeta)$ to obtain a cardinally equivalent expression of utility as $\tilde{u}_{i,j,t} = \bar{\delta}_{i,j,t}/(1 - \sigma_\zeta) + \tilde{\epsilon}_{i,j,t}$, so that utility is expressed as a location shifter plus an i.i.d. extreme value type 1 (T1EV) random variable with scale 1 and location 0. From here, we follow the notation of (Moraga-González, Sándor, and Wildenbeest 2023). Under this re-scaled utility measure, The reservation value under simultaneous search is defined as:

$$\tilde{r}_{i,j,t} = \bar{\delta}_{i,j,t}/(1 - \sigma_\zeta) + H_0^{-1}(\tilde{c}_{i,j,t}) \tag{A4}$$

Where $c_{i,j,t}$ is the user-specific search cost. Let $w_{i,j} = \min(\tilde{r}_{i,j}, \tilde{u}_{i,j})$. This is the minimum of the reservation value associated with searching product j, and actual utility derived from the product. Moraga-González, Sándor, and Wildenbeest (2023) show that under the sequential search environment, the optimal search strategy is to choose the product with the highest $w_{i,j}$.[2]

We assume in the main section of the paper that the distribution of search costs same distribution as Moraga-González, Sándor, and Wildenbeest (2023), therefore by Proposition 1 of in that paper, the distribution of $\tilde{w}_{i,j}$ follows a Gumbel distribution with location parameter $(\bar{\delta}_{i,j,t} - \mu)/(1 - \sigma_\zeta)$ and scale 1.[3]

[2]Formally, in (Moraga-González, Sándor, and Wildenbeest 2023), there are multiple products within each "search group", so they find it optimal to search the group with the highest w, then choose the best product in that group. In our context, we assume all consumers pay an equal search cost to search one product, so this second stage choice is degenerate, as there is one product in each group.

[3]We express the search location parameter μ in terms of the original utility space.

Since a T1EV variable with location δ and scale σ, multiplied by a constant c, is T1EV with location $c \cdot \delta$ and scale $c \cdot \sigma$, it follows that the random variable $(1-\sigma_\zeta)w_{i,j}$ is distributed according to a T1EV distribution with location $(\bar{\delta}_{i,j,t} - \mu)$ and scale $(1 - \sigma_\zeta)$. Since the mapping $f(w) = (1 - \sigma_\zeta)w$ is monotonic, it must also be the case that the optimal search strategy is to choose the product with the highest $(1 - \sigma_\zeta)w_{i,j}$.

Finally, since constant terms are additively separable from a T1EV distribution, we can express $(1 - \sigma_\zeta)w_{i,j}$ as

$$(1 - \sigma_\zeta)w_{i,j} \sim \left(\delta_{j,t} - \alpha_i p_{j,t} + \vec{\beta}_i \vec{X}_j - \mu\right) + \zeta_{i,j,t} + (1 - \sigma_\zeta)\eta_{i,j,t} \tag{A5}$$

Where $\eta_{i,j,t}$ is a T1EV random variable with location 0 and scale 1. Because ζ is a conjugate to the T1EV by Theorem 2.1 of (Cardell 1997), The distribution of $\zeta_{i,j,t} + (1 - \sigma_\zeta)\eta_{i,j,t}$ is also Gumbel with location 0 and scale 1, which implies that $(1 - \sigma_\zeta)w_{i,j}$ is distributed as T1EV with location $\delta_{j,t} - \alpha_i p_{j,t} + \vec{\beta}_i \vec{X}_j - \mu$ and scale 1. The probability that any particular $(1 - \sigma_\zeta)w_{i,j}$ is largest is given by the max-stability property of the T1EV distribution.

$$Pr(i \text{ choose } j) = Pr((1-\sigma_\zeta)w_{i,j} > \max_{k \neq j}(1-\sigma_\zeta)w_{i,k}) = \frac{\exp(\delta_{j,t} - \alpha_i p_{j,t} + \vec{\beta}_i \vec{X}_j - \mu)}{1 + \sum_{k \in J_t} \exp(\delta_{k,t} - \alpha_i p_{k,t} + \vec{\beta}_i \vec{X}_k - \mu)} \tag{A6}$$

We can then integrate over the distribution of consuemr preferences α_i, β_i to get market-level shares, which concludes the derivation.

D Derivation of Search Likelihood

Consider the normalized utility presented in Section C, where we divide utility by $(1 - \sigma_\zeta)$. Let $\delta_{i,j,t} = \delta_{j,t} - \alpha_i p_{j,t} + \beta_i X_j$, so that $E[\tilde{u}_{i,j,t}] = (\delta_{i,j,t} + \zeta_{i,j,t})/(1 - \sigma_\zeta)$ and reservation values are given by Equation A4. Under simultaneous search, the optimal search strategy is to search products in decreasing order of their reservation value. Our inequalities derived for search behavior require us to compare reservation values across searched and unsearch products; therefore, the relevant probability to consider is:

$$Pr(r_{i,j} > r_{i,k}) = Pr(r_{i,j} - r_{i,k} > 0) = Pr(-\frac{\Delta\zeta_{i,j,k}}{(1-\sigma_\zeta)} - \Delta H^{-1}_{i,j,k} < \frac{\Delta\delta_{i,j,k}}{(1-\sigma_\zeta)}) \quad \text{(A7)}$$

Where $\Delta\delta_{i,j,k} = \delta_{i,j,t} - \delta_{i,k,t}$, $\Delta\zeta_{i,j,k} = \zeta_{i,j,t} - \zeta_{i,k,t}$, and $\Delta H^{-1}_{i,j,k} = H_0^{-1}(\tilde{c}_{i,j,t}) - H_0^{-1}(\tilde{c}_{i,k,t})$. Given a guess of the latent parameters, The first component, $\Delta\delta_{i,j,k}$ is constant and serves to simply shift the CDF threshold. So to understand the global properties of the distribution, we ignore it. According to Cardell (1997), the difference of two gumbel conjugate with scale σ_ζ is logistic with scale σ_ζ. Therefore, the distribution of $-\Delta\zeta_{i,j,k}/(1-\sigma_\zeta)$ is simply a logistic distirbution with location 0 and scale $\sigma_\zeta/(1-\sigma_\zeta)$. The distribution of $-\Delta H^{-1}_{i,j,k}$ is unknown. In this appendix, we will simulate from it, and evalulate whether the logistic approximation for reservation values (which would be exact if $\Delta H^{-1}_{i,j,k}$ is zero) is a valid approximation.

The search cost distribution used in Moraga-González, Sándor, and Wildenbeest (2023) sets search costs equal to zero with probability $\exp(-\mu)$. This presents a problem, as $H_0^{-1}(0)$ is undefined, and implies a reservation value of $r_{i,j} = \infty$ (e.g. the product is always searched). This also implies that reservation utilities are equal for two products with zero search costs (e.g. they are both always searched). This implies the following probability that one product is searched and the other is not:

$$\begin{aligned}
&Pr(\text{search } j, \text{do not search } k) = Pr(r_{i,j,t} > r_{i,k,t}) \quad \text{(A8)}\\
&= Pr(c_{i,j,t} = 0, c_{i,k,t} > 0) + Pr(c_{i,j}, c_{i,k} > 0) \cdot Pr(r_{i,j,t} > r_{i,k,t} | c_{i,j,t}, c_{i,k,t} > 0)\\
&= Pr(c_{i,j,t} = 0) \cdot Pr(c_{i,k,t} > 0) + Pr(c_{i,j,t} > 0) \cdot Pr(c_{i,k,t} > 0) \cdot Pr(r_{i,j,t} > r_{i,k,t} | c_{i,j,t} > 0, c_{i,k,t} > 0)\\
&= \exp(-\mu)(1-\exp(-\mu)) + (1-\exp(-\mu))^2 \cdot Pr(-\frac{\Delta\zeta_{i,j,k}}{1-\sigma_\zeta} - \Delta H^{-1}_{i,j,k} < \frac{\Delta\delta_{i,j,k}}{1-\sigma_\zeta} | c_{i,j,t} > 0, c_{i,k,t} > 0)
\end{aligned}$$

Where the second equality follows from the fact that search costs are independent. Notice that μ is not a parameter of our search model, so it is fixed during optimization and does not impact the likelihood. Therefore, maximizing the likelihood of our search model is equivalent to maximizing:

$$Pr(-\frac{\Delta\zeta_{i,j,k}}{1-\sigma_\zeta} - \Delta H^{-1}_{i,j,k} < \frac{\Delta\delta_{i,j,k}}{1-\sigma_\zeta} | c_{i,j,t} > 0, c_{i,k,t} > 0) \tag{A9}$$

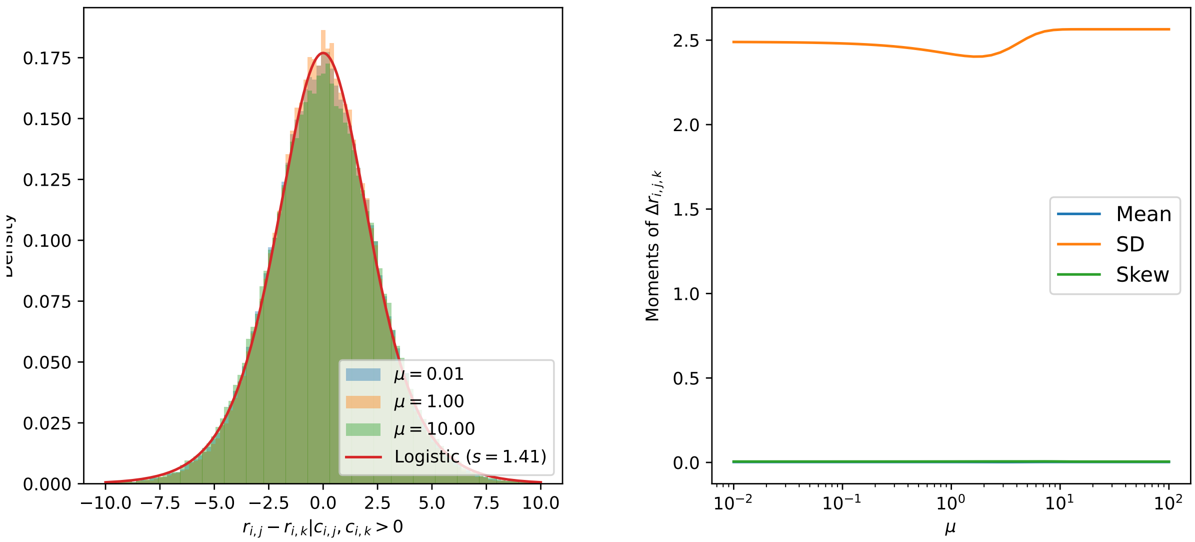
Since ζ is independent of search costs, its distribution is unchanged, conditional on $c_{i,j,t} > 0$. The CDF of $H_0^{-1}(c_{i,j})$ conditional on $c_{i,j} > 0$ is:

$$\begin{aligned} F_H(z|\mu) = Pr(H_0^{-1}(c) \leq z|\mu, c > 0) &= Pr(c \geq H_0(z)|\mu, c > 0) \\ &= 1 - Pr(c \leq H_0(z)|\mu, c > 0) \\ &= 1 - \frac{S(H_0(z)|\mu) - S(0|\mu)}{1 - S(0|\mu)} \\ &= \frac{F_\epsilon(z+\mu) - F_\epsilon(z)}{(1 - F_\epsilon(z))(1 - \exp(-\mu))} \end{aligned} \tag{A10}$$

Where $F_\epsilon(x) = \exp(-\exp(-x))$ is the CDF of a standard T1EV distribution. This suggests the search cost component of the reservation values, $H_0^{-1}(c_{i,j,t})$, is similar in shape to a (scaled) truncated T1EV distribution. We sample from this distribution by taking a uniform sample $U \sim [0, 1]$, then finding the nonlinear root z satisfies $F_H(z|\mu) = U$. We then use these simulated draws to sample from the distribution of reservation values (conditional on positive search costs).

Figure A7 displays simulations of reservation values across values of μ, the location of search costs. For this exercise, we normalize the scale of $\Delta\zeta$ to be 1 (e.g. $\sigma_\zeta = .5$), so that values of μ, which index the search cost component of reservation values, are relative to the dispersion of ζ. In Panel (a), we plot the distribution of $\Delta r_{i,j,k}$, the difference in reservation values across two products, conditional on positive search costs, for three values of μ (100,000 draws each). For comparison, we also show the PDF of a logistic distribution with identical variance with $\mu = 10$. The logistic approximation is quite accurate, irrespective of the choice of μ.

In Panel (b), we show the moments of the simulated $\Delta r_{i,j,k}$, across values of μ. The mean and skewness are close to zero across the parameter space. The variance increases



(a) Simulated $\Delta r_{i,j,k}$ (b) Moments of Distribution across μ

Figure A7: Simulated search cost Components and reservation values

siightly $\mu = 5$, but is mostly constant, even up to implausibly large values of μ such as 100. It appears that a logistic distribution with scale $\approx s = 1.4$ is a good approximation to the distribution of $\Delta r_{i,j,k}$, irrespective of the value of μ. This suggests that our logistic approximation will be accurate even for extreme values of search costs, and there will be little loss of information in our usage of a quasi likelihood.

References

Cardell, N. S. (1997): "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory*, 13(2), 185–213.

Dogru, T., A. Erdogan, and M. Kizildag (2018): "Marriott Starwood merger: what did we learn from a financial standpoint?," *Journal of Hospitality and Tourism Insights*.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2023): "Consumer search and prices in the automobile market," *The Review of Economic Studies*, 90(3), 1394–1440.