Title
Essays on Quantitative Marketing

Author
Zhang, Fan

Publication Date
2023

Peer reviewed|Thesis/dissertation

Essays on Quantitative Marketing

By

Fan Zhang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Przemyslaw Jeziorski, Co-Chair
Professor J.Miguel Villas-Boas, Co-Chair
Professor Kei Kawai

Spring 2023

Essays on Quantitative Marketing

Abstract

Essays on Quantitative Marketing

by

Fan Zhang

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Przemyslaw Jeziorski, Co-Chair
Professor J.Miguel Villas-Boas, Co-Chair

The dissertation has three chapters. In the first chapter, I estimate consumer search cost with purchase outcome data. I analyze a model where consumers search sequentially for the best option based on advertised price and partial product information. I verify the model's ability to recover the structural parameters by a numerical experiment. Then using Nielsen Consumer Panel Data, I estimate the product-specific search costs for the top five brands in 32oz refrigerated yogurt market. I find that after controlling for the prices, the private label brand has the lowest search cost. Counterfactual analysis shows that eliminating the search cost increases overall purchase and decreases the price sensitivity. Incorrectly ignoring search frictions leads to an overestimation of own-price elasticity.

In the second chapter, we state conditions under which choice data suffices to identify preferences when consumers may not be fully informed about attributes of goods. Our approach can be used to test for full information, to forecast how consumers will respond to information, and to conduct welfare analysis when consumers are imperfectly informed. In a lab experiment, we successfully forecast the response to new information when consumers engage in costly search. In data from Expedia, our method identifies which attribute was not immediately visible to consumers in search results, and we then use the model to compute the value of additional information.

In the third chapter, we study consumers' variety-seeking preferences and explore their implications for targeted marketing using proprietary data from a food delivery platform. We document that a substantial fraction of consumers have variety-seeking preferences. Consumers, on average, are willing to pay 20% more to switch to a different seller. In the counterfactual analysis, we find that optimizing rankings by taking into account variety-seeking preferences increases revenue, consumer welfare, and purchase probability. Furthermore, we find that consumers' variety-seeking preferences soften price competition. Optimal targeted pricing implies an increase in prices for rival sellers' consumers and a decrease in prices for the sellers' own consumers.

Contents

Acknowledgment

I thank my advisors Przemyslaw Jeziorski, J.Miguel Villas-Boas, and Kei Kawai for their continuous support, encouragement, and guidance throughout my PhD years. I also thank my coauthors Carol Hengheng Lu, Zhijie Lin, Giovanni Compiani, and Jason Abaluck for their invaluable collaboration and insightful contributions. This work also benefited from helpful discussions from the marketing and economic groups at UC Berkeley. I am grateful to my parents and friends for their love, accompany, and support.

Chapter 1

Estimation of Consumer Search Cost with Purchase Outcome Data

1.1 Introduction

The assumption that consumers have full information about all available options is usually violated in the real world. In such cases quantifying consumer's search cost is important for recovering true consumer preferences, evaluating firm's market power and making pricing and advertising decisions. Existing empirical research on consumer search overwhelmingly use data intensive estimation method, including click stream data (e.g. Bronnenberg et al., 2016; Chen and Yao, 2016), consideration set data (e.g. Kim et al., 2010; Honka, 2014), view-rank data (Kim et al., 2010), or at least long panel data (Seiler, 2013).

When such intensive data is not available, researchers mainly use standard discrete choice differentiated products model that assumes consumers are perfectly aware of their choices. However, ignoring search costs can leads to biased estimation of demand. Dong et al. (2018) shows that ignoring search costs leads to an overestimation of product intercepts by 30%. De los Santos et al. (2012) reports that incorrectly assuming consumer's full information leads to an underestimation of own-price elasticities.

The contribution of this paper is to recover search cost and consumer preference with only purchase outcome data. I study a model in which consumers sequentially search for the best product based on partial product information and advertised prices. I utilize an elegant necessary and sufficient condition that fully summarizes consumer's search outcomes uncovered by Choi et al. (2018a). I apply the condition to discrete choice models to estimate the extent to which search frictions distort consumers' purchase decisions, i.e.,how a consumer's purchase decision under sequential search differs from that under perfect information. I implement a numerical experiment to verify the model's ability to recover the structural parameters from simulated purchase outcome data. The estimates are significant and close to true parameters.

I use Nielsen Consumer Panel Data and focus on the 32 oz refrigerated yogurt market in one retailer. I focus on the 32oz Yogurt market for three reasons. First, most consumers purchase only one brand in one week because of the big size, which makes the market fit into the discrete choice model framework. Second, yogurt products are not storable, which makes it clearer to identify the Not Buy observations. Otherwise the reason why consumers didn't buy the products in a trip during a certain week is ambiguous, i.e., whether it's because they have some storage at home or they choose to not consume the product in this week is uncertain. Third, there are many dimensions of information to search regarding refrigerated yogurt products. There are many different brands, flavors, fat percentages, and calories,etc. The assumption that consumers are aware of all the available products in the market is very likely to be violated.

I find that after controlling for prices, among the top five brands in the market, the private label brand has the lowest search cost. One reason for this can be availability. The private label brand is owned by the retailer and is usually available in all stores under the retailer. The private label brand products are usually placed at the most prominent position on the shelf among products of the same category. The inventory and replacement management is usually more efficient than other brands not owned by the retailer. I also examine the out-of-sample fit of the model. I evaluate the fit in terms of market shares, and find that the model predictions fit the holdout data reasonably well. The largest difference between the data and prediction is within 0.005.

Counterfactual analysis shows that with price being pre-search information, eliminating search cost leads to more purchase overall, and a less price-sensitive market demand. This is intuitive: consumers rely less on their pre-search information, price, as they search more about other product features when search costs decrease, thus less reactive to prices. Incorrectly assuming consumers have full information will lead to an overestimation of own-price elasticities. Note that this is different from the existing literature in which price is the main characteristic to search(Honka, 2014). In that case consumers who sample only a subset of the products are more reactive to price changes as they search more when search costs decrease. The general intuition here is that a higher search cost prevents consumers from making fully informed decisions and makes them rely more on their pre-search information and thus more sensitive to it.

Other Related Literature The most related theoretical work on consumer sequential search is Choi et al. (2018a), in which they utilize Weitzman (1979)'s results on optimal search behavior, including *order of search* and *stopping rule* , to further discover a sufficient and necessary condition, which fully summarize the optimal *search outcome*. In this paper I apply the theoretical result to purchase outcome data to estimate consumer search frictions.

On the empirical side, Abaluck and Adams (2017) also deviate from the existing data intensive estimation method in a different search framework with consideration set models. They prove that utility and consideration set probabilities can be separately identified with only the search outcome data. Their identification result builds on the insight that imperfect consideration breaks the symmetry between cross-characteristic responses.

The economic assumptions behind consumer search model and preference heterogeneity model are fundamentally different, i.e.,whether goods are demanded because they are high-utility or because they are more easily to get known. In this paper I focus on the *estimation* of sequential search cost with only purchase outcome data, instead of trying to formally prove *identification* of search model from preference heterogeneity model with purchase outcome data.

The structure of the rest of the paper is as follows: section 1.2 introduces the model followed by a discussion of the estimation approach and identification. Section 1.3 presents the data, including details of dataset construction and descriptive statistics. Section 1.4 presents the estimation results, out-of-sample fit and counterfactual prediction. Section 1.5 concludes.

1.2 Model

1.2.1 Main assumptions

Before going into the notations, I discuss the main assumptions regarding consumer behavior. In this paper I assume: (i) consumers observe advertised prices before search; (ii) regarding the correlation structures for consumer's known and hidden values besides observed prices, I assume both are independent across products (iii) consumers use a sequential search approach.

Different from traditional consumer search models, in which consumers search to resolve uncertainty regarding prices, I assume prices are observable before search. With the development of online shopping and electronic devices, most big stores have their websites, mobile applications or weekly printed advertisements. The prevalence of such inventions significantly lowered the cost of collecting price information. Now it is common to check prices online and visit stores only to get hands-on information and finalize a purchase. In such important environment, where prices are part of pre-search information, prices affect each seller's demand not only through their effects on consumers' final purchase decisions, but also through their effects on consumer search behavior.

Existing literature adopt different assumptions about the correlation structure for consumers' known and hidden values. For example, Armstrong and Zhou (2011) assume that both are perfectly negatively correlated between the products. In this paper, I follow Choi et al. (2018a) and assume that both are independent.

Besides the sequential search framework I concentrate on in this paper, there is another strand of literature focusing on simultaneous search(Chade and Smith, 2006), where the number of search is predetermined. There have been papers testing which search model is happening in the real world, and the results are mixed Bronnenberg et al. (2016) finds evidence consistent with sequential search with online digital camera search data, whereas De los Santos et al. (2012) and Honka and Chintagunta (2016) find evidence in favor of simultaneous search. I assume consumers search sequentially,i.e., the number of search is not fixed and is affected dynamically by the result of search in the earlier stage.

1.2.2 Formal Setup

Formally, there are J sellers in the market, each indexed by $i = \{1, ..., J\}$, and a unit mass of consumers. Each consumer demands one unit among all products. The sellers simultaneously announce prices. Consumers observe those prices and then search optimally. I denote by $p_i \in R_+$ seller i's price. Let $\mathbf{p}$ denote the price vector for all sellers (i.e., $\mathbf{p} = (p_1, ..., p_J)$).

A consumer's random utility for seller i's product is given by $\tilde{V}_i = V_i + Z_i$. The first component V_i represents the consumer's prior value for product i, while the second component Z_i is the residual part that is revealed to the consumer only when she visits seller i and inspects his product. Let $\mathbf{v} = (v_1, ..., v_J)$ and $\mathbf{z} = (z_1, ..., z_J)$ denote the realization of a consumer's value profile for each component.

The products are horizontally differentiated. I assume that V_i and Z_i are drawn from the distribution functions F_i and G_i, respectively, identically and independently across consumers and products (and independently each other), where both F_i and G_i have full support over the real line and continuously differentiable density f_i and g_i, respectively.

Search is costly, but recall is costless. Specifically, each consumer must visit seller i

and discover her match value z_i in order to be able to purchase product i. She needs to incur search cost $s_i(> 0)$ on her first visit to seller i. She can purchase the product immediately or recall it at any point during her search. Each consumer can leave the market at any point and take an outside option u_0.

A consumer's ex post utility depends on her value for the purchased product , its price p_i and her search history. Let M be the set of sellers a consumer visits. If she purchases product i (in M), then her ex post utility is equal to

$$U(v_i, z_i, p_i, M) = v_i + z_i - p_i - \sum_{j \in M} s_j \tag{1.1}$$

If she does not purchase and takes an outside option, then her ex post utility is equal to

$$U(M) = u_0 - \sum_{j \in M} s_j \tag{1.2}$$

Each consumer is risk neutral and maximizes her expected utility. Given prices $\mathbf{p}$ and prior values, each consumer faces a sequential search problem. She decides in which order to search the products and, after each visit, whether to stop,in which case she chooses which product to purchase, if any, among those she has inspected so far, or search another product.

1.2.3 Consumer Behavior

Choi et al. (2018a) have the following two results about consumer's optimal search behavior.

Proposition 1. *Given $\mathbf{p} = (p_1, ..., p_J)$ and $\mathbf{v} = (v_1, ...v_J)$, the consumer's optimal search strategy is as follows: for each i, let z_i^* be the value such that*

$$s_i = \int_{z_i^*}^{\infty} (1 - G_i(z_i))dz_i \tag{1.3}$$

(i)Search order: the consumer visits the sellers in the decreasing order of $v_i + z_i^* - p_i$ (i.e., she visits seller i before j if $v_i + z_i^* - p_i > v_j + z_j^* - p_j$).

(ii)Stopping: let M be the set of sellers the consumer has visited so far. She stops, and takes the best available option by the point, if and only if

$$\max\{u_0, \max_{i \in M} v_i + z_i - p_i\} > \max_{j \notin M}\{v_j + z_j^* - p_j\}$$

Lemma 1. *Let $w_i \equiv v_i + \min\{z_i, z_i^*\}$ for each i. Given $\mathbf{p}, \mathbf{v}$ and $\mathbf{z}$, the consumer purchases product i if and only if $w_i - p_i > u_0$ and $w_i - p_i > w_j - p_j$ for all $j \neq i$.*

In order to utilize Lemma 1 [1], let H_i denote the distribution function for the new random variable $W_i = V_i + \min\{Z_i, z_i^*\}$, that is,

$$H_i(w_i) \equiv \int_{-\infty}^{z_i^*} F_i(w_i - z_i)dG_i(z_i) + \int_{z_i^*}^{\infty} F_i(w_i - z_i^*)dG_i(z_i) \tag{1.4}$$

[1] Lemma 1 holds even if prices are not observable to consumers before search, as long as consumers have correct beliefs about prices in equilibrium.

The distribution function H_i crucially depends on s_i. If s_i tends to 0, then z_i^* becomes arbitrarily large and, therefore, H_i becomes the convolution of G_i and F_i. If s_i explodes, then z_i^* approaches negative infinity, in which case H_i depends only on F_i.

Lemma 1 implies that the demand function for each seller can be derived as in standard discrete-choice models. A consumer purchases product i if and only if his effective utility for product i, $w_i - p_i$, exceeds the outside option u_0 and the corresponding utility for each other product, $w_j - p_j$. Therefore, the measure of consumers who purchase product i is given by

$$D_i(\mathbf{p}) = \int_{u_0+p_i}^{\infty} (\prod_{j\neq i} H_j(w_i - p_i + p_j))dH_i(w_i) \tag{1.5}$$

1.2.4 Simulation

Normalize the outside option u_0 at zero. Suppose there are N individuals and J products in total.To apply the theoretical results into a discrete choice model, denote the utility of individual n from consuming product i by $U_{ni} = X'_{ni}\beta + v_{ni} + z_{ni}$, where X_{ni} denotes the vector of exogenous variables, v_{ni} and z_{ni} denote the pre-search and post-search taste shock, respectively, and w_{ni} is defined correspondingly. Then based on Lemma 1, the probability of consumer n choosing product i is

$$P_{ni} = \int_{u_0-X'_{ni}\beta}^{\infty} (\prod_{j\neq i} H_j(w_{ni} + X'_{ni}\beta - X'_{nj}\beta))dH_i(w_{ni}) \tag{1.6}$$

Denote the vector of effective values by $\mathbf{w} = (w_{n1}, ..., w_{nJ})$, then the above choice probability can be expressed as

$$P_{ni} = \int_{u_0-X'_{ni}\beta}^{\infty} I(w_{ni} + X'_{ni}\beta > w_{nj} + X'_{nj}\beta, \forall j \neq i)dH(\mathbf{w}) \tag{1.7}$$

1.2.5 Identification

In this section I verify identification within the sequential search framework. I implement a simulation exercise to test whether we are able to recover the structural parameters of sequential search model with a purchase outcome dataset of certain size .

I simulate the behavior of 10,000 consumers. I consider a model with $J = 5$ products and fix distributions of random taste shocks, F_i and G_i at standard normal distribution from now on. Column 1 of Table 1.3 reports the true values of structural parameters used to generate the dataset. The prices of products are drawn from uniform distribution $U[0, 10]$, independent across consumers and products. Table 1.1 shows the purchase outcome in the simulated sample. 7,387 consumers choose the outside option, the remaining 2,613 purchase one of the five products. Table 1.2 shows the number of products searched by the consumers in the simulated sample. Among the 10,000 consumers, 5,928 search none, 3,746 search 1, 316 search 2, 9 search 3 and the remaining 1 searches 4 products. Table 1.1 and 1.2 shows that there is enough variation regarding purchase outcome and search behavior in the simulated sample. The similarity of the simulated data with available data helps us to understand whether the available data suffices to recover the structural parameters.

Table 1.3 presents the results of the simulation exercise. Overall, coefficients are precisely estimated; standard errors are small, and all estimates lie within two standard errors from the truth. I conclude that we are able to successfully recover structural parameters of the search model in the simulated sample[2].

Table 1.1: **Simulated Purchase Outcome.**

Alternative	Count	Percent
Product 1	738	7.38%
Product 2	602	6.02%
Product 3	496	4.96%
Product 4	404	4.04%
Product 5	373	3.73%
Not Buy	7387	73.87%

Table 1.2: **Number of Products Searched.** The table reports the search behavior of the simulated 10,000 consumers,i.e., how many products are searched before the consumers make a final decision.

Number of Product searched	Count	Percent
0	5928	59.28%
1	3746	37.46%
2	316	3.16%
3	9	0.09%
4	1	0.01%

Table 1.3: **Estimates from Simulation Exercise.** The table reports estimates obtained from the simulated sample (column 2) together with the true values of parameters used to generate the sample (column1). The simulated sample contains 10,000 consumers.

	Truth	Estimates	Standard Errors	T-Statistics
β	-0.6	-0.6088	0.0070	-87.1471
s_1	0.1	0.0870	0.0066	13.0909
s_2	0.2	0.1977	0.0001	3791.4242
s_3	0.3	0.2986	0.0102	29.2638
s_4	0.4	0.3891	0.0271	14.3524
s_5	0.5	0.4723	0.0235	20.0540

1.3 Data

I use the Nielsen Consumer Panel data in the year of 2016 and focus on the 32 oz refrigerated yogurt market. With one household-week as one observation unit, the sample contains 47,315 observations of 1,642 households' purchase outcome.

[2]The simulation is done by logit smoothed A-R estimator. The hessian is calculated by the BFGS updating procedure.

I focus on one retailer instead of one store to have reasonably large sample size, assuming that stores of the same retailer have similar display rules to follow and thus have approximately the same search cost for the same product. I see products with different flavors and fat percentages within the same brand as the same product, assuming that Chobani strawberry flavor yogurt has the same search cost with the vanilla ones. Details about the collapsing of multiple trips of one household in one week and multiple flavors within one brand are provided in Appendix A.0.1. I focus on the top five brands in the market:Chobani, CTL BR [3], Stonyfield, Dannon and Yoplait.

There are two important aspects of data construction. One is to recover the prices of all alternatives when I only observe the prices of the ones bought. The other is to determine what count as Not Buy observations.

For the first aspect, I calculate the store-week average price of all products, and use those to recover the prices unobserved in the trips with the same store-week combination. When there is no observation of one product within a store-week, I use the weekly average price of the product to recover that. More details are provided in the Appendix A.0.1. By doing so I assume that the price for a specific product in the same store in the same week is approximately the same, and that the prices for the same product in all stores under the same retailer is close. DellaVigna and Gentzkow (2017) document evidence of uniform pricing in US retail stores. Using the Nielsen Retailer Scanner and Consumer Panel dataset,they find that most US food, drugstore, and mass merchandise chains charge nearly-uniform prices across stores, despite wide variation in consumer demographics and the level of competition.

For the second aspect, I treat a trip without any purchase of the five products by one household-week as a Not Buy observation. I utilize the structure of panel data to be able to follow one household and observe their trip with no purchase of the five product. Meanwhile I break the panel structure of the data in two ways: one is that I don't follow the household every week, i.e., I have different consumer populations for different weeks in order to get rid of the potential multi-homing problem. A household with no trip to the retailer in a certain week is likely to have purchased the yogurt from another retailer, or haven't finished the last one they bought, or have been travelling outside the city. The other is that I assume there is no correlation among the observations from the same household. Note that I only assume the households don't have a persistent preference for brands, which is justified further in Appendix A.0.2, but meanwhile I allow for the preference for flavors, fat percentage and other product features. The brand varieties of the five brands are similar, i.e., they all have some basic flavors, such as plain,strawberry and vanilla, some basic fat percentage, like low-fat and no-fat. I allow for the possibility that a household prefers strawberry to vanilla, no-fat to low-fat, but assume that they don't have a persistent preference for Chobani strawberry low-fat yogurt to the Yoplait one.

1.3.1 Descriptive Statistics

Table 1.4 presents the purchase outcome of the 47,315 observations in the sample. Note that under the current specification of the Not Buy observations, the Not Buy observations takes more than 80% of the sample. A large portion of these Not Buy observations are from the households who only purchased one of the five products once in 2016. Although including these households leads to a high proportion of Not Buy observations , I

[3] CTR BR stands for the private label brands which are masked in Nielsen data.

choose not to drop them for two reasons. First, they account for a nontrivial percentage (more than 10% for each brand) of the total demand. Second, it's difficult to find a proper threshold of purchase times to define which households should be dropped. If the households who purchased once in the market should be dropped, then it's ambiguous whether the ones who purchased twice should be kept.

Table 1.4: **Purchase Outcome**

Alternative	Count	Percentage
Chobani	457	0.97
CTL BR	2,605	5.51
Dannon	1,435	3.03
Stonyfield	479	1.01
Yoplait	252	0.53
Not Buy	42,087	88.95
Total	47,315	100.00

Table 1.5 presents the descriptive statistics of the five top sellers I focus on.The average price is \$3.14 and I observe the purchase of them in 5,228 household-week observations from the specific retailer. The private label brand has the lowest average price of \$2.42 and highest market share of 49.83%. The minimum prices for Dannon and Stonyfield are zero because the coupon values are deducted from the tagged price. Chobani has the highest average price of \$5.45, while yoplait has the lowest market share of 4.82%.

Table 1.5: **Descriptive Statistics**

Product	Average Price	Std Dev	Max price	Min price	Total Purchase	Market share
Chobani	5.45	0.90	11.53	2.49	457	8.74%
CTL BR	2.42	1.04	0.69	8.37	2605	49.83%
Dannon	3.41	1.30	10.98	0	1435	27.45%
Stonyfield	4.19	1.34	13.98	0	479	9.16%
Yoplait	2.94	0.67	5.9	0.67	257	4.82%
Total	3.14	1.45	13.98	0	5228	100%

1.4 Results

1.4.1 Estimation Results

Table 1.6 presents the estimation results. All estimates are significant. The price sensitivity is negative. The private label brand has the lowest search cost of \$1.52 and Yoplait has the highest search cost of \$2.42.

Note that in our model, I assume that the consumers need to visit the store to search for additional information about sell by date, flavors, calories, fat percentage,etc. They also have to make the trip to purchase the product. So the search cost also includes transportation cost. People have to travel further to search and purchase the products that are only available in some stores. The result that the private label brand has the lowest search cost is intuitive: it is usually available in almost all stores of the retailer

Table 1.6: **Estimation Results.** Table present the estimation results from real dataset of size 47,315.

Coefficients	Estimates	Standard Errors	T-Statistics
Price Sensitivity	-0.0298	0.0055	-5.4603
Chobani	2.1422	0.0118	181.7785
CTL BR	1.5269	0.0046	329.5635
Dannon	1.7636	0.0036	493.4059
Stonyfield	2.1828	0.0137	159.5937
Yoplait	2.4191	0.0045	535.8005

and displayed in the most attractive position of the shelves. Consumers usually don't need to squat down to reach it at the bottom of the shelf. It also usually has some advantage in promotion and advertisement position. Also note that although Chobani has the relatively high average price(5.45) compared with Yoplait(2.94), it is estimated to have lower search costs(2.14) compared with Yoplait(2.41), and eventually higher market share. This can be a result of higher availability, or more advertisements or promotions to attract attention of the Chobani products. Supply and replacement speed, inventory management can also affect search cost. [4]

1.4.2 Model fit

I examine model fit by comparing several key predictions with their empirical counterparts. For this purpose, I split the sample of 47,315 observations into a training and a holdout sample. The training sample includes 24,000 randomly selected observations, whereas the remaining 23,315 observations constitutes the holdout sample.

Table 1.7 presents the estimation results of the training sample. I find that parameter estimates for the training sample are similar to those based on the full sample. Table 1.8 reports the simulated predictions using estimation result of the training sample and their empirical counterparts for the holdout sample.

Table 1.7: **Estimation Results of training sample.** The training sample includes 24,000 randomly selected observations.

Coefficients	Estimates	Standard Errors	T-Statistics
Price Sensitivity	-0.0329	0.0045	-7.3596
Chobani	2.1039	0.0281	74.7468
CTL BR	1.5212	0.0150	101.1172
Dannon	1.7568	0.0181	96.8775
Stonyfield	2.1676	0.0129	168.3591
Yoplait	2.4125	0.0260	92.8691

[4] Note that a full information traditional brand choice model will conclude that the retailer private label brand is demanded more because it provides higher utility. This is unconvincing as the brands like Chobani focus on yogurt products, and it is more likely that they make better yogurt than private label brands like Safeway Signature. Also, a future test of this can be to do the estimation on other retailers and see whether there are similar patterns, if so, it's hard to believe that every grocery retailer just makes better yogurt than branded products like Chobani.

Table 1.8: **Out-of-sample Fit of the Search Model.** I estimate the model on a training sample, which includes 24,000 randomly selected observations. The holdout sample consists of the remaining 23,315 observations

Alternative	Data	Simulate
Chobani	0.0084	0.0108
CTL BR	0.0549	0.0552
Dannon	0.0304	0.0305
Stonyfield	0.0099	0.0103
Yoplait	0.0047	0.0059
Not Buy	0.8917	0.8874

I evaluate the fit in terms of market shares for purchase behavior and find that the model predictions fit the holdout data reasonably well. The biggest difference between the real data and the simulated prediction is within 0.005.

1.4.3 Counterfactual Predictions

In this section I use the estimation results to do counterfactual predictions. Table 1.9 presents the counterfactual prediction results if all search costs are eliminated. The proportion of Not Buy option decreases substantially, and the distribution of purchase outcome becomes flatter,i.e., the market demand becomes less price-sensitive.

The counterfactual prediction results are consistent with our intuition. The existence of search cost makes consumers rely more on pre-search information,which in the current set up is the price. This is different from most current literature. In traditional literature, consumers search to resolve uncertainty about price. The existence of search cost makes them less reactive to prices, so ignoring search cost leads to an underestimation of elasticity because econometricians ascribe the unresponsiveness to price changes to low price elasticity. Whereas in our model, price is observed before search, so search cost makes consumers put too much weight on pre-search price information and thus more sensitive to price. Ignoring search cost leads to an overestimation of elasticity, as econometricians ascribe the responsiveness to price changes to high price elasticity. As shown in the table, the market shares under search cost setup is more sensitive to prices compared with the counterfactual prediction where there is no market friction.

Also, an implication of this on firm behavior is that a decrease in search cost can actually increase market prices, as consumers become less sensitive to prices. Choi et al. (2018a) provide sufficient conditions that guarantee the existence and uniqueness of pure-strategy market equilibrium and show that a reduction in search costs increases market prices.

Table 1.9: **Counterfactual Analysis**

Product	Actual	Market Share Simulated	Counterfactual
Chobani	0.0097	0.0102	0.1805
CTL BR	0.0551	0.0554	0.1993
Dannon	0.0303	0.0307	0.1930
Stonyfield	0.0101	0.0102	0.1884
Yoplait	0.0053	0.0059	0.1960
Not Buy	0.8895	0.8876	0.0427

Table 1.10: **Counterfactual Analysis: Search Behavior.**

Number of Products Searched	Simulate	Counterfactual
0	88.73%	0
1	11.25%	0
2	0.02%	0
5	0	100%

1.5 Conclusion

In this paper I estimate consumer search cost with purchase outcome data. I apply a theoretical result which fully summarizes consumer purchase outcome with a sufficient and necessary condition into a discrete choice framework.I verify the identification of the model by a numerical experiment. I am able to recover the structural parameters with the simulates sample. All parameters are significant and close to the true value.

I apply the estimation to real grocery purchase data, and find that after controlling for the price, the private label brand has lowest search cost. The low search cost may be a result of broader availability in stores, more efficient inventory management and faster replacement speed. I evaluate the out of sample fit of the model in terms of market shares and find that the model predictions fit the holdout data well. The estimates are significant and close to the truth. Counterfactual analysis suggest that if there is no market friction and consumers are fully aware of all products, there will be more purchases and that the market share will be less reactive to prices.

There are several interesting future research directions. One is to work out a more formal proof of identification to differentiate the search model with preference heterogeneity model. Abaluck and Adams (2017) provides important insights about this. In the current version I assume that the hidden values,i.e., the post-search information, are also unobservable to the researcher. An extension of this is to include some product characteristic that is unobservable to the consumers before search, but observable to the researchers. Then we may make more progress on identification from the asymmetric cross-derivative of demand on such characteristic. For example, consider the characteristic of sell-buy-date. In a model with full information, symmetry would require that switching decisions be equally responsive to an increase in the days towards sell-buy-date of one specific purchased good by a week or a decrease in the days towards sell-buy-date of all rival goods by a week. Suppose instead that there is search frictions and consumers only have information about a subset of the products. Now, switching decisions will be less responsive to changes in the days towards sell-buy-date of rival goods but more responsive to that of the purchased. In this way we can probably elicit more information from purchase outcome data. Other directions include to further analyze the full model and consider the oligopoly market structure and firm price competition.

Chapter 2

A Method to Estimate Discrete Choice Models that is Robust to Consumer Search

[1]

2.1 Introduction

When consumers purchase cars, houses, food, insurance, schooling and much else, they are often imperfectly informed about the attributes of relevant products in ways that substantially alter their choices Allcott and Knittel (2019); Woodward and Hall (2012); Abaluck and Gruber (2011); Allcott et al. (2019); Hastings and Weinstein (2008). Given this, models which assume full information may generate wrong conclusions about welfare and cannot be used to assess how choices would respond to more information. However, despite the emergence of behavioral economics as a major subfield of economic analysis, most work in applied economics continues to assume that choices are fully informed. We count 350 articles published in the AER, QJE, JPE, ECTA or ReStud since 2015 that estimate discrete choice models. Of these 350, 315 (90%) assume that consumers are fully informed.[2]

We believe this occurs for at least three reasons. First, for some positive purposes, it is irrelevant whether choices are informed since all that is required is to estimate how demand responds to price (although welfare evaluation still requires preferences) Berry and Haile (2014).[3] Second, the data necessary to directly measure consumers' beliefs is often unavailable, and even when it is available, survey data is often viewed with suspicion Gul and Pesendorfer (2008). Third, choice data alone does not suffice to separately identify preferences and beliefs without further assumptions Manski (2002). Structural search models in which consumer beliefs can be identified (e.g., Ursu (2018)) require assumptions regarding whether consumers take into account option value, whether they solve an optimal stopping problem or "satisfice," distributional assumptions about

[1]This chapter is based on work joint with Jason Abaluck and Giovanni Compiani.

[2]The list of papers and their classification is available upon request from the authors.

[3]For instance, price elasticities are sufficient to predict equilibrium prices after a counterfactual merger between two firms. But even among the 126 articles in our survey that conduct welfare analyses and thus must take a stand on whether consumers are informed, 109 (86.5%) assume full information without testing this assumption.

prior beliefs and search costs, and whether choices are simultaneous or sequential, among others.[4]

In this paper, we state what we believe to be plausible sufficient conditions under which choice data alone suffices to recover preferences whether consumers are fully or only partially informed. The approach relies on what we call *visible utility*, the component of utility visible to consumers prior to search (but not to the econometrician). In our baseline model, we impose that consumers search in decreasing order of visible utility — a condition we make precise below. We show that if this condition is satisfied, along with a few additional mild restrictions, there is a function of choice probabilities which recovers preferences whether consumers are fully or partially informed. Our approach does not require the researcher to fully specify a structural search model beyond the visible utility assumption. Specifically, no additional assumptions about option value, optimization vs. satisficing, simultaneous or sequential search, or distributional assumptions about beliefs and preferences are necessary for identification of preferences.

Recovering preferences under partial information has many applications. First, one can forecast the impact of informing consumers about attributes of goods prior to conducting such interventions, and compute the associated welfare benefits.[5] Second, our approach can inform firms' advertising strategies by, e.g., identifying product attributes that consumers care about but might not be currently aware of. Third, in settings where one would otherwise assume full information, the visible utility assumption provides a generalization which allows for both full and partial information, thus permitting a more realistic normative evaluation of choices.[6] Finally, given preferences recovered by our approach, we show that it is possible to identify other primitives of interest (notably, the distribution of search costs) under a maintained model of search.

One can think of our approach as a data-driven method of isolating consumers who maximize utility. Consider the example of consumers purchasing items in a grocery store: nutritional information is accessible, but at some cost. Consumers may fail to maximize utility if they do not pay the cost to examine labels. In this case, visible utility represents utility from all non-nutrient sources, e.g. a combination of prices and perceived taste. Our assumption states that if you bother to check the nutrition label for good j, you will first check the label for a good j' that you would otherwise prefer were it equally nutritious. This assumption implies that consumers who search the most nutritious good *always choose the good that maximizes utility among all options* (which is not necessarily the most nutritious good). To see this, note that if some other good has higher utility than the most nutritious good, it must have higher visible utility and thus is searched and then chosen by the consumer. Further, only consumers who search the most nutritious good are sensitive to nutrient content for that good. Therefore, by looking at the sensitivity of choices to the nutrient content of the most nutritious good we are able to isolate

[4]The empirical literature suggests that canonical assumptions in all of these cases are often rejected by the data (respectively, Gabaix et al. (2006), Schwartz et al. (2002), Jindal and Aribarg (2018), Honka and Chintagunta (2016)), limiting the applicability of structural search models.

[5]If an information intervention also reduces search costs, then the welfare gains via better choices given by our approach can be viewed as a lower bound to the total increase in welfare.

[6] For example, Allcott et al. (2019) propose taxing sugar-sweetened beverages to promote long-term health. A cost of this proposal might be that these foods are more desirable on other dimensions (e.g., tastiness), and conventional models would imply this if consumers appear willing to pay for high calorie foods. Allowing for imperfect information might reveal that consumers *prefer* low calorie foods once they are informed (e.g., for their physical appearance). In this case, the policy would be a win-win rather than one where health benefits must be weighed against short-term tastes.

consumers that behave *as if* they were fully informed; standard arguments then recover their preferences.

To spell things out in more detail, consider first a J-good model with linear utility $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ where $\alpha > 0$ and $\beta > 0$.[7] In the text, we extend this result to allow vector-valued x_j and z_j, as well as random coefficients and nonparametric utility. Suppose that consumer i observes x_j and ϵ_{ij} for all goods, but needs to engage in search to observe z_j. On the other hand, the researcher observes x_j, z_j, and choice probabilities s_j, but not ϵ_{ij}. With full information, we have $s_j = P(U_{ij} \geq U_{ij'}\ \forall j' \neq j)$ and we could estimate marginal rates of substitution using $\frac{\partial s_j}{\partial z_j} / \frac{\partial s_j}{\partial x_j} = \beta/\alpha$; in other words, β/α is identified by whether the choice probability for good j is more sensitive to z_j or x_j. If the underlying model is a search model in which consumers are informed about z_j only for some alternatives, then the standard approach will suffer from attenuation bias: $\left|\frac{\partial s_j}{\partial z_j} / \frac{\partial s_j}{\partial x_j}\right| \leq |\beta/\alpha|$. Some consumers will be insensitive to z_j variation not because they don't value it, but because they are not aware of it; thus, the observed sensitivity of choices to z_j will understate consumers' valuation of z_j. For each individual i and good j, we define *visible utility* as $VU_{ij} \equiv x_j\alpha + \epsilon_{ij}$. We call this quantity "visible utility" because it defines the utility that i receives from good j based only on x_j and ϵ_{ij}, the attributes of goods that consumers can observe without engaging in search (our main result also holds when ϵ_{ij} is only visible to consumers conditional on search, and so is not part of visible utility). In our baseline case, we assume that consumers search in decreasing order of visible utility. In other words, consumers always search first the goods that look more desirable given the information available to them. Note that the econometrician cannot tell the order of search since we do not observe ϵ_{ij}; this implies that observationally identical consumers can search in different orders.

The visible utility assumption is consistent with a broad class of search models. For example, in a Weitzman (1979) search model where the priors and search costs are the same across goods (but the latter vary across consumers), it is optimal to search the good in decreasing order of visible utility and decide whether to search the next good by comparing the expected benefits with search costs. Alternatively, consumers may myopically decide whether to continue searching by comparing utility in hand with expected utility of the next good (the "directed cognition" model of Gabaix et al. (2006)), consumers might engage in "satisficing," i.e. searching in order of visible utility and stopping whenever utility in hand is good enough, or they might simultaneously search all goods with visible utility above a certain threshold and then give up. In many cases, the underlying search process is simply unknown; in these cases, the conventional approach is to assume full information (potentially leading to biased estimates). Our approach is more general, allowing for full information, as well as a range of partial information models.

Our main result (for the case with linear utility and no random coefficients) is that, under the visible utility assumption and other conditions we make precise below, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \beta/\alpha$ for $j \neq 1$, where good 1 is defined as the good with the largest value of the hidden attribute z (which, again, is known to the econometrician but not necessarily to the consumer). This expression holds for any models where consumers search according to our assumptions above, including the full information case; it is thus robust to whether consumers are fully or partially informed under our assumptions. The main downside of our approach relative to the full information assumption is that it is more demanding of the data, but this may be tolerable in large datasets typical of modern empirical work.

[7]We will show that the signs of α and β are identified, so assuming they are positive is without loss.

The intuition for our result parallels the nutrition label example above: the expression $\frac{\partial s_1}{\partial z_1}$ is only non-zero for consumers who maximize utility, and consumers who maximize utility respond to attributes of rival goods (z_j and x_j) in proportion to their preferences. α is also separately identified from choice data alone,[8] and so our approach fully identifies preferences, not just marginal rates of substitution, and can be used for welfare analysis.

How general is this result? Using additional derivatives of the choice probability function, we can recover nonparametric utility functions $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$. Additionally, the approach extends to random coefficients on product characteristics. Specifically, letting $U_{ij} = x_j\alpha_i + z_j\beta_i + \epsilon_{ij}$, we can recover the distribution of random coefficients (α_i, β_i) over a known grid. With a sufficiently long panel and time-invariant preferences, $U_{ijt} = v_i(x_j, z_j) + \epsilon_{ijt}$, we can recover individual-specific, nonparametric utility functions $v_i(x_j, z_j)$. Thus, we can allow for a similar degree of unobserved heterogeneity as other constructive results on discrete choice demand with full information.[9] Berry and Haile (2009) and Berry and Haile (2014) also provide related results on nonparametric demand estimation. Their focus is on recovery of the conditional distribution of utilities rather than the structural parameters of utility; the latter are essential for our task of assessing whether consumers are informed about relevant attributes. We also consider cases where the visible utility assumption is *not* satisfied, such as models with search costs varying across goods. We extend our model to allow for cases where (i) search costs vary with observables (e.g., rank on a webpage), (ii) consumers form expectations about z based on x, (iii) search reveals unobservable information, and (iv) either x or z is endogenous and valid instruments are available.

Our identification proof lends itself naturally to estimation and testing. If one can nonparametrically estimate choice probabilities as a function of product attributes, then our results can be used to directly recover preferences. We also suggest an alternative parametric approach to estimate cross-derivatives that works well in simulations for larger numbers of goods. Given estimates of choice probabilities, one can use our result to test for full information by checking whether our "search-robust" estimates of preferences are equal to the conventional estimates based on first derivatives. This implies that one does not need to take an *a priori* stance on whether or not the attribute z is uncovered only after searching a good. That hypothesis can be tested provided that the data contains attributes x that can be assumed to be part of visible utility. Additionally, our model is overidentified; in the case of homogeneous, linear preferences, for example, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ will be equal for all goods $j \neq 1$. We also show that the assumptions in our model imply nontrivial bounds on choice probabilities that can be checked in the data.

We conduct two data analyses to validate our approach. First, we attempt to recover preferences in a lab experiment where individuals engage in costly search. Individuals choose from sets of three books with visible titles, authors, genre, star ratings and prices, but hidden discounts that can only be observed at some cost, with no other constraints on

[8]When z_j is the same for all goods, consumers maximize utility (although they themselves do not necessarily know this), so conventional methods suffice to recover α.

[9]Fox and Gandhi (2016) provide identification results for more general models allowing for both nonlinearity and flexible heterogeneity but these results are non-constructive and assume utility maximization; we use their result to identify parameters in corner cases where consumers maximize utility, but they otherwise are difficult to adapt to the more general case where choice probabilities need not maximize utility. This is in contrast to the constructive methods in Fox et al. (2012), who recover distributions satisfying the "Carleman condition," which implies that the distribution of preferences is uniquely characterized by its moments. Alternatively, we recover weights for distributions supported on a known and fixed grid, in line with the approach of Fox et al. (2011).

search. For each individual, we also observe treatments where consumers choose given full information. As expected, conventional logit estimates using the costly search data give attenuated coefficients on the discount variable relative to the full information case. By contrast, our search-robust estimates successfully recover full-information preferences. We show that our model successfully predicts the impact of an information intervention and permits an accurate welfare evaluation *before the intervention is conducted*. Estimated choice probabilities also satisfy bounds implied by the visible utility assumption.

Second, we apply our method to data from Expedia where consumers search for hotels. Many attributes of hotels, such as price or star rating, are immediately visible in search results. One attribute, hotel location desirability within a city, is only visible if you "search" a hotel by clicking through to find more information. We show that our approach correctly identifies location desirability as a hidden attribute, and we use the model to compute the benefits to making location desirability visible without search. Note that, in this non-experimental setting, consumers are free to search in whichever way they prefer, i.e. we are not enforcing any part of our model by design. The fact that the approach continues to work provides evidence that it can be successfully applied to observational data.

Our result relates to several existing literatures. An important theoretical and empirical literature models consumers as choosing from a possibly strict subset of the options available, their "consideration set."[10] This paper considers the complementary problem of imperfect information at the level of attributes rather than goods.[11] A growing literature, including Mehta et al. (2003), Kim et al. (2010), Honka and Chintagunta (2016), Kim et al. (2017), Ursu (2018) and Gardete and Hunter (2020), models consumers as searching products in order to uncover some of their attributes. We specifically consider the case when an attribute is observed by the econometrician and (potentially) not observed by consumers. In this case, we show that preferences can be recovered under our assumptions without the explicit structural search models used in the existing literature.[12,13] Another related literature studies whether consumers make informed choices by comparing the choices of regular consumers to that of a more informed subgroup. Bronnenberg et al. (2015) ask whether pharmacists make similar prescription drug choices

[10]Roberts and Lattin (1991), Goeree (2008), Conlon and Mortimer (2013) and Gaynor et al. (2016) — among others — estimate preferences when consumers may only consider some alternatives. Manzini and Mariotti (2014) establish that one can recover consideration probabilities as well as preferences if the data contains choices from every possible subset of the feasible set of goods. Abaluck and Adams (2017) show that identification can be achieved even without this type of variation under certain models of consideration set formation. Barseghyan et al. (2021) study partial identification of a general model with heterogeneous consideration sets.

[11]The recent theoretical literature on this question includes Branco et al. (2012), Ke et al. (2016) and Gabaix (2019).

[12]There is one special case where the problem of imperfect information about attributes has been addressed in the existing literature. This is the case in which all attributes can be expressed in dollar terms. For example, consumers should not care whether a health insurance plan saves them $100 in premiums or out of pocket costs (see Abaluck and Gruber (2011)), or whether a light bulb saves them money in upfront costs or shelf life (as in Allcott and Taubinsky (2015)). If one dollar-equivalent attribute is assumed to be visible to consumers, it can provide a benchmark for how consumers should respond to a hidden dollar-equivalent attribute. However, in many cases, attributes cannot easily be translated into dollars without first estimating consumer preferences. In these cases, our results still allow one to recover preferences given imperfectly informed consumers.

[13]Ericson et al. (2015) consider the related problem of inferring risk preferences separately from risk types using insurance choices. Their model differs from ours in that, in the special case they consider, the covariate "risk type" is not observed by the econometrician either.

to consumers, Handel and Kolstad (2015) ask whether better informed consumers make different health insurance choices, and Johnson and Rehavi (2016) study whether physicians treat differently when their patients are other physicians. Rather than identifying informed consumers, our paper develops a data-driven way of identifying consumers who maximize utility (despite not necessarily searching all goods) and whose choices can thus be used to recover preferences.

Section 3.4 lays out our formal framework and proves our identification results, Section 2.3 considers several empirically important extensions such as endogenous attributes, Section 3.5 provides details of estimation, Sections 2.5 and 2.6 report results from our experiment and Expedia application, respectively, Section 2.7 discusses the (counterfactual) questions that can be addressed using our approach, and Section 2.8 concludes.

2.2 Model and Identification

There are $J \geq 2$ goods indexed by $j = 1, ..., J$ with attributes x_j observed by consumers and the econometrician and attribute z_j observed by the econometrician but not necessarily by consumers.[14,15] We assume that x_j and z_j are continuously distributed. Further, in order to keep the analysis simple and focus on the intuition underlying our key results, we make a few simplifying assumptions. Each of these assumptions can be relaxed, as we discuss below. First, we let x_j be scalar for all j; our results immediately extend to the case of vector-valued x_j's at the cost of some extra notation. Second, we focus on the case where z_j is also a scalar and we let good 1 be the good with the largest value of z_j (we consider the case with multivariate z_j in Appendix B.9.2).[16] Third, we assume that the utility that individual i derives from good j is linear in x_j, z_j and an idiosyncratic shock ϵ_{ij} that is observed by consumers prior to search (we allow for nonparametric utility in Section 2.3.1). Fourth, we focus on the case where both x_j and z_j are exogenous (Section 2.3.3 discusses how to deal with endogeneity). We formalize these assumptions as follows.

Assumption 1. *The utility that individual i derives from good j is $U_{ij} = \alpha x_j + \beta z_j + \epsilon_{ij}$, where x_j, z_j are scalars, and $\mathbf{x} = (x_1, \ldots, x_J)$ and $\mathbf{z} = (z_1, \ldots, z_J)$ are independent of $\epsilon_\mathbf{i} = (\epsilon_{i1}, \ldots, \epsilon_{iJ})$. The consumer observes x_j, ϵ_{ij} for all j prior to search, but needs to search good j to uncover z_j.*

We use the term *visible utility* to indicate the component of utility that is known to the consumer before engaging in search, and denote it $VU_{ij} = \alpha x_j + \epsilon_{ij}$.

Next, we state the assumptions that characterize the class of search models we consider.

Assumption 2. *(i) Consumer i searches goods in decreasing order of VU_{ij}.*

(ii) Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(x_j, \epsilon_{ij}, \bar{u}) \geq 0$ where g_i is decreasing in $\bar{u}$.[17]

(iii) Consumers choose the good which maximizes utility among searched goods.

[14] Our model also permits the more general case where attributes are potentially both good and individual-specific, but we write x_j and z_j rather than x_{ij} and z_{ij} for notational simplicity.

[15] Since our model only requires variation in x and z for two goods, any of the remaining $J - 2$ goods may be taken to be the outside option.

[16] The definition of good 1 requires ruling out ties among the top two values of z_j.

[17] Assumption (*ii*) can be weakened to allow the function g_i to depend on a good-specific unobservable, such as search costs; however, good-specific search costs may lead to violations of Assumption (*i*). In Section 2.3, we extend our model to permit search costs to vary across goods with observable factors.

(iv) Only the value of z_j is unknown to consumers prior to search, and search fully reveals z_j.

We discuss these conditions at length in Section 2.2.2. To briefly clarify, Assumption (i) states that consumers search goods in order of the component of utility visible to them without search (although not entirely visible to the econometrician). We view this as the strongest restriction in our model; Section 2.3 considers two relaxations that are relevant for empirical work. Assumption (ii) states that consumers decide whether or not to search a good based on their utility in hand and the visible utility of the good they are considering searching. This rules out, for example, a sequential search protocol whereby one stops searching after discovering a good with large z irrespective of utility in hand. Further, Assumption (ii) also accommodates simultaneous search models in which consumers decide which goods to uncover based on visible utilities and then proceed to jointly search them. In this case, utility in hand is not a well-defined object and the function g_i does not vary with its second argument. We subscript the function g by i to emphasize that the function may depend on any individual (unobserved) heterogeneity in utility or search. For example, in a Weitzman search model, the stopping rule would depend on consumer i's reservation value, which in turn depends on i's search cost. Assumption (iii) states that consumers must search a good before choosing it.

Assumption (iv)—implicit in the model already—states that the econometrician observes all the information which is revealed by search, and that search is fully informative about the hidden attribute.

We pause here to highlight that Assumption 2 accommodates several commonly used models of search.

Example 1 (Sequential Search)**.** *Suppose that consumers search sequentially and consumer i must pay a cost c_i every time she uncovers the z attribute for a good. Further, assume that the consumer has the same prior F_z for all goods. Then, following Weitzman (1979), the consumer will rank goods according to their reservation value rv'_{ij} defined implicitly by*

$$c_i = \int_{rv'_{ij}}^{\infty} \left(u - rv'_{ij}\right) dF_{U_{ij}}(u) = \int_{rv_i}^{\infty} \beta_i \left(t - rv_i\right) dF_z\left(t\right) \tag{2.1}$$

where $rv_i \equiv \frac{rv'_{ij} - \alpha_i x_j - \epsilon_{ij}}{\beta_i}$ and the last step follows from a change of variable. We can interpret rv_i as the reservation value in units of z. To see this, note that consumer i ranks goods according to the visible utility $x_j\alpha_i + \epsilon_{ij}$ and for each good j' she chooses to uncover $z_{j'}$ if and only if the maximum utility secured so far is lower than $x_{j'}\alpha_i + rv_i\beta_i + \epsilon_{ij'}$. Once she stops searching, she maximizes utility among the searched goods. Thus, Assumption 2 is satisfied with $g_i\left(x_j, \epsilon_{ij}, \bar{u}\right) = x_j\alpha_i + rv_i\beta_i + \epsilon_{ij} - \bar{u}$.

Example 2 (Directed Cognition Model)**.** *As in the model of Gabaix et al. (2006), suppose that consumers rank goods in terms of expected utility*[18] *and myopically check whether searching the next good is worth the cost. The directed cognition model has the same g_i function as the Weitzman model,*[19] *but the order of search (and which goods are ultimately searched) may differ.*

[18] Note that we may assume without loss that $E(z_j) = 0$ for all j since the mean value of the hidden attribute (known by rational consumers before search) is subsumed by visible utility.

[19] The result that consumers in the fully rational Weitzman model decide whether to continue searching "as if" they were myopic is one of the main insights of Weitzman (1979).

Example 3 (Satisficing). *Suppose that consumer i searches in order of visible utility and stops whenever utility in hand is above a threshold τ_i. Then, Assumption 2 is satisfied with $g_i(x_j, \epsilon_{ij}, \bar{u}) = \tau_i - \bar{u}$.*

Example 4 (Full Information). *The full information model is subsumed within the previous example by letting $\tau_i = \infty$ for all i.*

Example 5 (Simultaneous Search). *Suppose that consumer i simultaneously searches all goods that have visible utility above a threshold $\tilde{\tau}_i$. Then, Assumption 2 is satisfied with $g_i(x_j, \epsilon_{ij}, \bar{u}) = \alpha x_j + \epsilon_{ij} - \tilde{\tau}_i$.*

Our results will not require the researcher to take a stand on the specific model of search that consumers follow (provided that our assumptions are met). Therefore, as illustrated by the examples above, the approach will be agnostic as to whether consumers search sequentially or simultaneously, are forward-looking or myopic and have biased or unbiased beliefs, among other things. In contrast, fully specifying a structural model requires one to take a stand on each of these dimensions.

Throughout the rest of the paper, we assume without loss that $\beta > 0$, i.e. we treat z_j as an attribute that customers value in good j.[20] We are now ready to state and prove a lemma that is at the core of our results.

Lemma 2. *Let Assumptions 1 and 2 hold. If consumer i searches good 1 (i.e., the good with the highest value of z), then i chooses the utility-maximizing good.*

Proof. If good 1 is searched but utility is not maximized, then for some unsearched j, $U_{ij} > U_{i1}$. Since $z_1 > z_j$, it must be that good $VU_{ij} > VU_{i1}$. But by Assumption (i), this implies that good j is searched, which is a contradiction. □

Note that Lemma 2 does *not* imply that good 1 always maximizes utility if it is searched. Rather, it implies that if good 1 is searched, the utility-maximizing good will also be searched (whether it is good 1 or not) and thus the consumer will choose that good. The lemma also does *not* mean that consumers searching good 1 are fully informed (in a search model they typically will not be), but just that those consumers act *as if* they were fully informed.

Lemma 2 will have far-reaching implications. To understand why, it will be convenient to define the choice probability for good j as:

$$s_j \equiv P\left(\left\{U_{ij} = \max_k U_{ik} \text{ for } k \in \mathcal{G}_i\right\} \cap \{j \in \mathcal{G}_i\}\right) \tag{2.2}$$

where $\mathcal{G}_i$ denotes the set of searched goods for individual i. Note that this probability is computed by integrating over any individual-specific unobserved heterogeneity in utility or search. Therefore, s_j is a function of $\mathbf{x} \equiv [x_1, \ldots, x_J]$ and $\mathbf{z} \equiv [z_1, \ldots, z_J]$, but we will often omit the dependence from the notation. Throughout the paper, the sources of unobserved heterogeneity will vary with the specific models we consider, so the symbol P will denote integrals over different distributions depending on the context.

[20] Conditional on Assumption (i), this is without loss, since Assumption 2 implies that an increase in U_{ij} can only induce consumer i to switch from not choosing j to choosing j, but never vice versa. Thus, by the chain rule, the sign of β is identified by the sign of $\frac{\partial s_j}{\partial z_j}$, where s_j is the choice probability function for good j from the data.

Now, Lemma 2 implies that z_1 only impacts choice probabilities for individuals who maximize utility. Therefore, looking at $\frac{\partial s_1}{\partial z_1}$ will isolate individuals who maximize utility and allow us to recover preferences using standard arguments.

In other words, the probability that good 1 is chosen is the probability that good 1 is utility-maximizing minus the probability of the only type of mistakes that consumers searching good 1 can make, i.e. failing to search good 1 *even though it is utility-maximizing.* Failing to search good 1 requires that there exists some other good j with $VU_{ij} \geq VU_{i1}$ and utility high enough that $g_i(x_1, \epsilon_{i1}, U_{ij}) \leq 0$. The other type of mistake, i.e. choosing good 1 when it is not utility-maximizing, is ruled out and thus does not feature in (**??**).

We will now show our key result, i.e. that the preference parameters α and β are identified from the second derivatives of function s_1. To keep the notation simple, we focus on the special case with with $J = 2$ goods. The result immediately extends to the case with $J \geq 2$ goods.

Lemma 3. *Let Assumptions 1 and 2 hold. Further, assume that* $\frac{\partial^2 s_1}{\partial z_1 \partial x_2}(\mathbf{x}^*, \mathbf{z}^*) \neq 0$ *for some* $(\mathbf{x}^*, \mathbf{z}^*)$*,* s_1 *is twice differentiable. Then,*

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2}(\mathbf{x}^*, \mathbf{z}^*) \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_2}(\mathbf{x}^*, \mathbf{z}^*) = \frac{\beta}{\alpha} \tag{2.3}$$

In addition, α *is identified by focusing on markets with* $z_j = z$ *for all* j *and thus* β *is also identified.*

Proof. First, we prove equation (2.3). In order to ease notation, we often suppress the subscript i in what follows. As above, good 1 is defined as the good with the highest value of z_j. Further, we let $\beta > 0$ without loss.[21] Using Lemma 2, the probability of choosing good 1 can be written as:

$$s_1 = P(U_1 > U_2) - P\left(\{U_1 > U_2\} \cap \{\mathcal{G} = \{2\}\}\right) \tag{2.4}$$

where, as above, $\mathcal{G}$ denotes the set of searched goods. This follows because (i) if good 1 is utility-maximizing, you will always choose it unless you search only good 2; and (ii) you only choose good 1 if it is utility-maximizing, since otherwise, good 2 must have higher visible utility, meaning it must be searched (and chosen) if good 1 is searched.

Let $\tilde{u}_j \equiv x_j \alpha + z_j \beta$, so that $U_{ij} = \tilde{u}_j + \epsilon_{ij}$, and let $(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^*, \mathbf{z}^*)$. Our goal will be to show that both z_2 and x_2 only impact $\frac{\partial s_1}{\partial z_1}$ via $\tilde{u}_2$. This, in turn, implies that $\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \frac{\partial^2 s_1}{\partial z_1 \partial \tilde{u}_2} \frac{\partial \tilde{u}_2}{\partial z_2}$ and $\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial^2 s_1}{\partial z_1 \partial \tilde{u}_2} \frac{\partial \tilde{u}_2}{\partial x_2}$, and the result in equation (2.3) follows. To establish this, note that we can write:

$$\begin{aligned} P\left(\{U_1 > U_2\} \cap \{\mathcal{G} = \{2\}\}\right) &= \\ P(\{U_1 > U_2\} \cap \{VU_2 > VU_1\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}) &= \\ P(\{U_1 > U_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}) - P(\{VU_1 > VU_2\} \cap \{g(x_1, \epsilon_1, U_2) \leq 0\}) & \end{aligned} \tag{2.5}$$

where the second line follows since $VU_1 > VU_2$ implies $U_1 > U_2$. The second term on the last line of display (2.5) is not a function of z_1. The first term is only a function of x_2 and z_2 via $\tilde{u}_2$. This, together with equation (2.4), is sufficient to show that both z_2 and x_2 only impact $\frac{\partial s_1}{\partial z_1} = \frac{\partial P(U_1 > U_2)}{\partial z_1} - \frac{\partial P(\{U_1 > U_2\} \cap \{\mathcal{G} = \{2\}\})}{\partial z_1}$ via $\tilde{u}_2$, thus proving equation (2.3).

[21] This is without loss, since the sign of β is immediately identified from the data (footnote 20).

Finally, we show that we can identify α using standard techniques by looking at choice sets where $z_j = z$ for all j. To see this, note that when $z_j = z$ for all j then consumers maximize utility if and only if they maximize visible utility. Since by assumption they always search the good with the highest visible utility, it follows that they maximize utility. Thus, one can pin down α by looking at how the choice probabilities vary with $\mathbf{x}$ conditional on $z_j = z$ for all j, just like in the full information case.[22] Given (2.3) and α, identification of β follows immediately. $\square$

Finally, we show that in many models of interest the conventional way of identifying preferences based on the ratio of *first* derivatives leads to understating consumers' taste for z. For this result, we further assume that the function $g_i(x_j, \epsilon_{ij}, \bar{u})$ is weakly increasing in x_j. This condition is satisfied in all the search model considered above (Examples 1–5) when the coefficient on x in utility is positive and corresponds to the mild requirement that consumers are (weakly) more prone to searching a good the higher the value of x for that good.

Lemma 4. *Let Assumptions 1 and 2 hold. Further, assume that $g_i(x_j, \epsilon_{ij}, \bar{u})$ is weakly increasing in x_j. Then,*

$$\left| \frac{\partial s_j}{\partial z_j} \Big/ \frac{\partial s_j}{\partial x_j} \right| \leq \left| \beta / \alpha \right|. \tag{2.6}$$

This shows that standard discrete choice models that assume full information—such as multinomial logit or probit—will typically suffer from attenuation bias under our assumptions.

2.2.1 Alternative Approaches and Support Assumptions

So far we have not focused on the support assumptions required for identification. These are nonetheless essential to understand our contribution. Alternative approaches to identification exist which differ principally in requiring much stronger support assumptions.

For instance, one could assume that the data exhibits "at-infinity" variation to effectively go back to a setting that is analogous to full information. As the visible utility for a subset of goods grows to infinity (minus infinity), the probability of searching those goods goes to one (zero) under reasonable assumptions on the search process. Using this, one could identify preferences using conventional arguments. However, in practice, it is often implausible that any goods are searched with probability close to 1, so this strategy would require substantial parametric extrapolation.

In contrast, our proof requires much more plausible support assumptions. There is always a good which maximizes z_j (or our weighted index in the vector-valued case, see Appendix B.9.2). To recover preferences in the homogeneous linear case, we only need

[22] There is one subtle exception to this argument. Suppose there is an outside option with utility normalized to 0, and we wish to identify a fixed effect which gives the utility of all inside goods relative to the outside good. In this case, consumers do not necessarily maximize utility when $z_j = z$ for all goods because consumers may decide to search *none* of the inside goods, and they may do so even when the outside good has lower utility than some of the inside goods if search costs are sufficiently high (in other words, an outside option may violate our assumption that consumers must search a good before they choose it). When consumers search none of the inside goods, it is never possible to separately identify whether consumers do not value the inside goods or have high search costs to examine any of the inside goods. It is possible to say something about the utility of consumers who are induced to search at least one of the inside goods when price is low enough (for example), but parametric assumptions are needed to make claims about the utility of consumers who never search any of the inside goods.

sufficient variation to estimate second derivatives of s_1 at a single point. As one would expect, flexibly recovering a nonparametric utility function or nonparametric random coefficients distribution requires considerably more variation and data, since it involves estimating higher order derivatives of choice probabilities, as we show in Section 2.3. Still, it remains less demanding than "at-infinity" identification. We further discuss these challenges in Section 3.5.

2.2.2 Discussion of Search Model Assumptions

To reiterate, we consider search models satisfying the following assumptions:

1. Consumer i searches goods in decreasing order of VU_{ij}.
2. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(x_j, \epsilon_{ij}, \bar{u}) \geq 0$ where g_i is decreasing in $\bar{u}$.
3. Consumers choose the good which maximizes utility among searched goods.
4. Only the value of z_j is unknown to consumers prior to search, and search fully reveals z_j.

As discussed above, there are several microfoundations for the first assumption. For example, in the Weitzman (1979) search model, consumers search goods in order of reservation utility, which is a function of the visible attributes of those goods, the distribution of the hidden attribute z_j, and search costs. If z_j is i.i.d. across goods and consumers have the same search cost for all goods, then it follows that consumers will search in order of visible utility (see Example 1). There are at least three reasons this might fail in the Weitzman (1979) model: first, there may be more uncertainty about the hidden attribute for some goods than others, and this might lead individuals to search such goods first. Second, unobservables might be correlated across goods, so that, e.g., learning good news about good 1 might cause one to positively update about good 2 and choose to search it before good 3 even if $VU_{i3} > VU_{i2}$. Third, search costs might vary across goods, meaning that consumers prefer to search first goods with lower search costs even if the payoff is potentially lower.

While the restriction that priors be i.i.d. and search costs be constant across goods is sufficient for Assumption (i) (the first assumption above), this is not necessary. Priors may be heterogeneous but consumers may be unsophisticated and fail to take into account option value, as in the directed cognition model studied in Gabaix et al. (2006). Consumers searching for a laptop online may enter some attributes into a search function and look at the items which rank highly according to those attributes without regard for whether a lower item is worth searching first because its value is more uncertain despite its lower average utility. Such examples also raise the natural concern that in many settings, factors like the order in which items appear in search may impact search costs separately from visible utility. Applications in the marketing literature often allow search costs to vary with observable attributes, such as the position of a good in search (e.g., Ursu (2018)). In Section 2.3.4, we extend our main result to allow for these violations of our visible utility assumption by considering cases where some observable attributes impact search but not utility. We can also relax the i.i.d. priors assumption by allowing consumers to form beliefs about the hidden attribute as a function of observed attributes.

Specifically, in Section 2.3.5, we extend our approach to the case where beliefs about z_j are a linear function of observables.

Our second assumption on search is that consumers search good j if and only if $g_i(x_j, \epsilon_{ij}, \bar{u}) \geq 0$ where $\bar{u}$ is utility in hand; we also impose the natural restriction that one is (weakly) less likely to search as $\bar{u}$ increases. This assumption is satisfied in most search models we are aware of in the literature, including Weitzman search, satisficing, simultaneously searching all goods with visible utility above a threshold, random search, and directed cognition. One exception is a model in which consumers simultaneously search the top K goods in terms of visible utility prior to engaging in search. This model would violate the assumption because the function g_i that determines whether i searches good j cannot be written only as a function of x_j and ϵ_{ij} since it will depend on the visible utility of all goods. We show in section 2.3.7 that our methods can be extended to accommodate one version of this model based on Honka et al. (2017). We also investigate the robustness of our approach to a violation of this assumption in the simulations of Section 3.5.

Our third assumption, that consumers choose the good which maximizes utility among searched goods, embeds two separate ideas: the first is that consumers do not choose a good they have not searched, and the second is that they maximize utility given the information available. This is natural in contexts such as e-commerce, where consumers typically have to open a product's page in order to add it to their carts. The assumption that consumers maximize utility given the information available can also be relaxed. One could specify a positive utility function that allows for consumer errors; as long as consumers maximize that positive utility function, the weight that they would attach to the hidden attribute given full information will be revealed. It is then up to the researcher whether to take this weight as the normative benchmark or whether to use some external standard.

The fourth assumption again nests two pieces. The first is that only the value of z_j is unknown prior to search. A consumer who clicks through to the product information page of an Amazon product might learn information about the attributes of a good ("the battery is compatible with USB-c"), but they also might learn information not observable to the econometrician ("one reviewer said the battery exploded into flames"). In section 2.3.6, we show that our results continue to hold if the ϵ component of utility is revealed only conditional on search (as in Kim et al. (2010) and Ursu (2018)). The second piece of the fourth assumption is that search reveals all information about the hidden attribute. This assumption is natural in settings where z_j is fully observed to the econometrician, as in our case. This is not always plausible: if the hidden attribute is "school-value added," a consumer who searches more may learn about test scores and graduation rates, but these are (imperfect) signals of the underlying variable. There is a literature on consumer (Bayesian) learning which models more explicitly the case when search is not fully informative (see Erdem and Keane (1996), Ackerberg (2003a), Crawford and Shum (2005), among others).

While our assumptions are not without bite, they subsume a range of search protocols. Still, one might wonder if they will hold in empirically relevant settings. We address this in two ways. First, in the next subsection, we show that the assumptions on the search process can be tested based on the same data required for estimation of the model. Second, we apply our approach to data from a lab experiment (Section 2.5) as well as observational data from Expedia (Section 2.6), and show that our method is able to successfully identify the attributes that are not immediately visible to consumers.

2.2.3 Testing Search Model Assumptions with and without Observable Search

Our analysis so far has proceeded as if search were not observed; that is, we observe final choices as a function of $\mathbf{x}$ and $\mathbf{z}$ but we do not observe which specific goods were searched. Datasets increasingly contain some information on what is searched: for example, in online clickstream data, one observes not only which product was purchased, but also which products were clicked on en route to purchase (e.g., Ursu (2018)). In many settings, it is plausible to assume that such clicks reveal which products were searched.

Can preferences be identified without resorting to our approach or an explicit search model in these cases? One might assume that our identification results would be unnecessary in such cases; given data on which products were searched, perhaps preferences can be estimated conditional on search without any of the assumptions we require here. However, this is not generally the case because the unobservable component of utility may also drive the search decision. One example would be if search depends on ϵ. In such cases, goods with undesirable observables that are searched likely have an especially high realization of ϵ. Thus, it will appear from conditional choice probabilities as though the observable attributes are not so bad when in practice, individuals dislike those attributes but this dislike is offset by a large ϵ. A second reason unobservable components of utility might impact search is if preferences are unobservably heterogeneous (random coefficients). Even if search does not depend on ϵ, preferences cannot generally be recovered using only conditional choices unless IIA is satisfied.[23] Thus, with heterogeneous preferences, the existing literature requires specifying a search model in order to estimate preferences even when search is observed. Our approach avoids the need to do this under the assumptions we have outlined.

Once our approach is used to identify preferences, clickstream data can be used to conduct additional overidentifying tests if we assume that the distribution of ϵ_{ij} is known. In the linear case, visible utility is given by $VU_{ij} = \alpha x_j + \epsilon_{ij}$. As shown in Lemma 3, examining choices with equal values of the hidden attribute is sufficient to identify α. Given α, the known distribution of ϵ_{ij}, and the number of goods searched $|\mathcal{G}_i|$, we can thus compute:

$$P(j \in \mathcal{G}_i|\mathbf{x}, \mathbf{z}) = \sum_k P(|\mathcal{G}_i| = k|\mathbf{x}, \mathbf{z}) P(j \in \mathcal{G}_i||\mathcal{G}_i| = k, \mathbf{x}, \mathbf{z}) \tag{2.7}$$

since the first probability on the RHS is observed and the second is pinned down by the model assumptions (specifically, the fact that with k goods searched, those k goods must be the k goods with the highest visible utility). Checking (2.7) against the observed search probabilities provides a test of the model.

[23] To see why heterogeneous preferences create a problem, imagine products have quality ratings from 1-5. There are two types of consumers, one type that cares about quality and one type that does not. The type that cares about quality is indifferent about quality over the 4-5 range, but values quality over the 1-4 range sufficiently that quality differences outweigh any other differences observable to consumers. Suppose that quality is observable to consumers (x) but price is only observed conditional on search (z). Quality conscious consumers only search goods with quality of at least 4. Other consumers will search all goods. If we estimate preferences conditional on search, we will wrongly conclude that no one cares about quality: quality conscious consumers don't care about quality given the goods they have searched (quality ranging from 4-5) and non-quality conscious consumers don't care about quality at all. To estimate preferences correctly, we would have to jointly model the decision of which goods to search and preferences conditional on searching.

Even when we do not observe auxiliary information on which goods are searched, the assumptions in our model can be jointly tested by checking whether the observed choice probabilities are consistent with bounds implied by the estimated preferences and assumed search rule. To construct an upper-bound on choice probabilities, note that a good j cannot be chosen if there is an alternative good with higher visible utility and higher utility. Thus, we have:

$$s_j(\mathbf{x}, \mathbf{z}) \leq 1 - P(U_{ik} \geq U_{ij} \text{ and } VU_{ik} \geq VU_{ij} \text{ for some } k) \tag{2.8}$$

The latter probability can be directly computed from knowledge of preferences and the distribution of ϵ. To construct a lower-bound, note that the probability of choosing good j is at least as large as the probability that good j maximizes both utility and visible utility. That is:

$$s_j(\mathbf{x}, \mathbf{z}) \geq P(U_{ij} \geq U_{ik} \text{ and } VU_{ij} \geq VU_{ik} \text{ for all } k) \tag{2.9}$$

Once again, this probability can be computed given knowledge of preferences and the distribution of ϵ. We can then check whether our estimated choice probabilities are consistent with these bounds.

Finally, our model is overidentified. For example, in the case of linear utility and homogeneous preferences that we have focused on so far, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \beta/\alpha$ for all alternative goods $j \neq 1$ and values of $(\mathbf{x}, \mathbf{z})$ at which the derivative in the denominator is nonzero. This provides a number of overidentifying restrictions which could be used to further test the model.

2.2.4 Testing for Full Information

Our results suggest a natural test for full information. Under the null hypothesis of full information, $s_j = P\left(U_{ij} \geq U_{ik} \ \forall k\right)$ and therefore:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_j} \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_j} = \frac{\partial s_k}{\partial z_j} \Big/ \frac{\partial s_k}{\partial x_j} = \frac{\beta}{\alpha} \tag{2.10}$$

for all $j \neq 1$ and all k. On the contrary, when consumers are unaware of z_j for some goods, then the ratios of first derivatives need not be equal to the ratios of the second derivatives. For example, Lemma 4 showed that $\frac{\partial s_1}{\partial z_1} / \frac{\partial s_1}{\partial x_1} \leq \frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ in the class of search models we consider. Since both the ratios of first derivatives and the ratios of second derivatives in (2.10) are estimable from the data, this immediately leads to a test based on the discrepancy between the two sets of ratios. More specifically, given estimators of the share functions, one can compute a Wald test-statistic based on the discrepancy between the two sets of ratios and reject the null hypothesis of full information if the statistic exceeds a critical value.

Note that this test is valid even if our assumptions on the search process fail to hold since with full information the two sets of ratios will be equal regardless. When our assumptions on the search process do hold, we expect the test to have power, since the first derivative ratio will be attenuated relative to the true preferences, which are recovered by the cross-derivative ratio.

One may wonder whether this testing approach continues to be valid in the case where consumers are heterogeneous in their preferences for the x_j and z_j attributes (we consider identification of that model in Section 2.3.2). Specifically, one might worry that, when consumers have heterogeneous preferences, the ratio of first derivatives might be attenuated relative to the ratio of second derivatives even under the null hypothesis of full information. In Appendix B.10, we provide verifiable sufficient conditions that rule this out and therefore guarantee the validity of our test for the mixed logit model.

2.3 Extensions

In this section, we consider several extensions to the baseline model. We consider each extension separately, although in principle, one could estimate models combining several such extensions. For example, Kim et al. (2010) estimate a search model in which only attributes unobservable to the econometrician are revealed during search, and in which some observables impact search but not utility.

2.3.1 Nonparametric utility

We start by extending Lemma 2 to the case with nonparametric utility U_{ij}. Without loss of generality, we can write: $U_{ij} = a_{ij}(x_j) + b_{ij}(x_j, z_j)$ where $b_{ij}(x_j, 0) = 0$ (to see this, define $b_{ij}(x_j, z_j) = U_{ij}(x_j, z_j) - U_{ij}(x_j, 0)$). The term $a_{ij}(x_j)$ is the component of utility that is known to the consumer before engaging in search, and thus corresponds to what we defined as visible utility, VU_{ij}. We make the following assumptions on the utility function.

Assumption 3. *(i) For all i and j, U_{ij} is strictly monotonic in z_j.*
(ii) For all i, the function $b_{ij}(x_j, z_j)$ is not alternative-specific, i.e. $b_{ij}(x_j, z_j) = b_i(x_j, z_j)$ for all j, and continuous in its first argument.

The class of utility functions satisfying Assumption 7 is broad and subsumes most specifications commonly used in empirical work as special cases, including logit with possibly nonlinear-in-characteristics utilities[24] and mixed-logit. For instance, in a mixed-logit model, one may specify $U_{ij} = \alpha_i x_j + \beta_i z_j + \epsilon_{ij}$. To map this specification into our notation, let $a_{ij}(x_j) = \alpha_i x_j + \epsilon_{ij}$, and $b_i(x_j, z_j) = \beta_i z_j$.

Lemma 5. *Let Assumptions 2 and 7 hold, and let $x_j \in [\bar{x} - \eta, \bar{x} + \eta]$ for all j, for some $\eta > 0$ sufficiently small. If consumer i searches good 1 (i.e., the good with the highest value of z), then i chooses the utility-maximizing good.*

Proof. If good 1 is searched but utility is not maximized, then for some unsearched j, $U_{ij} > U_{i1}$. Since $z_1 > z_j$, by monotonicity, $b_i(\bar{x}, z_1) > b_i(\bar{x}, z_j)$. By continuity of b_i in its first argument, this implies that for η sufficiently small, $b_i(x_1, z_1) \geq b_i(x_j, z_j)$.[25] Given this, $U_{ij} > U_{i1}$ implies $VU_{ij} > VU_{i1}$. But by Assumption (i), this implies that good j is searched, which is a contradiction. $\square$

[24] We allow for nonlinearities subject to Assumption (i) being satisfied.

[25] More formally, by continuity, for all $\delta > 0$ there exists $\eta > 0$ such that if $|x_1 - x_j| < 2\eta$, then $b_i(x_j, z_j) - b_i(x_1, z_j) \leq \delta$. Therefore, we have:

$$\begin{aligned} b_i(x_j, z_j) &= b_i(x_1, z_j) + b_i(x_j, z_j) - b_i(x_1, z_j) \\ &\leq b_i(x_1, z_j) + \delta \\ &\leq b_i(x_1, z_1) \end{aligned}$$

where the last inequality follows by choosing $\delta \equiv \frac{b_i(x_1,z_1) - b_i(x_1,z_j)}{2}$.

Next, we generalize Lemma 3. To this end, we consider the case where utility takes the form

$$U_{ij} = v(x_j, z_j) + \epsilon_{ij}$$

for an unknown function v. In what follows, we use x and z to denote generic arguments of v.

Theorem 2. *Let Assumption 2 hold and utility be given by $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ with v increasing in both arguments and infinitely differentiable. Further, assume that $\frac{\partial^2 s_1}{\partial z_1 \partial x_{j^*}}(\mathbf{x}^*, \mathbf{z}^* \neq 0$ for some $(\mathbf{x}^*, \mathbf{z}^*)$ and $j^* \neq 1$, s_1 is infinitely differentiable and $\epsilon_{\mathbf{i}} \perp (\mathbf{x}, \mathbf{z})$. Then, v is identified up to an additive constant.*

This theorem applies to a broad class of utility functions. The cost of this level of generality is that it requires the share function s_1 to be infinitely differentiable. However, the marginal rates of substitution are recovered under much weaker differentiability requirements.

Corollary 3. *Let Assumption 2 hold and utility be given by $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ with v increasing and differentiable in both arguments. Further, assume that s_1 is twice differentiable and $\epsilon_{\mathbf{i}} \perp (\mathbf{x}, \mathbf{z})$ and assume that $v(\cdot)$ is identifiable from fully informed choices. Then, the marginal effects $\frac{\partial v}{\partial z}$ and $\frac{\partial v}{\partial x}$ can be identified using only second derivatives. Specifically, (i) marginal rates of substitution, $\frac{\partial v}{\partial z} / \frac{\partial v}{\partial x}$, can be recovered using:*

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_j}(\mathbf{x}, \mathbf{z}) \Big/ \frac{\partial^2 s_1}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) = \frac{\partial v}{\partial z}(x, z) \Big/ \frac{\partial v}{\partial x}(x, z) \tag{2.11}$$

for all $j \neq 1$ such that $\frac{\partial^2 s_1}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) \neq 0$, and (ii) $\frac{\partial v}{\partial x}$ can be identified from choices where $z_j = z$ for all j.

Finally, we consider the case where the researcher has access to long panel data. This allows for even more flexibility in the specification of utility. In particular, we consider the case where

$$U_{ijt} = v_i(x_{jt}, z_{jt}) + \epsilon_{ijt}.$$

Note that we now let the function v be consumer-specific. This case closely parallels the proof for Theorem 2. Now, rather than observing only $s_j(\mathbf{x}, \mathbf{z})$, the choice probabilities for each alternative as a function of the attributes, panel data allows us to observe $s_{ij}(\mathbf{x}, \mathbf{z})$, the choice probabilities for each individual as $(\mathbf{x}, \mathbf{z})$ vary over a long period of time. Given these, the following result holds:

Theorem 4. *Let Assumption 2 hold and utility be given by $U_{ijt} = v_i(x_{jt}, z_{jt}) + \epsilon_{ijt}$ with v_i increasing in both arguments and infinitely differentiable. Further, assume that $\frac{\partial^2 s_{i1}}{\partial z_1 \partial x_{j^*}}(\mathbf{x}^*, \mathbf{z}^*) \neq 0$ for some $(\mathbf{x}^*, \mathbf{z}^*)$ and $j^* \neq 1$, s_{i1} is infinitely differentiable and $\epsilon_{\mathbf{i}} \perp (\mathbf{x}, \mathbf{z})$. Then, v_i is identified up to an additive constant.*

Corollary 5. *Let Assumption 2 hold and utility be given by $U_{ijt} = v_i(x_{jt}, z_{jt}) + \epsilon_{ijt}$ with v_i increasing and differentiable in both arguments. Further, assume that s_{i1} is twice differentiable, $\epsilon_{\mathbf{i}} \perp (\mathbf{x}, \mathbf{z})$, and assume that $v(\cdot)$ is identifiable from fully informed choices. Then, the marginal effects $\frac{\partial v_i}{\partial z}$ and $\frac{\partial v_i}{\partial x}$ can be identified using only second derivatives. Specifically: marginal rates of substitution, $\frac{\partial v_i}{\partial z} / \frac{\partial v_i}{\partial x}$, can be recovered using:*

$$\frac{\partial^2 s_{i1}}{\partial z_1 \partial z_j}(\mathbf{x}, \mathbf{z}) \Big/ \frac{\partial^2 s_{i1}}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) = \frac{\partial v_i}{\partial z}(x, z) \Big/ \frac{\partial v_i}{\partial x}(x, z)$$

for all $j \neq 1$ *such that* $\frac{\partial^2 s_{i1}}{\partial z_1 \partial x_j}(\mathbf{x}, \mathbf{z}) \neq 0$*, and* $\frac{\partial v_i}{\partial x}$ *can be identified from choices where* $z_j = z$ *for all* j*.*

The proofs of these two results exactly parallel the arguments for Theorem 2 and Corollary 3.

2.3.2 Random coefficients

The cases considered so far assume either that we have panel data or that all individual heterogeneity is additively separable. Due to the difficulty of separately identifying preferences and search as well as more practical difficulties with estimation, most empirical structural search models that we are aware of do not allow for non-separable unobserved heterogeneity (see, e.g., Ursu (2018), Honka et al. (2017)).

Of course, we would like to understand both from a theoretical perspective whether the assumption of separable heterogeneity is required for identification and from a practical perspective whether our results are applicable in such cases. The canonical case of non-separable heterogeneity that has been studied in the literature and for which constructive identification results exist is that of the linear random coefficients model. We maintain linearity and impose two additional assumptions.

Assumption 4. *(i) Utility is given by* $U_{ij} = x_j \alpha_i + z_j \beta_i + \epsilon_{ij}$*.*

(ii) The coefficients α_i *and* β_i *take values on a known finite support, i.e.* $\alpha_i \in \{\alpha_1, \ldots, \alpha_{K_\alpha}\}$ *and* $\beta_i \in \{\beta_1, \ldots, \beta_{K_\beta}\}$*. Further, the elements of* $\{\beta_1, \ldots, \beta_{K_\beta}\}$ *all have the same sign and, without loss, we assume that they are positive.*

(iii) The distribution of $\epsilon_\mathbf{i}$ *is known (or independently identified) and the three random vectors* $\epsilon_\mathbf{i}, (\alpha, \beta)$ *and* $(\mathbf{x}, \mathbf{z})$ *are mutually independent.*

Assumption (*ii*) follows Fox et al. (2011) and a recent strand of empirical papers (e.g., Nevo et al. (2016)) in assuming that the random coefficients are supported on a finite and known grid of points. Given the restriction that $\{\beta_1, \ldots, \beta_{K_\beta}\}$ all have the same sign, assuming that they are positive is without loss (see footnote 20). Assumption (*iii*) maintains knowledge of the distribution of all unobservables other than the random coefficients, consistent with recent papers on identification and estimation of demand (e.g. Fox et al. (2012), Fox et al. (2016)).

Theorem 6. *Let Assumptions 2 and 4 hold. If the market share of good 1 is* $K_\alpha K_\beta -$*th order differentiable, then the probability weights* $\tilde{\pi}_{k_\alpha, k_\beta}$ *for* $k_\alpha = 1, \ldots, K_\alpha$*,* $k_\beta = 1, \ldots, K_\beta$ *are identified.*

As it is clear from the statement of Theorem 6, allowing for heterogeneity across consumers in preferences for attributes typically requires taking derivatives of order higher than two. Thus, identifying heterogeneous preferences is more demanding of the data. When the sample size does not allow for direct application of our result, a natural approach is to impose more structure by specifying a structural search model. Theorem 6 may then be used to establish nonparametric identification of preferences within the specified model of search.[26]

Finally, we note that Theorem 6 focuses on recovering the entire distribution of the random coefficients. If the goal is simply to test whether consumers have full information,

[26] See Section 2.7 for more on this point.

then taking second-order derivatives turns out to be sufficient under certain conditions, which we spell out in Appendix B.10.

Taken together, the utility specifications considered in this and the last section are comparable in generality to existing constructive identification results for preferences in standard full information discrete choice models, such as Fox et al. (2012).

2.3.3 Endogenous attributes

So far, we have assumed that the observed product attributes are independent of all unobservables. This is restrictive, especially in settings in which product attributes — notably price — are chosen by firms who might know more about preferences or product attributes than is captured by the observed data. As highlighted by a large literature (e.g., Berry et al. (1995)), this typically leads to correlation between the attributes chosen by firms and product-level unobservables.

Here we consider an extension of our model that allows for endogenous product attributes. We specify the utility that consumer i gets from good j as

$$U_{ij} = \alpha x_j + \beta_i z_j + \lambda_i p_j + \xi_j + \epsilon_{ij} \tag{2.12}$$

where p_j denotes the endogenous characteristic and ξ_j is a product-specific characteristic that is known by consumers before search, but is not observed by the researcher.[27] If firms also know ξ_j when choosing p_j, then the two will typically be correlated, thus leading to endogeneity of p_j. We consider both the case where p_j is part of visible utility and that in which consumers need to search good j to uncover p_j (as well as possibly other non-endogenous attributes z_j). If p_j is price, the first scenario corresponds to settings such as e-commerce where typically price is visible on the results page and does not require any further clicking by the user. On the other hand, the second scenario covers cases in which price is itself the object of consumer search (there is a large literature on this, particularly in relation to the often observed price dispersion for relatively homogeneous goods; see, e.g., Stahl (1989), Hong and Shum (2006) and Hortaçsu and Syverson (2004)). We show identification of preferences for each of these two cases. To this end, we introduce two mutually exclusive variants of assumption (*ii*). Let $\delta_j = \alpha x_j + \xi_j$ for all j.

Assumption 5. *(i) The attribute p_j is part of the visible utility of good j. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(\delta_j, \epsilon_{ij}, p_j, \bar{u}) \geq 0$ where g_i is decreasing in $\bar{u}$.*

(ii) The attribute p_j is uncovered by consumers only upon searching good j. Conditional on having utility $\bar{u}$ in hand, consumer i searches j if and only if $g_i(\delta_j, \epsilon_{ij}, \bar{u}) \geq 0$ where g_i is decreasing in $\bar{u}$.

Like Assumption (*ii*), Assumption 5 states that consumers decide whether to search good j based on utility in hand and the visible utility of j. In Appendix B.9.3, we

[27]Note that the utility specification in (2.12) allows for random coefficients on both z_j and p_j, but not on x_j. This is stronger than needed, since the identification argument below only requires that x_j and ξ_j enter the demand functions via a linear index. Thus, another possible specification is

$$U_{ij} = \tilde{\alpha}_i \left(\alpha x_j + \xi_j\right) + \beta_i z_j + \lambda_i p_j + \epsilon_{ij}$$

The latter is weaker, but also less common in the discrete choice literature, so we focus on model (2.12) in what follows.

invoke results from Berry and Haile (2014) to show that these assumptions suffice for nonparametric identification of the choice probability functions provided we have valid instruments (in a sense we make precise in the Appendix). Once the choice probability functions are identified, one may apply our results in Section 2.3.2 to identify the distribution of the preference parameters α, β_i and λ_i.

2.3.4 Allowing for variables affecting search but not utility

One important case in which the visible utility assumption (i) is likely to fail is when factors exist which impact search costs but not utility. An example might be search position for online purchases. Arguably, search position impacts the order in which people search but has no direct impact on utility conditional on searching Ursu (2018). In this case, consumers might first search items with higher search position even if they do not have higher visible utility. For example, if we randomly assign search order, this is likely to impact choices even though we are not changing the utility of each item conditional on search. A second example is if we observe advertising expenditures for each good and believe that advertising entices consumers to search advertised goods.

Our model can be extended to deal with cases where the factors which impact search but not utility are observable and the sign of their impact on search probabilities is known (such as position in search). Denote the variable which perturbs search but not utility by r_j, suppose that r_j is observed and that higher values of r_j make a good weakly more likely to be searched. Now, rather than assuming that goods are searched based on VU_{ij} alone, we assume that goods are searched based on $m(VU_{ij}, r_j)$ where m is strictly increasing in both VU_{ij} and r_j. We show in Appendix B.9.4 that a version of our identification argument continues to hold provided we see sufficient variation in product attributes conditional on search position.

2.3.5 Allowing for consumers' expectations on z to depend on x

Another reason why the visible utility assumption (i) might fail is that consumers could form expectations about z based on x. For instance, if x is price and z is quality, consumers might infer that more expensive products tend to be higher quality. As a consequence, if they value quality to a sufficient degree relative to price, they may search a high-priced product and not search a low-priced product even if the former has a lower visible utility than the latter.

In our proofs so far, we have not made any explicit assumption about whether consumers update about z_j given x_j, but such updating is likely to lead to violations of the visible utility assumption if not explicitly modeled. We now show that we can identify preferences given consumer beliefs about z_j given x_j in a linear model. Further, under additional assumptions, we will show that we can identify β/α, the relative value of the hidden attribute, even when beliefs are unknown. In other words, we can do so without taking a stand on whether consumers have rational expectations and form beliefs based on the empirical relationship between z_j and x_j or naively update. Consider the linear model $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ and re-write it as

$$\begin{aligned} U_{ij} &= x_j\alpha + (z_j - E(z_j|x_j))\beta + E(z_j|x_j)\beta + \epsilon_{ij} \\ &= \beta\gamma_0 + x_j(\alpha + \beta\gamma_1) + \tilde{z}_j\beta + \epsilon_{ij} \end{aligned}$$

where the second equality assumes that consumers use the linear projection $E(z_j|x_j) = \gamma_0+\gamma_1 x_j$ and we let $\tilde{z}_j \equiv z_j - E(z_j|x_j)$. Visible utility is then given by $\beta\gamma_0 + x_j(\alpha+\beta\gamma_1) + \epsilon_{ij}$ and consumers learn the deviation from their expectation on z_j, $\tilde{z}_j$, upon searching. Note that γ_0 is not identified, but also does not generally impact choices since it enters utility as an additive constant.[28]

In Appendix B.9.5, we show that given γ_1, we can recover β and α using an analog of our usual approach. When γ_1 is not observed, we can still identify β/α if we know its sign and assume that we observe goods with the largest value of z_j and the smallest value of x_j. The quantity β/α is not sufficient to simulate choices with full information, since we cannot tell how responsive consumers would be to x_j were choices fully-informed. However, it is sufficient to identify the relative value placed on the hidden attribute as well as to conduct tests for full information as in Section 2.2.4.

2.3.6 Unobservables revealed by search

So far, we have focused on the case where the attribute(s) z revealed by searching a good are entirely observed by the researcher. However, it is easy to imagine settings in which the data does not capture all of the information that consumers acquire through search. Indeed, the existing literature often models search as the process whereby the idiosyncratic preference shocks—ϵ_{ij} in our notation—are revealed (e.g., Kim et al. (2010), Ursu (2018), Moraga-González et al. (2021)). To accommodate this, we consider a modification of our model where the shock ϵ_{ij} only becomes known to consumer i upon searching good j (along with z_j). In other words, consumers know x_j for all j prior to search and decide whether to acquire ϵ_{ik} and z_k for any given good k through search. This means that, in Assumption 2, VU_{ij} is now equal to αx_j and Assumption (iv) is dropped.

Given this setup, we show in Appendix B.9.6 that the ratio of second derivatives $\frac{\partial^2 s_j}{\partial z_j \partial z_k} \Big/ \frac{\partial^2 s_j}{\partial z_j x_k}$ recovers $\frac{\beta}{\alpha}$ provided that one chooses good k to be the good with the highest value of x (note that j need not be the good with the highest value of z here). Thus, our approach can be extended to deal with the possibility that search reveals unobservables.

2.3.7 The $K-$rank Simultaneous Search Model

As noted above, our main model allows for consumers to choose which goods to search in one simultaneous step. However, one form of simultaneous search that is not accommodated is that in which a consumer optimally chooses the number K of goods to uncover and then proceeds to simultaneously search the top K in terms of visible utility (e.g., Honka et al. (2017)). Our framework from Section 3.4 does not subsume this model since in this case the decision of whether or not to search good j depends not only on the visible utility of good j, but on the visible utility of all other goods as well, thus violating Assumption (ii).

In Appendix B.9.7, we show via an alternative argument that the usual second-derivative ratio from equation (2.3) still identifies $\frac{\beta}{\alpha}$ in the two-good $K-$rank model.

[28]There is one exception to the above claim, which is the case when there is an outside option for which the x and z attributes are not defined, so that a systematic bias in beliefs about the distribution of z_j given x_j would change the relative value of all the inside goods relative to that outside option. This might mean that the relative utility of the outside option cannot be separately identified from γ_0; the model could still be estimated, but the normative interpretation of fixed effects might change.

We also show in our simulation results that our method succeeds in a model where consumers search the top K goods (with K varying randomly across consumers).

2.4 Estimation

Our identification results show that preferences can be recovered given knowledge of the choice probability function for good 1, denoted by $s_1(\mathbf{x}, \mathbf{z})$. We now discuss how s_1 can be estimated from data on choices and product attributes. Note that the model implies the following conditional moment restrictions

$$E\left(y_j - s_j\left(\mathbf{x}, \mathbf{z}\right) | \mathbf{x}, \mathbf{z}\right) = 0 \quad \forall j, \tag{2.13}$$

where y_j is a dummy variable equal to 1 if a consumer chooses good j.[29] Thus, methods designed to estimate conditional moment restriction models can be used. Of course, the performance of an estimator will depend on how flexibly it captures the derivatives that identify preferences in our approach.

Here, we consider two approaches to estimating $s_1(\mathbf{x}, \mathbf{z})$: (i) an approximation via Bernstein polynomials which is viable when the number of goods and attributes is small; and (ii) a "flexible logit" model which is more ad hoc, but scales better as the number of goods increases. As in much of the demand estimation literature, a good in our model is defined by the collection of attributes observable to the econometrician (potentially including good fixed effects); in other words, different products with the same attributes count as the same good. Thus, estimation of choice probabilities and their derivatives does not require that all consumers have identical products in their choice sets, or even that the same products are available to many different consumers (unless product fixed effects are of interest). What we need is sufficient variation in attributes to flexibly estimate the mapping from the product attributes to choices.

Throughout this section, we focus on the linear homogeneous case of $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$. Lemma 3 shows that β/α can be recovered from $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ for $j \neq 1$. Relative to conventional estimation of linear homogeneous discrete choice models, our approach is more demanding of the data, requiring estimation of second derivatives for a specific good. In return, this allows us to be more agnostic about the underlying information structure. As discussed above, the model with linear, homogeneous preferences is the current standard in the empirical literature on search (e.g., Mehta et al. (2003), Honka and Chintagunta (2016) and Ursu (2018); Kim et al. (2010) is a notable exception in that they allow for random coefficients). In more general non-linear or random coefficients models, our identification arguments require recovery of higher-order derivatives and thus might not directly translate into viable estimation strategies in small to medium sample sizes or with a large number of goods. In these cases, the best way forward might be to parametrically specify a full structural search model and estimate it via standard methods, e.g. MLE. We would then view our identification results as providing reassurance that preferences are indeed identified, something that had not been formally established in the literature (see Section 2.7 and Appendix B.12 for more on this). Additionally, given the parametric structure imposed by such a model, one can construct

[29] Here, we focus on the case where data on individual-level choices are available, as in the experiment of Section 2.5 and the application of Section 2.6. However, our identification approach could also be applied to aggregate (i.e., market share) data as long as one can consistently estimate the share functions s_j.

moments using the derivatives in Section 2.2.3 and use them to estimate parameters and conduct overidentifying tests.

2.4.1 Approximation via Bernstein polynomials

Following Compiani (2022), one can approximate the demand function via Bernstein polynomials. This allows the researcher to impose natural restrictions via linear (and thus easy-to-enforce) constraints on the coefficients to be estimated. Specifically, the class of models considered in this paper satisfies standard monotonicity restrictions in $\mathbf{x}$ and $\mathbf{z}$ (s_j increasing in x_j and z_j and decreasing in $\mathbf{x}_{-j}$ and $\mathbf{z}_{-j}$). In addition, one can consider other constraints, such as exchangeability across goods, which requires demand to only depends on the attributes of the goods, but not their identity.[30] Exchangeability is satisfied if the unobservables entering demand (e.g., preference parameters and shocks, as well as search costs) have the same distribution across goods. We impose both monotonicity and exchangeability in the nonparametric results reported below. The purpose of these restrictions is twofold. First, they discipline the estimation routine in the sense that they help obtain reasonable estimates of quantities of interest (e.g., negative price elasticities). Second, they help partially alleviate the curse of dimensionality that arises as the number of goods increases. The coefficients in the Bernstein approximation of s_j can be estimated by minimizing a GMM objective function based on the restrictions in (2.13) subject to the constraints. More details on the implementation of the estimator can be found in Compiani (2022). In Appendix B.3, we report results from numerous simulations with a variety of data generating processes which suggest that this estimation approach performs well with a small number of goods (2 or 3) when the assumptions of our model are satisfied; it also consistently outperforms standard logit estimates in simulations where the visible utility assumption is violated.

2.4.2 "Flexible Logit"

As the number of goods increases, nonparametric methods face a curse of dimensionality, and thus it becomes necessary to place some parametric structure on the problem. In this section, we develop one such parametric approximation which performs well in simulations for a larger number of goods.

As discussed in more detail in Appendix B.11, conventional full-information models typically impose strong restrictions on the structure of the derivatives of choice probabilities. We would like to estimate a model of s_1 which is sufficiently flexible that ratios of first-derivatives differ from ratios of second cross-derivatives, as will generally occur if consumers engage in search. To allow for this additional flexibility, we let the mean utility for good 1 depend directly on attributes of rival goods as follows:

$$v_1 = \tilde{v}(x_1, z_1) + b_1 z_1 + \sum_{k \neq 1} \left(\gamma_k w_{z1k} z_k + \gamma_{2k} w_{x1k} x_k + w_{z2k} \delta_k z_k z_1 + w_{x2k} \delta_{2k} x_k z_1 \right) \quad (2.14)$$

where $\tilde{v}(x, z)$ is a differentiable function of x and z, w_{z1k}, w_{x1k}, w_{z2k} and w_{x2k} are known weights, and b_1, γ_k, γ_{2k}, δ_k and δ_{2k} are coefficients to be estimated. Further, we let $v_k = \tilde{v}(x_k, z_k)$ for $k \neq 1$. In Appendix B.11, we describe one way of choosing the weights which we find works well in simulations, and for which the ratio of second derivatives

[30] See Compiani (2022) for a formal definition of exchangeability.

(which recovers β/α) is a particularly convenient function of model parameters. We note that the parameters in (2.14) do not have a causal interpretation (i.e., we are not positing that the actual utility of good 1 depends on the attributes of good k for $k \neq 1$). Instead, (2.14) is simply a flexible function of $(\mathbf{x}, \mathbf{z})$ that captures the second derivatives of s_1 well. In Appendix B.3, we show that the flexible logit performs extremely well in simulations for a variety of data generating processes. For three DGPs satisfying the assumptions of our model, conventional logit estimates are biased, but flexible logit confidence intervals include the true values. For a fourth DGP violating the assumptions of our model, flexible logit has a small bias with a large number of goods, but is consistently less biased than the standard logit estimates.

2.5 Experimental Validation

Our identification proof and simulation results show that preferences can be estimated regardless of whether consumers are fully informed, provided consumers search in a way that is consistent with our assumptions. Of course, the theorem does not tell us whether those assumptions are likely to be satisfied in practice.

In this section, we test in a lab experiment whether we can recover preferences in a setting where consumers engage in costly search. Unlike in our simulations, the search protocol is unknown to us and not restricted to satisfy the assumptions of our model. We nonetheless show that we are able to correctly recover preferences using our "search-robust" estimation technique.

2.5.1 Set-up

We selected 1,000 books for sale on Amazon Kindle chosen from a wide variety of genres. For each book, we observe its average rating on the site "Goodreads.com" as well as the average rating from Amazon.com, the number of reviews on Goodreads, and the price of the book for Amazon Kindle.

In our experiment, conducted via Mechanical Turk, each participant made 40 choices from sets of 3 randomly selected books. For all books, participants could see a photo of the cover, the title, author and genre, as well as the Goodreads rating and the number of ratings. Prices were randomized to integers from \$11-\$15 (equally likely). All books were then further discounted by an integer amount from \$0-\$10 (equally likely). All users were given a \$25.00 bank at the start of each choice, from which any costs incurred were deducted. There were a total of 93 participants, yielding 3,720 choices.

The discount is our key variable of interest. For 10 of the 40 choices, users could see all discounts and thus could see the net price of all options at no cost. For 30 of the 40 choices, discounts were hidden and users had to pay a cost to see the discount for any given book.[31] The cost per click was constant for each user across the 30 choices, and randomly chosen from $\{\$0.10, \$0.25, \$0.35, \$0.50\}$. For the 30 choices with hidden information, users could only choose books after they clicked to reveal the discount and had to choose at least one book. One of the 40 choices made by each user was randomly chosen to be realized, and users received the chosen book as well as any money left over from the original \$25.00.

[31] The full information and costly information choice situations were randomly ordered, so that the 10 "full information" choices were intermixed with the costly information choices.

Figure 2.1: Lab Experiment: Sample Product Selection Screen

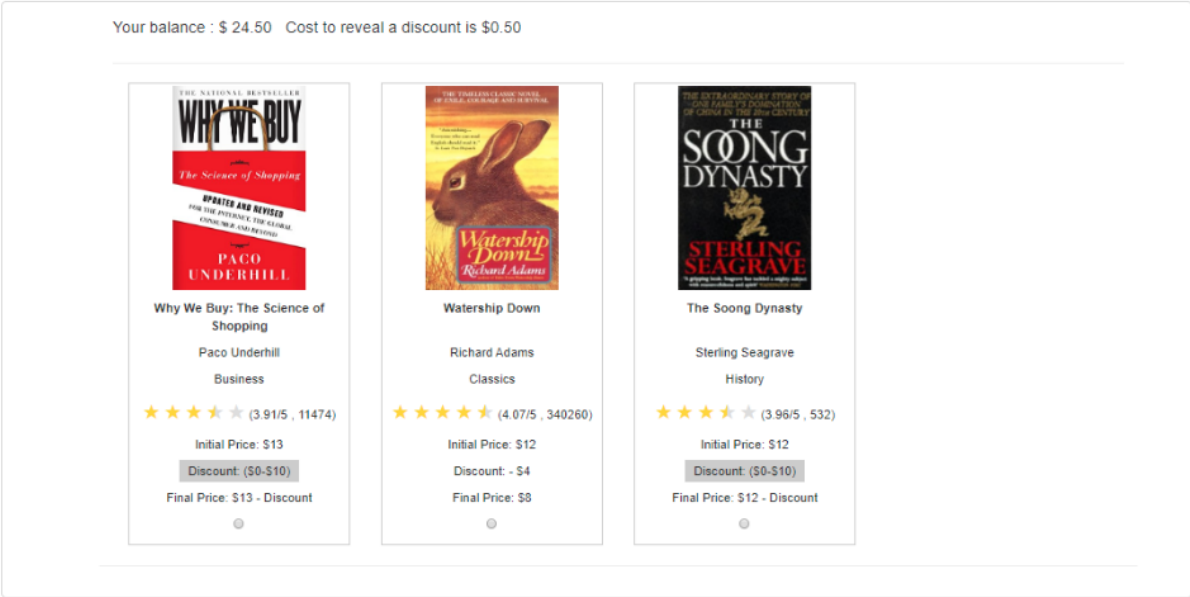


Figure 2.1 shows a sample product selection screen from a choice where discounts were hidden. In this case, the user clicked to reveal the discount of the second book and could either choose that book or continue by revealing the discounts for additional books. Note that the user could search books in any order she wished. The 10 choices where all information is revealed are our benchmark for the "truth." The goal is then to test whether the relative weight on discounts and prices that we estimate in the cases where discounts are costly to observe matches the relative weight we see when discounts are visible to everyone. Further, because both discounts and prices are in dollar terms, and because they are randomized (and so not signals of quality), there is a second benchmark: if consumers are rational, the weight on discounts and prices should be equal.

In other words, we will model choices using a linear utility specification as in our baseline model from Section 3.4:

$$U_{ij} = price_j \cdot \alpha_1 - discount_j \cdot \beta + rating_j \cdot \alpha_2 + \epsilon_{ij} \tag{2.15}$$

where ϵ_{ij} is i.i.d. type-I extreme value[32] and accounts for any aspects of consumers taste for books (based on the title, image, author or genre) not summarized by the price, discount and rating variables. Fully informed and rational consumers should have $\alpha_1 = \beta$. Our goal will be to show that we can recover these fully informed preferences using the choices of beneficiaries for whom revealing discounts is costly.

2.5.2 Estimation Results

Columns 1 and 2 of Table 2.1 show results from estimating a standard logit model on consumer choices for the 10 choice situations (per consumer) where all information is revealed (Full Info) and the 30 choice situations where consumers must pay to reveal

[32]We will compare nonparametric estimates based on our approach to estimates from a conventional logit model. Only the latter requires distributional assumptions on ϵ_{ij}.

information (Costly Info), respectively. With full information, consumers place equal weight on prices and (negative) discounts, so they pass our test of rationality. In other words, they care only about the final price of the product. By contrast, when discounts are costly to reveal, the coefficient on the discount variable in the standard logit model is attenuated (the "Costly Info" column). This is because consumers are insensitive to variation in discounts for books they do not search. The ratio of the two coefficients is 0.986 in the full information treatment and 0.683 in the costly information treatment.

Table 2.1: Standard Logit and Cross-Derivative Estimation Results

	Standard Approach		**Our Approach**
Variable	**Full Info**	**Costly Info**	**Costly Info**
Price	-0.386***	-0.302***	-0.387***
	(0.038)	(0.018)	(0.032)
Discount (-)	-0.376***	-0.206***	-0.399
	(0.020)	(0.009)	-
Rating	0.591***	0.421***	0.584***
	(0.190)	(0.099)	(0.161)
Discount (-) / Price	0.986***	0.683***	1.032***
	(0.093)	(0.044)	(0.102)
N	930	2790	2790

Note: The table shows estimation results from a standard logit model estimated on the full information and costly information treatments in columns 1 and 2, and Bernstein polynomials estimation of the cross-derivative ratio on the costly information treatment in column 3. The minus sign indicates that discount multiplied by -1 so that the coefficient on discount should equal that of price with full information. Standard errors on the ratio of the discount and price coefficients are computed using 250 bootstrap draws.*** denotes significance at the 1% level, ** at 5% level, and * at 10%.

Following Section 2.4.1, we estimate the demand function $s_1(\mathbf{x}, \mathbf{z})$ via Bernstein polynomials. Specifically, we use the tensor product of univariate Bernstein polynomials, one for each argument of the s_1 function.[33] Further, we impose the natural constraint that s_1 be decreasing in the price of book 1 and the discount of books 2 and 3, and increasing in the discount of book 1 and the price of books 2 and 3. The main result of this procedure is an estimate of β/α_1, which we obtain by dividing a trimmed mean (across choices) of $\frac{\partial^2 s_1}{\partial discount_1 \partial discount_j}$ by a trimmed mean of $\frac{\partial^2 s_1}{\partial discount_1 \partial price_j}$ for all $j \neq 1$, and then averaging across j.[34] The estimate is 1.032, which is close to the corresponding number from column 1. In addition to estimating β/α_1, we need to directly recover the α coefficients. Consistent with Lemma 3, we compute these by estimating a logit model using only choice sets

[33]We use univariate polynomials of degree three for the arguments z_1, x_2, z_2 and of degree two for the remaining arguments. The total degree of the approximation is 21.

[34]Specifically, for each second derivative, we take the mean over values in the interquartile range. As is often the case in nonparametric estimation, trimming helps obtain less noisy estimates.

where the variance of the discount across goods is in the bottom quintile. The results are reported in column 3 of Table 2.1, along with the value of β implied by our estimates of α_1 and β/α_1. The confidence intervals include the full information values. In other words, using data only on choices when information is costly, we successfully recover informed preferences. Further, the confidence interval is sufficiently tight to exclude the standard logit estimates in the costly information treatment.

Having recovered all preference parameters, we can compute how information will change behavior and choice quality. Using only data on choices when search is costly, our model predicts that, on average, full information consumers would save \$0.66 per choice from choosing books with lower discounts. The corresponding number in the data is \$0.69 per choice situation, since consumers in the costly information treatment average discounts of \$6.24, while consumers in the full information treatment average discounts of \$6.93.[35] In other words, we can accurately predict how consumers will respond to information provision *before* the information is provided. We can also compute the dollar equivalent welfare benefits of providing consumers with information. To do so, we take our estimates from column 1 as the normative preferences (i.e., as the correct metric to compute consumer welfare) and calculate by how much welfare changes when consumers go from making partially uninformed choices to fully informed choices using the approach in Appendix B.13. We then repeat this exercise using the estimates from column 3 as the normative preferences. We estimate an average welfare gain of \$0.18 per choice based on column 1 and of \$0.15 based on column 3. Thus, our model again yields results that are quite close to those coming from the "true" fully informed choices in the data.[36]

As in most real-world settings, visible utility is not observable to the econometrician in our experiment: while we can see attributes of the goods in question, we do not know how individuals will weigh these attributes, nor do we know their preferences for specific genres or book titles and images. The assumption that consumers search according to the visible utility assumption is substantive and could be violated in numerous ways: users might always reveal discounts for the lowest priced book first or they might search in the order in which books are displayed. Nonetheless, our "robust" estimation approach succeeds in recovering the preferences consumers reveal with full information. In Appendix B.7, we report results from the test discussed in Section 2.2.3, showing additionally that the estimated choice probabilities lie within the upper and lower bounds implied by the visible utility assumption. Thus, we fail to reject the assumption.

2.6 Field Validation

Our lab experiment demonstrates one setting where our approach correctly identifies hidden attributes and forecasts how consumers will respond to information about hidden attributes. Of course, this leaves open the question of whether we can identify preferences in real-world settings with a larger number of goods, where search costs are implicit and potentially heterogeneous, and where we (as experimenters) cannot strictly control the information available to consumers.

In this section, we investigate these issues using publicly available data from a leading

[35]We look at differences in discounts as opposed to final prices paid since the latter are essentially the same in the full information and in the costly information conditions.

[36]Note that the benefits are smaller than the increase in discounts because information induces consumers to be more responsive to discounts, sacrificing some value on unobservable factors.

online travel agency, Expedia Ursu (2018).[37] The data we use includes transactions from 54,648 consumers over an eight-month period between November 1, 2012, and June 30, 2013.[38] At the time, consumers would search for a hotel in a given city, and Expedia would present a list of available options. On the list, consumers observe a range of hotel characteristics including the price per night, whether the hotel is on promotion or part of a chain, star rating, and the review score.

One attribute, location, is not visible to consumers in search results but is only visible *after* consumers click on the hotel and can see a map showing the exact location of the hotel within the city. The dataset contains a measure of location desirability, but this measure is not visible to consumers in search results. We thus ask: can our model correctly recover that location is a hidden attribute, whereas other attributes are directly visible in search results?

Table 2.2 provides summary statistics. Hotels on average charge $162 per night, with 3.5 stars and a review score of 4 out of 5. 64% of the hotels belong to a chain, and 34% of the hotels display a promotion. Location attractiveness is a score ranging from 0 to 7 designed by Expedia to measure how centrally a hotel is located, what amenities surround it, and other aspects of location desirability. The average hotel has a location score of 3.26. Figure 2.2 illustrates how hotel characteristics appear to consumers in search. Note that information on features like price, stars, review score, and promotion flag are saliently displayed. However, consumers do not observe the detailed map unless they click, and the quantitative location score is not shown. In order to evaluate the attractiveness of a location, consumers need to spend time and effort to examine the map.

Figure 2.2: Hotel Characteristics Shown in the Search Impression

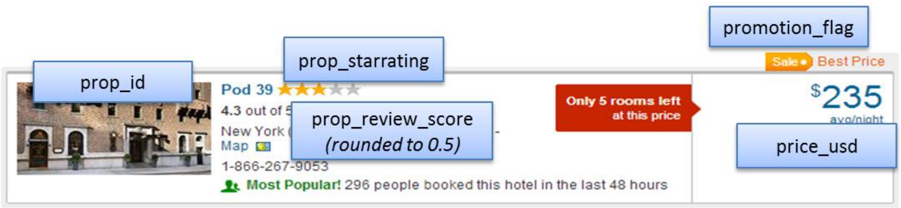


Table 2.2: Summary Statistics

	Observations	Mean	Median	SD	Min	Max
Price ($)	546,480	162.11	139.57	92.94	10	1000
Stars	546,480	3.42	3.00	0.91	0	5
Review Score	546,480	4.01	4.00	0.71	0	5
Chain	546,480	0.64	1.00	0.48	0	1
Location Score	546,480	3.26	3.22	1.45	0	7
Promotion	546,480	0.34	0.00	0.47	0	1

As in Section 3.4, we consider a homogeneous linear specification for utility:

$$U_{ij} = \mathbf{\tilde{x}_j} \cdot \tilde{\alpha} + \mathbf{x_j} \cdot \alpha + z_j \cdot \beta + \epsilon_{ij} \tag{2.16}$$

[37] The data is available for download at `https://www.kaggle.com/c/expedia-personalized-sort/data`.

[38] This is the final dataset. Appendix B.14.1 contains details of data cleaning and variable descriptions.

where ϵ_{ij} is type-I extreme value. $\tilde{\mathbf{x}}_\mathbf{j}$ is the vector of attributes which we always model as visible, including the chain, promotion, and position dummies.

Table 2.3: Model Specifications

Model	$\mathbf{x_j}$	z_j
I	Price, Stars, Review Score	Location Score
II	Stars, Review Score, Location Score	Price
III	Price, Stars, Location Score	Review Score
IV	Price, Review Score, Location Score	Stars

For the four remaining variables — stars, review score, location score, and price — we estimate four models with one of them as z_j, and the other three as $\mathbf{x_j}$. Table 2.3 shows the model specifications. We estimate each model using the "flexible logit" approach described in Section 2.4.2. The model is overidentified, so we construct an estimator of β using a weighted average of estimates from different moments based on bootstrapped variance estimators. More specifically, taking Model I as an example, for each bootstrap draw and each variable in $\mathbf{x_j}$, we obtain an estimate of $\frac{\beta}{\alpha_k}$, for $k \in$ {price, stars, review score}, and then examine choice sets where variation in z_j is limited to estimate α_k and thus obtain the implied $\beta_k = \frac{\beta}{\alpha_k} \cdot \alpha_k$.[39] We repeat the estimation for 250 bootstrap samples, and compute the empirical variance for each implied $\hat{\beta}_k$, denoted $var(\hat{\beta}_k)$. Then we calculate the weighted average estimate $\hat{\beta} = \sum_k w_k \cdot \hat{\beta}_k$, of each repetition, where the weights, $w_k = \frac{1/var(\hat{\beta}_k)}{\sum_k (1/var(\hat{\beta}_k))}$, are proportional to the inverse of variance so that we put less weight on less informative estimates. We use the estimates from the original dataset as point estimates, and the estimates from the bootstrapped samples to construct confidence intervals.[40]

Figure 2.3 shows the estimates from standard logit and flexible logit for each candidate z variable.[41] Location score is the only variable where we see clear evidence that standard logit is attenuated relative to flexible logit, which is consistent with the fact that location is not immediately available to consumers in the results page. Thus, standard logit tends to underestimate how much consumers value location. On the contrary, for the visible attributes – price, review score and star rating – we find that the flexible logit confidence intervals include the standard logit estimates, and the differences between flexible logit and standard logit estimates are not significantly different from zero.[42] Therefore, our approach correctly identifies that all attributes except location are immediately visible on the results page.

Given the evidence that location is a hidden attribute, we use our estimates of consumer preferences to compute how information about location will change behavior and

[39] Specifically, using Lemma 3, we compute these estimates $\hat{\alpha}_k$ by estimating a logit model using only the choice sets where the variance of the z is in the bottom decile.

[40] We use bias corrected confidence intervals (with acceleration parameter $a = 0$) to account for the skewness of distribution. Details of the construction of confidence intervals are in Appendix B.8.2.

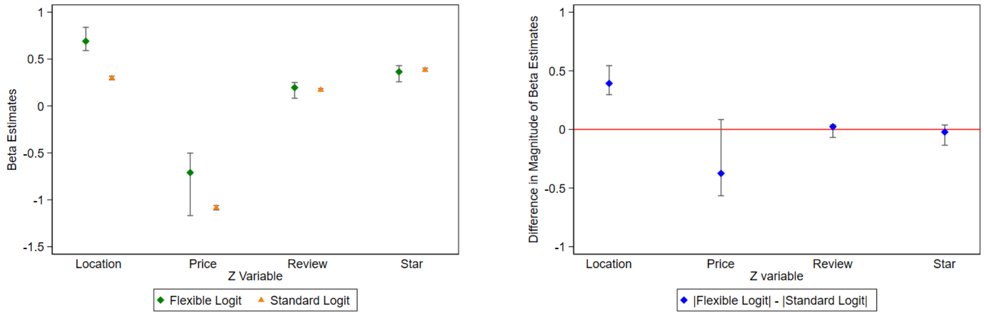
[41] Values are reported in Appendix B.14.2.

[42] For price, the standard logit point estimate is somewhat larger in magnitude than the flexible logit estimate, although the flexible logit confidence interval is wide. This occurs because price is the variable with the most variation in the data and thus the most precise estimates of $\frac{\beta}{\alpha_k}$ are obtained when $k =$ price. When price is treated as the z variable, it of course cannot be used as an x variable and therefore the flexible logit estimates of the price coefficient tend to be more noisy.

Figure 2.3: Estimation Results

(A) β Estimates

(B) Difference in Magnitude of β Estimates



Note: In Panel A, we report 95% confidence intervals for the coefficient β for different choices of z variable. In each case, we normalize the coefficient by multiplying it by the standard deviation of the variable. In Panel B, we report 95% confidence intervals for the difference between the absolute value of the normalized β estimate from flexible logit and the absolute value of the corresponding standard logit estimate.

choice quality. Table 2.4 shows the counterfactual results when we make the location score information visible to all consumers. The average location score among transacted hotels increases from 3.32 in the data to 3.40 in the counterfactual scenario where location is fully visible. Further, using the approach in Appendix B.13, we compute the welfare benefits of providing consumers with location score information. We estimate an average welfare gain of $1.93 per choice, which is 1.4% of the average transaction price.

Table 2.4: Counterfactual Results

	Status quo	Counterfactual
Average Value for the Transacted Item		
Price ($)	143.52	136.90
Stars	3.47	3.42
Review Score	4.03	4.03
Chain	0.64	0.64
Promotion	0.40	0.40
Position	4.55	4.91
Location Score	3.32	3.40
Welfare Difference per Choice ($)	-	1.93

Note: Average value of different attributes for the transacted item in the data (first column) and in the counterfactual scenario where consumers have full information on location (second column). The last row reports the average welfare change from the status quo to the counterfactual.

2.7 Which Counterfactuals Can Our Approach Address?

In the experiment and empirical application, we showed that our approach can be used to assess the impact of information interventions on consumer behavior and welfare. In this section, we discuss more broadly the class of counterfactual questions that can be addressed using our method. Additionally, we discuss how our results can be used to aid in estimation of search costs given a fully specified search model or for specification testing after such a search model is estimated.

2.7.1 Applications without Recovering Search Costs

Benefits of Full Information One important class of counterfactuals asks: how would consumers choose if search costs were reduced? The most natural counterfactuals in our baseline case involve directly informing consumers about the hidden attribute. These counterfactuals are natural in our setting because the hidden attribute is observable to the econometrician.[43] In these cases, knowing preferences is sufficient to simulate how information would impact choices without a structural search model, as we demonstrate in our lab and field experiments. In settings like Hastings and Tejeda-Ashton (2008) or Allcott and Taubinsky (2015) where experimenters fully-inform consumers about attributes of goods which were previously accessible at a financial or cognitive cost, our approach can be used to forecast the impact and value of interventions before they are conducted. Of course, our method quantifies the welfare gains from more informed choices, but not the gains directly stemming from reduced search costs. In this sense, the estimated increase in welfare can be viewed as a lower bound the total gains from an information intervention. Estimating the reduction in search costs requires either fully specifying a search model and recovering the cost distribution (see the next subsection) or using some auxiliary data on, e.g., time spent searching and value of time.

Advertising and Product Design As a second related example, consider a firm trying to understand which features to emphasize in the advertising of a product. Conditional on visible attributes, our results could be used to identify features that consumers value but are not currently always aware of. The firm could use this insight to optimize its advertising strategy, as well as to inform the design of new products (see, e.g., Bagwell (2007), Becker and Murphy (1993a)).

Normative Evaluation of Choices In many counterfactuals where limited information or search costs are not the primary object of interest, one nonetheless is concerned to accurately value attributes of goods. In footnote 6, we give the example of a tax on sugar-sweetened beverages. An alternative example is a subsidy for environmentally friendly automobiles. To evaluate such a subsidy, one would conventionally estimate demand and cost parameters in the automobile market Berry et al. (1995). If the market were otherwise competitive and efficient, the subsidy might distort choices (creating deadweight

[43]This can be contrasted with cases where information is only partial and so some search costs likely remain. For example, when unobservables are revealed by search (as in Section 2.3.6), some information consumers learn upon search is not observable to the econometrician, so informing consumers about the observable component would not eliminate the need to search.

loss) but have offsetting externalities. If, however, some consumers are unaware of differences in energy efficiency, the subsidy might redirect consumers to the products they would otherwise value if they had more information, meaning that it is both privately and socially desirable. Our methods can be used to recover whether, prior to imposing the subsidy, consumers are informed about differences in energy efficiency.

2.7.2 Applications with Search Costs

We focused above on applications where search costs do not need to be recovered. However, our model can also be used to identify search costs given preferences and an underlying structural search model. In Appendix B.12, we give an explicit example of how search costs can be recovered in a Weitzman model once preferences are known. Intuitively, when preferences are known, we know how consumers would respond to the hidden attribute with zero search costs, and thus we can trace out the distribution of search costs from the observed responsiveness of choice probabilities to the hidden attribute. There are several reasons search costs might be of interest.

Welfare with Structural Search Costs A full normative evaluation of an information intervention might directly include search costs: information may benefit consumers both by helping them make better choices and by helping them make choices more easily, and search costs quantify the value of making choices more easily. Note that structural search costs may be the wrong object to use for normative evaluation even if a structural search model performs well as a positive model of choices. For example, if consumers spend one hour choosing insurance plans and we estimate that they act *as if* they have search costs of \$1,000 per plan, this does not imply that they are made \$1,000 better off by eliminating the need to search. Search behavior may be well-described by a model with large search costs even if consumers' willingness to pay to avoid search is substantially less than the costs implied by any given model. Back of the envelope estimates of search costs based on survey data or other information on the time consumers spend choosing may often be more credible and less prone to misspecification than structural estimates (e.g. Kling et al. (2008)).

Counterfactuals with Non-Zero Search Costs Search costs may also be of interest for counterfactuals where the choice environment is altered in ways that change search behavior without eliminating search entirely. As we emphasize above, eliminating search entirely is a reasonable counterfactual in our setting where search uncovers objective information that is available to the econometrician. However, other counterfactuals may be of interest, such as changing the order in which items are presented to consumers in search. Modeling explicitly how these changes would impact search costs for different goods, and thus which goods are chosen, would require an explicit search model.

Validating Parametric Models A final reason to estimate a full structural search model is to impose parametric restrictions on the data necessary for estimation in finite samples. Our identification proof shows that, in principle, these parametric restrictions are unnecessary for identification. This is confirmed by the simulation results we presented for models with linear utility and homogeneous preferences over observables. However, when the coefficients on attributes are heterogeneous—something the empirical search literature typically rules out—estimation of the higher-order derivatives of choice probabilities necessary for nonparametric identification (see Section 2.3.2) may not be possible given the data available. In such cases, a natural approach is to specify a structural search model with random coefficients in order to place some parametric structure on these higher-order derivatives. This requires taking an explicit stand on the underlying search model. Nonetheless, once the model has been estimated and preferences recovered, the results in Section 2.2.3 can be used to conduct specification tests. If these tests reject, an alternative search model may fit the data better.

2.8 Conclusion

We prove that it is possible to estimate preferences using only data on attributes and choices in cross-sectional or panel data even when consumers must search to acquire information about product attributes. This result holds in a broad class of search models. The functions of choice probabilities which identify preferences in our model are "robust" in the sense that they work in both full information and search models. Further, our results can be used to test whether consumers are fully or only partially informed about a given attribute.

Because our conditions allow preferences to be recovered when consumers are imperfectly informed, our results allow a wide range of inquiries that are impossible using conventional methods. Prior to conducting an information intervention, choice data can be used to estimate counterfactually how consumers would choose were they fully informed. If preferences are not informed, the preferences consumers would have if they were informed can be used to conduct a more defensible welfare analyses.

Preferences can (sometimes) be identified in structural search models, but such models require making many explicit assumptions about how consumers search. Do consumers consider option value or are they myopic? Do they solve an optimal stopping problem or search until they find a good enough option? Is search sequential or simultaneous? If search costs vary across consumers, what is their statistical distribution? Our approach attempts to avoid these complexities by instead relying on a sufficient condition satisfied in a broad class of search models that can be falsified by the available data via bounds on choice probabilities and overidentification tests.

In many settings, one can conceive of reasons that the visible utility assumption would fail, but it must be assessed relative to the alternatives. The vast majority of empirical work currently makes the often dubious assumption that consumers are fully informed about all attributes of products. Even if one lacks contextual information to support the visible utility assumption, our approach is much weaker than the standard assumption of full information and may be preferable in settings where preferences are needed to conduct welfare analysis. The main downside of our approach relative to full information is statistical power, but this concern is less relevant given rich microdata which is increasingly available. In settings where one would otherwise make many untested structural assumptions about search, visible utility may be more parsimonious and leads to clear, testable predictions.

Our assumptions are sufficient for identification but not necessary. This raises several questions for future research: are there other conditions aside from the visible utility assumption which permit analogous data-driven identification of consumers who maximize utility? Are there necessary and sufficient conditions for preferences to be recoverable from choice data when consumers have partial information?[44]

Increasingly, empirical analyses relax the assumption that consumers make informed choices. Typically, behavioral welfare analysis is done using auxiliary data, restrictions on preferences, or by testing whether consumers choose differently when provided with information. Despite this, absent data to the contrary, the default assumption in most economic analysis remains that consumers make informed choices. Our result suggests this need not be the case. Even with no auxiliary data, researchers can use observed choices both to test whether choices are informed and to recover what preferences would be were consumers more informed. This removes the (often compelling) excuse that while consumers may not be informed, assuming informed choices is the only constructive way to proceed given the data available.

[44]This question parallels Falmagne (1978)'s derivation of necessary and sufficient conditions for choice probabilities to be rationalizable by utility maximization given full information. Our question differs by relaxing the full information assumption.

Chapter 3

Variety Seeking in High-Frequency Consumption: New Implications for Targeted Marketing

[1]

3.1 Introduction

Targeted marketing has become a widespread business practice with the arrival and rise of the Internet. Firms now have access to an unprecedented amount of individual consumer data such as demographics (Ansari et al., 2000) and purchase history (Rossi et al., 1996; Villas-Boas, 1999), and then target them with marketing activities based on the information. The quality of the targeted marketing strategy depends on which information firms use. Rossi et al. (1996) find that even a short purchase history of consumers can improve the profitability of marketing more than using demographic information, highlighting the importance of utilizing purchase history data. Using purchase history makes it possible to identify and utilize persistent consumer heterogeneity that is not captured by demographic information.[2] Another potential benefit of using purchase history is to account for the dynamics in demand or serial correlation of demand across different time periods. Commonly studied sources of dynamics in demand include inventory effects (Hendel and Nevo, 2013; Johnson et al., 2013; Gabel and Timoshenko, 2022) and intertemporal price discrimination in durable goods markets (Gowrisankaran and Rysman, 2012).

In this paper, we highlight consumers' variety-seeking preferences as an important source of dynamics in demand that can be estimated from purchase history data and utilized for targeted marketing. Specifically, we emphasize the dynamics in demand resulting from consumers' negative state dependent preferences, which result in a decrease in demand following a recent purchase. Failing to account for such dynamics in demand in targeted marketing practices can decrease consumer welfare and the effec-

[1] This chapter is based on joint work with Carol Hengheng Lu and Zhijie Lin.

[2] This application has been the focus of a large literature on personalized pricing, including Besanko et al. (2003); Pancras and Sudhir (2007); Howell et al. (2016); Morozov et al. (2021); Smith et al. (2021); Donnelly et al. (2021).

tiveness of firms' marketing policies.[3,4] We document a substantial fraction of consumers with variety-seeking preferences, and then study the implications of consumers' variety-seeking preferences for targeted ranking and competitive targeted pricing. To the best of our knowledge, this is the first paper to study the implications of state dependence for targeted ranking and competitive targeted pricing.

We use proprietary data from a peer-to-peer (P2P) homemade food delivery platform. Our context provides several advantages that enable us to examine variety-seeking preferences and their implications for targeted marketing. First, food is consumed frequently with clearly defined daily demand. This eliminates the typical issues of long purchase cycles, durability, and stockpiling with consumer packaged goods. Second, since the platform is a P2P homemade food platform, the products are uniquely available through this channel; the outside options are guaranteed to exclude consumption of these products. In traditional packaged goods markets, we can not distinguish between consumers having access to the same product from other unobserved channels and no purchase. Moreover, most of the orders are workday lunch orders. This removes the potential for intrahousehold heterogeneity (Che et al., 2003) to confound measuring variety-seeking preferences with scanner data. Finally, we observe marketing variables including pricing and ranking in our data, which enable us to estimate the effect of these marketing instruments and conduct corresponding counterfactual analysis when we optimize them taking into consideration consumers' variety-seeking preferences.

We first report motivating evidence on consumers' variety-seeking preferences. In our data, the probability that a consumer switches to a different seller from their previous seller choice is 81%, which stands in stark contrast to the switching probability found in the existing literature. To better account for alternative driving forces of switching probability besides variety-seeking preferences, such as choice set size and heterogeneous preferences, we use a permutation-based test and show that consumers in our data switch across different kitchens more frequently than the switch frequency in a random reshuffle benchmark, in which the choice sequences are randomly reshuffled with market shares fixed. This provides strong suggestive evidence of consumers' variety-seeking preferences.

We then model consumers' discrete demand for kitchens allowing for negative state dependence. We estimate a random coefficient multinomial logit model to measure variety-seeking preferences controlling for heterogeneous preferences, product attributes, and marketing variables. We find that, on average, consumers exhibit variety-seeking preferences at the kitchen level and consumers are willing to pay 19.9% more to switch to a different seller. We find rich heterogeneity in consumers' state-dependent preference: 67.2% of consumers have variety-seeking preferences, and 32.8% of consumers have inertia. We then rule out alternative sources of state dependence that result from information gained from consumption instead of changes in utility, including learning and new product discovery.

We use our demand estimates to study the implications of variety-seeking preferences for two important aspects of targeted marketing: targeted ranking and targeted pricing.

[3]See Appendix C.5.2 for examples of personalized recommendations that push recommendations on favorite sellers ordered very recently. Such practices emphasize persistent heterogeneous preferences, but have the hazard of potentially neglecting the dynamics in demand driven by state dependent preferences.

[4]Another example of inefficient targeted advertising because of neglect of negative state dependence is retargeting: `https://www.digitaltrends.com/features/wh-do-you-see-ads-stuff-already-bought/` shows a case where a recent past consumption is followed by a decrease in demand, and failing to acknowledge this in retargeting advertising decreases both profits and consumer welfare.

For targeted ranking, we examine profits, welfare, and purchase probability under several different counterfactual ranking schemes. The ranking algorithm in the data does not consider consumers' personal purchase history, and is based on price, distance, monthly sales, and ratings. In order to study the effect of personalized ranking, we examine two alternative ranking schemes: optimal ranking and suboptimal ranking. Under optimal ranking, the platform ranks kitchens in decreasing order of expected utility, where the expected utility is calculated incorporating the variety-seeking preference. Under suboptimal ranking, the platform ranks kitchens in decreasing order of expected utility, where the variety-seeking term is omitted in the expected utility calculation. We use random ranking as a benchmark to evaluate the performance of each ranking algorithm. We measure the effect of variety-seeking preferences on targeted ranking by the following ratio, $\frac{Optimal-Suboptimal}{Optimal-Random}$, where the numerator is the difference between the optimal and the suboptimal rankings and the denominator is the difference between the optimal and random rankings. This ratio measures variety-seeking effect as a percentage of the total ranking effect. We find that optimizing ranking algorithms with variety-seeking preferences comprises 18.2% of the revenue improvement, 14.2% of the consumer welfare improvement, and 18.9% of the purchase probability improvement out of the total ranking effect.

We next use our demand estimates to study how variety-seeking preferences affect competitive targeted pricing. Theoretical research generates ambiguous predictions on the effect of variety-seeking preference on price competition (Seetharaman and Che, 2009; Sajeesh and Raju, 2010; Zeithammer and Thomadsen, 2013), which suggest that the effect of variety-seeking preferences on price competition is an empirical question. We start from studying how optimal targeted pricing strategies affect equilibrium prices, profits, and consumer welfare at the estimated variety-seeking level. We find that optimal targeted pricing results in lower prices to own consumers and higher prices to rival kitchens' consumers. In general, the average transaction price level increases by 6.82%, total profits increase by 5.18%, and consumer welfare decreases by 9.12%. We then study more general targeted pricing implications on price competition when we change the level of consumer variety-seeking preferences. We demonstrate the directions of pressures on prices resulting from the kitchens' current period and future profit incentives when facing variety-seeking consumers. We find that variety-seeking preferences soften price competition and leads to a higher average equilibrium price compared to the no state dependence case. Average equilibrium prices and profits increase with the level of variety seeking, and consumer welfare decreases with the level of consumer variety seeking. Moreover, the prices to a kitchen's own consumers decrease with variety-seeking level while the prices to rival kitchens' consumers increase with variety-seeking level. These findings have important implications for competitive targeted pricing and suggest that kitchens could increase profits by offering their own consumers coupons for quantity discounts. However, under current common industry practice, popular food delivery platforms do not enable merchants to implement this type of state-based targeted pricing.[5]

This paper contributes to several strands of literature. It adds to the literature on consumers' state-dependent preferences. There is a large literature on positive state dependence in economics and business studies (Dubé et al., 2010; Osborne, 2011; MacKay and Remer, 2019; Kong et al., 2022), mainly using consumer scanner data from grocery shopping. We find little work on consumers' variety-seeking preferences in economics. Meanwhile, variety seeking has been an important topic in marketing. McAlister (1982) studies variety-seeking preferences for soft drinks using a dynamic attribute satiation

[5]Section C.5.1 provides more details about the current targeted pricing industry practice.

model.[6] Subsequent research on variety seeking mostly uses stochastic choice models (Givon, 1984; Kahn and Raju, 1991; Trivedi et al., 1994; Chintagunta, 1998), or a hybrid version of stochastic choice models and utility-based models (Bawa, 1990; Seetharaman and Chintagunta, 1998; Chintagunta, 1999; Park and Gupta, 2011; Smith et al., 2020). In this paper we use a utility-based choice model, to better accommodate the variation in substantial marketing variables, choice sets, product attributes, and unobserved heterogeneity. Using a utility-based choice model also enables us to do counterfactual predictions and welfare analysis. Combined with our high-frequency food consumption data including various marketing variables, we generate novel implications of variety-seeking preferences on personalized ranking and pricing on online platforms.

There are different definitions of variety in the existing literature (Van Herpen and Pieters, 2002). The definition of variety seeking in our project is negative state dependence: a disutility from consuming the same product consecutively. We study consumers' preferences for dynamic variety over time where the order of the choice sequences matters. This is different from some other preferences for variety in the existing literature, such as the size, dispersion level, and diversity of the choice set (Brynjolfsson et al., 2003; Broda and Weinstein, 2006; Quan and Williams, 2018; Datta et al., 2018; Ershov, 2020; Holtz et al., 2020; Natan, 2020). For these definitions of variety, the order of the choice sequence does not matter, and consumers' choices are aggregated along the time dimension.

With the development of digital platforms and more individual level ranking data available, there is active literature focusing on improving ranking algorithms from different perspectives (Jeziorski and Segal, 2015; Yoganarasimhan, 2020; Compiani et al., 2021; Derakhshan et al., 2022). Our paper is the first paper to optimize the ranking algorithm with state-dependent utility. We highlight consumers' variety-seeking preferences as an important element in ranking that will affect consumer welfare, purchase probability, and platform revenue. For the pricing implications of state dependence, Che et al. (2003) study the forward-looking pricing behavior in markets with state dependence within a two-period model. Dubé et al. (2008) and Dubé et al. (2009) study the pricing implications of switching cost with infinite horizon, under monopoly and oligopoly market structures, respectively.[7] Zhang and Krishnamurthi (2004) and Zhang and Wedel (2009) study the implications of consumers' state dependent preferences for targeted pricing in a three-period off-equilibrium analysis, where the target brand is allowed to conduct targeted pricing while keeping other brands' prices fixed. We complement this literature by studying an infinite horizon price competition game with variety-seeking consumers, allowing the kitchens to set different prices to own and rival's consumers. Our paper is the first paper to study competitive targeted pricing with state-dependent utility.

The rest of the paper proceeds as follows: Section 3.2 describes the data and background, Section 3.3 discusses some motivating evidence of variety-seeking patterns in our data, Section 3.4 and 3.5 provides details of the structural model and estimation, Sections 3.6 discusses potential confounding factors and alternative explanations. Section 3.7 reports results from counterfactual analysis of targeted ranking and targeted pricing, and Section 3.8 concludes.

[6] McAlister and Pessemier (1982) provides a review of the different psychological factors behind the variety-seeking behavior, such as satiation, boredom, experience seeking, or thrill and adventure seeking. In this project, we do not try to differentiate these psychological factors; instead, we summarize them together by negative state dependence preference.

[7] Theoretical work on price competition with switching cost include Farrell and Klemperer (2007) and Villas-Boas (2015).

3.2 Data

We use proprietary data from a P2P online food ordering platform for this analysis. The platform, the first P2P homemade food delivery platform in China, was founded in Beijing in 2014, connecting households offering home-cooked food with nearby consumers. Until 2017 there were more than a million buyers who ordered food from sellers via the mobile app, and twelve thousands sellers who were mostly homemakers and the retired.[8]

The data include detailed records of all orders placed by buyers on the platform in Beijing from November 2016 to January 2017. For each order, we observe the identity and location of the buyer and seller, time of order, order fee, dishes ordered, delivery fee, coupon offered, and rating and review for the order. We also observe the ranking of kitchens shown on consumers' homepages and the final purchase outcome. We focus on the frequent users who ordered more than 4 times per month with a duration on the platform of more than 3 months.[9]

Because food consumption happens frequently on a daily basis, a natural definition of one time period is one calendar day. Even though we select the frequent users of the platform for analysis, there are still 74% of periods when consumers did not order from the platform. In those 74% periods when consumers chose the outside option, 27% are periods when consumers opened the platform;p in those cases, we observe the choice sets in their rank lists. For the other 47% of periods, we do not observe any platform usage. The periods when consumers opened the app and checked out the kitchen list are important from the platform's perspective even though consumers did not order from the platform, since the pricing and ranking strategies in these periods may have a substantive effect on the extensive margin. For the periods when consumers did not open the platform, consumers could obtain food from restaurants, cooking at home, or other food delivery platforms. We assume that there was no demand for the platform during these periods when consumers did not open the platform and drop these periods from the analysis. Meanwhile, we keep the periods when consumers opened the platform but eventually chose the outside option. We assume that there was potential demand for the platform in these cases and the outside option is different from the kitchens on the platform. The following example 3.1 illustrates how we treated the two kinds of outside options when constructing the choice panel from the full panel. In the example, A, B, C, D represent different kitchens on the platform, and O represents the outside option.

$$ABA \quad \underbrace{OOOOO}_{\text{use platform, no order}} \quad OOOCA \quad \underbrace{OOOOOOOOOO}_{\text{no platform usage, no order}} \quad AD \longrightarrow ABAOOOOOOCAAD \tag{3.1}$$

The final sample contains 6,629 consumers, 6,664 kitchens, and 125,646 orders. One consumer on average placed 18.95 orders during the three-month period, and ordered 2.91 times a week conditional on ordering that week, mostly on weekdays and for lunch. There are on average 14.04 kitchens available nearby for each consumer per day, and an average consumer ordered from 9.28 distinct sellers throughout the three-month period. Detailed summary statistics are shown in Table 3.1.

[8]Appendix C.1 provides more background information about the market structure of the food delivery industry, consumers' daily and weekly order trends, app display, and kitchen cuisine distribution.

[9]These frequent users take 10% of the consumers and 50% of the orders and revenues.

Table 3.1: Descriptive Statistics

	Mean	Std. Dev.	10 Pctile	Median	90 Pctile
Consumer-Level Data					
No. Platform Uses	40.03	15.22	22	38	62
No. Orders	18.95	8.20	13	16	30
No. Orders per Week	1.91	0.80	1.09	1.70	3.01
No. Orders per Week Conditional on Ordering	2.91	0.94	2	3	4
Prob. of Week with Ordering	0.83	0.13	0.67	0.83	1
Gender (M=1)	0.50	0.50	0	0	1
No. Kitchens (Daily Average)	14.04	5.70	6	15	31
Distinct Kitchens Ordered	9.28	5.21	3	7	15
Distinct Cuisines Ordered	3.65	1.91	1	3	6
Kitchen-Level Data					
Rating (1-5)	4.89	0.15	4.49	4.87	5
Menu Size	25.32	9.81	3	23	49
Average Dish Price (CNY)	25.21	9.45	14.29	23.32	29.15
Average Price of a Basket of Food (CNY)	50.23	17.32	26.19	50.86	58.49
No. Orders per Week	18.76	7.64	1.72	10.40	48.25
Delivery Radius (km)	2.27	1.31	2	3	9
No. Days Available per Week	3.96	2.06	1	4.23	6.46
No. Choice Sets Daily Conditional on Availability	17.61	15.96	3.67	13.07	37.01
Order-Level Data					
Order Payment (CNY)	45.46	22.37	22	48	68
Delivery Fee (CNY)	2.89	0.77	0	3	3
Coupon Value Applied (CNY)	4.36	4.77	0	4	15
Distance (km)	1.68	0.99	1.54	2.43	2.76
No. Dishes per Order	2.34	1.51	1	2	4
No. Consumers	6,629				
No. Sellers	6,664				
No. Orders	125,646				
No. Platform Usage	265,375				

Note: The table presents summary statistics of our sample. We focus on frequent users who ordered more than 4 times per month with at least 3 months of tenure on the platform. We report buyers and sellers' activities from November 2016 to January 2017.

3.3 Motivating Evidence

In this section, we provide motivating evidence of consumers' variety-seeking preferences. As variety seeking is about the demand dynamics across different time periods, we study the switching probability as a measure of consumer choice variety. We find that in our data consumers stay with the same kitchen they chose in the last period only 19% of the time, for 36% of the time they switched to a different kitchen on the platform, and for 45% of the time they switched off the platform.[10] The switching probability of 81% in our data stands in stark contrast to that found in the existing literature, which normally finds a repurchase probability of 80-90% (Dubé et al., 2010). However, the switching probability is also affected by the size of choice sets and the relative magnitude of consumers' persistence preferences. A switching probability of 81% has different implications when consumers face only two options each period, and the case when consumers face hundreds of options. Switching away from a favorite kitchen is also more informative than switching away from a less preferred kitchen. Thus in order to have a clearer evaluation of the switching frequency, we construct a permutation based benchmark to compare the switching frequency we observed in the data with that when the order of choices does

[10]As the products on our platform are homemade food only available through the platform channel, instead of standardized branded food, it is reasonable to assume that people consumed different products from the ones on the platform when they chose the outside option.

Table 3.2: Example of Permutation Test

Observed Choice Sequence	Switching Probability	Permutations	Switching Probability
		CBAAA	0.5
		ABAAC	0.75
ABACA	1	BCAAA	0.5
		ACABA	1
		CABAA	0.75
		...	...

Note: The table illustrates how the permutation test is implemented with an example. Given an observed choice sequence in the data, we hold fixed the market shares and randomly reshuffle the order of choices. The distribution of switching probability from the permutations recovers the distribution of switching probability under the null hypothesis that there is no state-dependent utility and the order of choices does not matter.

not matter given the observed market shares.[11]

To account for consumers' persistent kitchen preference, we construct a permutation based switching probability benchmark by randomly shuffling the order sequence of each consumer 100 times while holding fixed the market shares for each kitchen ordered. Table 3.2 presents an example of how the permutations are implemented.

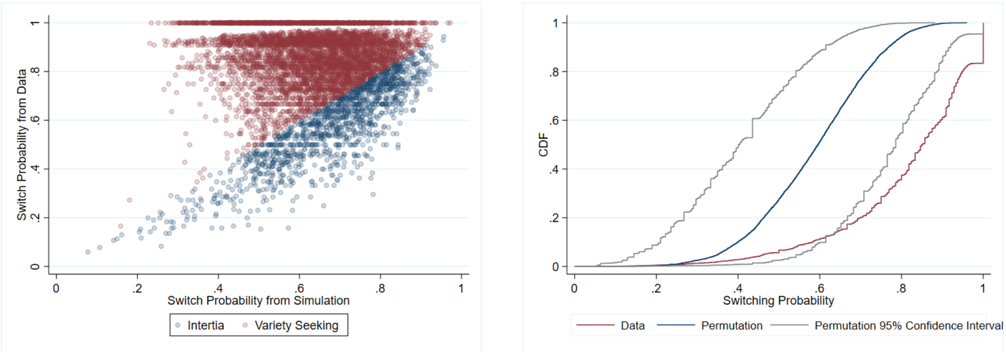
We compare the switching probability observed in the data and the average switching frequency in the randomly shuffled benchmark. The distribution is shown in Figure 3.1. Panel (A) presents the scatter plot with each dot representing one consumer. For consumers with actual switching probability greater (less) than the randomly reshuffled benchmark level, they switch more (less) frequently in the data than the case when the order of choices does not matter, and exhibit variety-seeking preferences (inertia). A majority of consumers exhibit variety-seeking preferences (79%), and a minority of consumers have inertia (21%). The figure also suggests that there is heterogeneity in consumers' state dependence preferences, and we formally model the heterogeneity through a random coefficient model in the structural estimation. Panel (B) presents the cumulative distribution function of the switching probability in data, in the random reshuffled benchmark, and the 95% confidence bands of the reshuffled sequences. For most parts of the distribution the observed switch probability lies well outside the 95% confidence bands of the random reshuffled distribution, suggesting that the switch probability observed in the data is not simply from randomness in choices. The mean switch probability in the random reshuffled benchmark is 64%, with a 95% confidence interval of [47%, 79%], which suggests that the consumers switches 17% more than a simple random choice. The average switching probability across consumers indicates variety seeking with a significance level $p < 0.01$.[12]

[11] Appendix C.2.1 presents details of the statistical test. Similar permutation based test is used by Miller and Sanjurjo (2018). By permutations, we preserve the randomness in choices and the number of orders for each chosen kitchen in the choice sequences while removing the state dependence. If the observed switching probability is not significantly higher than the random permutation switching probability, then we can't reject the null hypothesis that consumers don't have state dependent preferences. Such tests of checking for an association that should be present if the design is flawed but not otherwise are called placebo tests by some existing literature (Larsen et al., 2021). See Eggers et al. (2021) for a review of placebo tests.

[12] We conduct this pooled test with the average of standardized switching probability $\bar{p} = \frac{1}{N}\sum_i \tilde{p}^i$, where, for each consumer, the switching probability $\tilde{p}^i$ is normalized by shifting its mean and scaling its standard deviation under H_0. In this case: $H_0 : \mathbb{P}(y_t \neq s_t | s_t \neq 0) = p^i$ for all t, i, where y_t denotes current period choice and s_t denotes previous period choice. We stratify the permutations by consumer

Figure 3.1: Motivating Evidence

(A) Switching Probability Distribution (B) Switching Probability CDF



Note: Panel (A) shows the switching frequency observed in the data relative to that in the randomly reshuffled benchmark, where the order of choices does not matter given the observed market share. One dot represents one consumer. The 45-degree line represents the cases when consumers switch as frequently as the randomly reshuffled benchmark. The consumers above the 45-degree line switch more frequently in the data than the random reshuffle benchmark and exhibit variety-seeking preferences, whereas the consumers below the 45-degree line switch less frequently than the randomly shuffled benchmark and exhibit inertia. Panel (B) presents the cumulative distribution function of the switching probability in data, in the random reshuffled benchmark, and the 95% confidence bands of the reshuffled sequences.

The analysis above is based on stochastic choice models and does not account for price, availability, changing of the choice set, and other product attributes such as cuisine type and menu size, or time-related effects such as day of the week, week of month, and holiday effects. To formally control for these confounding factors, to measure consumers' state dependent preferences, and to further conduct counterfactual predictions, we need structural modeling.

3.4 Structural Model

We introduce the structural model in this section. We assume that consumers maximize their utility given the available choice set, which we observe from the search result data.[13] The indirect utility that consumer i derives from kitchen j on platform in period t is

$$u_{ijt} \quad = \quad \beta_{ij} - \alpha_i p_{jt} + \eta r_{ijt} + \gamma_i \mathbb{1}(s_{it} = j) + \gamma^C \mathbb{1}(s_{it}^c = c_j) + \psi \cdot X_{ijt} + \epsilon_{ijt}. \qquad (3.2)$$

The utility from choosing the outside option is $u_{i0t} = \gamma^O \mathbb{1}(s_{it} = 0) + \epsilon_{i0t}$.

We use $\beta_{ij} \sim N(\mu_j, \sigma_j^2)$ to capture the match value of consumer i and kitchen j. The mean μ_j captures the unobserved quality of the kitchen, and the variance allows for heterogeneity across different consumers' tastes, which is centered around the mean quality. Different kitchens are allowed to have different mean quality and dispersion

to recover the sampling distribution of average switching probability under null hypothesis. Moreover, a Kolmogorov-Smirnov test comparing the actual CDF of switching probability to the CDF of the mean switching probability from permutations yields a p¡0.01.

[13]We keep the periods when consumers ordered from kitchens on the first page of search results, which account for 92% of all orders. We drop the purchases on pages beyond the first page because of the potential selection problem: consumers who search beyond the first page are more likely to be searching for a specific niche kitchen, and the selection effect will bias the estimates of position cost.

of tastes. Price sensitivity is captured by $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$, we use random coefficient to allow for heterogeneity across consumers. Since our main focus is on the state dependence preference, we use a position cost to capture the ranking effect instead of imposing a full structural search model: r_{ijt} is the rank of kitchen j in consumer i's rank list in period t, and η represents the position cost from choosing a lower-ranked kitchen.[14]

We have three state dependent terms capturing consumers' state dependent preferences on different levels. In each period t, consumer i has chosen one option in the last period, captured by state variable $s_{it} \in \mathcal{J}_{it}$. The choice set $\mathcal{J}_{it}$ includes the outside option (represented by option index zero) and the observed kitchen list for consumer i in period t. If the consumer chose option j in the last period, $s_{it} = j$, and subsequently purchases product, $k \in \mathcal{J}_{it}$, then his or her state becomes, $s_{i(t+1)} = k$.[15] The state variable s_{it}^c represents the cuisine type consumer ordered last period. The cuisine type of kitchen j is c_j. The coefficient $\gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2)$ captures the state dependence when consumer i chooses the same kitchen on the platform. The coefficient γ^O captures the state dependence when consumer i stays off the platform, and γ^C captures state dependence for the cuisine.[16] A positive state dependence coefficient suggests inertia, and a negative state dependence coefficient corresponds to variety seeking. The covariate vector X_{ijt} includes the menu size, cuisine, day of the week, week of the month and holiday fixed effects. The idiosyncratic taste shock ϵ_{ijt} is drawn from type I extreme value distribution.

3.5 Estimation

This section provides our parametric assumptions, estimation strategies, and identification inference.

3.5.1 Likelihood Function

We use the simulated maximum likelihood method to estimate the model. We assume that ϵ_{ijt} is distributed Type I Extreme Value and is independently and identically distributed (i.i.d) across consumers, time, and kitchens. We use U_{ijt} to represent the deterministic part of the utility, i.e., $u_{ijt} = U_{ijt} + \epsilon_{ijt}$. Then the probability of consumer i choosing option j is

[14] See Ansari and Mela (2003),Ghose et al. (2014), and De los Santos and Koulayev (2017) for examples of a similar method to capture the ranking position effect. Partial information models in which previous period choice has a higher probability of entering the consideration set, or a lower search cost, will imply that our estimate of variety seeking is conservative, since the information frictions work in favor of the previous period choice because memory and salience of the recent choice is strongest among all options. Appendix C.3.6 reports results on the out of sample fit of this model specification.

[15] As we study high-frequency food consumption, our definition of the state variable when consumers chose the outside option is different from common practice in the literature, such as Dubé et al. (2010), which often assume that the state variable does not change when the consumer the chooses outside option. In the context of our data, it is more reasonable to assume that consumers have lunch every day and have demand for the platform on the days when they open the platform, and that the outside option is different from the kitchens on the platform. As our platform was the only P2P platform at the time, the products were exclusively available through the channel of the platform. This is an advantage compared with platforms where the sellers are restaurants that can be multi-homing, such as Uber Eats and DoorDash.

[16] γ^O is included mainly for the purpose of separating the state dependence parameter for kitchens on the platform and the outside option, since there are sequences of consecutive outside option choices in our data that could confound the estimates of γ_i if we treat the outside option as one kitchen.

$$\mathrm{P}_{ij} = \frac{\exp\left(U_{ijt}\right)}{\sum_{k\in\mathcal{J}_{it}} \exp\left(U_{ikt}\right)}. \tag{3.3}$$

The likelihood function for a sequence of consumer i' purchase decisions $\mathbf{y_i^*}$ is

$$\mathcal{L}_i\left(\mathbf{y_i^*} \mid \theta\right) = \frac{1}{R}\sum_{r=1}^{R}\left(\Pi_{t\in T}\Pi_{j\in\mathcal{J}_{it}}\left(\frac{\exp\left(U_{ijt}^r\right)}{\sum_{k\in\mathcal{J}_{it}}\exp\left(U_{ikt}^r\right)}\right)^{y_{ijt}^*}\right), \tag{3.4}$$

where r is the index for draws of random coefficients. We use 100 draws in the simulations.

3.5.2 Identification

The main challenge of evaluating past choices' effect on current choices is to disentangle consumers' heterogeneous preferences and structural state dependence. To solve this problem, we first control for the unobserved heterogeneity by adding the random coefficients β_{ij}, which captures both vertical heterogeneity across kitchens by the mean quality and also horizontal taste heterogeneity by the dispersion level of taste distribution.[17] We assume that the unobserved heterogeneity is stable over time, which is a reasonable assumption since we focus on a relatively short time period. Thus the persistent unobserved heterogeneity is well captured by the random coefficient β_{ij}. The identification of the state dependence parameters separately from heterogeneous preferences then comes from the variation of purchase probability right after a previous purchase induced by a rich set of frequent exogenous shocks to availability, prices, and rankings over time in our data.

Availability The special composition of the supply side of the P2P platform provides exogenous shocks to kitchens' availability. Since the kitchen owners on the platform are mostly homemakers and retired workers, they make flexible decisions of availability based on personal life arrangements, such as childcare and travel. The median number of orders per kitchen per week is 10; it is reasonable to assume that the kitchens are not making sophisticated strategic decisions on availability given the sales volume, potential profit margin, and the lifestyle of kitchen owners.

Price The variation in prices in our data comes from differences across kitchens, different choices of menu items, menu changes, delivery fees, and promotional discounts. The potential endogeneity of time-persistent price variation across different kitchens is captured by β_{ij}. We discuss the price variation over time caused by menu changes, delivery fees, and promotional discounts below. We do not model consumers' dish-level choices. Instead, we observe the kitchen's menus over time and construct the price variable based on the average price of a basket of food at the restaurant. Specifically, we calculate a weighted average of menu prices, with each item price weighted by the aggregate popularity. Sometimes kitchens change their menu by adding or deleting dishes and adjusting prices. When kitchens need to change menus permanently, they need to

[17] As there are 6,664 kitchens in our data, it is intractable to allow for a different distribution for each kitchen. To deal with this dimensionality problem, we follow Bonhomme et al. (2019), Bonhomme et al. (2022) and use kmeans method to group all kitchens into 16 clusters based on observed average purchase probability, prices, and kitchen characteristics. Kitchens in each cluster share the same distribution of β_{ij}. The grouping brings approximation error but reduces incidental parameter bias. We restrict the approximation error by following the data-driven rule provided in Bonhomme et al. (2022). The reduction in incidental parameter bias comes from two sources: utilizing more informative moments about the latent types, and regularization.

schedule an in-person check with the platform before the change is implemented. As other active kitchens also need to be checked in person at least every three months, kitchens can not control or predict the timing of menu changes, therefore we assume that the menu changes are exogenous. We control for the menu size to capture the potential diversity variation from menu changes. Beyond the menu prices (tax included), we added the delivery fee. Kitchens set a delivery radius within which they deliver food to consumers and set a delivery fee.[18]Thus variation in delivery fee over time is generated by variation in consumer' geographic location and distance to the kitchens, and is assumed to be orthogonal to food consumption utility. The platform has platform-level promotions such as Tuesday half-price, free delivery on Friday, and coupons on every 17th day of the month. These promotions are provided by the platform to all consumers and could be potentially correlated with demand. We controlled for day of week, week of month, and holiday fixed effects to deal with the potential endogeneity problem. The frequent temporary platform level sales, delivery fee variations, and kitchen level menu adjustments provide exogenous variation in prices after controlling for day of the week, week of the month, and holiday fixed effects, which identify consumers' price sensitivity.

Rank The ranking algorithm in our data is not personalized based on individual consumption history. Instead, the default ranking is based on distance, price, monthly sales, and ratings. The time-persistent unobserved heterogeneity is captured by the random coefficient β_{ij}. The remaining factors that vary over time and could affect rank are availability, new entries, distance, ratings, price, and monthly sales. Since our platform is a P2P platform, some kitchens may be unavailable on certain days and change the relative ranking positions of the choice set. There can also be some new kitchens entering the market, which could change the ranking positions of existing kitchens. We assume these changes are exogenous for the reasons discussed above. Geographic distance between consumers and kitchens varies as consumers change location during the period. We assume that geographic distance does not enter utility and provides an exogenous shock to rankings over time that identifies the position effect. Monthly sales volume is an aggregate variable that captures the population choice in past periods. The variation in monthly sales over time comes mostly from kitchen availability decisions. We don't include it in the utility function and assume that it is exogenous after controlling for the kitchen fixed effects. The rating variable has very limited variation across kitchens and across time periods. The median rating is 4.87. The median of the ratio of kitchens' overtime rating standard deviation to mean rating is 0.07%. We don't include rating in the utility function as the limited variation makes it less informative given kitchen fixed effects. The frequent temporary kitchen availability changes, new entries, and geographic distance changes provide exogenous variation in ranks after controlling for day of week, week of month, and holiday fixed effects, which identify consumers' position cost. We perform a robustness check on rank endogeneity by adding distance and ratings in the utility function and discuss more details on potential rank endogeneity in Section 3.6.3.

[18]The sellers can choose to deliver by themselves, use the platform's delivery service, or use their own third-party delivery service. In all cases, the consumer needs to pay the delivery fee. If sellers choose to deliver by themselves, they do not need to pay extra delivery cost to platform or third-party delivery service. We do not observe great variation in delivery fee across orders, and we assume the heterogeneity in delivery fee is driven by supply side cost heterogeneity such as transportation cost of the sellers.

3.5.3 Estimation Results

The estimation results are in column (1) of Table 3.3. The price coefficient is negative and significant.[19] The implied position cost is 3.9% of the average payment per order. We find γ_i has a negative mean -0.1981 and a standard deviation of 0.4450. That suggests that 67.2% of the consumers are variety seeking, with $\gamma_i < 0$; 32.8% of the consumers have inertia, with $\gamma_i > 0$. The heterogeneity level estimated from structural estimation is similar to what we see in the reduced form evidence in figure 3.1 panel (B). The ratio of the mean of γ_i to α_i represents consumers' willingness to pay for variety seeking, which is 19.9% of the average payment per order. We do not find significant levels of state dependence at the outside option level or at the cuisine level.[20] Section C.2.2 provides the results of permutation tests on the switching probability of cuisine types. We find that a majority of consumers(64.5%) switch cuisines less frequently in the data than in the random reshuffle benchmark, and the CDF of cuisine switching probability in data is within the 95% confidence interval of the permutation CDF. The results are consistent with the positive and insignificant cuisine level state dependence estimate we get in the structural estimation.

[19] The average own price elasticity is 1.3, which is inelastic compared with estimates using data from related industries such as Natan (2020). The difference is likely from the monopoly nature of the platform. In Natan (2020), there is competition on both the platform and seller level, buyers and sellers can both be multihoming. Whereas in our setting, the platform is a monopoly in the homemade food subcategory, and the food is exclusively available from this channel, which enables the platform to price at the more inelastic region of the demand curve.

[20] The lack of evidence of variety seeking on the cuisine level can result from categorization method. Figure C.1.5 presents the distribution of kitchen cuisines in our data, which is based on geographic regions. The estimation result suggest that at the regional cuisine level, the persistent preference dominates, and consumers are switching across different kitchens within their favorite cuisine to seek variety.

Table 3.3: Estimation Results.

		(1) Baseline	(2) Learning	(3) New Product	(4) Rank Robustness
State Dependence –Kitchen level (γ_i)	μ_γ	-0.1981***	-0.1796***	-0.1847***	-0.1713***
		(0.0503)	(0.0487)	(0.0502)	(0.0491)
	σ_γ	0.4450***	0.7829***	0.6979***	0.6797***
		(0.0635)	(0.0609)	(0.0595)	(0.0732)
Price (α_i)	μ_α	-0.0219***	-0.0213***	-0.0226***	-0.0207***
		(0.0036)	(0.0032)	(0.0051)	(0.0042)
	σ_α	0.0107***	0.0105***	0.0102***	0.0101***
		(0.0011)	(0.0013)	(0.0022)	(0.0019)
Rank		-0.0391***	-0.0328***	-0.0348***	-0.0346***
		(0.0041)	(0.0041)	(0.0043)	(0.0043)
State Dependence -Outside option	γ^O	0.0705	0.0692	0.0623	0.0819
		(0.0430)	(0.0461)	(0.0427)	(0.0560)
State Dependence –Cuisine	γ^C	0.0616	0.0705	0.0429	0.0438
		(0.0531)	(0.0481)	(0.0530)	(0.0492)
Cumulative Experience	γ^E		-0.0023		
			(0.0019)		
New Kitchens	γ^N			-0.0051	
				(0.0255)	
Distance					-0.0016*
					(0.0009)
Rating					0.0013
					(0.0018)

Note: The table shows estimation results after controlling for cuisine, day of week, week of month, holiday, menu size, distance, rating, monthly sales, and kitchen fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Column (1) is the baseline model. Column (2)-(4) are robustness checks. Column (2) tests for learning with the model specified in section 3.6.1. Column (3) tests for new product discovery with the model specified in section 3.6.2. Column (4) tests for robustness to the potential rank endogeneity problem.

3.6 Robustness Checks

In this section, we investigate several alternative models that could be sources of the state dependence we identified, including learning and new product discovery. We do not impose specific structural models of learning or new product discovery which would involve some strong structural assumptions on consumer behavior. Rather, we focus on aspects of consumer behavior that differentiate learning or new product discovery explanations from variety seeking and that can be directly observed in our data. We also discuss and test for additional sources of potential rank endogeneity.[21]

[21] In Appendix C.3 we provide more robustness discussions about initial conditions bias, rank dummies, inter-purchase time span, variety seeking on the dish level, and spurious state dependence from negative autocorrelation in unobserved shocks. We also report results from a test on model out-of-sample fit.

3.6.1 Learning

One alternative explanation of the negative state dependence we identify is consumer learning. Consumers can switch away from the recent choice not because of the disutility of repurchase, but from extra information or newly resolved uncertainty from consumption, especially at the beginning stage of platform experience (Ackerberg, 2003b; Narayanan and Manchanda, 2009; Osborne, 2011). To investigate this possibility, we define consumer-kitchen level consumption experience as the cumulative number of a consumer's orders of a kitchen, E_{ijt}. We interact the state dependence variable with the experience variable to compare the learning and variety-seeking models. Specifically, the indirect utility that consumer i derives from kitchen j on platform in period t is

$$u_{ijt} = \beta_{ij} - \alpha_i p_{jt} + \eta r_{ijt} + \gamma_i \mathbb{1}(s_{it} = j) + \gamma^E \mathbb{1}(s_{it} = j) E_{ijt} + \phi^E E_{ijt} + \gamma^C \mathbb{1}(s^c_{it} = c_j) + \psi \cdot X_{ijt} + \epsilon_{ijt}. \tag{3.5}$$

The utility from choosing the outside option is $u_{i0t} = \gamma^O \mathbb{1}(s_{it} = 0) + \epsilon_{i0t}$.

Under learning, the interaction term should reduce state dependence as brand experience accumulates, i.e., $\gamma^E > 0$. This will generate a convergence pattern in consumers' choice processes, which is prevalent in learning models: they will switch away less from more frequently ordered kitchens, since more information is gained and more uncertainty is resolved, and eventually behave in accordance with a standard choice model with no uncertainty. In contrast, in variety-seeking models, accumulative experience does not affect the state dependence. Table 3.3 column (2) shows the estimation results. The coefficient of the interaction term is close to zero and insignificant, which suggests that the variety-seeking preference we identify does not decrease with accumulative experience between consumers and kitchens.

3.6.2 New Product Discovery

In Section 3.6.1 we show that negative state dependence does not decrease with the accumulative experience between consumers and kitchens, which suggests non-convergence in choice behavior and serves as evidence against learning models. However, the non-convergence can also come from constant new entries throughout the time period; consumers could be trying out new options as an investment for discovering a potentially good choice for the future (Ershov, 2020). To examine this possibility we estimate the following specification, in which we add a variable indicating the number of new kitchens in consumer i's choice set in period t, N_{it}, where a new kitchen is defined by the first time the kitchen appears in the consumer's choice set.[22] Under new product discovery, the interaction term should increase state dependence as more new products in the choice set will encourage consumers to switch away from former purchase, i.e., $\gamma^N < 0$. However, in variety-seeking models, the state dependence should not be affected by the number of new products available in the choice set. Specifically, the indirect utility that consumer i derives from kitchen j on platform in period t is

$$u_{ijt} = \beta_{ij} - \alpha_i p_{jt} + \eta r_{ijt} + \gamma_i \mathbb{1}(s_{it} = j) + \gamma^N \mathbb{1}(s_{it} = j) N_{it} + \phi^N N_{it} + \gamma^C \mathbb{1}(s^c_{it} = c_j) + \psi \cdot X_{ijt} + \epsilon_{ijt}. \tag{3.6}$$

[22] An alternative definition of a new kitchen is a kitchen from which the consumer has never ordered. However, since there are kitchens that consumers do not like and will never order from, this definition would mix new kitchens and old disliked kitchens together.

The utility from choosing the outside option is $u_{i0t} = \gamma^O \mathbb{1}(s_{it} = 0) + \epsilon_{i0t}$.

The results are in column (3) of Table 3.3. We do not find a significant coefficient for the interaction term, and thus conclude that the variety seeking we identified is not from new product discovery or investing in a promising option for the future. The robustness checks on learning and new product discovery correspond to the explore vs exploit tradeoff in models with information frictions. Appendix C.6 provides more details about the difference between variety-seeking behavior and the explore vs exploit tradeoff in a multi-armed bandit context.[23]

3.6.3 Rank Endogeneity

We observe the app version history of the platform and observe the factors included in the ranking algorithm. The platform provides to consumers a default comprehensive ranking that is based on four factors: price, distance, monthly sales, and ratings. We do not include distance and ratings in the baseline utility specification. A potential endogeneity problem is that if time-varying geographic distance and ratings enter both utility and ranking, then it will bias our position effect estimates away from zero. To investigate this possibility we also estimate a model with distance and ratings entering utility. Table 3.3 column (4) shows the results. The distance coefficient is not significant at the 5% level, and the rating coefficient is not significant. The rank effect estimate does not change much. The insignificance of the rating coefficient is from a lack of variation both across kitchens and over time. The insignificance of the distance coefficient suggests that within the delivery distance limit, geographic distance does not affect consumers' utility substantively. Thus variation over time in rank associated with geographic location changes is exogenous.

3.7 Counterfactual

In this section, we use our demand estimates to study the managerial implications of variety seeking for targeted ranking and targeted pricing.

3.7.1 Targeted Ranking

In this section, we examine the implications of variety-seeking preferences for targeted ranking. We use our demand estimates to compare consumer welfare, platform revenue, and purchase probability under several different ranking algorithms shown in Table 3.4.

For each ranking algorithm, we hold the choice set, prices, and all variables in X_{ijt} fixed as we observed in the data, and only change the ranking positions of kitchens. The random ranking benchmark ranks kitchens randomly.[24] The platform ranking uses the default ranking we observed in the real dataset. The suboptimal ranking ranks kitchens in decreasing order of expected utility, where the expected utility is calculated without the variety-seeking term. The optimal ranking ranks kitchens in the decreasing order of

[23] Other sources of information frictions include searching. However, search costs belong to forces that would motivate consumers to stay with the choices with which they are familiar with and from which they order repeatedly, which is inconsistent with our observation of frequent switching.

[24] Specifically, the results are averaged across multiple randomly drawn rankings.

Table 3.4: Ranking Algorithms

	Ranking Algorithm
Random Ranking	Rank kitchens in a random order.
Platform Ranking	The platform ranking observed in the data.
Suboptimal Ranking	Rank kitchens in decreasing order of expected utility, without considering variety seeking effect ($\gamma = 0$).
Optimal Ranking	Rank kitchens in decreasing order of expected utility, considering variety seeking effect ($\gamma < 0$).

expected utility with the variety-seeking term. The optimal ranking optimizes ex ante consumer welfare.[25]

We assume that platform can potentially estimate consumers' individual specific state-dependent preferences and rank kitchens corresponding to their personalized utility index.

We measure the total effect of ranking by the difference between optimal ranking and the random ranking benchmark. The difference between the optimal ranking and suboptimal ranking represents the effect of considering consumer variety-seeking preferences in ranking. Then the ratio $\frac{Optimal-Suboptimal}{Optimal-Random}$ shows how much of the improvement from random ranking to optimal ranking is due specifically to inclusion of variety-seeking preferences.

Counterfactual prediction results are shown in Figure 3.2. We simulate choices under each given ranking algorithm 50 times to get the mean and standard deviation of corresponding market outcome variables. We find that optimizing the ranking algorithm with variety-seeking preferences takes up 18.2% of the revenue improvement, 14.2% of the consumer welfare improvement, and 18.9% of the purchase probability improvement, out of the total ranking effect. We further decompose the welfare gain and find that consumers benefit both from better matches (higher utility) and lower position costs. Specifically, we find that, on average, 8% of the increase in welfare comes from lower position costs, with the optimal ranking decreasing the average position purchased by 0.2 positions.[26] Most of the welfare gain (92%) is from consumers purchasing a more desirable seller under the optimal ranking, which was ranked lower under the suboptimal ranking such that the consumers choose the outside option or a worse match ranked at a higher position.[27] We also find that the two utility-based rankings are both better than the platform ranking.[28]

[25]Appendix C.4.1 provides detailed discussion of the optimality of rankings. Note that the optimal ranking of consumer welfare used in the analysis does not necessarily optimize platform revenue, for which there is no simple index like utility to follow and one need to exhaust all the $J!$ potential combinatorics to compare revenues.

[26]Note that in the comparison between optimal and suboptimal ranking, the average position where a transaction happens is moved upwards by at most 1 position, since the utility index will be different for at most one option (the previous period choice) for each choice set.

[27]Appendix C.4.2 provides more discussion of the decomposition of gains from suboptimal ranking to optimal ranking. Appendix C.4.3 provides results when platform presents heterogeneous state-based targeted ranking based on different consumers' state-dependent preferences.

[28]In our sample this could result from the naivety of ranking algorithm the start-up platform is using in the early stage. Moreover, existing literature consistently document that utility-based rankings

Figure 3.2: Targeted Ranking Counterfactual Results

Note: The figure shows the percentage change in welfare, revenue and purchase probability of Platform Ranking, Suboptimal Ranking, and Optimal Ranking relative to the Random Ranking benchmark. We simulate choices under each given ranking algorithms for 50 times to get the mean and standard deviation of corresponding variables.

3.7.2 Targeted Pricing

In this section, we use the demand estimates to investigate the implications of consumers' variety-seeking preferences for competitive targeted pricing. Given the large number of consumers and sellers in our data, it is intractable to estimate a full dynamic pricing game model. Instead, we use the data to calibrate parameters in consumer preference, and simulate the optimal pricing strategy in a duopoly model, in order to understand the pricing implications under variety seeking. Consumers always choose with variety-seeking preferences, and we check the profit and welfare difference between optimal targeted pricing, where kitchens' policy functions are gained from the real state dependence parameter, and suboptimal pricing, where the kitchens do not consider the state dependence preference when making pricing decisions. We also analyze the comparative statics when the level of variety seeking varies.

The model consists of two kitchens competing for consumers with variety-seeking preferences by pricing heterogeneous products. Each kitchen sets a pricing policy to maximize the discounted sum of profits over an infinite horizon. Demand is derived from a population of consumers who make discrete choices from two products and an outside option. For simplicity, we drop the consumer specific index. In each period t, a consumer has chosen one option in the last period, $s_t \in \mathcal{J} = \{0, 1, 2\}$. If the consumer chose option j in the last period, $s_t = j$, and purchases product, $k \in \mathcal{J}$, then his or her state becomes, $s_{t+1} = k$. Conditional on price p_{jt} and the consumer's current state s_t, the utility from

increase consumer welfare and purchase probability.(Compiani et al., 2021; Ursu, 2018; De los Santos and Koulayev, 2017; Chen and Yao, 2017; Ghose et al., 2014, 2012). Results on how a utility-based ranking affects revenues are mixed. Ghose et al. (2014) find that it yields the largest total revenue among the rankings considered, whereas Ursu (2018) find that revenues decrease in three out of the four destinations in her sample.

the product j at time t is as follows:

$$u_{jt} = \beta_j + \alpha p_{jt} + \gamma 1\{s_t = j\} + \epsilon_{jt}. \tag{3.7}$$

The preference parameters are from our demand estimation. We assume that the random utility component, ϵ_{jt} is i.i.d. Type I extreme value distributed. We study heterogeneous products and use the mean estimates μ_j of the estimates of β_j of the two most popular kitchen groups. Price is targeted based on state:

$$p_{jt} = \begin{cases} p_{jt}^{own}, & \text{if } s_t = j. \\ p_{jt}^{rival}, & \text{if } s_t \neq j. \end{cases} \tag{3.8}$$

At any point in time, the market is summarized by the distribution of consumers over the states. Let $x_{jt} \in [0,1]$ be the fraction of consumers who chose j in the last period. Let $U_j(s_t, p_t)$ denote the deterministic component of the utility index, such that $u_{jt} = U_j(s_t, p_t) + \epsilon_{jt}$. The consumer's choice probability has the following logit form:

$$P_j(s_t, \mathbf{p_t}) = \frac{exp[U_j(s_t, p_{jt})]}{\sum_{k \in \mathcal{J}} exp[U_k(s_t, p_{kt})]}. \tag{3.9}$$

We assume that a seller is able to observe whether s consumer ordered from its kitchen in the last period or not, and potentially sets different prices for its own consumers ($s_t = j$) and its rival seller's consumers ($s_t \neq j$). We obtain the respective aggregate demand by summing demand over consumer states. Demand from own consumers and rival's consumers for product j are:

$$D_j^{own}(\mathbf{x_t}, \mathbf{p_t}) = x_{jt} P_j(j, \mathbf{p_t}) \tag{3.10}$$

$$D_j^{rival}(\mathbf{x_t}, \mathbf{p_t}) = \sum_{k \in \mathcal{J}/\{j\}} x_{kt} P_j(k, \mathbf{p_t}) \tag{3.11}$$

The distribution of consumer states, x_t, summarizes all current period payoff relevant information for the kitchen and describes the state of the market. The transition of the aggregate state can be derived from the transition probabilities of the individual states. Conditional on a price vector $\mathbf{p_t}$, we can define a Markov transition matrix $\mathbf{Q(p_t)}$ with the following elements:

$$Q_{jk}(\mathbf{p_t}) = P_j(k, \mathbf{p_t}). \tag{3.12}$$

The whole state vector evolves according to the Markov chain:

$$\mathbf{x_{t+1}} = \mathbf{Q}(\mathbf{p_t})\mathbf{x_t}. \tag{3.13}$$

The evolution of the state vector is deterministic, and we denote the transition function by f, $\mathbf{x_{t+1}} = f(\mathbf{x_t}, \mathbf{p_t})$. Time is discrete, $t = 0, 1, ...$ Conditional on all product prices and the state of the market, x_t, kitchen j's current period profit function is

$$\pi_j(\mathbf{x_t}, \mathbf{p_t}) = D_j^{own}(\mathbf{x_t}, \mathbf{p_t}) \cdot (p_j^{own} - c_j) + D_j^{rival}(\mathbf{x_t}, \mathbf{p_t}) \cdot (p_j^{rival} - c_j) \tag{3.14}$$

where c_j is the marginal cost of production, which does not vary over time. We use historical ingredient prices during December 10 to 20, 2016[29] to estimate the marginal cost of production, and adjust the variance of the logit distribution so that the equilibrium price at the estimated variety-seeking level equals the observed mean price in our data. Kitchens compete in prices and choose Markovian strategies, $\sigma_j : \mathbf{X} \rightarrow R^2$, that depend on the current payoff-relevant information, summarized by $\mathbf{x}$. This assumption rules out strategies that depend on historical prices and potential collusion among sellers. Kitchens discount the future using the common factor β, $0 < \beta < 1$. For a given profile of strategies, $\sigma = (\sigma_1, \sigma_2)$, the present discounted value of profits is $\sum_{t=0}^{\infty} \beta^t \pi_j[x_t, \sigma(x_t)]$. Conditional on a profile of competitor's strategies, σ_{-j}, kitchen j chooses a pricing strategy that maximizes its expected value. Associated with a solution to this problem is kitchen j's value function, which satisfies the Bellman equation

$$V_j(\mathbf{x}) = \max_{\mathbf{p_j} \geq 0} \{\pi_j(\mathbf{x}, \mathbf{p}) + \beta V_j[f(\mathbf{x}, \mathbf{p})]\}, \forall \mathbf{x} \in \mathbf{X}. \tag{3.15}$$

We use Markov Perfect Equilibrium (MPE) as our solution concept. In pure strategies, MPE is defined by a pricing strategy for each kitchen, σ_j^*, and an associated value function, V_j, such that

$$V_j(\mathbf{x}) = \max_{\mathbf{p_j} \geq 0} \{\pi_j[\mathbf{x}, \mathbf{p_j}, \sigma^*_{-\mathbf{j}}(\mathbf{x})] + \beta V_j\{f[\mathbf{x}, \mathbf{p_j}, \sigma^*_{-\mathbf{j}}(\mathbf{x})]\}\}, \tag{3.16}$$

for all states, $\mathbf{x}$, and kitchens. In each subgame starting at $\mathbf{x}$, the kitchen's strategy is a best response to the strategies its competitor choose.[30,31]

We calculate steady-state prices as follows. We begin by computing the equilibrium pricing strategies of each kitchen. Then, we choose an arbitrary initial state for the period $t = 0$ and calculate the corresponding sequence of equilibrium price levels and state vectors for periods $t = 1, 2, ...$ We stop this process after convergence of the state vector and corresponding equilibrium prices to fixed values occurs.

Results

Optimal Targeted Pricing We study the effect of state-based targeted pricing strategy by comparing the performance of two pricing schemes: uniform pricing and targeted pricing. Table 3.5 presents the details of each pricing strategy. In both pricing strategies, consumers always choose with variety-seeking preferences. Uniform pricing corresponds to the case where kitchens do not consider consumers' variety-seeking preferences and do not do targeted pricing based on it. Optimal targeted pricing corresponds to the case when kitchens consider consumers' variety-seeking preferences and set different prices to their own and rival's consumers. Table 3.6 shows the percentage changes of the equilibrium steady state prices, profits, and consumer welfare. With optimal targeted pricing, kitchens set lower prices for their own consumers and higher prices for the rival seller's consumers; this strategy is similar to a quantity discount for existing customers. Optimal pricing increases average transaction prices by 6.8%, profits by 5.2%, and decreases consumer welfare by 9.1%.

[29]Historic prices of ingredients come from the National Bureau of Statistics website: `http://www.stats.gov.cn/tjsj/zxfb/201612/t20161226_1445711.html`

[30]Appendix B1 of Dubé et al. (2009) proves the existence of a pure strategy equilibrium of a similar model.

[31]Appendix C.4.4 discusses the procedure for the numerical solution to the dynamic programming

Table 3.5: Uniform Pricing v.s. Targeted Pricing

Demand	Supply	
Consumers choose with VS preferences($\gamma = \hat{\gamma} < 0$)	(1)Uniform Pricing (Status quo)	Sellers set the same price to all consumers ($\gamma = 0, p_j^{own} = p_j^{rival}$)
	(2)Targeted Pricing	Sellers set targeted prices based on last period choice ($\gamma = \hat{\gamma} < 0, p_j^{own} \neq p_j^{rival}$)

Note: The table presents details of how we compare the performance of two pricing schemes in the price competition equilibrium: uniform pricing and targeted pricing.

Table 3.6: Optimal Pricing

	% Change with Targeted Pricing
Price_{1own}	-0.70
Price_{1rival}	9.42
Price_{2own}	-0.18
Price_{2rival}	9.42
Average Price	6.82
Total Profits	5.18
Platform Revenue	1.68
Consumer Welfare	-9.12

Note: Percentage change is calculated by formula $\frac{Targeted-Uniform}{Uniform}$ of the corresponding variable. Average price is the weighted sum of prices, where weights are the corresponding choice probabilities.

This optimal pricing counterfactual predicts that it is profitable to do targeted pricing based on consumers variety-seeking preferences. However, the platform we study and most popular food delivery platforms do not have targeted pricing based on recent orders. The common existing industry practices in targeted pricing are focused on new consumer acquisition, past consumer retention, and whole platform promotion. Appendix C.5.1 provides more details about the current industry practice for targeted pricing.

Comparative Statics The effect of variety-seeking preferences on price competition is theoretically ambiguous, and empirical work is necessary to pin down the prediction. Besides the comparison between optimal targeted pricing and uniform pricing, we also study how the optimal targeted pricing levels change with the variety-seeking level. We compute equilibrium prices for a range of variety seeking achieved by scaling the mean value of γ_i gained from demand estimation $s \in \{0, \frac{1}{3}, \frac{1}{2}, 1, 2, 3\}$.[32] Figure 3.3 illustrates the

problem.

[32]The corresponding level of variety seeking is the mean estimate of γ_i multiplied by the scale factor $\gamma = \hat{\gamma} \cdot s$. The variety-seeking level we estimate from the original data corresponds to a scale factor equal to one. The equilibrium when consumers don't have state-dependent preferences corresponds to a scale

key pricing incentives in our model. The x-axis is the level of variety-seeking preferences represented by a scale factor. Similar to the traditional harvest vs invest tradeoff in a switching cost model, there is also a current period vs future incentive difference under variety seeking. Panel (A) and (C) in Figure 3.3 show kitchen 1's prices for its own and rival's consumers, respectively. Panel (B) and (D) show kitchen 2's prices for its own and rival's consumers. To decompose kitchens' current and future period incentives, in each graph we plot the equilibrium price levels for both cases when kitchens are forward-looking ($\beta = 0.98$) and myopic ($\beta = 0$). Then the difference between the price levels when $\gamma = 0$ and the myopic line is the current period incentive for kitchens, and the difference between the price levels when kitchens are myopic and those when kitchens are forward-looking reflects the future incentive of kitchens. In terms of the current period incentive, for consumers who purchased from the kitchen in the last period, the kitchen has the incentive to decrease the price to keep them from switching to the competitor. For consumers who did not purchase from the kitchen in the last period, the kitchen has the incentive to increase the price to exploit the market power induced by consumers' variety-seeking preferences. In terms of future incentives, the tendency of switching from consumers' variety-seeking preferences discourages kitchens from competing for them and imposes upward pressure on prices. To understand this more formally, we can write down seller j's profit-maximizing first-order conditions based on equation 3.16:

$$\underbrace{\frac{\partial \pi_j}{\partial p_j^{own}}}_{\text{Current profits}} + \underbrace{\beta \frac{\partial V_j\{f[\mathbf{x}, \mathbf{p_j}, \sigma^*_{-\mathbf{j}}(\mathbf{x})]\}}{\partial p_j^{own}}}_{\text{Future distribution of consumer states}} = 0, \tag{3.17}$$

$$\underbrace{\frac{\partial \pi_j}{\partial p_j^{rival}}}_{\text{Current profits}} + \underbrace{\beta \frac{\partial V_j\{f[\mathbf{x}, \mathbf{p_j}, \sigma^*_{-\mathbf{j}}(\mathbf{x})]\}}{\partial p_j^{rival}}}_{\text{Future distribution of consumer states}} = 0. \tag{3.18}$$

The prices affect the current period profits, and meanwhile affects the future distribution of consumer states. Table 3.7 provides a summary of the direction of the effect of consumers' variety-seeking preferences on kitchens' pricing incentive. When kitchens are very differentiated, it is possible that variety seeking can intensify price competition, and the driving force comes from the more popular kitchens trying to maintain a large own consumer base.[33] But in our demand estimate range, it turns out that kitchens are not differentiated enough and consumers' variety-seeking preferences softens kitchen price competition.

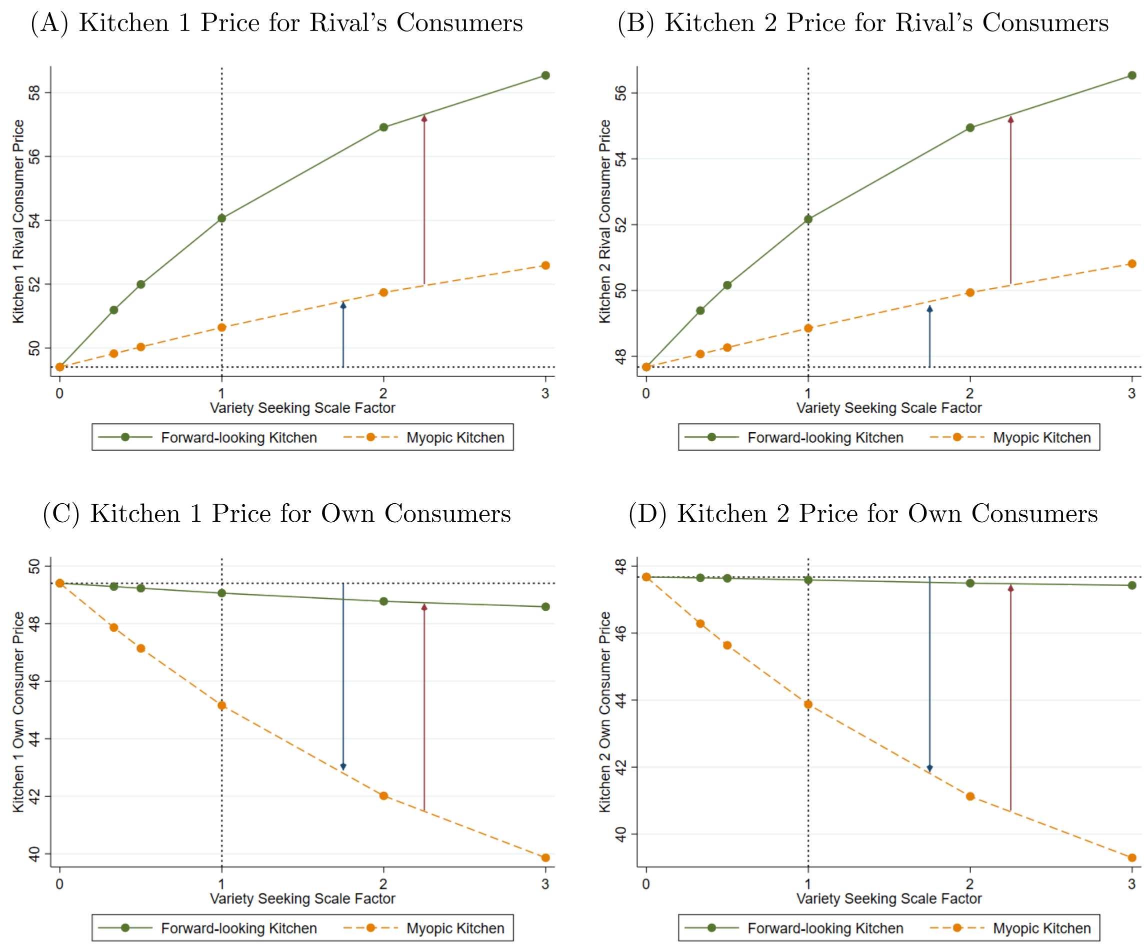
In order to study variety-seeking preference's effect on equilibrium profit and welfare level and eliminate the difference between "staying cost" and "switching bonus" models, which is due to the differential impact on the outside good market share under changes in γ, we calculate the difference of the steady state profit and welfare relative to a pure demand response benchmark.

Denote the steady state equilibrium price, market shares, profit, and welfare level as $\{\mathbf{p}_\gamma, \mathbf{x}_\gamma(\mathbf{p}_\gamma), \pi(\mathbf{p}_\gamma, \mathbf{x}_\gamma(\mathbf{p}_\gamma)), w(\mathbf{p}_\gamma, \mathbf{x}_\gamma(\mathbf{p}_\gamma), \gamma)\}$. The equilibrium profits are defined by the

factor of zero.

[33] Appendix C.4.5 provides more intuition for how product differentiation and variety-seeking level affect equilibrium price competition level.

Figure 3.3: Steady State Equilibrium Pricing



Note: The figure shows the steady state equilibrium prices for different levels of variety-seeking preferences. The x-axis is the scale factor, and the corresponding variety-seeking level equals the scale factor times the mean estimate of γ_i. Panel (A) and (C) show Kitchen 1's prices for its rival's and its own consumers, respectively. Panel (B) and (D)shows Kitchen 2's prices for its rival's its own consumers. In each graph we plot the equilibrium price levels for cases when kitchens are forward-looking ($\beta = 0.98$) and myopic ($\beta = 0$). The difference between the price level at $\gamma = 0$ and the myopic kitchen price level reflects the kitchens' current period pricing incentives (blue arrow), and the difference between myopic and forward-looking kitchen prices captures the kitchens' future pricing incentives (red arrow).

Table 3.7: Variety Seeking's Effect on Kitchens' Pricing Incentives

	Price for Own Consumer	Price for Rival's Consumer
Current Period	–	+
Future	+	+

Note: The table summarizes the effect of consumers' variety-seeking preferences on kitchen's pricing incentives for own and rival's consumers in the current period and future. '+' suggests an upward pressure on price levels, and '–' suggests a downward pressure on prices.

total profits of all firms

$$\pi(\mathbf{p}_\gamma, \mathbf{x}_\gamma(\mathbf{p}_\gamma)) = \sum_{j \in \mathcal{J}/\{0\}} \pi_j(\mathbf{p}_\gamma, \mathbf{x}_\gamma(\mathbf{p}_\gamma))$$

Table 3.8: Decomposition of Welfare and Profit Effects of Variety-Seeking Preferences

Decomposition	Market Outcome
(0) No VS	$\{\pi(\mathbf{p}_0, \mathbf{x}_0(\mathbf{p}_0)), w(\mathbf{p}_0, \mathbf{x}_0(\mathbf{p}_0), 0)\}$
(1) VS, demand-preference	$\{\pi(\mathbf{p}_0, \mathbf{x}_0(\mathbf{p}_0)), w(\mathbf{p}_\gamma, \mathbf{x}_0(\mathbf{p}_0), \hat{\gamma})\}$
(2) VS, demand-choice	$\{\pi(\mathbf{p}_0, \mathbf{x}_{\hat{\gamma}}(\mathbf{p}_0)), w(\mathbf{p}_0, \mathbf{x}_{\hat{\gamma}}(\mathbf{p}_0), \hat{\gamma})\}$
(3) VS, supply and equilibrium	$\{\pi(\mathbf{p}_{\hat{\gamma}}, \mathbf{x}_{\hat{\gamma}}(\mathbf{p}_{\hat{\gamma}})), w(\mathbf{p}_{\hat{\gamma}}, \mathbf{x}_{\hat{\gamma}}(\mathbf{p}_{\hat{\gamma}}), \hat{\gamma})\}$

Note: The table shows decomposition of equilibrium analysis when consumers have varying levels of variety-seeking preferences. Counterfactual (0) corresponds to the steady state equilibrium when consumers don't have state-dependent preferences. Starting from this equilibrium, counterfactual (1), (2), and (3) denote the market outcome when (1) only consumers change state-dependent preferences to $\gamma = \hat{\gamma}$, (2) consumers changes preferences and choices when sellers don't change prices, (3) sellers change prices and consumers change choices responding to the new price levels.

The equilibrium welfare are defined by the sum of ex-ante consumer welfare from consumer with all states.

$$w(\mathbf{p}_\gamma, \mathbf{x}_\gamma(\mathbf{p}_\gamma), \gamma) = -\frac{1}{\alpha} \sum_{j \in \mathcal{J}} x_j \cdot \left[ln \left(\sum_{k \in \mathcal{J}} exp(U_k(j, p_k)) \right) \right]$$

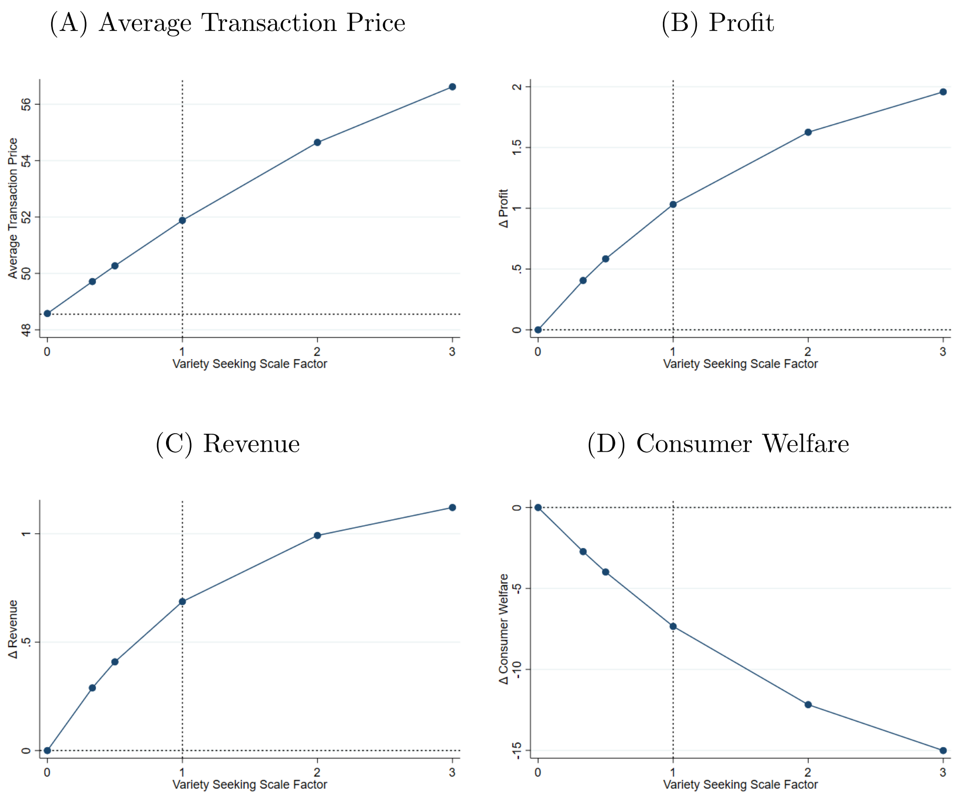
where x_j denotes the share of consumers with $s_t = j$, and p_k denotes product k's price for consumers with $s_t = j$. Table 3.8 presents the decomposition of welfare and profit effects of changing the levels of consumers' variety-seeking preference. Starting from the steady state equilibrium when consumers do not have state-dependent preference ($\{\pi(\mathbf{p}_0, \mathbf{x}_0(\mathbf{p}_0)), w(\mathbf{p}_0, \mathbf{x}_0(\mathbf{p}_0), 0)\}$), as γ becomes negative, we decompose the changes in steady state equilibrium into three channels. First, before prices and choices change, consumer welfare decreases because of the "staying cost" imposed boredom. Second, before prices change, as γ decreases the utility from consumption, consumers will switch to the outside option, which further decreases profits and increases consumer welfare. Third, kitchens change price levels given the new consumer state dependent preferences and consumers respond to the changes. The channel of changes in equilibrium that we are interested is the third channel, which captures the effect of consumers' variety-seeking preferences on price competition level. The first two channels are pure demand side responses from changes in γ and are susceptible to the model specification of "staying cost" and "switching bonus".[34] This challenge of welfare and profit analysis when consumer preferences change is related to the challenge in welfare analysis of persuasive advertising discussed in Becker and Murphy (1993b).

To tackle this problem, we use the pure demand response benchmark to separate the price competition effects on profits and consumer welfare from the pure demand response. The profit and welfare in the pure demand response benchmark for each variety-seeking level are calculated by equations 3.10 and 3.11, where prices and initial state variables are at the $\gamma = 0$ steady state equilibrium level, i.e., the pure demand response benchmark profit and welfare are $\pi(p_0, \gamma)$ and $\pi(p_0, \gamma)$, respectively. Then the price competition effect at each variety-seeking level is $\Delta\pi = \pi(p_\gamma, \gamma) - \pi(p_0, \gamma)$, $\Delta u = u(p_\gamma, \gamma) - u(p_0, \gamma)$.[35]

[34] In this paper we used the "staying cost" specification in accordance with boredom and satiation. The "switching bonus" specification would be more consistent with thrill seeking. We do not try to identify the different psychological factors behind negative state dependent preference in this project.

[35] Dubé et al. (2009) adopts a different approach to solve this problem by changing the intercept of

Figure 3.4: Steady State Equilibrium



Note: The figure shows steady state equilibrium average transaction price, profit, revenue, and consumer welfare. The average transaction price is the weighted average of prices by transaction probability. Profits, revenue, and welfare are measured in CNY, and are calculated relative to the pure demand response benchmark to measure the pure price competition effect.

Figure 3.4 shows how steady state prices, profits, and consumer welfare change with the level of consumers' variety-seeking preferences. We find that as the variety-seeking level increases, equilibrium average transaction prices and profits increase, whereas consumer welfare decreases. To summarize, we find that consumers' variety-seeking preferences soften price competition.

3.8 Conclusion

In this paper, we document consumers' variety-seeking preferences in high-frequency consumption. We show that a substantial fraction of consumers have negative state dependent preference, after controlling for potential confounding factors. We also further ruled out alternative sources of structural state dependence, including learning and new product discovery. We find that consumers, on average, are willing to pay 19.9% more for switching to a different seller. There is heterogeneity in consumers' state dependence preference, 32.8% of consumers have inertia, whereas 67.2% of consumers exhibit variety-

the outside option in order to keep the outside option market share fixed. This method also removes the extensive margin demand response to price changes in price competition in the third channel. We adopt an alternative method here to capture the extensive margin effect of price competition level.

seeking preferences. This finding has important managerial implications for targeted marketing.

Using the demand estimates, we examine the managerial implications of variety seeking on targeted ranking and targeted pricing. For targeted ranking, we examine the profits, welfare, and purchase probability of several different ranking schemes. We find that optimizing ranking algorithm with variety-seeking preferences takes up 18.2% of the revenue improvement, 14.2% of the consumer welfare improvement, and 18.9% of the purchase probability improvement, out of the total ranking effect.

For targeted pricing, we study the effect of optimal pricing with variety seeking and the comparative statics of the effect of consumers' variety-seeking preferences on price competition. We find that optimal targeted pricing increases prices and profits, and decreases consumer welfare. The comparative statics analysis shows that the average equilibrium prices and profits increase with consumers' variety-seeking level, whereas consumer welfare decreases with it. In general, variety seeking softens price competition.

Although we only do counterfactual analysis on pricing and ranking in this paper, consumers' variety-seeking preferences have more general managerial implications in other areas, including targeted advertising, personalized recommendation systems (Ansari et al., 2000; Ansari and Mela, 2003), positioning (Sajeesh and Raju, 2010; Zeithammer and Thomadsen, 2013), and the distributional channel.[36] The managerial implications of consumer variety-seeking preferences are also not limited to the food delivery industry. Cadario and Morewedge (2021) document the phenomenon that many people eat the same breakfast every day, yet seek variety for lunch and dinner, and attribute this difference to a psychological driver: variance in the pursuit of hedonic and utilitarian goals across meals. This is also helpful to understand the direction of state dependence we might expect in different industries. In utilitarian choice situations, such as work-day breakfast, health care, gasoline, and grocery products, former research has found evidence of inertia. In hedonic choice situations, such as lunch, dinner, music, movies, podcast, magazine, art, and resorts, where people care more about entertainment and pleasure, it is important to consider consumers' variety-seeking preferences.

[36] Cosguner et al. (2017) and Cosguner et al. (2018) study dynamic pricing models with switching cost in a distribution channel. The implications of variety seeking in a vertical relationship are also important to investigate.

Bibliography

Abaluck, J. and A. Adams (2017). What do consumers consider before they choose? identification from asymmetric demand responses. Technical report, National Bureau of Economic Research.

Abaluck, J. and J. Gruber (2009). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program (working paper# 14759). *Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www. nber. org/papers/w14759. doi 10*, w14759.

Abaluck, J. and J. Gruber (2011). Choice inconsistencies among the elderly: evidence from plan choice in the medicare part d program. *The American economic review 101*(4), 1180–1210.

Ackerberg, D. (2003a). Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. *International Economic Review 44*.

Ackerberg, D. A. (2003b). Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review 44*(3), 1007–1040.

Allcott, H. and C. Knittel (2019). Are consumers poorly informed about fuel economy? evidence from two experiments. *American Economic Journal: Economic Policy 11*(1), 1–37.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019). Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics 134*(3), 1557–1626.

Allcott, H. and D. Taubinsky (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review 105*(8), 2501–38.

Ansari, A., S. Essegaier, and R. Kohli (2000). Internet recommendation systems.

Ansari, A. and C. F. Mela (2003). E-customization. *Journal of marketing research 40*(2), 131–145.

Armstrong, M. (2017). Ordered consumer search. *Journal of the European Economic Association 15*(5), 989–1024.

Armstrong, M. and J. Zhou (2011). Paying for prominence. *The Economic Journal 121*(556), F368–F395.

Bagwell, K. (2007). Chapter 28 The Economic Analysis of Advertising. *Handbook of Industrial Organization 3*, 1701–1844.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021). Heterogeneous choice sets and preferences. *Econometrica 89*(5), 2015–2048.

Bass, F. M., M. M. Givon, M. U. Kalwani, D. Reibstein, and G. P. Wright (1984). An investigation into the order of the brand choice process. *Marketing Science 3*(4), 267–287.

Bawa, K. (1990). Modeling inertia and variety seeking tendencies in brand choice behavior. *Marketing Science 9*(3), 263–278.

Becker, G. S. and K. M. Murphy (1993a). A Simple Theory of Advertising as a Good or Bad. *The Quarterly Journal of Economics 108*(4), 941–964.

Becker, G. S. and K. M. Murphy (1993b). A simple theory of advertising as a good or bad. *The Quarterly Journal of Economics 108*(4), 941–964.

Berry, S., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica 81*, 2087–2111.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. and P. A. Haile (2009). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Technical report, National Bureau of Economic Research.

Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Besanko, D., J.-P. Dubé, and S. Gupta (2003). Competitive price discrimination strategies in a vertical channel using aggregate retail data. *Management Science 49*(9), 1121–1138.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–739.

Bonhomme, S., T. Lamadon, and E. Manresa (2022). Discretizing unobserved heterogeneity. *Econometrica 90*(2), 625–643.

Branco, F., M. Sun, and J. M. Villas-Boas (2012). Optimal search for product information. *Management Science 58*(11), 2037–2056.

Broda, C. and D. E. Weinstein (2006). Globalization and the gains from variety. *The Quarterly journal of economics 121*(2), 541–585.

Bronnenberg, B. J., J.-P. Dubé, M. Gentzkow, and J. M. Shapiro (2015). Do pharmacists buy bayer? informed shoppers and the brand premium. *The Quarterly Journal of Economics 130*(4), 1669–1726.

Bronnenberg, B. J., J. B. Kim, and C. F. Mela (2016). Zooming in on choice: How do consumers search for cameras online? *Marketing Science 35*(5), 693–712.

Brynjolfsson, E., Y. Hu, and M. D. Smith (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management science 49*(11), 1580–1596.

Cadario, R. and C. K. Morewedge (2021). Why do people eat the same breakfast every day? goals and circadian rhythms of variety seeking in meals. *Appetite*, 105716.

Chade, H. and L. Smith (2006). Simultaneous search. *Econometrica 74*(5), 1293–1307.

Che, H., S. Seetharaman, and K. Sudhir (2003). Pricing behavior in markets with state-dependence in demand. *Available at SSRN 442580*.

Chen, Y. and S. Yao (2016). Sequential search with refinement: Model and application with click-stream data. *Management Science 63*(12), 4345–4365.

Chen, Y. and S. Yao (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science 63*(12), 4345–4365.

Chintagunta, P. K. (1998). Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science 17*(3), 253–270.

Chintagunta, P. K. (1999). Variety seeking, purchase timing, and the "lightning bolt" brand choice model. *Management Science 45*(4), 486–498.

Choi, M., A. Y. Dai, and K. Kim (2018a). Consumer search and price competition. *Econometrica 86*(4), 1257–1281.

Choi, M., A. Y. Dai, and K. Kim (2018b). Consumer search and price competition. *Econometrica 86*(4), 1257–1281.

Compiani, G. (2019). Market counterfactuals and the specification of multi-product demand: a nonparametric approach. *Working Paper*.

Compiani, G. (2022). Market counterfactuals and the specification of multi-product demand: a nonparametric approach.

Compiani, G., G. Lewis, S. Peng, and W. Wang (2021). Online search and product rankings: A double index approach. *Available at SSRN 3898134*.

Conlon, C. T. and J. H. Mortimer (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics 5*(4), 1–30.

Cosguner, K., T. Y. Chan, and P. Seetharaman (2017). Behavioral price discrimination in the presence of switching costs. *Marketing Science 36*(3), 426–435.

Cosguner, K., T. Y. Chan, and P. Seetharaman (2018). Dynamic pricing in a distribution channel in the presence of switching costs. *Management Science 64*(3), 1212–1229.

Crawford, G. S. and M. Shum (2005). Uncertainty and Learning in Pharmaceutical Demand. *Econometrica 73*(4), 1137–1173.

Datta, H., G. Knox, and B. J. Bronnenberg (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science 37*(1), 5–21.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review 102*(6), 2955–80.

De los Santos, B. and S. Koulayev (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science 36*(4), 542–564.

DellaVigna, S. and M. Gentzkow (2017). Uniform pricing in us retail chains. Technical report, National Bureau of Economic Research.

Derakhshan, M., N. Golrezaei, V. Manshadi, and V. Mirrokni (2022). Product ranking on online platforms. *Management Science*.

Dong, X., I. Morozov, S. Seiler, and L. Hou (2018). Estimation of preference heterogeneity in markets with costly search.

Donnelly, R., F. J. Ruiz, D. Blei, and S. Athey (2021). Counterfactual inference for consumer choice across many product categories. *Quantitative Marketing and Economics*, 1–39.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2009). Do switching costs make markets less competitive? *Journal of Marketing research 46*(4), 435–445.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics 41*(3), 417–445.

Dubé, J.-P., G. J. Hitsch, P. E. Rossi, and M. A. Vitorino (2008). Category pricing with state-dependent utility. *Marketing Science 27*(3), 417–429.

Eggers, A. C., G. Tuñón, and A. Dafoe (2021). Placebo tests for causal inference. *Working paper*.

Erdem, T. and M. P. Keane (1996). Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science 15*(1), 1–20.

Ericson, K. M., P. Kircher, J. Spinnewijn, and A. Starc (2015). Inferring risk perceptions and preferences using choice from insurance menus: theory and evidence. Technical report, National Bureau of Economic Research.

Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Statistical Science*, 676–685.

Ershov, D. (2020). Consumer product discovery costs, entry, quality and congestion in online markets. *Unpublished manuscript*.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology 18*(1), 52–72.

Farrell, J. and P. Klemperer (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization 3*, 1967–2072.

Fox, J. T. and A. Gandhi (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *The RAND Journal of Economics 47*(1), 118–139.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics 2*, 381–418.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2012). The random coefficients logit model is identified. *Journal of Econometrics 166*(2), 204–212.

Fox, J. T., K. i. Kim, and C. Yang (2016). A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics 195*(2), 236–254.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 261–343. Elsevier.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review 96*(4), 1043–1068.

Gabel, S. and A. Timoshenko (2022). Product choice with large assortments: A scalable deep-learning model. *Management Science 68*(3), 1808–1827.

Gardete, P. and M. Hunter (2020). Guiding consumers through lemons and peaches: An analysis of the effects of search design activities.

Gaynor, M., C. Propper, and S. Seiler (2016). Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review*.

Ghose, A., P. G. Ipeirotis, and B. Li (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science 31*(3), 493–520.

Ghose, A., P. G. Ipeirotis, and B. Li (2014). Examining the impact of ranking on consumer behavior and search engine revenue. *Management Science 60*(7), 1632–1654.

Givon, M. (1984). Variety seeking through brand switching. *Marketing Science 3*(1), 1–22.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5), 1017–1074.

Good, P. (2013). *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of political Economy 120*(6), 1173–1219.

Gul, F. and W. Pesendorfer (2008). The case for mindless economics. *The foundations of positive and normative economics: A handbook 1*, 3–42.

Hamilton, W., Z. Ying, and J. Leskovec (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems 30*.

Handel, B. R. and J. T. Kolstad (2015). Health insurance for" humans": Information frictions, plan choice, and consumer welfare. *American Economic Review 105*(8), 2449–2500.

Hastings, J. S. and L. Tejeda-Ashton (2008). Financial literacy, information, and demand elasticity: Survey and experimental evidence from mexico. Technical report, National Bureau of Economic Research.

Hastings, J. S. and J. M. Weinstein (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics 123*(4), 1373–1414.

Hendel, I. and A. Nevo (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review 103*(7), 2722–51.

Holtz, D., B. Carterette, P. Chandar, Z. Nazari, H. Cramer, and S. Aral (2020). The engagement-diversity connection: Evidence from a field experiment on spotify. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 75–76.

Hong, H. and M. Shum (2006). Using price distributions to estimate search costs. *RAND Journal of Economics 37*(2), 257–275.

Honka, E. (2014). Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics 45*(4), 847–884.

Honka, E. and P. Chintagunta (2016). Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science 36*(1), 21–42.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *RAND Journal of Economics 48*(3), 611–646.

Hortaçsu, A. and C. Syverson (2004). Product differentiation, search costs, and competition in the mutual fund industry: a case study of s&p 500 index funds. *Quarterly Journal of Economics 119*(2), 403–456.

Howell, J. R., S. Lee, and G. M. Allenby (2016). Price promotions in choice models. *Marketing Science 35*(2), 319–334.

Jain, A., I. Liu, A. Sarda, and P. Molino (2019). Food discovery with uber eats: Using graph learning to power recommendations. *Uber engineering.*.

Jeziorski, P. and I. Segal (2015). What makes them click: Empirical analysis of consumer demand for search advertising. *American Economic Journal: Microeconomics 7*(3), 24–53.

Jindal, P. and A. Aribarg (2018). The value of price beliefs in consumer search.

Johnson, E. M. and M. M. Rehavi (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy 8*(1), 115–41.

Johnson, J., G. J. Tellis, and E. H. Ip (2013). To whom, when, and how much to discount? a constrained optimization of customized temporal discounts. *Journal of Retailing 89*(4), 361–373.

Kahn, B. E. and J. S. Raju (1991). Effects of price promotions on variety-seeking and reinforcement behavior. *Marketing Science 10*(4), 316–337.

Ke, T. T., Z.-J. M. Shen, and J. M. Villas-Boas (2016). Search for information on multiple products. *Management Science 62*(12), 3576–3603.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010). Online demand under limited consumer search. *Marketing science 29*(6), 1001–1023.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2017). The probit choice model under sequential search with an application to online retailing. *Management Science 63*(11), 3911–3929.

Kling, J. R., S. Mullainathan, E. Shafir, L. Vermeulen, and M. V. Wrobel (2008). Misperception in choosing medicare drug plans. *Unpublished manuscript*.

Kong, X., J.-P. Dubé, and O. Daljord (2022). Non-parametric estimation of habitual brand loyalty. *Available at SSRN 4070747*.

Koren, Y., R. Bell, and C. Volinsky (2009). Matrix factorization techniques for recommender systems. *Computer 42*(8), 30–37.

Larsen, B., C. H. Lu, and A. L. Zhang (2021). Intermediaries in bargaining: Evidence from business-to-business used-car inventory negotiations. Technical report, National Bureau of Economic Research.

MacKay, A. and M. Remer (2019). Consumer inertia and market power. *Available at SSRN 3380390*.

Manski, C. F. (2002). Identification of decision rules in experiments on simple games of proposal and response. *European Economic Review 46*(4-5), 880–891.

Manzini, P. and M. Mariotti (2014). Stochastic choice and consideration sets. *Econometrica 82*(3), 1153–1176.

McAlister, L. (1982). A dynamic attribute satiation model of variety-seeking behavior. *Journal of Consumer Research 9*(2), 141–150.

McAlister, L. and E. Pessemier (1982). Variety seeking behavior: An interdisciplinary review. *Journal of Consumer research 9*(3), 311–322.

Mehta, N., S. Rajiv, and K. Srinivasan (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science 22*(1), 58–84.

Miller, J. B. and A. Sanjurjo (2018). Surprised by the hot hand fallacy? a truth in the law of small numbers. *Econometrica 86*(6), 2019–2047.

Miller, J. B., A. Sanjurjo, et al. (2014). A cold shower for the hot hand fallacy. Technical report, Working paper.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2021). Consumer search and prices in the automobile market.

Morozov, I., S. Seiler, X. Dong, and L. Hou (2021). Estimation of preference heterogeneity in markets with costly search. *Marketing Science 40*(5), 871–899.

Narayanan, S. and P. Manchanda (2009). Heterogeneous learning and the targeting of marketing communication for new products. *Marketing science 28*(3), 424–441.

Natan, O. R. (2020). Choice frictions in large assortments. Technical report.

Nevo, A., J. L. Turner, and J. W. Williams (2016). Usage-Based Pricing and Demand for Residential Broadband. *Econometrica 84*(2), 411–443.

Osborne, M. (2011). Consumer learning, switching costs, and heterogeneity: A structural examination. *Quantitative Marketing and Economics 9*(1), 25–70.

Pancras, J. and K. Sudhir (2007). Optimal marketing strategies for a customer data intermediary. *Journal of Marketing research 44*(4), 560–578.

Park, S. and S. Gupta (2011). A regime-switching model of cyclical category buying. *Marketing Science 30*(3), 469–480.

Quan, T. W. and K. R. Williams (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics 49*(4), 877–913.

Roberts, J. H. and J. M. Lattin (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, 429–440.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1996). The value of purchase history data in target marketing. *Marketing Science 15*(4), 321–340.

Sajeesh, S. and J. S. Raju (2010). Positioning and pricing in a variety seeking market. *Management science 56*(6), 949–961.

Schwartz, B., A. Ward, J. Monterosso, S. Lyubomirsky, K. White, and D. R. Lehman (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of personality and social psychology 83*(5), 1178.

Seetharaman, P. and H. Che (2009). Price competition in markets with consumer variety seeking. *Marketing Science 28*(3), 516–525.

Seetharaman, P. and P. Chintagunta (1998). A model of inertia and variety-seeking with marketing variables. *International Journal of Research in Marketing 15*(1), 1–17.

Seiler, S. (2013). The impact of search costs on consumer behavior: A dynamic approach. *Quantitative Marketing and Economics 11*(2), 155–203.

Simonov, A., J.-P. Dubé, G. Hitsch, and P. Rossi (2020). State-dependent demand estimation with initial conditions correction. *Journal of Marketing Research 57*(5), 789–809.

Smith, A. N., I. Aggarwal, and S. Seiler (2021). Optimal price targeting.

Smith, A. N., Y. Li, S. Hasegawa, N. Terui, and G. M. Allenby (2020). Integrated factor models of variety seeking dynamics.

Stahl, D. (1989). Oligopolistic pricing with sequential consumer search. *American Economic Review 79*(4), 700–712.

Trivedi, M., F. M. Bass, and R. C. Rao (1994). A model of stochastic variety-seeking. *Marketing Science 13*(3), 274–297.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science 37*(4), 530–552.

Van Herpen, E. and R. Pieters (2002). The variety of an assortment: An extension to the attribute-based approach. *Marketing Science 21*(3), 331–341.

Villas-Boas, J. M. (1999). Dynamic competition with customer recognition. *The Rand Journal of Economics*, 604–631.

Villas-Boas, J. M. (2015). A short survey on switching costs and dynamic competition. *International Journal of Research in Marketing 32*(2), 219–222.

Wang, Y., Y. Ning, I. Liu, and X. X. Zhang (2018). Food discovery with uber eats: recommending for the marketplace.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, 641–654.

Wijsman, R. A. (1985). A useful inequality on ratios of integrals, with application to maximum likelihood estimation. *Journal of the American Statistical Association 80*(390), 472–475.

Woodward, S. E. and R. E. Hall (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review 102*(7), 3249–76.

Yoganarasimhan, H. (2020). Search personalization using machine learning. *Management Science 66*(3), 1045–1070.

Zeithammer, R. and R. Thomadsen (2013). Vertical differentiation with variety-seeking consumers. *Management Science 59*(2), 390–401.

Zhang, J. and L. Krishnamurthi (2004). Customizing promotions in online stores. *Marketing science 23*(4), 561–578.

Zhang, J. and M. Wedel (2009). The effectiveness of customized promotions in online and offline stores. *Journal of marketing research 46*(2), 190–206.

Appendix A

Appendix for Chapter 1

A.0.1 More details about Data Construction

To collapse the different products within the same brand, I calculate the average price. For example, if in one observation I observe a purchase of Chobani strawberry yogurt and vanilla yogurt, I see them as the same product Chobani and calculate the average price by dividing the total price paid (after subtracting any coupon value applied) by quantity.

I drop the consumers who never purchase any of the top five 32oz yogurt. Then I collapse a household's multiple trips in one week into one to see one household-week as an observation unit. If a household made multiple trips with no purchase of the five brands, I collapse the multiple trips into one. If a household made multiple trips in a week, including trips with purchase of one of the five products and trips with purchase of none, I collapse the trips into one observation with the purchase of the corresponding product.

Table A.1 presents the details of data construction.N represents the situation that the household made no trip to the retailer in the specific week. 1-5 represent the five products. 0 denotes the case when the household made a trip to the retailer but purchase none of the five products. Only the colored cells are taken as observations.The light gray cells are seen as Not Buy observations.

Table A.1: **Data Construction.** N represents the situation that the household made no trip to the retailer in the specific week. 1-5 represent the five products. 0 denotes the case when the household made a trip to the retailer but purchase none of the five products. Only the colored cells are taken as observations.The light gray cells are seen as Not Buy observations.

Household \ Week	1	2	3	4
1	blue!25 1	light-gray 0	N	N
2	blue!25 2	N	blue!25 3	blue!25 5
3	blue!25 4	blue!25 1	N	light-gray 0
4	N	N	N	blue!25 1

Table A.2shows an example of how the unobserved prices are recovered. In this example of observations in one store in one week, p_1 for household 2 and 3 are recovered by household 1's observed price. p_2 for household 1 is recovered by an average of household 2 and household 3's observed prices. As no one bought product 3 in this store in the specific week, p_3 for all three households are recovered by the average of observed p_3 in all other stores of the retailer.

Table A.2: **Price Recovery.** In this example of observations in one store in one week, p_1 for household 2 and 3 are recovered by household 1's observed price. p_2 for household 1 is recovered by an average of household 2 and household 3's observed prices. p_3 for all three households are recovered by the average of observed p_3 in all other stores of the retailer.

Household	p_1	p_2	p_3
1	observed		
2		observed	
3		observed	

A.0.2 Choice Persistence Examination

I use the panel data to track one household and gain proper Not Buy observations, but one potential problem of using the panel data is the potential choice persistence problem. This section is to show that choice persistence problem is limited in our data.

Table A.3: **Distribution of Observations.** The variable *Trips* denotes the total number of trips in which the household purchased one of the five products in 2016. The variable *Brands* denotes the total number of different brands the household purchased in 2016.

Trips \ Brands	1	2	3	4	Total
1	21383	0	0	0	21383
2	5908	2783	0	0	8691
3	2584	1314	142	0	4040
4	1512	1013	423	0	2948
5	680	960	190	0	1830
6	915	521	119	0	1555
7	641	361	104	0	1106
8	368	258	80	0	706
9	355	70	83	0	508
10	518	365	0	0	883
11	216	185	111	0	512
≥ 12	1978	830	326	19	3153
Total	37,058	8,660	1,578	19	47,315

Table A.3 presents the detailed structure of the 47,315 observations. Correspondingly, table A.4 presents the detailed structure of the 1,641 households. The variable *Trips* denotes the total number of trips in which the household purchased one of the five products in 2016. The variable *Brands* denotes the total number of different brands the household purchased in 2016. The cells in table A.3 and A.4 reports the number of observations and

Table A.4: **Distribution of Households.** The variable *Trips* denotes the total number of trips in which the household purchased one of the five products in 2016. The variable *Brands* denotes the total number of different brands the household purchased in 2016.

Trips \ Brands	1	2	3	4	Total
1	808	0	0	0	808
2	206	99	0	0	305
3	86	44	4	0	134
4	50	34	12	0	96
5	24	28	6	0	58
6	27	17	3	0	47
7	19	12	3	0	34
8	12	8	3	0	23
9	10	3	2	0	15
10	14	10	0	0	24
11	6	5	3	0	14
≥ 12	50	23	9	1	83
Total	1,312	283	45	1	1,641

households, respectively. Table A.3 shows that although a large number of observations (37,058) are from the households who only bought one brand in the year of 2016, most of them (21,383) are from those who only purchased 32 oz yogurt once. Only 1,978 observations are from the households who purchased more than 12 times but always bought the same brand. Similarly, in table A.4 note that although a large number of households (1312) are from the households who only bought one brand in the year of 2016, most of them (808) only purchased 32 oz yogurt once. Only 50 households purchased more than 12 times but always bought the same brand.

The above analysis of the detailed data structure shows that the households who exhibit the choice persistence behavior only account for a small portion (3.05%) of the sample households, and the observations they contribute only account for a small portion (4.18%) of the whole sample. Thus the potential choice persistence problem is limited in our sample.

Appendix B

Appendix for Chapter 2

B.1 Additional Proofs

In this appendix, we collect the proofs not included in the main text. Throughout, we let $\mathcal{J} \equiv \{1, \ldots, J\}$ and use the notational convention $\frac{\partial f}{\partial x^0}(x) = f(x) \forall x$ for any function f. We also often drop the i subscript for notational simplicity.

B.1.1 Identifying good 1 when z_j is vector-valued in the linear homogeneous case

For simplicity, the results in the main text are for the case where z_j is scalar-valued for all goods j. This implies that one can label good 1 as the good with the highest value of z without loss of generality. As we have noted, if there are multiple z attributes per good, then our results apply if the data contains one choice set where one good is preferable to all other goods on each of the z attributes. This is not without loss.

We now show how to relax this restriction in the linear homogeneous case of Lemma 3. Let z_{kj} be the $k-$th hidden attribute of good j and let β_k be the associated preference parameter. As above, let $\tilde{u}_j = \alpha x_j + \beta z_j$. By Assumption 2, we can write $s_j = f_{s_j}(\tilde{u}_1, \ldots, \tilde{u}_J, x_1, \ldots, x_J)$ for all j and thus $\frac{\partial s_j}{\partial z_{kj}} = \frac{\partial f_{s_j}}{\partial \tilde{u}_j} \beta_k$, implying $\frac{\partial s_j}{\partial z_{kj}} / \frac{\partial s_j}{\partial z_{k'j}} = \beta_k / \beta k'$ for all k, k'. This means that we can compare the hidden component of utility across goods. Specifically, letting $\beta_1 > 0$ without loss, we have that, for any pair of goods j and j', $\sum_k \beta_k z_{kj} \geq \sum_k \beta_k z_{kj'}$ if and only if $z_{1j} - z_{1j'} + \sum_{k>1} \frac{\beta_k}{\beta_1}(z_{kj} - z_{kj'}) \geq 0$. Since the l.h.s. of the last inequality is identified, we can rank goods based on their non-visible utility. Lemma 3 then applies by defining good 1 as the good with the highest value of $\sum_k \beta_k z_{kj}$. Note that such a good always exists in any choice set (excluding ties) since $\sum_k \beta_k z_{kj}$ is scalar-valued.

(k_α, k_β), $a_{k,n,\tilde{n}} \equiv \frac{\partial^{1+n} P_6}{\partial v_1 \partial v_2^n}(\mathbf{0}) \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1}$, $b_{k,n,\tilde{n}} \equiv \alpha_{k_\alpha}^{n-\tilde{n}} \beta_{k_\beta}^{\tilde{n}+1}$ are known scalars, and $f_{k,n} \equiv \frac{\partial^{1+n} P_7^S\left(\mathbf{0}, \mathbf{0}, 0; \alpha_{k_\alpha}, \beta_{k_\beta}\right)}{\partial v_1 \partial v_2^n}$ and $\pi_k \equiv \tilde{\pi}_{k_\alpha, k_\beta}$ are unknown scalars.

Setting $n = K - 1$ and stacking the equations corresponding to $\tilde{n} = 0, \ldots, K - 1$, we get

$$q = A\pi + B(f * \pi)$$

where q is a known column $K-$vector, A, B are known $K-$by$-K$ matrices, and $f * \pi$ denotes the column vector given by the element-by-element product of $(f_{1,K-1} \ldots f_{K,K-1})$

and $\pi \equiv (\pi_1, \ldots, \pi_K)'$. We re-write this system of equations in a way that highlights which objects depend on $\mathbf{z} \equiv (z_1, \ldots, z_J)$ as follows

$$q(\mathbf{z} = \mathbf{0}) \quad = \quad A\pi + B(f(\mathbf{z} = \mathbf{0}) * \pi)$$

Note that A depends on z only through $z_1 - z_j$ (i.e. it exhibits a lack of nominal illusion property) and we leave that dependence implicit. Now consider increasing z_j by Δz for all j relative to the baseline $\mathbf{z} = \mathbf{0}$. Then we can write

$$q(\mathbf{z} = \mathbf{\Delta z}) \quad = \quad A\pi + B(f(\mathbf{z} = \mathbf{\Delta z}) * \pi)$$

Combining the last two systems, we get

$$q(\mathbf{z} = \mathbf{\Delta z}) - q(\mathbf{z} = \mathbf{0}) \quad = \quad B\left[(f(\mathbf{z} = \mathbf{\Delta z}) - f(\mathbf{z} = \mathbf{0})) * \pi\right]$$

If B is full rank,[1] we obtain identification of $(f(\mathbf{z} = \mathbf{\Delta z}) - f(\mathbf{z} = \mathbf{0})) * \pi$. Also, note that, for all k, $\lim_{\Delta z \to 0} \frac{f_{k,K-1}(\mathbf{z}=\mathbf{\Delta z}) - f_{k,K-1}(\mathbf{z}=\mathbf{0})}{\Delta z}$ is the directional derivative of $f_{k,K-1}$ in the direction $\mathbf{1} = (1, \ldots, 1)$ and thus is equal to $\sum_{j=1}^{J} \frac{\partial f_{k,K-1}}{\partial z_j} (\mathbf{z} = \mathbf{0})$ if $f_{k,K-1}$ is differentiable.

Therefore, we can write

$$\lim_{\Delta z \to 0} \frac{q(\mathbf{z} = \mathbf{\Delta z}) - q(\mathbf{z} = \mathbf{0})}{\Delta z} \quad = \quad B\left[\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j} (\mathbf{z} = \mathbf{0})\right) * \pi\right]$$

Because the lhs is identified, this shows that we can identify $\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j} (\mathbf{z} = \mathbf{0})\right) * \pi$.

Next, for $j \in \mathcal{J}$, we can take another derivative wrt z_j and write

$$q_{(j)}(\mathbf{z} = \mathbf{0}) = A_{(j)}\pi + B_{(j)} \left(\frac{\partial f}{\partial z_j}(\mathbf{z} = \mathbf{0}) * \pi\right) \tag{B.1}$$

for known K−by−K matrices $A_{(j)}, B_{(j)}$ and a known column K−vector $q_{(j)}(\mathbf{z} = \mathbf{0})$. Note that $B_{(j)} = B$ for all $j \in \mathcal{J}$ and so we can write

$$\sum_{j=1}^{J} q_{(j)}(\mathbf{z} = \mathbf{0}) = \left(\sum_{j=1}^{J} A_{(j)}\right) \pi + B\left[\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j} (\mathbf{z} = \mathbf{0})\right) * \pi\right] \tag{B.2}$$

From above, $\left(\sum_{j=1}^{J} \frac{\partial f}{\partial z_j} (\mathbf{z} = \mathbf{0})\right) * \pi$ is identified. This implies that π is identified if the matrix $\sum_{j=1}^{J} A_{(j)}$ is invertible.[2]

B.1.2 Endogenous attributes

Here, we show how to extend our results to the case where some product attributes are endogenous (Section 2.3.3). Letting $\delta = (\delta_1, \ldots, \delta_J)$, we may write the share of good j as

$$s_j = \sigma_j (\delta, \mathbf{z}, \mathbf{p}) \tag{B.3}$$

[1] Note that this condition is immediately verifiable since the points in the support of α and β are chosen by the researcher.

[2] Again, this condition is immediately verifiable given the support points for α and β chosen by the researcher.

for some function σ_j. Repeating this for all j and stacking the equations, we obtain a demand system of the form

$$\mathbf{s} = \sigma\left(\delta, \mathbf{z}, \mathbf{p}\right) \tag{B.4}$$

where $\mathbf{s} = (s_1, \ldots, s_J)$. We also define the share of the outside option as $s_0 \equiv 1 - \sum_{j=1}^{J} s_j$, with associated function $\sigma_0\left(\delta, \mathbf{z}, \mathbf{p}\right)$. We establish nonparametric identification of this demand system by invoking results from Berry and Haile (2014) (henceforth, BH).[3] Specifically, the results in BH yield identification of $(\xi_j)_{j=1}^{J}$ for every unit (individual or market) in the population. This means that all the arguments of σ are known, which immediately implies (nonparametric) identification of σ itself. Once σ is identified, one may apply our results in Section 2.3.2 to identify the distribution of the preference parameters α, β_i and λ_i. Note that, while knowledge of σ is sufficient for several counterfactuals of interest (e.g., computing equilibrium prices after a potential merger or tax), the preference parameters are required to predict how choices and welfare would change if consumers were given full information, among other things. In this sense, our approach complements the identification results in BH within the class of search models we consider.

To prove identification of σ, we first note that model (B.26) satisfies the index restriction in BH's Assumption 1. Second, we assume that we have excluded instruments $\mathbf{w}$ which, together with the exogenous attributes, satisfy the following mean-independence restriction

$$E\left(\xi_j | \mathbf{x}, \mathbf{z}, \mathbf{w}\right) = 0 \quad \text{for all } j \tag{B.5}$$

almost surely (Assumption 3 in BH) and assume that the instruments shift the endogenous variables (market shares and endogenous attributes $\mathbf{p}$) to a sufficient degree (as in BH's Assumption 4). Finally, we verify that the demand system satisfies the "connected substitutes" restriction defined in BH's Assumption 2. To this end, we prove the following result.

Lemma 6. *Let utility be given by* (2.12) *with ϵ_i supported on $\mathbb{R}^J$ and let Assumptions (i), (iii), (iv), and either (i) or (ii) hold. Then, for all $j, k = 1, \ldots, J$ with $j \neq k$, σ_j is (i) strictly increasing in δ_j and (ii) strictly decreasing in δ_k.*

Proof. First, assume that p_j is part of the visible utility of good j and fix (δ_j, p_j, z_j) for all j. To prove claim (i), we show that an increase in δ_j can only induce a consumer to switch from not choosing j to choosing j but never vice versa, and that a positive mass of consumers will switch to choosing j. To see this, consider the case where consumer i initially searches j, which happens if and only if $g_i\left(\delta_j, \epsilon_{ij}, p_j, U_{ik}\right) \geq 0$ for all k such that $VU_{ik} \geq VU_{ij}$. Let $\Delta \geq 0$ be the change in δ_j. Since g_i is increasing in its first argument, we have $g_i\left(\delta_j + \Delta, \epsilon_{ij}, p_j, U_{ik}\right) \geq 0$ for all k such that $VU_{ik} \geq VU_{ij} + \Delta$ and thus i will still search j. Moreover, since g_i is decreasing in its last argument, if $g_i\left(\delta_k, \epsilon_{ik}, p_k, U_{ij}\right) \leq 0$ for some k such that $VU_{ik} \leq VU_{ij}$ (i.e. if k is initially not searched), then $g_i\left(\delta_k, \epsilon_{ik}, p_k, U_{ij} + \Delta\right) \leq 0$ (i.e. k is also not searched after the change in δ_j), which means that the set of goods searched by i never becomes larger. Next, note that if $U_{ij} \geq U_{ik}$ for all k in the set of searched goods $\mathcal{G}_i$, then $U_{ij} + \Delta \geq U_{ik}$ for all $k \in \mathcal{G}_i$. Further, since ϵ_i is supported on all of $\mathbb{R}^J$, there is a positive mass of consumers for which $U_{ik} \geq U_{ij}$ for some $k \in \mathcal{G}_i$, but $U_{ij} + \Delta \geq U_{ik}$ for all $k \in \mathcal{G}_i$. An analogous argument proves claim (ii).

[3] See also Berry et al. (2013).

Since the argument above does not rely on the fact that p_j is part of the visible utility of good j, the conclusion also holds for the case in which p_j is only uncovered upon searching good j. $\square$

Lemma 9 implies that the goods are connected substitutes in δ (see Definition 1 in BH), which in turn allows us to prove identification of σ by invoking Theorem 1 in BH.[4] Since Lemma 9 holds under either Assumption (*i*) or (*ii*), we obtain identification of preferences both in the case where p_j is part of the visible utility of good j and in the case where p_j is only uncovered upon searching j. Moreover, Theorem 1 of BH implies that one can invert the demand system σ for the indices δ and write

$$\alpha x_j + \xi_j = \sigma_j^{-1}(\mathbf{s}, \mathbf{z}, \mathbf{p}) \tag{B.6}$$

for all j. Equations (B.29) and (B.28) naturally lead to a nonparametric instrumental variable approach to estimate σ_j^{-1} (and thus σ_j).[5]

B.1.3 Identification when Observables Impact Search but not Utility

Here, we state and prove the results described in Section 2.3.4. We make the following assumptions:

Assumption 6. *(i) If consumer i searches j, then i also searches all j' s.t. $m(VU_{ij'}, r_{j'}) \geq m(VU_{ij}, r_j)$, where m is strictly increasing in both arguments;*

(ii) There is at least one good $j \neq 1$ such that $r_j > r_1$;

(iii) The support of $(\mathbf{x}, \mathbf{z}) \Big| (r_1, \ldots, r_J)$ has positive Lebesgue measure for all $(r_1, \ldots,$ r_J.

(iv) The search model admits a discrete choice representation that also satisfies the independence of irrelevant alternatives (IIA) property.

Assumption (*iii*) is substantive: for identification purposes, we consider variation in product characteristics holding fixed product search position. In practice, search position is likely to vary as a function of observables (e.g. products are sorted in order of price). However, because of the discrete nature of search position, we are likely to see variation conditional on search position and this is the variation we will use to identify our model. Assumption (*iv*) requires that consumers' search behavior can be represented as a standard discrete choice model satisfying IIA. As shown in Armstrong (2017), the Weitzman (1979) sequential search model (see Example 1) can be represented as a discrete choice model where consumers maximize product-specific indices defined as the minimum between the utility and the reservation value for each product. Then, Assumption (*iv*) is satisfied by letting the ϵ_{ij} be Gumbel distributed.

Violations of the visible utility assumption due to search position will cause Lemma 8 to no longer hold as stated: the good with the highest value of z_j can be searched, another good j' may have higher utility (and thus higher visible utility), but good j' may not be searched because it has lower search position. However, an extension of Lemma 8 will still hold in this case, which then allows us to prove identification of preferences.

[4] Note that the proof of Theorem 1 in BH only uses the fact that goods are connected substitutes in δ, not in $-\mathbf{p}$.

[5] Compiani (2022) proposes to approximate σ_j^{-1} using Bernstein polynomials. We use a similar approach in Section 3.5 to estimate the demand function for the case without endogeneity.

Lemma 7. *Let Assumptions 7, (ii)-(iv), and 8 hold and let $x_j \in [\bar{x}-\eta, \bar{x}+\eta]$ for all j, for some $\eta > 0$ sufficiently small. Then, if consumer i searches good 1 (i.e. the good with the highest value of z), then i chooses the good which maximizes utility among all goods with $r_j \geq r_1$.*

Proof. Suppose there was a good j with $r_j \geq r_1$ and $U_{ij} > U_{i1}$ that consumer i does not search. We can follow the proof of Lemma 8 to show that $VU_{ij} > VU_{i1}$. By Assumption (i), this implies that good j is searched, which is a contradiction. □

In words, if higher search position only makes a good more likely to be searched, then goods with higher visible utility *and* higher search position will always be searched if good 1 is searched. Given this Lemma, we can apply a modification of the identification argument in Theorem 2 after conditioning on the subset of goods with higher search position than good 1 (defined as usual as the good with the largest value of z_j):

Theorem 7. *Let the assumptions of Lemma 10 hold and let utility be given by $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ with v increasing in both arguments and infinitely differentiable. Further, assume that $\frac{\partial^2 s_1}{\partial z_1 \partial x_{j^*}}(\mathbf{x}^*, \mathbf{z}^*) \neq 0$ for some $(\mathbf{x}^*, \mathbf{z}^*)$ and $j^* \neq 1$, s_1 is infinitely differentiable and $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$. Then, v is identified up to an additive constant.*

Proof. Let $R = \{j : r_j \geq r_1\}$. Under Assumption (iv), the choice probability for good 1 conditional on consumers choosing in R, denoted $s_{1|R}$, is equal to the choice probability for good 1 if consumers only faced R as their choice set. Further, by Lemma 10, the only mistake a consumer can make when faced with choice set R is to fail to search good 1 when it is in fact the good with the highest utility in R.

This argument exactly parallels the argument, except now we have additionally used the fact that $U_1 \geq U_j$ for all $j \in R$, since (i) if $j \in \mathcal{S}$, then $m(VU_1, r_1) \geq m(VU_j, r_j)$ implies $VU_1 \geq VU_j$, which in turn implies $U_1 \geq U_j$; (ii) if $j \notin \mathcal{S}$, then $g(x_1, \epsilon_1, U_j) \geq 0 \geq g(x_1, \epsilon_1, U_k)$ for all $k \in \mathcal{S}$ implies $U_j \leq U_k$. Note that $P^{\mathcal{S}}_{5new,2}$ does not depend on z_1 and $P^{\mathcal{S}}_{5new,1}(\mathbf{v}, x_1, \mathbf{r})$ only depends on x_j and z_j via v_j for $j \neq 1$. In practice, this means that, when estimating the model, one needs to take R as the choice set faced by consumers and drop those consumers that choose products outside R.

□

B.1.4 Identification of a model where consumers form expectations on z_j based on x_j

Here, we state and prove the results described in Section 2.3.5. Given γ_1, we can identify the ranking of goods in terms of $\tilde{z}$ and we label good 1 as the good with the largest value of $\tilde{z}$. Then, an argument analogous to that in Lemma 3 yields identification of $\frac{\beta}{\alpha+\beta\gamma_1}$.[6] We can also recover $\alpha + \beta\gamma_1$ in a manner that parallels our usual identification of α (Lemma 3). When $\tilde{z}_j = \tilde{z}$ for all j, consumers who search based on our visible utility assumption always maximize utility, and thus we can directly estimate $\alpha + \beta\gamma_1$ as the coefficient on x_j for those consumers (we provide a formal proof of this in the next subsection). Therefore, this gives separate identification of β and α given γ_1.

When γ_1 is unknown, we can identify β/α if we know its sign and make a further support assumption. Suppose that the sign of γ_1 is known (e.g. higher priced goods have

[6] In Lemma 3, we showed identification of $\frac{\beta}{\alpha}$ by taking derivatives of s_1 w.r.t. z_1, z_2, x_2. Similarly, here we obtain identification of $\frac{\beta}{\alpha+\beta\gamma_1}$ by taking derivatives of s_1 wrt $\tilde{z}_1, \tilde{z}_2, x_2$.

weakly higher quality). Without loss, we assume $\gamma_1 > 0$. In addition, suppose that there exist choice sets in which a good has both the highest value of z and the lowest value of x. Even when γ_1 is unknown, this good is known to maximize $\tilde{z}$; thus, we can label it by 1. Note that we cannot differentiate with respect to $\tilde{z}$ as in the case above since γ_1 and thus $\tilde{z}$ is unknown. However, with good 1 defined appropriately, Corollary 3 shows that cross-derivatives with respect to z_1, z_j, x_j for $j \neq 1$ identify β/α (specifically, consumers who search the good with the highest value of $\tilde{z}$ will always maximize utility, and so their sensitivity to x_j and z_j identifies their true preferences).

Identification of $\alpha + \beta\gamma_1$

Note that if $\tilde{z}_j = 0$ for all j, then consumers always maximize utility. Thus, seeing how choice probabilities change with x conditional on $\tilde{z}_j = 0$ for all j should help identify $\alpha+\beta\gamma_1$. Because the event $\tilde{z}_j = 0$ involves x_j, we need to differentiate choice probabilities with respect to x_j on the envelope satisfying the condition $\tilde{z}_j = 0$ for all x_j. Formally, fix any $j \in \mathcal{J}$ and choose (x_k, z_k) so that $z_k = \gamma_0 + \gamma_1 x_k$ (which implies $\tilde{z}_k = 0$) for all $k \neq j$. For every $\delta > 0$, let $\epsilon(\delta) \equiv \gamma_0 + (x_j + \delta)\gamma_1 - z_j$, so that $z_j + \epsilon(\delta) - E(z_j | x_j + \delta) = 0$. Note that $\epsilon(\delta)$ is known to the econometrician. Thus, evaluating the last display at $\mathbf{x} = \mathbf{0}$ yields identification of $(\alpha + \beta\gamma_1)$ under a parametric assumption on ϵ_i.

B.1.5 Unobservables revealed by search

Here, we show that the ratio of second derivatives in (2.3) robustly identifies $\frac{\beta}{\alpha}$ in the model where ϵ_{ij} is revealed to consumer i only upon searching good j (Section 2.3.6).

Order goods in increasing order of x. Then, for $j = 1, \dots, J$,

$$
\begin{aligned}
s_j &= \sum_{k=1}^{j} P\left(\{U_j \geq U_{j'}\ \forall\, j' \in \{k, \dots, J\}\} \cap \{\text{search exactly } k, \dots, J\}\right) \\
&\quad \{g(x_h, \epsilon_h, U_j) \leq 0\ \forall h = 1, \dots, k-1\}) \\
&\equiv \sum_{k=1}^{j} P_j^{(k)}\left(\tilde{\mathbf{u}}, \mathbf{x}_{-J}\right),
\end{aligned}
$$

where $\tilde{u}_j = \alpha x_j + \beta z_j$ and $\tilde{\mathbf{u}} = (\tilde{u}_1, \dots, \tilde{u}_J)$, as above. Thus,

$$
\begin{aligned}
\frac{\partial^2 s_j}{\partial z_j \partial z_J} &= \sum_{k=1}^{j} \frac{\partial^2 P_j^{(k)}}{\partial \tilde{u}_j \partial \tilde{u}_J} \beta^2 \\
\frac{\partial^2 s_j}{\partial z_j \partial x_J} &= \sum_{k=1}^{j} \frac{\partial^2 P_j^{(k)}}{\partial \tilde{u}_j \partial \tilde{u}_J} \alpha\beta
\end{aligned}
$$

So the ratio of the latter two derivatives identifies $\frac{\beta}{\alpha}$. (Note that the ratio of $\frac{\partial s_j}{\partial z_J}$ to $\frac{\partial s_j}{\partial x_J}$

for any j would also work). On the other hand,

$$\begin{aligned}\frac{\partial s_j}{\partial z_j} &= \sum_{k=1}^{j} \frac{\partial P_j^{(k)}}{\partial \tilde{u}_j}\beta \\ \frac{\partial s_j}{\partial x_j} &= \sum_{k=1}^{j} \left(\frac{\partial P_j^{(k)}}{\partial \tilde{u}_j} + \frac{1}{\alpha}\frac{\partial P_j^{(k)}}{\partial x_j} \right)\alpha\end{aligned} \tag{B.7}$$

Since $\frac{1}{\alpha}\frac{\partial P_j^{(k)}}{\partial x_j} \geq 0$, (B.30) implies that the ratio of first derivatives suffers from attenuation bias, i.e. $\frac{\frac{\partial s_j}{\partial z_j}}{\frac{\partial s_j}{\partial x_j}} \leq \frac{\beta}{\alpha}$.

B.1.6 $K-$rank model

Consider the simultaneous search model in Honka et al. (2017) with $J = 2$ goods. In this model, a consumer looks at the visible utilities and decides whether to search the good with the highest visible utility or search both goods. Searching a second good entails a cost c, constant across consumers. As usual, we denote by 1 the good with the highest value of z.

Note that consumer i searches 2 but not 1 if and only if $VU_{i2} > VU_{i1}$ and

$$E_{z_1,z_2}\left[\max\left\{VU_{i1} + \beta z_1, VU_{i2} + \beta z_2\right\}\right] - c < E_{z_2}\left[VU_{i2} + \beta z_2\right] \tag{B.8}$$

i.e.

$$E_{z_1,z_2}\left[\max\left\{VU_{i1} - VU_{i2} + \beta\left(z_1 - z_2\right), 0\right\}\right] - c < 0 \tag{B.9}$$

or $g_{sim}\left(VU_{i1} - VU_{i2}\right) < 0$ for an increasing function g_{sim}. Equation (2.4) then can be written as

$$\begin{aligned} s_1 &= P(U_1 > U_2) - P(\{U_1 > U_2\} \cap \{VU_2 > VU_1\} \cap \{g_{sim}(VU_1 - VU_2) < 0\}) \\ &= P_{1,sim} - P_{2,sim} \end{aligned} \tag{B.10}$$

Letting $\tilde{u}_j = \alpha x_j + \beta z_j$ and $\tilde{vu}_j = \alpha x_j$, we also have:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \beta^2 \left(\frac{\partial^2 P_{1,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} \right) \tag{B.11}$$

and

$$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \alpha\beta \left(\frac{\partial^2 P_{1,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} \right) \tag{B.12}$$

So, if $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} = 0$, then the ratio of (B.34) to (B.35) identifies $\frac{\beta}{\alpha}$. Note that the event in $P_{2,sim}$ is equivalent to the following set of inequalities: (i) $\epsilon_{i1} > \tilde{u}_2 - \tilde{u}_1 + \epsilon_{i2}$, (ii) $\epsilon_{i1} < \tilde{vu}_2 - \tilde{vu}_1 + \epsilon_{i2}$, (iii) $\epsilon_{i1} < g_{sim}^{-1}(0) + \tilde{vu}_2 - \tilde{vu}_1 + \epsilon_{i2}$, where $VU_{ij} = \tilde{vu}_j + \epsilon_{ij}$ and

$U_{ij} = \tilde{u}_j + \epsilon_{ij}$, as above. Then, letting $\tilde{\epsilon} = \epsilon_1 - \epsilon_2$, we have:

$$P_{2,sim} = \int_{\tilde{u}_2-\tilde{u}_1}^{\min(\tilde{vu}_2-\tilde{vu}_1, g_{sim}^{-1}(0)+\tilde{vu}_2-\tilde{vu}_1)} f_{\tilde{\epsilon}}(\tilde{\epsilon})d\tilde{\epsilon} = \int_{\beta(z_2-z_1)}^{\min(0,g_{sim}^{-1}(0))} f_{\tilde{\epsilon}}(\tilde{\epsilon})d\tilde{\epsilon}$$

Thus, $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} = 0$.

Finally, we show that the ratio of first derivatives leads to attenuation bias. This follows directly from

$$\begin{aligned}
\frac{\partial s_1}{\partial z_1} &= \beta\left(\frac{\partial P_{1,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{u}_1}\right) \\
\frac{\partial s_1}{\partial x_1} &= \alpha\left(\frac{\partial P_{1,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{vu}_1}\right)
\end{aligned}$$

and the fact that $\frac{\partial P_{2,sim}}{\partial \tilde{vu}_1} < 0$.

B.2 Testing for full information with heterogeneous preferences

In Section 2.2.4, we considered the problem of testing the null hypothesis of full information and showed that, in the case where the coefficients α and β are homogeneous across consumers, a valid test rejects the null when the ratios of first derivatives are attenuated relative to the ratio of second derivatives in (2.10). Here, we provide conditions under which the same test is valid in the case where one of the two coefficients is allowed to be heterogeneous.[7] We focus on the case where β is heterogeneous and z_j is a scalar; the argument for the case where α is heterogeneous (and x_j is a scalar) is analogous. We also assume that the ϵ_{ij} shocks are type-I extreme-value distributed and let $s_j(\tilde{\beta})$ be the market share of good j for consumers with $\beta = \tilde{\beta}$ under full information, i.e. $s_j(\mathbf{x}, \mathbf{z}; \tilde{\beta}) \equiv \frac{\exp(\alpha x_j + \tilde{\beta} z_j)}{\sum_{k=1}^{J} \exp(\alpha x_k + \tilde{\beta} z_k)}$.

We let $j = 2, k = k' = 1$ in equation (2.10), i.e. we consider the case where the test compares the ratio of second derivatives taken with respect to good 1 and 2 to the ratio of first derivatives taken with respect to good 1. Analogous sufficient conditions could be obtained for different choices of j, k, k'. Then, we want to show that

$$\frac{\int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta}{\alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta))dF_\beta} \geq \frac{-\int s_1(\mathbf{x}, \mathbf{z}; \beta)s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta^2 dF_\beta}{-\alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta)s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta}$$

where F_β denotes the distribution of β. We take a pair $(\mathbf{x}, \mathbf{z})$ such that $\frac{\partial s_1(\mathbf{x},\mathbf{z})}{\partial x_1} > 0$ and $\frac{\partial^2 s_1(\mathbf{x},\mathbf{z})}{\partial z_1 \partial x_2} > 0$ (both of which can be verified from the data), so that (B.13) holds if and

[7]The reason why we let only one of the coefficients be heterogeneous is that we leverage a result from the statistics literature that applies to ratios of one-dimensional integrals.

only if

$$
\begin{aligned}
- \quad & \alpha \int s_1(\mathbf{x},\mathbf{z};\beta)s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta dF_\beta \int s_1(\mathbf{x},\mathbf{z};\beta)(1-s_1(\mathbf{x},\mathbf{z};\beta))\beta dF_\beta \geq \\
- \quad & \int s_1(\mathbf{x},\mathbf{z};\beta)s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta^2 dF_\beta \alpha \int s_1(\mathbf{x},\mathbf{z};\beta)(1-s_1(\mathbf{x},\mathbf{z};\beta))dF_\beta
\end{aligned}
$$

Then, by Theorem 2 of Wijsman (1985), the desired inequality holds if (i) $\beta > 0$, and (ii) $\frac{\alpha}{\beta}$ and $\frac{-s_1(\mathbf{x},\mathbf{z};\beta)s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{s_1(\mathbf{x},\mathbf{z};\beta)(1-s_1(\mathbf{x},\mathbf{z};\beta))} = -\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)}$ are monotonic functions of β in the same direction. Since we assumed throughout that $\alpha > 0$, we want to show that $-\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)}$ decreases in β monotonically. After some algebra, we have that

$$
\frac{\partial\left[-\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)}\right]}{\partial\beta} < 0 \qquad \forall\beta
$$

Under these conditions, at the chosen values of $\mathbf{x},\mathbf{z}$, a valid test of the null of full information rejects when the ratio of first derivatives is sufficiently attenuated relative to the ratio of first derivatives. Note that the condition in (**??**) can be verified given the support of the distribution of β. For example, if β takes values on a finite grid of points (as in Section 2.3.2), then one needs to check whether (**??**) holds for all values in the grid. Finally, we emphasize that (**??**) is a sufficient, but in general not necessary condition, implying that the proposed test could be valid even if the restriction is not satisfied.

B.3 Simulations Results

To test the performance of our approach, we consider several simulations. In all simulations, we generate $N = 20,000$ choices with utility given by:

$$
U_{ij} = \alpha x_{ij} + \beta z_{ij} + \epsilon_{ij} \tag{B.13}
$$

with $\alpha = \beta = 1$, $x_{ij} \sim_{i.i.d} N(0,1)$, $z_{ij} \sim_{i.i.d.} N(0,1)$, and ϵ_{ij} i.i.d. Type 1 extreme value.

We simulate data from four data generating processes, three of which satisfy the assumptions of our theorem and one of which does not. These are:

1. Weitzman search, with search costs $c \sim LogNormal(-2, 2.25)$
2. Satisficing, searching in order of visible utility until utility-in-hand is at least $T \sim LogNormal(-0.35, 2.25)$
3. Search all goods with visible utility above a threshold given by $c \sim N(-1, 16)$ (if no goods are above the threshold, search and choose the good with the highest visible utility)
4. Randomly search $K \in \{1, \dots, J\}$ goods, where the searched goods are the K highest in terms of visible utility

DGPs 1-3 satisfy our assumptions. By contrast, DGP 4 violates Assumption (ii) because the decision of whether to search a good does not just depend on that good's visible utility, but on the visible utilities of all goods.

Bernstein Polynomial Simulation Results Table B.1 reports results from the Bernstein approximation of the cross-derivative ratio which identifies β/α. For comparison, we also report estimates of $\frac{\partial s_j/\partial z_j}{\partial s_j/\partial x_j}$, which would recover β/α with full information. In all cases, the estimates based on first-derivatives are attenuated relative to the true values. This occurs for the reason discussed in Section 3.4: consumer insensitivity to variation in z for goods that are not searched biases the coefficients towards zero. In contrast, the confidence intervals from Bernstein estimation of the cross-derivative ratio include the true values in DGPs 1-3, and are fairly precise for the $J = 3$ case. For DGP 4, where the assumptions of our model do not hold (see Section 2.3.7), the coefficient is attenuated for $J = 3$, although the point estimates remain much closer to the true values the first-derivative estimates.

Table B.1: Bernstein Approximation

	Number of Goods			
	2		3	
DGP	First-Derivatives	Cross-Derivatives	First-Derivatives	Cross-Derivatives
1	0.610	0.977	0.403	0.997
	(0.024)	(0.304)	(0.012)	(0.076)
2	0.691	1.280	0.361	0.935
	(0.024)	(0.538)	(0.014)	(0.068)
3	0.527	0.870	0.330	0.872
	(0.021)	(0.190)	(0.010)	(0.071)
4	0.444	0.801	0.206	0.626
	(0.018)	(0.301)	(0.010)	(0.075)

Note: Across all rows, the data the sample size is $N = 20,000$ and the data in each row is generated by the corresponding DGP described in the main text. In all cases, the true value is 1. Standard errors, obtained via 250 bootstrap repetitions, are reported in parentheses.

Flexible Logit Simulation Results For each of the DGPs described in Section 2.4.1, we consider simulations with $J \in \{2, 3, 5, 10\}$. We report estimates from the flexible logit model as well as the standard logit model. We bootstrap the standard errors using 250 repetitions.

Results from these simulations are reported in Table B.2. The table shows estimates of β/α from a conditional logit model with no adjustment for imperfect information, as well as the cross-derivative ratio estimates from the flexible logit model. In the standard logit model, the coefficient is attenuated, typically biased towards zero by 30-50%. The flexible logit model performs substantially better, with 95% confidence intervals including the true estimates in DGPs 1-3. Perhaps surprisingly, the flexible logit model also performs well for DGP 4; the confidence intervals include the true values for 2 and 5 goods, and have less bias than the standard logit model for 5 and 10 goods.

Table B.2: Estimator based on cross-derivatives ratio (flexible logit) vs standard logit

	Number of Goods							
	2		3		5		10	
DGP	Standard	Flexible	Standard	Flexible	Standard	Flexible	Standard	Flexible
1	0.6590	1.0214	0.6330	1.0671	0.6050	0.9633	0.5770	0.8986
	(0.0158)	(0.1208)	(0.0122)	(0.1259)	(0.0095)	(0.1254)	(0.0089)	(0.1053)
2	0.7403	0.9976	0.6194	1.0854	0.4587	1.0407	0.2909	1.0004
	(0.0162)	(0.1034)	(0.0135)	(0.1300)	(0.0102)	(0.1578)	(0.0083)	(0.2603)
3	0.5424	1.1177	0.5945	1.0286	0.6543	0.9017	0.7246	0.8822
	(0.0149)	(0.1716)	(0.0117)	(0.1469)	(0.0099)	(0.1071)	(0.0106)	(0.0733)
4	0.4543	1.1358	0.5568	0.9614	0.6691	0.8015	0.7887	0.8151
	(0.0140)	(0.1906)	(0.0118)	(0.1659)	(0.0105)	(0.1012)	(0.0104)	(0.0679)

Note: Across all rows, the data the sample size is $N = 20,000$ and the data in each row is generated by the corresponding DGP described in the main text. "Standard" refers to estimates of β/α from a conventional logit model, and "Flexible" refers to estimates from the flexible logit model. In all cases, the true value is 1. Standard errors, obtained via 250 bootstrap repetitions, are reported in parentheses

B.4 Derivation of Flexible Logit Weights and Choice Probabilities

To motivate our parametric approach to estimating $s_1(\mathbf{x}, \mathbf{z})$, note that standard full-information logit models typically impose strong restrictions on the structure of the derivatives of choice probabilities. Specifically, if $u_{ij} = v_j^* + \epsilon_{ij}$ and ϵ_{ij} is i.i.d. extreme value where v_j^* is a differentiable function of x_j and z_j, then for $q_j \in \{x_j, z_j\}$:

$$\begin{aligned}
\frac{\partial s_j}{\partial q_j} &= \frac{\partial s_j}{\partial v_j^*}\frac{\partial v_j^*}{\partial q_j} = \frac{\partial v_j^*}{\partial q_j} s_j(1-s_j) \\
\frac{\partial s_j}{\partial q_{j'}} &= \frac{\partial s_j}{\partial v_{j'}^*}\frac{\partial v_{j'}^*}{\partial q_{j'}} = -\frac{\partial v_{j'}^*}{\partial q_j'} s_j s_{j'} \\
\frac{\partial^2 s_j}{\partial z_j \partial q_{j'}} &= -\frac{\partial v_j^*}{\partial q_{j'}}\frac{\partial v_j^*}{\partial z_j} s_j s_{j'}(1-2s_j) \qquad \text{(B.14)}
\end{aligned}$$

for $j' \neq j$. Thus, in a conventional logit model, $\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \frac{\partial s_1}{\partial z_{j'}} / \frac{\partial s_1}{\partial x_{j'}} = \frac{\partial v_{j'}^*}{\partial z_{j'}} / \frac{\partial v_{j'}^*}{\partial x_{j'}}$ for all $j' \neq 1$, and this further equals $\frac{\partial s_1}{\partial z_1} / \frac{\partial s_1}{\partial x_1}$ when $\frac{\partial v_j^*}{\partial q_j} = \frac{\partial v_{j'}^*}{\partial q_{j'}}$ for all j, j'. We would like to estimate a model of s_1 which is sufficiently flexible that ratios of first-derivatives differ from ratios of second cross-derivatives, as will generally occur if consumers engage in search. To allow for this additional flexibility, we let the utility for good 1 depend directly on attributes of rival goods as follows:

$$v_1 = \tilde{v}(x_1, z_1) + b_1 z_1 + \sum_{k \neq 1} \left(\gamma_k w_{z1k} z_k + \gamma_{2k} w_{x1k} x_k + w_{z2k}\delta_k z_k z_1 + w_{x2k}\delta_{2k} x_k z_1\right) \quad \text{(B.15)}$$

where $\tilde{v}(x, z)$ is a differentiable function of x and z, w_{z1k}, w_{x1k}, w_{z2k} and w_{x2k} are known weights, and b_1, γ_k, γ_{2k}, δ_k and δ_{2k} are coefficients to be estimated. Further, we let $v_k = \tilde{v}(x_k, z_k)$ for $k \neq 1$.

In this section, we derive the relevant derivatives of choice probabilities for the flexible logit model described in the text and motivate our choice of weights. The weights w_{x1k}, w_{z1k}, w_{x2k} and w_{z2k} are chosen so that, given the logit functional form, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ can be constant across goods as our structural model implies when these weights are regarded as constant in derivatives. With these weights, we have the following derivatives (where we use the notation $\tilde{v}_j$ to refer to the function $\tilde{v}$ evaluated at (x_j, z_j):

$$
\begin{aligned}
\frac{\partial v_1}{\partial z_1} &= \frac{\partial \tilde{v}_1}{\partial z} + b_1 + \sum_{k \neq 1} (w_{z2k}\delta_k z_k + w_{x2k}\delta_{2k} x_k) \\
\frac{\partial s_1}{\partial x_1} &= \frac{\partial s_1}{\partial v_1}\frac{\partial v_1}{\partial x_1} = \frac{\partial \tilde{v}_1}{\partial x} s_1(1-s_1) \\
\frac{\partial s_1}{\partial z_1} &= \frac{\partial s_1}{\partial v_1}\frac{\partial v_1}{\partial z_1} = \frac{\partial v_1}{\partial z_1} s_1(1-s_1) \\
\frac{\partial s_1}{\partial x_{j'}} &= \frac{\partial s_1}{\partial v_{j'}}\frac{\partial v_{j'}}{\partial x_{j'}} + \frac{\partial s_1}{\partial v_1}\frac{\partial v_1}{\partial x_{j'}} = -\frac{\partial \tilde{v}_{j'}}{\partial x} s_1 s_{j'} + [w_{x1j'}\gamma_{2j'} + w_{x2j'}\delta_{2j'} z_1] s_1(1-s_1) \\
\frac{\partial s_1}{\partial z_{j'}} &= \frac{\partial s_1}{\partial v_{j'}}\frac{\partial v_{j'}}{\partial z_{j'}} + \frac{\partial s_1}{\partial v_1}\frac{\partial v_1}{\partial z_{j'}} = -\frac{\partial \tilde{v}_{j'}}{\partial z} s_1 s_{j'} + [w_{z1j'}\gamma_{j'} + w_{z2j'}\delta_{j'} z_1] s_1(1-s_1) \\
\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial x_{j'}}\frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1}\frac{\partial^2 v_1}{\partial z_1 \partial x_{j'}} \\
&= \frac{\partial v_1}{\partial z_1}(1-2s_1)\frac{\partial s_1}{\partial x_{j'}} + s_1(1-s_1) w_{x2j'}\delta_{2j'} \\
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial z_{j'}}\frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1}\frac{\partial^2 v_1}{\partial z_1 \partial z_{j'}} \\
&= \frac{\partial v_1}{\partial z_1}(1-2s_1)\frac{\partial s_1}{\partial z_{j'}} + s_1(1-s_1) w_{z2j'}\delta_{j'}
\end{aligned}
\tag{B.16}
$$

Thus, we have:

$$
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \frac{-\frac{\partial \tilde{v}_{j'}}{\partial z} + \gamma_{j'} + \delta_{j'}}{-\frac{\partial \tilde{v}_{j'}}{\partial x} + \gamma_{2j'} + \delta_{2j'}} \tag{B.17}
$$

Given a linear specification of $\tilde{v}$, $\tilde{v}(x_j, z_j) = x_j a_1 + z_j a_2$, this implies that the above ratio is a constant for each j'.

Estimation of the model with these weights is infeasible since the levels of the choice probabilities s_1 and s_k, as well as the derivatives $\partial v_1/\partial z_1$ are unknown ex ante and thus we do not know the weights. We estimate the model via a two-step process where s_1 and s_k are estimated using a standard logit model (where utility for each good is a linear function of x_j and z_j), these estimates are used to construct weights, and then the model in equation (B.15) is estimated treating these weights as constants.[8]

To recover estimates of β/α from the flexible logit model, we use the ratio in equation (B.38). With the linear specification of $\tilde{v}$, this ratio is given by $\frac{\beta}{\alpha} = \frac{-a_2 + \gamma_{j'} + \delta_{j'}}{-a_1 + \gamma_{2j'} + \delta_{2j'}}$. In cases where the identity of goods is not meaningful (e.g. "good 2" does not refer to the same

[8]Since $\partial v_1/\partial z_1$ is estimated imprecisely from the standard logit, when $1 + (\partial v_1/\partial z_1)(1-2s_1)z_1$ is close to 0 (leading to very large weights), we set $\partial v_1/\partial z_1 = 0$ when the former term falls below 1 in absolute value.

good across different choice sets and there are no alternative-specific fixed effects), we can further impose $\gamma_k = \gamma$, $\gamma_{2k} = \gamma_2$, $\delta_k = \delta$ and $\delta_{2k} = \delta_2$, which gives a single estimate of $\frac{\beta}{\alpha}$.

B.5 Recovery of Search Costs Given Preferences in the Weitzman Model

Suppose that utility is given by $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ and that consumers search sequentially according to the model of Weitzman (1979).

As shown in Armstrong (2017),[9] the optimal search strategy is for consumers to behave as if they were choosing among options in a static model with utilities given by $\tilde{U}_{ij} = x_j\alpha + \min\{z_j, rv_i\}\beta + \epsilon_{ij}$, where rv_i denotes i's reservation value in units of z (see Example 1). Thus, dropping i subscripts, ordering goods so that $z_1 \geq z_2 \geq \ldots \geq z_J$, we can write

$$\begin{aligned} s_1 &= P\left(x_1\alpha + \min\{z_1, rv\}\beta + \epsilon_1 \geq x_k\alpha + \min\{z_k, rv\}\beta + \epsilon_k \ \forall k\right) \\ &= P\left(\epsilon_k - \epsilon_1 \leq (x_1 - x_k)\alpha \ \forall k\right) P\left(rv \leq z_J\right) \\ &+ \sum_{t=0}^{J-2} \int P\left(\{\epsilon \in E_t\} \cap \{z_{J-t} \leq rv \leq z_{J-t-1}\}\right) dF_{rv}(rv) \\ &+ P\left(\epsilon_k - \epsilon_1 \leq (x_1 - x_k)\alpha + (z_1 - z_k)\beta \ \forall k\right) P\left(rv \geq z_1\right) \end{aligned}$$

where F_{rv} denotes the cdf of rv and the second equality assumes that search costs (and thus rv) are independent of ϵ. Therefore, we have

$$\frac{\partial s_1}{\partial z_1} = \left[\frac{\partial}{\partial z_1} P\left(\epsilon_k - \epsilon_1 \leq (x_1 - x_k)\alpha + (z_1 - z_k)\beta \ \forall k\right)\right] P\left(rv \geq z_1\right) \tag{B.18}$$

Given identification of (α, β) by the argument in Section 3.4, the first term on the rhs of (B.39) is identified given parametric assumptions on the distribution of ϵ. Thus, $P\{rv \geq z_1\}$ is identified. Repeating the argument for all z_1, one can trace out the entire distribution of rv. Since c, the search cost for consumer i, is a known transformation of rv,[10] the distribution of c is also identified.

Equation (B.39) also lends itself to a different argument that does not require making a parametric assumption on the distribution of ϵ, but instead relies on "at-infinity" variation. Note that the first term on the rhs of (B.39) is invariant to increasing all z_j's by the same amount. Thus, we can write

$$\frac{\frac{\partial s_1}{\partial z_1}(\mathbf{z} + \mathbf{\Delta})}{\frac{\partial s_1}{\partial z_1}(\mathbf{z})} = \frac{P\left(rv \geq z_1 + \Delta\right)}{P\left(rv \geq z_1\right)} \tag{B.19}$$

where $\mathbf{\Delta}$ is a $J-$vector with all elements equal to some Δ. Letting $\Delta \to -\infty$, the numerator on the rhs of (B.40) goes to 1, which yields identification of $P\left(rv \geq z_1\right)$.

[9] See also Choi et al. (2018b).

[10] This assumes that the prior F_z used by consumers in forming expectations are known to the researcher, as in the case where consumers have rational expectations and F_z coincides with the observed distribution of z across goods and/or markets.

Repeating the argument for all z_1, one can trace out the entire distribution of rv and recover the distribution of c as above.

B.6 Welfare Benefits of Information

Evaluating the welfare benefits of information requires three pieces: first, status quo choice probabilities; second, ex post choice probabilities with information; third, a normative utility function to evaluate choices. Status quo choice probabilities are identified nonparametrically from the data. To deal with the curse of dimensionality, we assume that status quo choice probabilities are well-approximated by the standard logit, estimated on existing choices (the flexible logit estimates could be used here instead). We assume that informed choices would be given by a logit model with the flexible logit estimate of β used in lieu of the standard logit estimate. We assume that this same logit model (with the flexible logit estimate of β) is also suitable for evaluating normative choices.

Appendix D of Abaluck and Gruber (2009) shows that dollar-equivalent consumer surplus in logit models where positive preferences (i.e., preferences describing potentially uninformed behavior) are given by β_{pos} and normative preferences (i.e., those relevant for welfare evaluations) are given by β_{norm} can be computed as:

$$E(CS_0) = -\frac{1}{\alpha_p}\left[\sum_k (x_k\beta_{norm} - x_k\beta_{pos})s_k(\beta_{pos}) + \ln\sum_k \exp(x_k\beta_{pos})\right]$$

where α_p is the (normative) marginal utility of income, estimated as the coefficient on price. Once consumers are informed and their preferences are β_{norm}, consumer surplus is given by the conventional log-sum formula:

$$E(CS_1) = -\frac{1}{\alpha_p}\ln\sum_k \exp(x_k\beta_{norm})$$

The change in consumer surplus from providing consumers with information is thus:

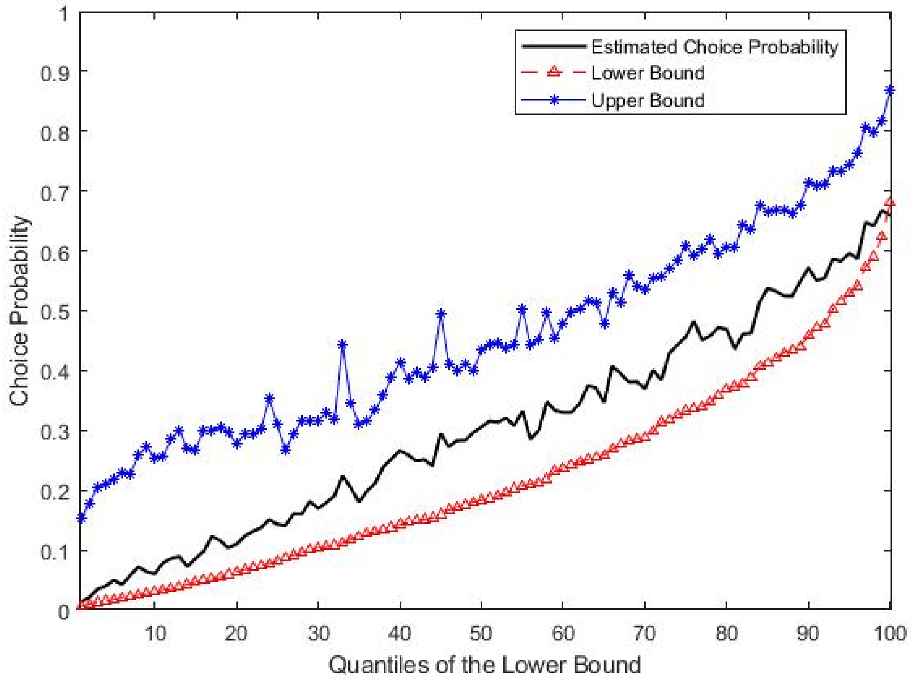
$$\Delta CS = -\frac{1}{\alpha_p}\left[\ln\sum_k \exp(x_k\beta_{norm}) - \ln\sum_k \exp(x_k\beta_{pos}) + \sum_k (x_k\beta_{pos} - x_k\beta_{norm})s_k(\beta_{pos})\right]$$

B.7 Testing the Visible Utility Assumption in the Laboratory Experiment

As discussed in Section 2.2.3, while the visible utility assumption cannot be verified directly, it can be tested along with the other restrictions of our model. One such test is to compute bounds on the choice probabilities implied by the model. Given our estimates of preferences and assumptions about the distribution of ϵ_{ij}, we can compute the upper and lower bounds described in Section 2.2.3 for each individual via simulation. We sort the data by the lower bound, bin the data into 100 quantiles, and graph in each quantile the mean of the upper and lower bounds, as well as the choice probabilities estimated via Bernstein polynomials.

Figure B.1 shows the results of this exercise. We can see that the bounds in the experimental data have some bite: the range between the lower bound and the upper bound ranges from 15 to 30 percentage points. The estimated choice probabilities in nearly all cases lie within this range. These probabilities thus appear broadly consistent with the visible utility assumption.

Figure B.1: Choice Probabilities, Upper and Lower Bounds from Visible Utility Assumption



B.8 Field Validation Details

B.8.1 Data Cleaning

The dataset from Kaggle.com contains 9,917,530 observations on a hotel-consumer level. We filtered out the following categories of observations.

First, the data set contains some errors in the price information. We removed search impressions that contain at least one observation for which the listed hotel price is below \$10 or above \$1000 per night, or the implied tax paid per night either exceeds 30% of the listed hotel price, or is less than \$1.

Second, we removed the search impressions where the consumer observed a hotel in position 5, 11, 17, 23. These positions usually correspond to "opaque offers" (Ursu (2018) provides a detailed description of this feature in the data).

Third, the original data set contains observations on more than 20,000 destinations, with a median of two search impressions per destination. We focused our attention on destinations with at least 50 search impressions.

Fourth, we kept the search impressions where all transactions happened within the top 10 positions excluding the opaque offer positions, and we only kept these top 10 hotels in these choice sets.

The final dataset then contains 54,648 choice sets and 546,480 observations. Table B.6 provides a detailed description of each variable.

Table B.3: Variable Description

Variable	Description
Price	Gross price in USD
Stars	Number of hotel stars
Review Score	User review score, mean over sample period
Chain	Dummy whether hotel is part of a chain
Location Score	Expedia's score for desirability of hotel's location
Promotion	Dummy whether hotel is on promotion

B.8.2 Confidence Interval Construction

Let $\hat{\beta}$ be the estimate from the original dataset. Let $n = 1, 2, ..., N$ denote the bootstrap samples, and $\hat{\beta}_n$ be the estimate from the nth bootstrap sample. In our case, we set $N = 250$.

Let $z_0 = \Phi^{-1}\{\#(\hat{\beta}_n \leq \hat{\beta})/N\}$, where $\#(\hat{\beta}_n \leq \hat{\beta})$ is the number of elements of the bootstrap distribution that are less than or equal to the estimate from the original dataset and Φ is the standard normal CDF. z_0 is known as the median bias of $\hat{\beta}$. Let $p_1 = \Phi\left(z_0 + \frac{z_0 - z_{1-\alpha/2}}{1-a(z_0 - z_{1-\alpha/2})}\right)$, $p_2 = \Phi\left(z_0 + \frac{z_0 + z_{1-\alpha/2}}{1-a(z_0 + z_{1-\alpha/2})}\right)$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the normal distribution. The bias-corrected and accelerated (BCa) method yields confidence intervals $[\beta^*_{p_1}, \beta^*_{p_2}]$, where β^*_p is the pth quantile of the bootstrap distribution $(\hat{\beta}_1, ..., \hat{\beta}_N)$. We use the bias-corrected (but not accelerated) method as a special case, i.e. we set $a = 0$.

B.8.3 Estimation Results

Table B.7 and B.8 show the detailed results from flexible logit and standard logit estimations. To facilitate comparisons, we report the coefficients multiplied by the standard deviation of the corresponding variable.

Table B.4: Estimation Results: Normalized β Estimates

z Variable	Standard Estimate	Standard CI	Flexible Estimate	Flexible CI
Location Score	0.298	(0.278, 0.317)	0.691	(0.591, 0.839)
Price	-1.085	(-1.109, -1.061)	-0.710	(-1.169, -0.503)
Review Score	0.172	(0.159, 0.185)	0.195	(0.082, 0.250)
Stars	0.386	(0.369, 0.403)	0.364	(0.258, 0.430)

Note: We report point estimates and 95% confidence intervals for the β coefficients for different choices of z variable. The first two columns report results from the standard logit model and the second two report results from the flexible logit approach.

Table B.5: Estimation Results: Difference in Magnitude of β Estimates

z Variable	Point Estimate	Confidence Interval
Location Score	0.393	(0.296, 0.544)
Price	-0.375	(-0.566, 0.085)
Review Score	0.023	(-0.069, 0.038)
Stars	-0.022	(-0.135, 0.038)

Note: For different choices of z variable, we report point estimates and 95% confidence intervals for the difference between the absolute value of the β coefficient estimate from the flexible logit approach and the absolute value of the estimate from the standard logit model.

B.9 Generalizations

B.9.1 Non-linear utility

Let individual i's utility from alternative j be denoted by $U_{ij}(x_j, z_j)$. In what follows, we often omit the dependence of U_{ij} on (x_j, z_j) unless it is necessary to avoid confusion. We can always write: $U_{ij} = a_{ij}(x_j) + b_{ij}(x_j, z_j)$ where $b_{ij}(x_j, 0) = 0$ (to see this, define $b_{ij}(x_j, z_j) = U_{ij}(x_j, z_j) - U_{ij}(x_j, 0)$). Since in our setting $a_{ij}(x_j)$ is the component of utility that is known to the consumer before engaging in search, we label it "visible utility," VU_{ij}. We make the following assumptions on the utility function.

Assumption 7. *(i) For all i and j, U_{ij} is strictly monotonic in z_j.*
(ii) For all i, the function $b_{ij}(x_j, z_j)$ is not alternative-specific, i.e. $b_{ij}(x_j, z_j) = b_i(x_j, z_j)$ for all j, and continuous in its first argument.

The class of utility functions satisfying Assumption 7 is broad and subsumes most specifications commonly used in empirical work as special cases, including logit with possibly nonlinear-in-characteristics utilities[11] and mixed-logit. For instance, in a mixed-logit model, one may specify $U_{ij} = \alpha_i x_j + \beta_i z_j + \epsilon_{ij}$. To map this specification into our notation, let $a_{ij}(x_j) = \alpha_i x_j + \epsilon_{ij}$, and $b_i(x_j, z_j) = \beta_i z_j$. As another example, consider the logit specification $U_{ij} = \alpha x_j + \beta z_j + \gamma x_j z_j + \epsilon_{ij}$. This is subsumed in our notation by letting $a_{ij}(x_j) = \alpha x_j + \epsilon_{ij}$, and $b_i(x_j, z_j) = \beta z_j + \gamma x_j z_j$.

Lemma 8. *Let Assumptions 7 and 2 hold and let $x_j \in [\bar{x} - \eta, \bar{x} + \eta]$ for all j, for some $\eta > 0$ sufficiently small. If consumer i searches good 1 (i.e. the good with the highest value of z), then i chooses the utility-maximizing good.*

Proof. If good 1 is searched but utility is not maximized, then for some unsearched j, $U_{ij} > U_{i1}$. Since $z_1 > z_j$, by monotonicity, $b_i(\bar{x}, z_1) > b_i(\bar{x}, z_j)$. By continuity of b_i in its first argument, this implies that for η sufficiently small, $b_i(x_1, z_1) \geq b_i(x_j, z_j)$.[12] Given

[11] We allow for nonlinearities subject to Assumption (i) being satisfied.

[12] More formally, by continuity, for all $\delta > 0$ there exists $\eta > 0$ such that if $|x_1 - x_j| < 2\eta$, then $b_i(x_j, z_j) - b_i(x_1, z_j) \leq \delta$. Therefore, we have:

$$\begin{aligned} b_i(x_j, z_j) &= b_i(x_1, z_j) + b_i(x_j, z_j) - b_i(x_1, z_j) \\ &\leq b_i(x_1, z_j) + \delta \\ &\leq b_i(x_1, z_1) \end{aligned}$$

where the last inequality follows by choosing $\delta \equiv \frac{b_i(x_1, z_1) - b_i(x_1, z_j)}{2}$.

this, $U_{ij} > U_{i1}$ implies $VU_{ij} > VU_{i1}$. But by Assumption (i), this implies that good j is searched, which is a contradiction. $\square$

To recover derivatives of v with respect to z, we will use Lemma 8. Specifically, we will take $x_j \in [\bar{x} - \eta, \bar{x} + \eta]$ for all j, where $\bar{x}$ and η are defined in Lemma 8, and use the fact that $\frac{\partial s_1}{\partial z_1}$ can be written as a function of terms which only depend on x_2 and z_2 via U_2. To formalize this, we let $\mathcal{J}_1 \equiv \{2, \ldots, J\}$, $v_j \equiv v(x_j, z_j)$ for all j, and $\mathbf{v} \equiv (v_1, \ldots, v_J)$. Similarly, we let $v_j^0 = v(x_j, 0)$ and $\mathbf{v}^0 = (v_1^0, \ldots, v_J^0)$.

The first equality follows from basic set algebra while the second follows from the fact that for all $j \in \mathcal{S}$ and all $k \in \mathcal{J}_1 \backslash \mathcal{S}$, (i) $VU_1 \geq VU_j$ implies $U_1 \geq U_j$ since $z_1 \geq z_j$ for all $j \in \mathcal{J}_1$; and (ii) $g_1(x_1, U_k) \geq 0 \geq g_1(x_1, U_j)$ implies $U_k \leq U_j$, which (together with the implication in (i)) implies $U_1 \geq U_k$. Thus the event $U_1 \geq U_k\ \forall k \in \mathcal{J}_1$ is implied by the other events inside the probability and can be dropped.

Note that $P_{5,2}^{\mathcal{S}}$ does not depend on z_1. Thus, omitting the function arguments, we have

$$\frac{\partial s_1}{\partial z_1} = \frac{\partial P_4}{\partial v_1}\frac{\partial v_1}{\partial z_1} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial P_{5,1}^{\mathcal{S}}}{\partial v_1}\frac{\partial v_1}{\partial z_1} \tag{B.20}$$

Differentiating again with respect to z_2 gives:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \frac{\partial^2 P_4}{\partial v_1 \partial v_2}\frac{\partial v_1}{\partial z_1}\frac{\partial v_2}{\partial z_2} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^2 P_{5,1}^{\mathcal{S}}}{\partial v_1 \partial v_2}\frac{\partial v_1}{\partial z_1}\frac{\partial v_2}{\partial z_2} \tag{B.21}$$

Differentiating equation (B.20) with respect to x_2 gives:

$$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\partial^2 P_4}{\partial v_1 \partial v_2}\frac{\partial v_1}{\partial z_1}\frac{\partial v_2}{\partial x_2} - \sum_{\mathcal{S} \subset \mathcal{J}_1, \mathcal{S} \neq \emptyset} \frac{\partial^2 P_{5,1}^{\mathcal{S}}}{\partial v_1 \partial v_2}\frac{\partial v_1}{\partial z_1}\frac{\partial v_2}{\partial x_2} \tag{B.22}$$

Combining (B.21) and (B.22), we obtain

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} / \frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \frac{\frac{\partial v_2}{\partial z_2}}{\frac{\partial v_2}{\partial x_2}} \tag{B.23}$$

Since this equation holds for all $(\mathbf{x}, \mathbf{z})$ such that $\frac{\partial^2 s_1}{\partial z_1 \partial x_2} \neq 0$ and we already showed that we can recover $\frac{\partial v}{\partial x}(0,0)$, we can also recover $\frac{\partial v}{\partial z}(0,0)$.

Next, note that, fixing $z_k = 0$ for all $k = 1, \ldots, J$ and $x_j = 0$ for all $j \neq 2$ in (B.20), we can write

$$\frac{\partial s_1}{\partial z_1} = k(l(x_2)) \tag{B.24}$$

where $l(x_2) : x_2 \mapsto v(x_2, 0)$. So by the chain rule we have that, for $n > 1$, $\frac{\partial^n s_1}{\partial z_1 \partial x_2^{n-1}}$ is a linear function of the $(n-1)-$th derivative of k with slope depending on the first derivative of l and intercept depending on derivatives of l and derivatives of k of order strictly less than $n - 1$. Further, by the above, all derivatives of l are known. Thus, we have a system of equations that can be uniquely solved for the derivatives of k by recursion.[13]

Next, we differentiate $\frac{\partial s_1}{\partial z_1}$ once with respect to z_2 and $n-2$ times with respect to x_2. Similar to the above, we can write

$$\frac{\partial s_1}{\partial z_1} = k(v(x_2, z_2)) \tag{B.25}$$

where now note that z_2 is no longer fixed at 0. Again by the chain rule we have that, for $n \geq 3$, $\frac{\partial^n s_1}{\partial z_1 \partial z_2 \partial x_2^{n-2}}$ evaluated at $(\mathbf{0}, \mathbf{0})$ is a linear function of $\frac{\partial^{n-1} v}{\partial z_2 \partial x_2^{n-2}}(0,0)$ with slope coefficient depending on $k'(v(0,0))$ and intercept depending on lower-order derivatives of v as well as derivatives of k.[14] Because all derivatives of k are known by the argument above, we can iteratively solve for $\frac{\partial^{n-1} v}{\partial z_2 \partial x_2^{n-2}}(0,0)$ for all $n \geq 3$.

The remaining terms in the Taylor expansion can be recovered by an analogous argument. Specifically, for any $n \geq 3, m \geq 2$, by differentiating (B.25) m times wrt z_2 and again $n-m-1$ times wrt x_2, one can write $\frac{\partial^n s_1}{\partial z_1 \partial z_2^m \partial x_2^{n-m-1}}$ as a linear function of $\frac{\partial^{n-1} v}{\partial z_2^m \partial x_2^{n-m-1}}(0,0)$ with known, nonzero slope and known intercept. This system can then be solved iteratively for $\frac{\partial^{n-1} v}{\partial z_2^m \partial x_2^{n-m-1}}(0,0)$ for all $n > m \geq 2$.

Therefore, we know all the coefficients in the Taylor-expansion of $v(x,z)$ except the constant $v(0,0)$, i.e. we can recover $v(x,z)$ up to a constant.

B.9.2 Identifying good 1 when z_j is vector-valued in the linear homogeneous case

For simplicity, the results in the main text are for the case where z_j is scalar-valued for all goods j. This implies that one can label good 1 as the good with the highest value of z without loss of generality. As we have noted, if there are multiple z attributes per good, then our results apply if the data contains one choice set where one good is preferable to all other goods on each of the z attributes. This is not without loss.

We now show how to relax this restriction in the linear homogeneous case of Lemma 3. Let z_{kj} be the $k-$th hidden attribute of good j and let β_k be the associated preference parameter. By Assumption 2, we can write $s_j = f_{s_j}(\tilde{u}_1, \ldots, \tilde{u}_J, x_1, \ldots, x_J)$ for all j and thus $\frac{\partial s_j}{\partial z_{kj}} = \frac{\partial f_{s_j}}{\partial \tilde{u}_j}\beta_k$, implying $\frac{\partial s_j}{\partial z_{kj}} \Big/ \frac{\partial s_j}{\partial z_{k'j}} = \beta_k / \beta_{k'}$ for all k, k'. This means that we can compare the hidden component of utility across goods. Specifically, letting $\beta_1 > 0$ without loss, we have that, for any pair of goods j and j', $\sum_k \beta_k z_{kj} \geq \sum_k \beta_k z_{kj'}$ if and only if $z_{1j} - z_{1j'} + \sum_{k>1} \frac{\beta_k}{\beta_1}(z_{kj} - z_{kj'}) \geq 0$. Since the l.h.s. of the last inequality is identified, we can rank goods based on their non-visible utility. Lemma 3 then applies by defining good 1 as the good with the highest value of $\sum_k \beta_k z_{kj}$. Note that such a good always exists in any choice set (excluding ties) since $\sum_k \beta_k z_{kj}$ is scalar-valued.

B.9.3 Endogenous attributes

Here, we show how to extend our results to the case where some product attributes are endogenous (Section 2.3.3). Letting $\delta = (\delta_1, \cdots, \delta_J)$, we may write the share of good j

[13] Here, we use the fact that, by assumption, the first derivative of l is nonzero.

[14] Note that $\frac{\partial^2 s_1}{\partial z_1 \partial x_2}(\mathbf{0}, \mathbf{0}) = k'(v(0,0))\, l'(0)$, so we have $k'(v(0,0)) \neq 0$ by assumption.

as

$$s_j = \sigma_j (\delta, \mathbf{z}, \mathbf{p}) \tag{B.26}$$

for some function σ_j. Repeating this for all j and stacking the equations, we obtain a demand system of the form

$$\mathbf{s} = \sigma (\delta, \mathbf{z}, \mathbf{p}) \tag{B.27}$$

where $\mathbf{s} = (s_1, \cdots, s_J)$. We also define the share of the outside option as $s_0 \equiv 1 - \sum_{j=1}^{J} s_j$, with associated function $\sigma_0 (\delta, \mathbf{z}, \mathbf{p})$. We establish nonparametric identification of this demand system by invoking results from Berry and Haile (2014) (henceforth, BH).[15] Specifically, the results in BH yield identification of $(\xi_j)_{j=1}^{J}$ for every unit (individual or market) in the population. This means that all the arguments of σ are known, which immediately implies (nonparametric) identification of σ itself. Once σ is identified, one may apply our results to identify the distribution of the preference parameters α, β_i and λ_i. Note that, while knowledge of σ is sufficient for several counterfactuals of interest (e.g., computing equilibrium prices after a potential merger or tax), the preference parameters are required to predict how choices and welfare would change if consumers were given full information, among other things. In this sense, our approach complements the identification results in BH within the class of search models we consider.

To prove identification of σ, we first note that model (B.26) satisfies the index restriction in BH's Assumption 1. Second, we assume that we have excluded instruments $\mathbf{w}$ which, together with the exogenous attributes, satisfy the following mean-independence restriction

$$E (\xi_j | \mathbf{x}, \mathbf{z}, \mathbf{w}) = 0 \quad \text{for all } j \tag{B.28}$$

almost surely (Assumption 3 in BH) and assume that the instruments shift the endogenous variables (market shares and endogenous attributes $\mathbf{p}$) to a sufficient degree (as in BH's Assumption 4). Finally, we verify that the demand system satisfies the "connected substitutes" restriction defined in BH's Assumption 2. To this end, we prove the following result.

Lemma 9. *Let utility be given by* (2.12) *with ϵ_i supported on $\mathbb{R}^J$ and let Assumptions (i), (iii), (iv), and either (i) or (ii) hold. Then, for all $j, k = 1, \cdots, J$ with $j \neq k$, σ_j is (i) strictly increasing in δ_j and (ii) strictly decreasing in δ_k.*

Proof. First, assume that p_j is part of the visible utility of good j and Fix (δ_j, p_j, z_j) for all j. To prove claim (i), we show that an increase in δ_j can only induce a consumer to switch from not choosing j to choosing j but never vice versa, and that a positive mass of consumers will switch to choosing j. To see this, consider the case where consumer i initially searches j, which happens if and only if $g_{ij} (\delta_j, p_j, U_{ik}) \geq 0$ for all k such that $VU_{ik} \geq VU_{ij}$. Let $\Delta \geq 0$ be the change in δ_j. Since g_{ij} is increasing in its first argument, we have $g_{ij} (\delta_j + \Delta, p_j, U_{ik}) \geq 0$ for all k such that $VU_{ik} \geq VU_{ij} + \Delta$ and thus i will still search j. Moreover, since g_{ij} is decreasing in its last argument, if $g_{ik} (\delta_k, p_k, U_{ij}) \leq 0$ for some k such that $VU_{ik} \leq VU_{ij}$ (i.e. if k is initially not searched), then $g_{ik} (\delta_k, p_k, U_{ij} + \Delta) \leq 0$ (i.e. k is also not searched after the change in δ_j), which means that the set of goods searched by i never becomes larger. Next, note that if

[15] See also Berry et al. (2013).

$U_{ij} \geq U_{ik}$ for all k in the set of searched goods $\mathcal{G}_i$, then $U_{ij} + \Delta \geq U_{ik}$ for all $k \in \mathcal{G}_i$. Further, since ϵ_i is supported on all of $\mathbb{R}^J$, there is a positive mass of consumers for which $U_{ik} \geq U_{ij}$ for some $k \in \mathcal{G}_i$, but $U_{ij} + \Delta \geq U_{ik}$ for all $k \in \mathcal{G}_i$. An analogous argument proves claim (ii).

Since the argument above does not rely on the fact that p_j is part of the visible utility of good j, the conclusion also holds for the case in which p_j is only uncovered upon searching good j. □

Lemma 9 implies that the goods are connected substitutes in δ (see Definition 1 in BH), which in turn allows us to prove identification of σ by invoking Theorem 1 in BH.[16] Since Lemma 9 holds under either Assumption (i) or (ii), we obtain identification of preferences both in the case where p_j is part of the visible utility of good j and in the case where p_j is only uncovered upon searching j. Moreover, Theorem 1 of BH implies that one can invert the demand system σ for the indices δ and write

$$\alpha x_j + \xi_j = \sigma_j^{-1}(\mathbf{s}, \mathbf{z}, \mathbf{p}) \tag{B.29}$$

for all j. Equations (B.29) and (B.28) naturally lead to a nonparametric instrumental variable approach to estimate σ_j^{-1} (and thus σ_j).[17]

B.9.4 Identification when Observables Impact Search but not Utility

Here, we state and prove the results described in Section 2.3.4. We make the following assumptions:

Assumption 8. *(i) If consumer i searches j, then i also searches all j' s.t. $m(VU_{ij'}, r_{j'}) \geq m(VU_{ij}, r_j)$, where m is strictly increasing in both arguments;*

(ii) There is at least one good $j \neq 1$ such that $r_j > r_1$;

(iii) The support of $(\mathbf{x}, \mathbf{z}) \Big| (r_1, \ldots, r_J)$ has positive Lebesgue measure for all $(r_1, \ldots, r_J)$.

(iv) The search model admits an Armstrong representation that also satisfies the IIA property.

Assumption (iii) is substantive: for identification purposes, we consider variation in product characteristics holding fixed product search position. In practice, search position is likely to vary as a function of observables (e.g. products are sorted in order of price). However, because of the discrete nature of search position, we are likely to see variation conditional on search position and this is the variation we will use to identify our model.

Violations of the visible utility assumption due to search position will cause Lemma 8 to no longer hold as stated: the good with the highest value of z_j can be searched, another good j' may have higher utility (and thus higher visible utility), but good j' may not be searched because it has lower search position. However, an extension of Lemma 8 will still hold in this case, which then allows us to prove identification of preferences.

Lemma 10. *Let Assumptions 7, (ii)-(iv), and 8 hold and let $x_j \in [\bar{x} - \eta, \bar{x} + \eta]$ for all j, for some $\eta > 0$ sufficiently small. Then, if consumer i searches good 1 (i.e. the good*

[16] Note that the proof of Theorem 1 in BH only uses the fact that goods are connected substitutes in δ, not in $-\mathbf{p}$.

[17] Compiani (2019) proposes to approximate σ_j^{-1} using Bernstein polynomials. We use a similar approach in Section 3.5 to estimate the demand function for the case without endogeneity.

with the highest value of z), then i chooses the good which maximizes utility among all goods with $r_j \geq r_1$.

Proof. Suppose there was a good j with $r_j \geq r_1$ and $U_{ij} > U_{i1}$ that consumer i does not search. We can follow the proof of Lemma 8 to show that $VU_{ij} > VU_{i1}$. By Assumption (i), this implies that good j is searched, which is a contradiction. $\square$

In other words, if higher search position only makes a good more likely to be searched, then goods with higher visible utility *and* higher search position will always be searched if good 1 is searched. Given this Lemma, we can apply a modification of the identification argument in Theorem 2 after conditioning on the subset of goods with higher search position than good 1 (defined as usual as the good with the largest value of z_j):

Theorem 8. *Let the assumptions of Lemma 10 hold and let utility be given by $U_{ij} = v(x_j, z_j) + \epsilon_{ij}$ with v increasing in both arguments and infinitely differentiable. Further, assume that $\frac{\partial^2 s_1}{\partial z_1 \partial x_{j^*}}(\mathbf{x}^*, \mathbf{z}^*) \neq 0$ for some $(\mathbf{x}^*, \mathbf{z}^*)$ and $j^* \neq 1$, s_1 is infinitely differentiable and $\epsilon_i \perp (\mathbf{x}, \mathbf{z})$. Then, v is identified up to an additive constant.*

Proof. Let $R = \{j : r_j \geq r_1\}$. Under Assumption (iv), the choice probability for good 1 conditional on consumers choosing in R, denoted $s_{1|R}$, is equal to the choice probability for good 1 if consumers only faced R as their choice set. Further, by Lemma 10, the only mistake a consumer can make when faced with choice set R is to fail to search good 1 when it is in fact the good with the highest utility in R.

This argument exactly parallels the argument, except now we have additionally used the fact that $U_1 \geq U_j$ for all $j \in R$, since (i) if $j \in \mathcal{S}$, then $m(VU_1, r_1) \geq m(VU_j, r_j)$ implies $VU_1 \geq VU_j$, which in turn implies $U_1 \geq U_j$; (ii) if $j \notin \mathcal{S}$, then $g_1(x_1, U_j) \geq 0 \geq g_1(x_1, U_k)$ for all $k \in \mathcal{S}$ implies $U_j \leq U_k$. Note that $P^{\mathcal{S}}_{5new,2}$ does not depend on z_1 and $P^{\mathcal{S}}_{5new,1}(\mathbf{v}, x_1, \mathbf{r})$ only depends on x_j and z_j via v_j for $j \neq 1$, so the remainder of the argument in Appendix **??** applies with s_1 replaced by $s_{1|R}$. In practice, this means that, when estimating the model, one needs to take R as the choice set faced by consumers and drop those consumers that choose products outside R.

$\square$

B.9.5 Identification of a model where consumers form expectations on z_j based on x_j

Here, we state and prove the results described in Section 2.3.5. Given γ_1, we can identify the ranking of goods in terms of $\tilde{z}$ and we label good 1 as the good with the largest value of $\tilde{z}$. Then, an argument analogous to that in Lemma 3 yields identification of $\frac{\beta}{\alpha+\beta\gamma_1}$. We can also recover $\alpha + \beta\gamma_1$ in a manner that parallels our usual identification of α (Lemma 3). When $\tilde{z}_j = \tilde{z}$ for all j, consumers who search based on our visible utility assumption always maximize utility, and thus we can directly estimate $\alpha + \beta\gamma_1$ as the coefficient on x_j for those consumers (we provide a formal proof of this in the next subsection). Therefore, this gives separate identification of β and α given γ_1.

When γ_1 is unknown, we can identify β/α if we know its sign and make a further support assumption. Suppose that the sign of γ_1 is known (e.g. higher priced goods have weakly higher quality). Without loss, we assume $\gamma_1 > 0$. In addition, suppose that there exist choice sets in which a good has both the highest value of z and the lowest value of x. Even when γ_1 is unknown, this good is known to maximize $\tilde{z}$; thus, we can label it by

1. Note that we cannot differentiate with respect to $\tilde{z}$ as in the case above since γ_1 and thus $\tilde{z}$ is unknown. However, with good 1 defined appropriately, Corollary 3 shows that cross-derivatives with respect to z_1, z_j, x_j for $j \neq 1$ identify β/α (specifically, consumers who search the good with the highest value of $\tilde{z}$ will always maximize utility, and so their sensitivity to x_j and z_j identifies their true preferences).

Identification of $\alpha + \beta\gamma_1$

Note that if $\tilde{z}_j = 0$ for all j, then consumers always maximize utility. Thus, seeing how choice probabilities change with x conditional on $\tilde{z}_j = 0$ for all j should help identify $\alpha + \beta\gamma_1$. Because the event $\tilde{z}_j = 0$ involves x_j, we need to differentiate choice probabilities with respect to x_j on the envelope satisfying the condition $\tilde{z}_j = 0$ for all x_j. Formally, fix any $j \in \mathcal{J}$ and choose (x_k, z_k) so that $z_k = \gamma_0 + \gamma_1 x_k$ (which implies $\tilde{z}_k = 0$) for all $k \neq j$. For every $\delta > 0$, let $\epsilon(\delta) \equiv \gamma_0 + (x_j + \delta)\gamma_1 - z_j$, so that $z_j + \epsilon(\delta) - E(z_j | x_j + \delta) = 0$. Note that $\epsilon(\delta)$ is known to the econometrician. Thus, evaluating the last display at $\mathbf{x} = \mathbf{0}$ yields identification of $(\alpha + \beta\gamma_1)$ under a parametric assumption on ϵ_i.

B.9.6 Unobservables revealed by search

Here, we show that the ratio of second derivatives in (2.3) robustly identifies $\frac{\beta}{\alpha}$ in the model where ϵ_{ij} is revealed to consumer i only upon searching good j (Section 2.3.6).

Order goods in increasing order of x. Then, for $j = 1, \ldots, J$,

$$\begin{aligned} s_j &= \sum_{k=1}^{j} P\left(\{U_j \geq U_{j'}\ \forall\, j' \in \{k, \ldots, J\}\} \cap \{\text{search exactly } k, \ldots, J\}\right) \\ &\equiv \sum_{k=1}^{j} P_j^{(k)}\left(\tilde{\mathbf{u}}, \mathbf{x}_{-J}\right) \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\partial^2 s_j}{\partial z_j \partial z_J} &= \sum_{k=1}^{j} \frac{\partial^2 P_j^{(k)}}{\partial \tilde{u}_j \partial \tilde{u}_J} \beta^2 \\ \frac{\partial^2 s_j}{\partial z_j \partial x_J} &= \sum_{k=1}^{j} \frac{\partial^2 P_j^{(k)}}{\partial \tilde{u}_j \partial \tilde{u}_J} \alpha\beta \end{aligned}$$

So the ratio of the latter two derivatives identifies $\frac{\beta}{\alpha}$. (Note that the ratio of $\frac{\partial s_j}{\partial z_J}$ to $\frac{\partial s_j}{\partial x_J}$ for any j would also work). On the other hand,

$$\begin{aligned} \frac{\partial s_j}{\partial z_j} &= \sum_{k=1}^{j} \frac{\partial P_j^{(k)}}{\partial \tilde{u}_j} \beta \\ \frac{\partial s_j}{\partial x_j} &= \sum_{k=1}^{j} \left(\frac{\partial P_j^{(k)}}{\partial \tilde{u}_j} + \frac{1}{\alpha} \frac{\partial P_j^{(k)}}{\partial x_j} \right) \alpha \end{aligned} \tag{B.30}$$

Since $\frac{1}{\alpha} \frac{\partial P_j^{(k)}}{\partial x_j} \geq 0$, (B.30) implies that the ratio of first derivatives suffers from attenuation bias, i.e. $\frac{\frac{\partial s_j}{\partial z_j}}{\frac{\partial s_j}{\partial x_j}} \leq \frac{\beta}{\alpha}$.

B.9.7 $K-$rank model

Consider the simultaneous search model in Honka et al. (2017) with $J = 2$ goods. In this model, a consumer looks at the visible utilities and decides whether to search the good with the highest visible utility or search both goods. Searching a second good entails a cost c, constant across consumers. As usual, we denote by 1 the good with the highest value of z.

Note that consumer i searches 2 but not 1 if and only if $VU_{i2} > VU_{i1}$ and

$$E_{z_1,z_2}\left[\max\left\{VU_{i1} + \beta z_1, VU_{i2} + \beta z_2\right\}\right] - c < E_{z_2}\left[VU_{i2} + \beta z_2\right] \tag{B.31}$$

i.e.

$$E_{z_1,z_2}\left[\max\left\{VU_{i1} - VU_{i2} + \beta\left(z_1 - z_2\right), 0\right\}\right] - c < 0 \tag{B.32}$$

or $g_{sim}\left(VU_{i1} - VU_{i2}\right) < 0$ for an increasing function g_{sim}. Equation (2.4) then can be written as

$$\begin{aligned} s_1 &= P(U_1 > U_2) - P(\{U_1 > U_2\} \cap \{VU_2 > VU_1\} \cap \{g_{sim}(VU_1 - VU_2) < 0\}) \\ &= P_{1,sim} - P_{2,sim} \end{aligned} \tag{B.33}$$

We also have:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_2} = \beta^2\left(\frac{\partial^2 P_{1,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2}\right) \tag{B.34}$$

and

$$\frac{\partial^2 s_1}{\partial z_1 \partial x_2} = \alpha\beta\left(\frac{\partial^2 P_{1,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{u}_2} - \frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2}\right) \tag{B.35}$$

So, if $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} = 0$, then the ratio of (B.34) to (B.35) identifies $\frac{\beta}{\alpha}$. Note that the event in $P_{2,sim}$ is equivalent to the following set of inequalities: (i) $\epsilon_{i1} > \tilde{u}_2 - \tilde{u}_1 + \epsilon_{i2}$, (ii) $\epsilon_{i1} < \tilde{vu}_2 - \tilde{vu}_1 + \epsilon_{i2}$, (iii) $\epsilon_{i1} < g_{sim}^{-1}(0) + \tilde{vu}_2 - \tilde{vu}_1 + \epsilon_{i2}$, where $VU_{ij} = \tilde{vu}_j + \epsilon_{ij}$ and $U_{ij} = \tilde{u}_j + \epsilon_{ij}$, as above. Then, letting $\tilde{\epsilon} = \epsilon_1 - \epsilon_2$, we have:

$$P_{2,sim} = \int_{\tilde{u}_2 - \tilde{u}_1}^{\min(\tilde{vu}_2 - \tilde{vu}_1, g_{sim}^{-1}(0) + \tilde{vu}_2 - \tilde{vu}_1)} f_{\tilde{\epsilon}}(\tilde{\epsilon}) d\tilde{\epsilon} = \int_{\beta(z_2 - z_1)}^{\min(0, g_{sim}^{-1}(0))} f_{\tilde{\epsilon}}(\tilde{\epsilon}) d\tilde{\epsilon}$$

Thus, $\frac{\partial^2 P_{2,sim}}{\partial \tilde{u}_1 \partial \tilde{vu}_2} = 0$.

Finally, we show that the ratio of first derivatives leads to attenuation bias. This follows directly from

$$\begin{aligned} \frac{\partial s_1}{\partial z_1} &= \beta\left(\frac{\partial P_{1,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{u}_1}\right) \\ \frac{\partial s_1}{\partial x_1} &= \alpha\left(\frac{\partial P_{1,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{u}_1} - \frac{\partial P_{2,sim}}{\partial \tilde{vu}_1}\right) \end{aligned}$$

and the fact that $\frac{\partial P_{2,sim}}{\partial \tilde{vu}_1} < 0$.

B.10 Testing for full information with heterogeneous preferences

In Section 2.2.4, we considered the problem of testing the null hypothesis of full information and showed that, in the case where the coefficients α and β are homogeneous across consumers, a valid test rejects the null when the ratios of first derivatives are attenuated relative to the ratio of second derivatives in (2.10). Here, we provide conditions under which the same test is valid in the case where one of the two coefficients is allowed to be heterogeneous.[18] We focus on the case where β is heterogeneous and z_j is a scalar; the argument for the case where α is heterogeneous (and x_j is a scalar) is analogous. We also assume that the ϵ_{ij} shocks are type-I extreme-value distributed and let $s_j(\tilde{\beta})$ be the market share of good j for consumers with $\beta = \tilde{\beta}$ under full information, i.e. $s_j(\mathbf{x}, \mathbf{z}; \tilde{\beta}) \equiv \frac{\exp(\alpha x_j + \tilde{\beta} z_j)}{\sum_{k=1}^{J} \exp(\alpha x_k + \tilde{\beta} z_k)}$.

We let $j = 2, k = k' = 1$ in equation (2.10), i.e. we consider the case where the test compares the ratio of second derivatives taken with respect to good 1 and 2 to the ratio of first derivatives taken with respect to good 1. Analogous sufficient conditions could be obtained for different choices of j, k, k'. We take a pair $(\mathbf{x}, \mathbf{z})$ such that $\frac{\partial s_1(\mathbf{x},\mathbf{z})}{\partial x_1} > 0$ and $\frac{\partial^2 s_1(\mathbf{x},\mathbf{z})}{\partial z_1 \partial x_2} > 0$ (both of which can be verified from the data), so that (B.13) holds if and only if

$$\begin{aligned} - \quad & \alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta) s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta \int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta))\beta dF_\beta \geq \\ - \quad & \int s_1(\mathbf{x}, \mathbf{z}; \beta) s_2(\mathbf{x}, \mathbf{z}; \beta)(1 - 2s_1(\mathbf{x}, \mathbf{z}; \beta))\beta^2 dF_\beta \alpha \int s_1(\mathbf{x}, \mathbf{z}; \beta)(1 - s_1(\mathbf{x}, \mathbf{z}; \beta)) dF_\beta \end{aligned}$$

Then, by Theorem 2 of Wijsman (1985), the desired inequality holds if (i) $\beta > 0$, and (ii) $\frac{\alpha}{\beta}$ and $\frac{-s_1(\mathbf{x},\mathbf{z};\beta)s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{s_1(\mathbf{x},\mathbf{z};\beta)(1-s_1(\mathbf{x},\mathbf{z};\beta))} = -\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)}$ are monotonic functions of β in the same direction. Since we assumed throughout that $\alpha > 0$, we want to show that $-\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)}$ decreases in β monotonically. After some algebra, we have that

$$\frac{\partial \left[-\frac{s_2(\mathbf{x},\mathbf{z};\beta)(1-2s_1(\mathbf{x},\mathbf{z};\beta))\beta}{1-s_1(\mathbf{x},\mathbf{z};\beta)} \right]}{\partial \beta} < 0 \qquad \forall \beta$$

Under these conditions, at the chosen values of $\mathbf{x}, \mathbf{z}$, a valid test of the null of full information rejects when the ratio of first derivatives is sufficiently attenuated relative to the ratio of first derivatives. Note that the condition in (**??**) can be verified given the support of the distribution of β. For example, if β takes values on a finite grid of points, then one needs to check whether (**??**) holds for all values in the grid. Finally, we emphasize that (**??**) is a sufficient, but in general not necessary condition, implying that the proposed test could be valid even if the restriction is not satisfied.

[18] The reason why we let only one of the coefficients be heterogeneous is that we leverage a result from the statistics literature that applies to ratios of one-dimensional integrals.

B.11 Derivation of Flexible Logit Weights and Choice Probabilities

In this section, we derive the relevant derivatives of choice probabilities for the flexible logit model described in the text. In this model:

$$v_1 = \tilde{v}(x_1, z_1) + b_1 z_1 + \sum_{k \neq 1} (\gamma_k w_{z1k} z_k + \gamma_{2k} w_{x1k} x_k + w_{z2k} \delta_k z_k z_1 + w_{x2k} \delta_{2k} x_k z_1) \quad \text{(B.36)}$$

and $v_k = \tilde{v}(x_k, z_k)$ for $k \neq 1$ where b_1, γ_k, γ_{2k}, δ_k and δ_{2k} are coefficients to be estimated which allow greater flexibility in how derivatives with respect to z_1 vary with attributes of rival goods. The weights w_{x1k}, w_{z1k}, w_{x2k} and w_{z2k} are chosen so that, given the logit functional form, $\frac{\partial^2 s_1}{\partial z_1 \partial z_j} / \frac{\partial^2 s_1}{\partial z_1 \partial x_j}$ can be constant across goods as our structural model implies when these weights are regarded as constant in derivatives. With these weights, we have the following derivatives (where we use the notation $\tilde{v}_j$ to refer to the function $\tilde{v}$ evaluated at (x_j, z_j)):

$$\begin{aligned}
\frac{\partial v_1}{\partial z_1} &= \frac{\partial \tilde{v}_1}{\partial z} + b_1 + \sum_{k \neq 1} (w_{z2k} \delta_k z_k + w_{x2k} \delta_{2k} x_k) \\
\frac{\partial s_1}{\partial x_1} &= \frac{\partial s_1}{\partial v_1} \frac{\partial v_1}{\partial x_1} = \frac{\partial \tilde{v}_1}{\partial x} s_1 (1 - s_1) \\
\frac{\partial s_1}{\partial z_1} &= \frac{\partial s_1}{\partial v_1} \frac{\partial v_1}{\partial z_1} = \frac{\partial v_1}{\partial z_1} s_1 (1 - s_1) \\
\frac{\partial s_1}{\partial x_{j'}} &= \frac{\partial s_1}{\partial v_{ij'}} \frac{\partial v_{ij'}}{\partial x_{j'}} + \frac{\partial s_1}{\partial v_{i1}} \frac{\partial v_{i1}}{\partial x_{j'}} = -\frac{\partial \tilde{v}_{j'}}{\partial x} s_1 s_{j'} + [w_{x1j'} \gamma_{2j'} + w_{x2j'} \delta_{2j'} z_1] s_1 (1 - s_1) \\
\frac{\partial s_1}{\partial z_{j'}} &= \frac{\partial s_1}{\partial v_{ij'}} \frac{\partial v_{ij'}}{\partial z_{j'}} + \frac{\partial s_1}{\partial v_{i1}} \frac{\partial v_{i1}}{\partial z_{j'}} = -\frac{\partial \tilde{v}_{j'}}{\partial z} s_1 s_{j'} + [w_{z1j'} \gamma_{j'} + w_{z2j'} \delta_{j'} z_1] s_1 (1 - s_1) \\
\frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial x_{j'}} \frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1} \frac{\partial^2 v_1}{\partial z_1 \partial x_{j'}} \\
&= \frac{\partial v_1}{\partial z_1} (1 - 2s_1) \frac{\partial s_1}{\partial x_{j'}} + s_1 (1 - s_1) w_{x2j'} \delta_{2j'} \\
\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} &= \frac{\partial^2 s_1}{\partial v_1 \partial z_{j'}} \frac{\partial v_1}{\partial z_1} + \frac{\partial s_1}{\partial v_1} \frac{\partial^2 v_1}{\partial z_1 \partial z_{j'}} \\
&= \frac{\partial v_1}{\partial z_1} (1 - 2s_1) \frac{\partial s_1}{\partial z_{j'}} + s_1 (1 - s_1) w_{z2j'} \delta_{j'}
\end{aligned} \quad \text{(B.37)}$$

If we define the weights: $w_{x1j'} = w_{z1j'} = \frac{s_{j'}}{1 - s_1}$ and $w_{x2j'} = w_{z2j'} = [\frac{z_1(1-s_1)}{s_{j'}} + \frac{(1-s_1)}{(\partial v_1 / \partial z_1)(1-2s_1) s_{j'}}]^{-1} = \frac{(1-2s_1) s_{j'}}{1-s_1} \left(\frac{1}{\partial v_1 / \partial z_1} + (1 - 2s_1) z_1 \right)^{-1}$, Thus, we have:

$$\frac{\partial^2 s_1}{\partial z_1 \partial z_{j'}} / \frac{\partial^2 s_1}{\partial z_1 \partial x_{j'}} = \frac{-\frac{\partial \tilde{v}_{j'}}{\partial z} + \gamma_{j'} + \delta_{j'}}{-\frac{\partial \tilde{v}_{j'}}{\partial x} + \gamma_{2j'} + \delta_{2j'}} \quad \text{(B.38)}$$

where $w_{z1j'} = w_{x1j'} = \frac{s_{j'}}{1 - s_1}$ and $w_{x2j'} = w_{z2j'}$. Given a linear specification of $\tilde{v}$, $\tilde{v}(x_j, z_j) =$

$x_j a_1 + z_j a_2$, this implies that the above ratio is a constant for each j'.

Estimation of the model with these weights is infeasible since the levels of the choice probabilities s_1 and s_k, as well as the derivatives $\partial v_1/\partial z_1$ are unknown ex ante and thus we do not know the weights. We estimate the model via a two-step process where s_1 and s_k are estimated using a naive logit model (where utility for each good is a linear function of x_j and z_j), these estimates are used to construct weights, and then the model in equation (2.14) is estimated treating these weights as constants.[19]

To recover estimates of β/α from the flexible logit model, we use the ratio in equation (B.38). With the linear specification of $\tilde{v}$, this ratio is given by $\frac{\beta}{\alpha} = \frac{-a_2+\gamma_{j'}+\delta_{j'}}{-a_1+\gamma_{2j'}+\delta_{2j'}}$. In cases where the identity of goods is not meaningful (e.g. "good 2" does not refer to the same good across different choice sets and there are no alternative-specific fixed effects), we can further impose $\gamma_k = \gamma$, $\gamma_{2k} = \gamma_2$, $\delta_k = \delta$ and $\delta_{2k} = \delta_2$, which gives a single estimate of $\frac{\beta}{\alpha}$.

B.12 Recovery of Search Costs Given Preferences in the Weitzman Model

Suppose that utility is given by $U_{ij} = x_j\alpha + z_j\beta + \epsilon_{ij}$ and that consumers search sequentially according to the model of Weitzman (1979).

As shown in Armstrong (2017),[20] the optimal search strategy is for consumers to behave as if they were choosing among options in a static model with utilities given by $\tilde{U}_{ij} = x_j\alpha + \min\{z_j, rv_i\}\beta + \epsilon_{ij}$, where rv_i denotes i's reservation value in units of z (see Example 1). Thus, dropping i subscripts, ordering goods so that $z_1 \geq z_2 \geq \ldots \geq z_J$, we can write

$$\begin{aligned}
s_1 &= P\left(x_1\alpha + \min\{z_1, rv\}\beta + \epsilon_1 \geq x_k\alpha + \min\{z_k, rv\}\beta + \epsilon_k \ \forall k\right) \\
&= P\left(\epsilon_k - \epsilon_1 \leq (x_1 - x_k)\alpha \ \forall k\right) P\left(rv \leq z_J\right) \\
&+ \sum_{t=0}^{J-2} \int P\left(\{\epsilon \in E_t\} \cap \{z_{J-t} \leq rv \leq z_{J-t-1}\}\right) dF_{rv}(rv) \\
&+ P\left(\epsilon_k - \epsilon_1 \leq (x_1 - x_k)\alpha + (z_1 - z_k)\beta \ \forall k\right) P\left(rv \geq z_1\right)
\end{aligned}$$

where F_{rv} denotes the cdf of rv and the second equality assumes that search costs (and thus rv) are independent of ϵ. Therefore, we have

$$\frac{\partial s_1}{\partial z_1} = \left[\frac{\partial}{\partial z_1} P\left(\epsilon_k - \epsilon_1 \leq (x_1 - x_k)\alpha + (z_1 - z_k)\beta \ \forall k\right)\right] P\left(rv \geq z_1\right) \qquad \text{(B.39)}$$

Given identification of (α, β), the first term on the rhs of (B.39) is identified given parametric assumptions on the distribution of ϵ. Thus, $P\{rv \geq z_1\}$ is identified. Repeating the argument for all z_1, one can trace out the entire distribution of rv. Since c, the search cost for consumer i, is a known transformation of rv,[21] the distribution of c is also

[19] Since $\partial v_1/\partial z_1$ is estimated imprecisely from the naive logit, when $1 + (\partial v_1/\partial z_1)(1 - 2s_1)z_1$ is close to 0 (leading to very large weights), we set $\partial v_1/\partial z_1 = 0$ when the former term falls below 1 in absolute value.

[20] See also Choi et al. (2018b).

[21] This assumes that the prior F_z used by consumers in forming expectations are known to the re-

identified.

Equation (B.39) also lends itself to a different argument that does not require making a parametric assumption on the distribution of ϵ, but instead relies on "at-infinity" variation. Note that the first term on the rhs of (B.39) is invariant to increasing all z_j's by the same amount. Thus, we can write

$$\frac{\frac{\partial s_1}{\partial z_1}(\mathbf{z}+\mathbf{\Delta})}{\frac{\partial s_1}{\partial z_1}(\mathbf{z})} = \frac{P(rv \geq z_1 + \Delta)}{P(rv \geq z_1)} \tag{B.40}$$

where $\mathbf{\Delta}$ is a $J-$vector with all elements equal to some Δ. Letting $\Delta \to -\infty$, the numerator on the rhs of (B.40) goes to 1, which yields identification of $P(rv \geq z_1)$. Repeating the argument for all z_1, one can trace out the entire distribution of rv and recover the distribution of c as above.

B.13 Welfare Benefits of Information

Appendix D of Abaluck and Gruber (2009) shows that dollar-equivalent consumer surplus in logit models where positive preferences (i.e., preferences describing potentially uninformed behavior) are given by β_{pos} and normative preferences (i.e., those relevant for welfare evaluations) are given by β_{norm} can be computed as:

$$E(CS_0) = -\frac{1}{\alpha_p}\left[\sum_k (x_k\beta_{norm} - x_k\beta_{pos})s_k(\beta_{pos}) + \ln\sum_k \exp(x_{ik}\beta_{pos})\right]$$

where α_p is the (normative) marginal utility of income, estimated as the coefficient on price. Once consumers are informed and their preferences are β_{norm}, consumer surplus is given by the conventional log-sum formula:

$$E(CS_1) = -\frac{1}{\alpha_p}\ln\sum_k \exp(x_k\beta_{norm})$$

The change in consumer surplus from providing consumers with information is thus:

$$\Delta CS = -\frac{1}{\alpha_p}\left[\ln\sum_k \exp(x_k\beta_{norm}) - \ln\sum_k \exp(x_k\beta_{pos}) + \sum_k (x_k\beta_{pos} - x_k\beta_{norm})s_k(\beta_{pos})\right]$$

B.14 Application

B.14.1 Data Cleaning

The dataset from Kaggle.com contains 9,917,530 observations on a hotel-consumer level. We filtered out the following categories of observations.

First, the data set contains some errors in the price information. We removed search impressions that contain at least one observation for which the listed hotel price is below

searcher, as in the case where consumers have rational expectations and F_z coincides with the observed distribution of z across goods and/or markets.

$10 or above $1000 per night, or the implied tax paid per night either exceeds 30% of the listed hotel price, or is less than $1.

Second, we removed the search impressions where the consumer observed a hotel on position 5, 11, 17, 23, i.e., the consumer did not have opaque offers (Ursu (2018) provides a detailed description of this feature in the data).

Third, the original data set contains observations on more than 20,000 destinations, with a median of two search impressions per destination. We focused our attention on destinations with at least 50 search impressions.

Fourth, we kept the search impressions where all clicks and transactions happened within the top 10 positions excluding the opaque offer positions, and we only kept these top 10 hotels in these choice sets.

The final dataset then contains 54,648 choice sets and 546,480 observations. Table B.6 provides a detailed description of each variable.

Table B.6: Variable Description

Variable	Description
Price	Gross price in USD
Stars	Number of hotel stars
Review Score	User review score, mean over sample period
Chain	Dummy whether hotel is part of a chain
Location Score	Expedia's score for desirability of hotel's location
Promotion	Dummy whether hotel is on promotion

B.14.2 Estimation Results

Table B.7 and B.8 shows the detailed results of flexible logit and naive logit estimations.

Table B.7: Estimation Results

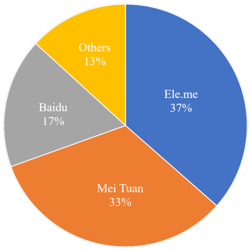
z Variable	Standard Estimate	Standard CI	Flexible Estimate	Flexible CI
Location Score	0.298	(0.278, 0.317)	0.691	(0.591, 0.839)
Price	-1.085	(-1.109, -1.061)	-0.710	(-1.169, -0.503)
Review Score	0.172	(0.159, 0.185)	0.195	(0.082, 0.250)
Stars	0.386	(0.369, 0.403)	0.364	(0.258, 0.430)

Table B.8: Estimation Results

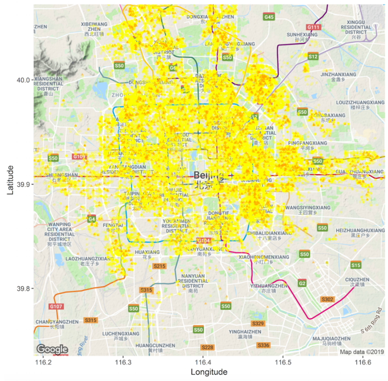
z Variable	Point Estimate	Confidence Interval
Location Score	0.393	(0.296, 0.544)
Price	-0.375	(-0.566, 0.085)
Review Score	0.023	(-0.069, 0.038)
Stars	-0.022	(-0.135, 0.038)

Figure C.1.1: Food Delivery Market

(A) Market Structure of the Food Delivery Market



(B) Buyer and Seller Geographic Distribution



Note: Panel (A) shows the market share of the food delivery market in China, 2017. The platform we are studying is a relatively niche platform. Panel (B) shows the geographic distribution of sellers and buyers on the may. The orange dots are the sellers and the yellow dots are the buyers.

Appendix C

Appendix for Chapter 3

C.1 Platform Background Details

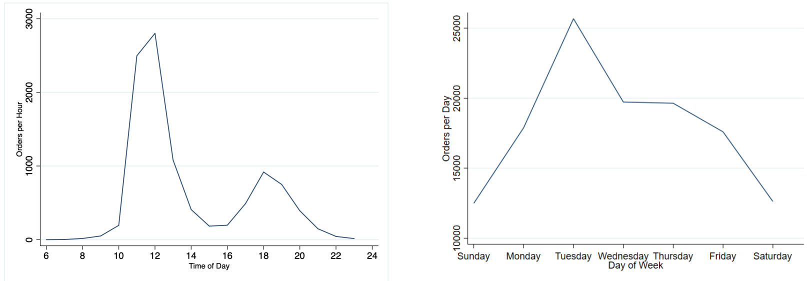
Market Structure Figure C.1.1 (A) shows the market share of the food delivery market in China in 2017. The total market value is $28 billion.[1] The platform we study has a relatively small market share and is a niche platform. Homemade food in China is believed to be cleaner and healthier than restaurants given the food security scandals. Figure C.1.1 (B) shows the buyer and seller distribution map in the city of Beijing. The orange dots are the sellers and the yellow dots are the buyers. In terms of delivery service, sellers can choose to deliver by themselves or use full-time third-party carriers. The delivery fees are equally shared by kitchens and consumers.

Daily Order Trend Figure C.1.2 shows the order trend during the day. Most orders are workday lunch orders. So we assume one consumption per period. We also observe the utensils in each order, and in 90% of our orders requested one set of utensils. This is also an advantage of our data since one major concern of using household level packaged good

[1] https://www.iimedia.cn/c400/60449.html

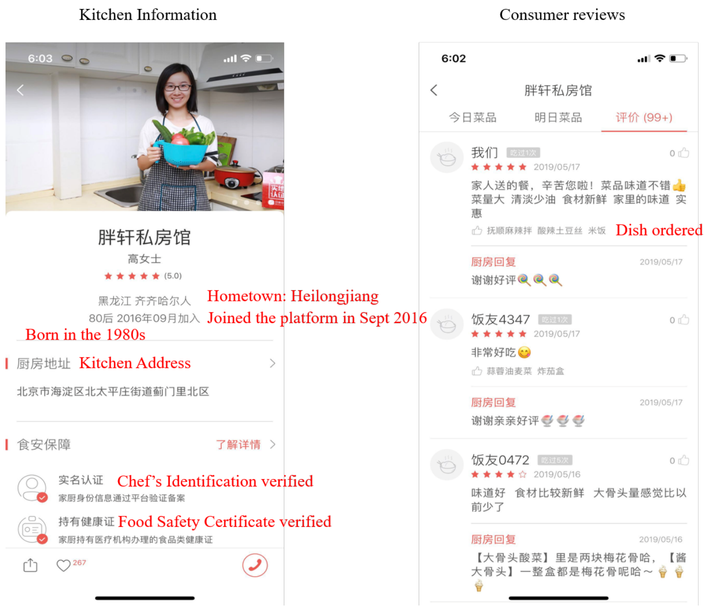
Figure C.1.2: Order Trend

(A) Order Trend during the Day (B) Order Trend during the Week



Note: The figure shows the timeline of orders during the day and the week. Panel (A) shows the number of orders per hour throughout the day. Panel (B) shows the number of orders per day on each day of the week. 80 % of orders are workday orders.

Figure C.1.3: Kitchen Information



Note: The figure shows the kitchen information page and the information included.

scanner data is the confound of intrahousehold heterogeneity: multiple family members could be using the same membership card to purchase grocery products.

App Display More information about the kitchen can be found in the app as in figure C.1.3. There are seller's demographic information like hometown and age, time to join the platform, kitchen's address, certificate, and consumer reviews. The homepage of the platform for the consumer side is shown in Figure C.1.4. On the homepage, there is a kitchen list showing information about monthly sales, ratings, delivery radiance, distance, address, seller's hometown (which is correlated with kitchen's cuisine type), and pictures of the dishes where the users can swipe left and right to browse the menu without clicking on the kitchen. After clicking into the detailed kitchen page, there is a dish list showing more detailed information, including dish prices, availability, sales, and reviews.

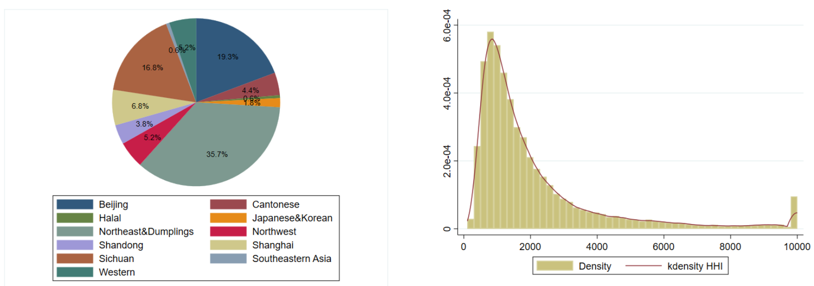
Cuisine Distribution We divide all kitchens into eleven different cuisines, where the distribution of each cuisine is in Figure C.1.5. The cuisine variable is constructed based

Figure C.1.4: Homepage

Note: The figure shows the app display details. On the homepage there is a kitchen list. After clicking on the kitchen, consumers can see more detailed dish lists.

Figure C.1.5: Motivating Evidence

(A) Distribution of Cuisine (B) Distribution of HHI



Note: Panel (A) pie chart shows the distribution of cuisines among kitchens in our data. The cuisine variable is constructed based on kitchen features, dish features, and hometown information of the kitchen owners. Panel (B) shows the distribution of HHI constructed seeing each consumer as a market. $HHI_i = \sum_j ms_{ij}^2 \cdot 10000$, where ms_{ij} is the market share of order from kitchen j among all of consumer i's orders.

on kitchen features, dish features, and hometown information of the kitchen owners.

HHI Distribution We compute the Herfindahl-Hirschman Index (HHI) as a measure of the dispersion level of consumers' chosen sets aggregated over time. Each consumer is seen as a market, and we define the index as $HHI_i = \sum_j ms_{ij}^2 \cdot 10000$, where ms_{ij} is the market share of orders from kitchen j among all consumer i's orders on the platform over time. The distribution of HHI_i in our data is shown in panel (B) of Figure C.1.5. The figure shows that most consumers' chosen sets are rather dispersed.

C.2 Permutation Test

C.2.1 Technical Proof

In this section we provides the detailed construction of permutation test.[2] The null hypothesis for the permutation test is that consumers have no state dependent preferences, and has an i.i.d fixed (unknown) probability of choosing each option from the choice set. The first step is to test for variety-seeking preferences in consumer i is to observe his/her order sequence and compute the switching probability from a kitchen on platform, $\hat{P}^i(y_t \neq j | s_t = j \neq 0)$, where s_t represents previous period choice and y_t is the current period choice. The second step is to compute this switching probability for each permutation of the observed order sequence: each of these permutations is equally likely because consumer i's probability of choosing each option is fixed under the null hypothesis. Note that the implicit assumption here is a stochastic choice model: for each period the probability of choosing each option is i.i.d across different periods. The set of switching probabilities and their respective relative frequencies computed in the second step constitutes the sampling distribution of switching probabilities under the null hypothesis conditional on the observed number of orders for each option. We use this distribution for statistical testing.

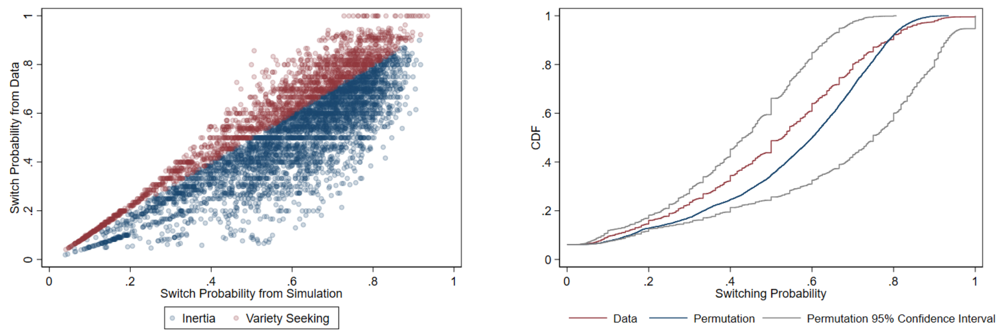
Let $\mathbf{Y}$ be an order sequence from consumer i, and $\mathbf{N_i}$ denote the numbers of orders for all chosen kitchens by consumer i. The null hypothesis is that the choice probabilities are i.i.d across periods with $P(Y_t = j) = p_i^j$. This implies that, conditional on the number of orders for each kitchen, $\mathbf{N_i}(\mathbf{Y}) = \mathbf{n_i}$, each permutation is equally likely. Variety-seeking hypothesis predicts that the switching probability $\hat{P}^i(y_t \neq s_t | s_t \neq 0)$ will be significantly larger than what one would expect by chance. Let $\hat{P}(\mathbf{Y})$ denote this switching probability for sequence $\mathbf{Y}$. For an observed sequence $\mathbf{y}$, with $\mathbf{N_i}(\mathbf{y}) = \mathbf{n_i}$ orders for each chosen kitchen on the platform, to test the null hypothesis at the α level, we check if $\hat{P}(\mathbf{y}) \geq c_{\alpha,\mathbf{n_i}}$, where the critical value $c_{\alpha,\mathbf{n_i}}$ is defined as the smallest c such that $\mathbb{P}(P(\mathbf{Y}) \geq c | H_0, \mathbf{N_i}(\mathbf{y}) = \mathbf{n_i}) \leq \alpha$, and the distribution $\mathbb{P}(P(\mathbf{Y}) \geq c | H_0, \mathbf{N_i}(\mathbf{y}) = \mathbf{n_i})$ is generated from the sampling distribution from permutations. For the quantity $\mathbb{P}(P(\mathbf{Y}) \geq c_{\alpha,\mathbf{n_i}} | H_0, \mathbf{N_i}(\mathbf{Y}) = \mathbf{n_i})$, it may be the case that, for some c^*, it is strictly greater than α for $c \leq c^*$. In this case, for any sequence with $\mathbf{N_i}(\mathbf{y}) = \mathbf{n_i}$, one cannot reject H_0 at an α level of significance.[3] From the ex ante perspective, a test of variety seeking at the α level of significance consists of a family of such critical values $\{c_{\alpha,\mathbf{n_i}}\}$. It follows immediately that $\mathbb{P}(reject | H_0) \leq \alpha$ because $\mathbb{P}(reject | H_0) = \sum_{\{\mathbf{n_i}\}} \mathbb{P}(P(\mathbf{Y}) \geq c_{\alpha,\mathbf{n_i}} | H_0, N_i^j(\mathbf{Y}) = \mathbf{n_i}) \mathbb{P}(\mathbf{N_i}(\mathbf{Y}) = \mathbf{n_i} | H_0) \leq \alpha$. Last, for

[2]Miller and Sanjurjo (2018) perform a similar permutation test as a nonparametric robustness test that is by construction invulnerable to conditional probability bias from small samples. Miller et al. (2014) showed that the runs and serial correlation tests, along with the conditional probability test all amount roughly to the same test, and moreover, that they are not sufficiently powered. Traditional research on state dependence use the runs tests to test for state dependence, such as the binomial runs test and multinomial runs test discussed in Bass et al. (1984). The permutation test is equivalent to the runs test in binomial choices. In multinomial choices, the multinomial runs test need an extra assumption that the probability of choosing a brand on each choice occasion is equal to the observed empirical frequency of choosing the brand, where as in the permutation test we do not need to make the extra assumption and only need the choice probabilities to be i.i.d across periods.

[3]Examples of such choice sequences are AAAA and ABCD, where $\mathbb{P}(P(\mathbf{Y}) \geq c_{\alpha,\mathbf{n_i}} | H_0, \mathbf{N_i}(\mathbf{Y}) = \mathbf{n_i})$ always equal to zero and one, respectively. In such cases we are not able to reject the null hypothesis that there is no state dependence and switching is simply due to randomness in choices.

Figure C.2.1: Motivating Evidence

(A) Switching Probability Distribution (B) Switching Probability CDF



Note: Panel (A) shows the switching frequency observed in the data relative to that in the randomly reshuffled benchmark, where the order of choices does not matter given the observed market share. One dot represents one consumer. The 45-degree line represents the cases when consumers switch as frequently as the randomly reshuffled benchmark. The consumers above the 45-degree line switch more frequently in the data than the random reshuffle benchmark and exhibit variety-seeking preferences, whereas the consumers below the 45-degree line switch less frequently than the randomly shuffled benchmark and exhibit inertia. Panel (B) presents the cumulative distribution function of the switching probability in data, in the random reshuffled benchmark, and the 95% confidence bands of the reshuffled sequences.

arbitrary test statistic $T(\mathbf{Y})$, the fact that the distribution of $\mathbf{Y}$ is exchangable conditional on $\mathbf{N_i}(\mathbf{Y}) = \mathbf{n_i}$ means that $\mathbb{P}(P(\mathbf{Y}) \geq c_{\alpha,\mathbf{n_i}} | H_0, \mathbf{N_i}(\mathbf{Y}) = \mathbf{n_i})$ can be approximated to arbitrary precision with Monte Carlo permutations of the sequence $\mathbf{y}$ (Ernst, 2004; Good, 2013).

C.2.2 Permutation Test for Cuisines

In this section we present the results of permutation test for cuisines. Figure C.2.1 presents the distributions of cuisine type switching probability observed in data and in permutations. In Panel (A) 64.5% of consumers are below the 45 degree line, which implies that a majority of consumers switch cuisines less frequently than in the randomization test. In Panel (B) the CDF of cuisine switching probabilities lies within the 95 percent confidence interval of the random permutations distributions, suggesting that they are not significantly different. This result is consistent with the positive and insignificant estimates of cuisine level state dependence. This result implies that heterogeneous preferences explain consumers' cuisine level choices better than state-dependent preferences.

C.3 More Robustness Checks

C.3.1 Initial Conditions Bias

For each consumer choice panel, we drop the first observation of purchase and keep the choice as the state variable for the next purchase occasion. The implicit assumption behind this practice is that the initial state is exogenous and independent of preference. Simonov et al. (2020) discuss the potential initial conditions bias in demand estimation with state dependence and show that this assumption could overestimate loyalty in small samples. On the one hand, our panel is relatively long with 40 choices per consumer on average, which reduces the small-sample bias. On the other hand, in the case that consumers are more likely to order the product that they prefer more in the initial period, treating the initial state as exogenous will attribute preference to loyalty and generate a conservative estimate of the variety-seeking preferences.

For orders put through channels other than the homepage listing, such as discover channel or favorite kitchen folder. We drop the specific period's choice situation from the consumer panel and use the observed choice to update the state variable for the next period. Following similar arguments as in the initial conditions bias discussion above, if we assume that in these orders put through channels other than the search listing, consumers are more likely to order the product that they prefer more, by treating the state from these periods as exogenous will generate a conservative estimate of variety-seeking preferences. The same is true for orders put on the pages other than the first page, for which we drop the choice situations and update the state variable for the next period. When consumers search and order beyond the first page, it is very likely that consumers are looking for a specific niche kitchen that they like. The same analysis of the direction of bias by treating states from these periods as exogenous applies. Orders of kitchens beyond the first page and through other channels take 17% of the total number of orders.

C.3.2 Dish-Level Variety Seeking

In the baseline model we do not consider the dish-level variety-seeking preferences, while in reality consumers could be switching across different dishes within the same kitchen. To test for this possibility we add an interaction term of the last kitchen choice with the menu size of the kitchen. Specifically, the indirect utility that consumer i derives from kitchen j on platform in period t is

$$u_{ijt} = \beta_{ij} - \alpha_i p_{jt} + \eta r_{ijt} + \gamma_i \mathbb{1}(s_{it} = j) + \gamma^M \mathbb{1}(s_{it} = j) M_{jt} + \gamma^C \mathbb{1}(s_{it}^c = c_j) + \psi \cdot X_{ijt} + \epsilon_{ijt}, \tag{C.1}$$

where M_{jt} denotes the menu size of kitchen j in period t. The utility from choosing the outside option is $u_{i0t} = \gamma^O \mathbb{1}(s_{it} = 0) + \epsilon_{i0t}$. If consumers are switching across different dishes within the same kitchen, we should expect consumers to switch less away from kitchens with a larger menu size, i.e., $\gamma^M < 0$. Column (2) of Table C.3.1 presents the result of this alternative specification. We find that γ^M is not significantly different from zero, which implies consumers do not switch less away from kitchens with larger menu sizes.[4] This could happen when consumers have one or two favorite dishes that they always ordered from a specific kitchen.

C.3.3 Interpurchase Time Span

When constructing the choice panel, we drop the days when consumers did not use the platform, which makes the interpurchase time spans different across different periods in term of calendar days. Intuitively, calendar day should be a natural period of satiation

[4]We do not model variety seeking on the dish level mainly for tractability reason. Future research could study state dependence on a finer level including dishes, ingredients, recipe, cooking methods. Natural language processing and graph classification could be helpful methods for such analysis.

Table C.3.1: Robustness Check

		(1) Baseline	(2) Menu Size	(3) Interpurchase Time Span	(4) Nonlinear Position Effect	(5) Negative Autocorrelation
State Dependence –Kitchen level (γ_i)	μ_γ	-0.1981***	-0.1876***	-0.2014***	-0.1873***	-0.2143***
		(0.0503)	(0.0323)	(0.0605)	(0.0529)	(0.0637)
	σ_γ	0.4450***	0.6754***	0.5062***	0.4614***	0.7544***
		(0.0635)	(0.0510)	(0.0687)	(0.0413)	(0.0698)
Price (α_i)	μ_α	-0.0219***	-0.0248***	-0.0267***	-0.0201***	-0.0197***
		(0.0036)	(0.0037)	(0.0061)	(0.0033)	(0.0021)
	σ_α	0.0107***	0.0121***	0.0118***	0.0193***	0.0171***
		(0.0011)	(0.0017)	(0.0034)	(0.0028)	(0.0019)
Rank	η	-0.0391***	-0.0417***	-0.0302***	–	-0.0314***
		(0.0041)	(0.0063)	(0.0059)	–	(0.0052)
State Dependence -Outside option	γ^O	0.0705	0.0621	0.0749	0.0928	0.0717
		(0.0430)	(0.0516)	(0.0712)	(0.0801)	(0.0641)
State Dependence –Cuisine	γ^C	0.0616	0.0492	0.0751	0.0824	0.0752
		(0.0531)	(0.0407)	(0.0506)	(0.0917)	(0.0514)
Menu Size	γ^M		-0.0039			
			(0.0027)			
Interpurchase Time Span	γ^T			0.0196*		
				(0.0112)		
Favorite Seller Unavailable in Previous Period	γ^F					0.0081
						(0.0063)

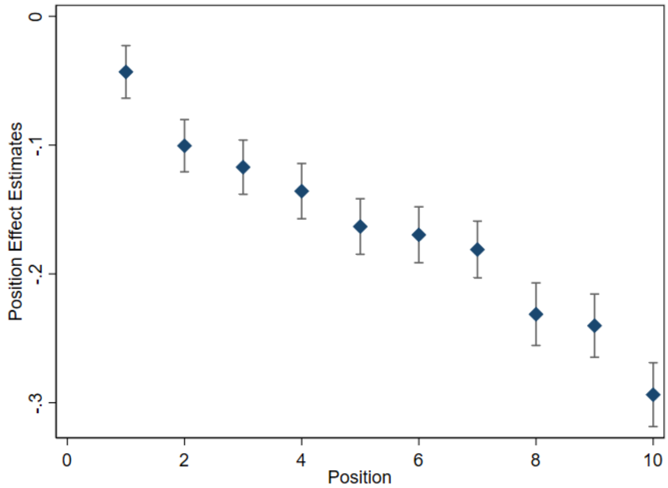
Note: The table shows estimation results after controlling for cuisine, day of week, week of month, holiday, menu size, distance, rating, monthly sales, and kitchen fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Column (1) is the baseline model. Column (2)-(3) are robustness checks. Column (2) tests for dish-level variety seeking. Column (3) tests for the effect of inter-purchase time span on variety-seeking level. Column (4) tests for robustness to the linear position effect problem.

over tastes and the variety-seeking preferences could decrease with the interpurchase time span across platform usage. To test this interpurchase time span effect, we add an interaction term of the last kitchen choice with the number of calendar days from the last consumption. Specifically, the indirect utility that consumer i derives from kitchen j on platform in period t is

$$u_{ijt} = \beta_{ij} - \alpha_i p_{jt} + \eta r_{ijt} + \gamma_i \mathbb{1}(s_{it} = j) + \gamma^T \mathbb{1}(s_{it} = j) T_{ijt} + \gamma^C \mathbb{1}(s_{it}^c = c_j) + \psi \cdot X_{ijt} + \epsilon_{ijt}, \quad \text{(C.2)}$$

where T_{ijt} denotes the interpurchase time span of consumer i with kitchen j until period t. The utility from choosing the outside option is $u_{i0t} = \gamma^O \mathbb{1}(s_{it} = 0) + \epsilon_{i0t}$. If the state dependent preferences decrease with interpurchase time span, we should have $\gamma^T > 0$. Column (3) of Table C.3.1 presents the result of this alternative specification. We find that γ^T is positive but insignificant at the 5% level, which implies consumers do not switch less away from kitchens with longer interpurchase span. This can result from the fact that our sample is mainly composed of the frequent users who ordered at least once a week, which limits the variation in interpurchase time span of consumptions.

Figure C.3.1: Nonlinear Position Effects



Note: The figure shows the estimation results from nonlinear position effects specification.

C.3.4 Nonlinear Position Effects

In the baseline model we specify the ranking position effect as linear, which can be restrictive. To check the robustness of position effect with respect to this functional form restriction, we estimate a model with ranking position dummies. Specifically, the indirect utility that consumer i derives from kitchen j on platform in period t is

$$u_{ijt} = \beta_{ij} - \alpha_i p_{jt} + \sum_{n=1}^{10} \eta_n \mathbb{1}(r_{ijt} = n) + \gamma_i \mathbb{1}(s_{it} = j) + \gamma^C \mathbb{1}(s^c_{it} = c_j) + \psi \cdot X_{ijt} + \epsilon_{ijt}, \quad \text{(C.3)}$$

where η_n denotes the position effect of the nth position in the kitchen list. The utility from choosing the outside option is $u_{i0t} = \gamma^O \mathbb{1}(s_{it} = 0) + \epsilon_{i0t}$. Column (3) of Table C.3.1 presents the result of this alternative specification. We find that nonlinear position effects do not alter other estimates substantially. Figure C.3.1 presents the position effect estimates. The first position do have a larger effect than the other estimates, which could be an effect from the mobile display design.[5] However, the overall functional form does not deviate from linear effect too much, which suggests that the linear specification we adopt in the baseline model does not deviate from the reality substantially.

C.3.5 Negative Autocorrelation

One potential source of spurious state dependence is from negative autocorrelation in the unobserved taste shocks. If the choice model errors are negatively autocorrelated, a past purchase will proxy for a large past and hence a small current random utility draw.[6] The policy implications for the two models will be different, since in the autocorrelated

[5] The mobile display design is shown in Figure C.1.4. Consumers can see the top two positions and half of the third position on the homepage.

[6] Unlike the positive autocorrelation case, which could capture the seasonality or temporal cravings in preference for food, this is more of a mathematical possibility than an intuitive preference possibility.

error case sellers can not use marketing instruments to influence future demand.[7] We test this by interacting the lagged choice with a dummy variable of whether the consumer's favorite kitchen was available in the previous period, where favorite kitchen is defined by the one that is ordered most during the time window we study. Specifically, the indirect utility that consumer i derives from kitchen j on platform in period t is

$$u_{ijt} = \beta_{ij} - \alpha_i p_{jt} + \eta r_{ijt} + \gamma_i \mathbb{1}(s_{it} = j) + \gamma^F \mathbb{1}(s_{it} = j)\mathbb{1}(k_i \notin \mathcal{J}_{i(t-1)}) \\ + \gamma^C \mathbb{1}(s_{it}^c = c_j) + \psi \cdot X_{ijt} + \epsilon_{ijt},$$

where k_i denotes consumer i's favorite kitchen.[8] Under variety seeking, whether the previous period choice was ordered when the consumer's favorite kitchen was available should not affect the state-dependent preferences. However, when ϵ_{ijt} are negatively autocorrelated, kitchens ordered when favorite kitchen was not available will have a lower $\epsilon_{ij(t-1)}$, which implies a higher ϵ_{ijt}. This should decrease the switching probability in the current period, i.e., consumers should switch less away from kitchens that are ordered when the favorite kitchen was not available.($\gamma^F > 0$) Column (5) of Table C.3.1 presents the result of this alternative specification. We find that γ^F is positive but insignificant, which implies consumers do not switch less away from kitchens ordered when favorite kitchen was not available. This could result from the fact that our sample is mainly composed of the frequent users who ordered at least once a week, which limits the variation in interpurchase time span of consumptions. Figure C.3.1 presents the position effect estimates. The first position do have a larger effect than the other estimates, which could be an effect from the mobile display design.[9] However, the overall functional form does not deviate from linear effect too much, which suggests that the linear specification we adopt in the baseline model does not deviate from the reality substantially.

C.3.6 Model Fit

We examine model fit by comparing several key predictions to their empirical counterparts. We split the sample of 6,629 consumers randomly into a 80% training sample (5303 consumers) and a 20% holdout sample (1326 consumers). We estimate the baseline model on the training sample and use the estimated model to predict several key moments in the holdout sample. Table C.3.2 presents the simulated predictions and the empirical counterparts from the holdout sample. We evaluate model out-of-sample fit in terms of purchase probability, switching probability, purchase position, and purchase probabilities by position. We find that the model predicts the holdout sample data reasonably well in terms of the key moments.

[7] From the firm's point of view, autocorrelation is equivalent to a more dispersed distribution of unobserved heterogeneity, whereas in the variety-seeking case, prices affect both current period demand and future distribution of consumer states, which gives rise to a nontrivial dynamic pricing problem. If sellers lower their current price, own consumer share will increase and sellers will face lower and more elastic demand in the next period.

[8] The favorite kitchen is defined as the kitchen with highest choice probability conditional on availability.

[9] The mobile display design is shown in Figure C.1.4. Consumers can see the top two positions and half of the third position on the homepage.

Table C.3.2: Model Out-of-Sample Fit

I. Key Moments		
	Data	Predicted
Purchase Probability	0.473	0.485
Purchase Position	3.464	3.445
Switching Probability	0.814	0.804
II. Purchase Probabilities by Position		
Position	Data	Predicted
1	0.145	0.154
2	0.118	0.117
3	0.0471	0.0465
4	0.0335	0.0356
5	0.0313	0.0332
6	0.0265	0.0291
7	0.0250	0.0237
8	0.0224	0.0223
9	0.0193	0.0190
10	0.0174	0.0170

Note: The table reports results from an out-of-sample validation exercise. We split the sample of 6,629 consumers randomly into a 80% training sample (5303 consumers) and a 20% holdout sample (1326 consumers). We estimate the baseline model on the training sample and use the estimated model to predict several key moments in the holdout sample.

C.3.7 Match Values Estimates

In this section we present the estimates for grouped match value distribution. Table C.3.3 presents the moments and estimates of mean and standard deviation of match value distribution (μ_j, σ_j).

Table C.3.3: Group Estimates

	market share	mean	std		market share	mean	std
Group1	0.1713	1.6214***	1.7180***	Group9	0.0375	0.9768***	1.7599***
		(0.0513)	(0.1248)			(0.0488)	(0.1318)
Group2	0.2366	1.6456***	1.6146***	Group10	0.0132	0.8294***	0.9891***
		(0.0560)	(0.1272)			(0.0500)	(0.1118)
Group3	0.0023	1.2295***	1.4480***	Group11	0.0032	0.7903***	0.6244***
		(0.0664)	(0.1265)			(0.0441)	(0.0801)
Group4	0.0014	1.4634***	1.3761***	Group12	0.1231	1.3608***	0.5332***
		(0.0569)	(0.1192)			(0.0445)	(0.0890)
Group5	0.0002	0.6513***	0.9674***	Group13	0.0198	0.5240***	0.3640***
		(0.0644)	(0.1289)			(0.0438)	(0.0817)
Group6	0.0322	0.4334***	1.1564***	Group14	0.1163	1.2758***	0.3465***
		(0.0526)	(0.1041)			(0.0424)	(0.0746)
Group7	0.1382	1.5843***	0.9602***	Group15	0.1041	1.0952***	0.2085***
		(0.0527)	(0.0892)			(0.0428)	(0.0785)
Group8	0.0005	0.2783***	0.9479***	Group16	0.0001	0.1155***	1.0085
		(0.0483)	(0.0973)			(0.0407)	(0.7846)

Note: The table reports grouping moments and match value distribution parameter estimates.

C.4 Counterfactual Details

C.4.1 Optimality of Utility-based Ranking

We prove the optimality of ranking kitchens in decreasing order of expected utility in this section. The ex ante consumer welfare is

$$E[CS] = -\frac{1}{\alpha}\left[ln \sum_{j\in\mathcal{J}} exp(v_j - c_j)\right]$$

where α is the price coefficient, v_j represents the deterministic part of utility excluding the position cost, and $c_j = \eta \cdot r_j, r_j = 1, 2, ..., J$ represents the position cost. Then the optimal ranking should maximize $E[CS]$ by assigning $c_1, c_2, ..., c_J$ to kitchens in the choice set, represented by $v_1, v_2, ..., v_J$. Without loss of generality, we assume that $v_1 > v_2 > ... > v_J, c_1 < c_2 < ... < c_J$.

Lemma 11. *Ranking kitchens in the decreasing order of expected utility v_j optimize ex ante consumer welfare, given $v_1, v_2, ..., v_J, c_1, c_2, ..., c_J$.*

Proof. Consider the two product case where $v_1 > v_2, c_1 < c_2$, we want to prove that the expected welfare from $\{v_1 - c_1, v_2 - c_2\}$ is greater than $\{v_1 - c_2, v_2 - c_1\}$. Denote $x_1 = v_1 - c_1, x_2 = v_2 - c_2, \tilde{x}_1 = min\{v_1 - c_2, v_2 - c_1\}, \tilde{x}_2 = max\{v_1 - c_2, v_2 - c_1\}$, then $x_1 + x_2 = \tilde{x}_1 + \tilde{x}_2, x_1 < \tilde{x}_1 < \tilde{x}_2 < x_2$. Because the exponential function is increasing and convex, we have

$$\begin{aligned} exp(x_2) - exp(\tilde{x}_2) &> exp(\tilde{x}_2) \cdot (x_2 - \tilde{x}_2) \\ &> exp(\tilde{x}_1) \cdot (\tilde{x_1} - x_1) \\ &> exp(\tilde{x}_1) - exp(x_1) \end{aligned}$$

Table C.4.1: Decomposition of Optimal Ranking Effect

Choice under Supoptimal Ranking	Choice under Optimal Ranking	Market Outcome			
		Consumer Welfare		Revenue	Purchase Prob.
		Utility	Position Cost		
Outside Option	Different seller	↑	↑	↑	↑
Last period choice	Different seller	↑	–	Uncertain	–
Different seller	Different seller	–	↓	Uncertain	–

Note: The table shows the different scenarios where considering variety-seeking effects in ranking algorithms could affect consumer welfare (through position cost and utility), platform revenue, and purchase probability.

Thus $E[CS_{\{v_1-c_1,v_2-c_2\}}] > E[CS_{\{v_1-c_2,v_2-c_1\}}]$.

Then given a general number of products J, we want to prove that the optimal ranking is $\{v_1 - c_1, v_2 - c_2, ..., v_J - c_J\}$. Starting from any ranking position assignment $R_1 = \{..., v_j - c_j, ..., v_k - c_k, ...\}$ where $\exists j, k, s.t., v_k > v_j, c_k > c_j$, we show that this assignment is dominated by reversing the position of the two kitchens while keeping the other positions fixed $R_2\{..., v_j - c_k, ..., v_k - c_j, ...\}$.

$$\begin{aligned} E[CS_{R_1}] = & \ -\tfrac{1}{\alpha}\left[ln(exp(v_j - c_j) + exp(v_k - c_k) + C)\right] \\ E[CS_{R_2}] = & \ -\tfrac{1}{\alpha}\left[ln(exp(v_j - c_k) + exp(v_k - c_j) + C)\right] \end{aligned}$$

where C denotes the sum of exponential funtions for the other positions. Then from the proof in the case when $J = 2$, we have $E[CS_{R_1}] < E[CS_{R_2}]$. We can continue this process until there is no higher expected utility assigned to lower ranking positions, and thus we prove the optimality of expected utility based ranking. □

Note that the utility based ranking need not optimize the platform revenue, since the platform could promote lower utility but higher price kitchens to increase revenue. There is not a simple index like in the utility case, and the ranking will face a tradeoff between price and choice probability. Appendix 6 in Compiani et al. (2021) provides more details on the intuition of relevant tradeoffs of revenue-maximizing ranking algorithms. The crucial value in the optimality of ranking algorithms in this case is the difference between option j's revenue, and the choice-probability weighted average revenue from all other options.[10]

C.4.2 Decomposition of Optimal Ranking Effect

Optimal ranking only generates different market outcomes from suboptimal ranking when the variety-seeking effects are high enough to either change the position of options or change the choice of consumers. Table C.4.1 summarizes different scenarios when considering variety-seeking effects in ranking algorithms generates different market outcomes, and shows how consumer welfare (through position cost and utility), platform revenue, and purchase probability are affected in each scenario.

[10]The industry ranking algorithm usually use the multi-objective optimization model. Food Discovery with Uber Eats: Recommending for the Marketplace `https://eng.uber.com/uber-eats-recommending-marketplace/` In this project, we focus on documenting, measuring variety seeking, and showing its importance through counterfactual analysis, instead of providing the exact engineering algorithms.

C.4.3 Heterogeneous State-Based Targeted Ranking

In this section we study the ranking effects allowing consumers to have heterogeneous state-dependent preferences, and platform presents personalized rankings to them accounting for it. We study this by taking 1000 draws from the estimated distribution of state-dependent preferences. For each draw of random coefficients, we simulate consumer choices 50 times to get the market outcome. For each market outcome, we calculate the ratio $\Delta S = \frac{Optimal-Suboptimal}{Optimal-Random}, S \in \{W, \Pi, q\}$ to measure the state-based targeted ranking effect on consumer welfare, platform revenue, and purchase probability. For each consumer i, $\Delta \tilde{S}$ is a function of the state-dependent preference γ_i, and the choice set effect C_i.[11] Equations C.4, C.5, and C.6 present the average ranking effect on the platform.

$$\Delta W = \int \Delta \sum_i W(\gamma_i, C_i) dF(\gamma_i), \tag{C.4}$$

$$\Delta \Pi = \int \Delta \sum_i \Pi(\gamma_i, C_i) dF(\gamma_i), \tag{C.5}$$

$$\Delta q = \int \frac{1}{N} \Delta \sum_i q(\gamma_i, C_i) dF(\gamma_i). \tag{C.6}$$

Figure C.4.1 presents the average ranking effects on consumer welfare, platform revenue, and purchase probability. We find that optimizing the ranking algorithm with variety-seeking preferences takes up 13.0% of the revenue improvement, 16.4% of the consumer welfare improvement, and 17.3% of the purchase probability improvement, out of the total ranking effect. Figure C.4.2 presents the detailed distribution of state-based targeted ranking effect on the platform level across simulations.

C.4.4 Numerical Solution for Price Competition

We use numerical methods to solve for the equilibrium of the pricing game. We first discretize each axis of the state space using a finite number of points. We then form a grid representing the whole state space from the Cartesian product of these points, excluding the support where the sum of market shares exceeds one. For each point in the grid, we store the value and policy functions of each competitor in the computer memory. For states outside the grid, we calculate the value and policy function using polynomial interpolation.

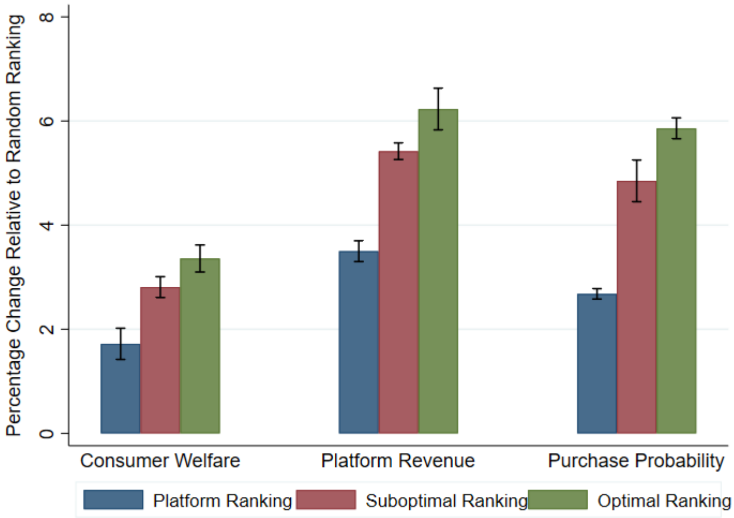
To solve for the equilibrium, we employ the following algorithm, which is an adaptation of policy iteration applied to the case of the games: start with some initial guess of the strategy profile, $\sigma^0 = (\sigma_1^0, \sigma_2^0)$, and then proceed along with the following steps:

1. For the strategy profile σ^n, calculate the corresponding value functions for each of the J kitchens. These value functions are defined by the Bellman equation, where the right hand side of the Bellman equation is maximized, given the current strategy profile of other kitchens. Update each kitchen's strategy using the Bellman equation. Denote the resulting new policies and value functions by σ_j^{n+1} and V_j^{n+1}.

2. If $n > 0$, check whether the value functions and policy functions satisfy the convergence criteria, $||V_j^n - V_j^{n-1}|| < \epsilon_V$ and $||\sigma_j^n - \sigma_j^{n-1}|| < \epsilon_\sigma$ for all kitchens j. If so,

[11] To understand the intuition why ΔS is a function of individual choice set C_i, consider an extreme example where there is no overlapping options of a consumer's choice set across different periods, then we the state-based targeted ranking will have no effect on this consumer's choices.

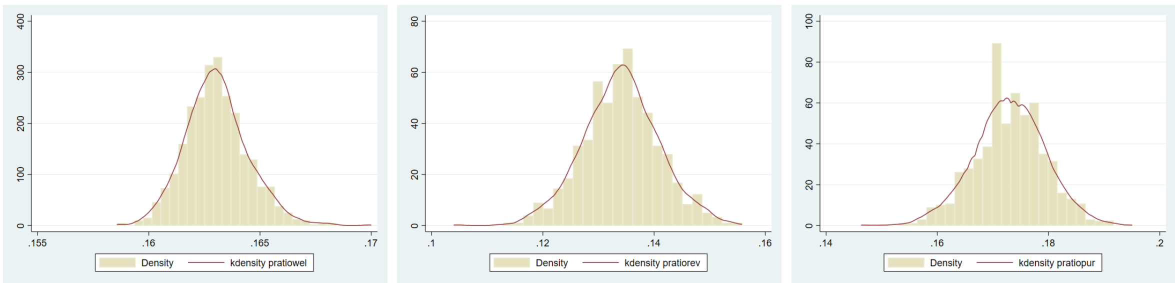
Figure C.4.1: Ranking Effects with Heterogeneous State-Based Targeting



Note: The figure shows the percentage change in average platform-level welfare, revenue and purchase probability of platform ranking, suboptimal ranking, and optimal ranking relative to the random ranking benchmark. We simulate choices under each given ranking algorithms for 50 times to get the mean and standard deviation of corresponding variables.

Figure C.4.2: Distribution of Ranking Effects with Heterogeneous State-Based Targeting

(A) Welfare (B) Revenue (C) Purchase Probability



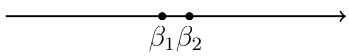
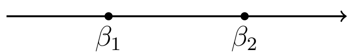
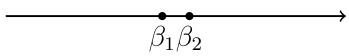
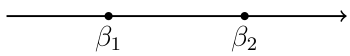
Note: The figure shows the detailed distribution of state-based targeted ranking effect on the platform level. Panel(A)-(C) presents the distribution of platform-level consume welfare, revenue, and purchase probability across 1000 simulations.

stop. If not, return to step 1.

C.4.5 Intuition for Variety Seeking Effect on Price Competition

In this section we provide more intuition on variety seeking effect on price competition. Seetharaman and Che (2009) find that variety seeking softens price competition and Sajeesh and Raju (2010) find that variety seeking intensifies price competition. Zeithammer and Thomadsen (2013) shows that when products are similar and variety seeking is strong, variety seeking softens price competition, whereas when products are dissimlar and variety seeking is weak, variety seeking intensifies price competition. This seems in conflict with the traditional wisdom from industrial organization that differentiation

Table C.4.2: Product Differentiation without Variety-Seeking Effect

	Similar, strong VS	Dissimilar, weak VS
Seller 1's consumer:		
Seller 2's consumer:		

Note: The table shows product differentiation level without variety seeking effect. The columns corresponds two cases when product are similar and variety seeking is relatively strong, and when product are dissimilar and variety seeking is relatively weak. The rows are products' values from consumers who purchased product 1, and 2 in the previous period, respectively.

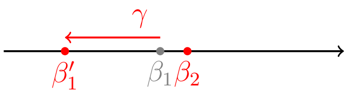
softens price competition. Here we discuss how we can understand this result intuitively and how it connects with our empirical finding. Table C.4.2 shows product differentiation level without variety seeking effect. The columns corresponds two cases when product are similar and variety seeking is relatively strong, and when product are dissimilar and variety seeking is relatively weak. The rows are products' values from consumers who purchased product 1, and 2 in the previous period, respectively.

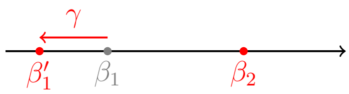
Table C.4.3 shows product differentiation level with variety seeking effect. For consumers who purchased the inferior product in the last period, variety seeking decrease the product value further and increase the differentiation level. For consumers who purchased the superior product in the last period, variety seeking decrease the product value of the superior product, and the effect on differentiation level depends on the relative magnitude of static differentiation from product value, and the dynamic differentiation from state dependence. When products are similar and variety seeking is strong, variety seeking increases overall differentiation and softens price competition, whereas when products are dissimlar and variety seeking is weak, variety seeking decreases overall differentiation and intensifies price competition.[12] [13] Our demand estimates suggest that we are in the case when the top two popular groups are relatively similar and consumers' variety-seeking preferences are relatively strong, so variety seeking softens price competition within a range around our demand estimate level.

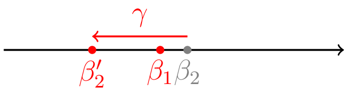
[12] Variety seeking intensifies price competition when products are dissimilar and there is one big popular seller maintaining own consumers who are choosing between consuming the superior product consecutively and switching towards the inferior product.

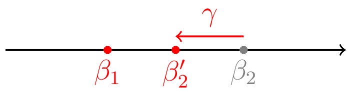
[13] This decomposition is also helpful for understanding the nonmonotonic relationship between state dependence level and price competition level. Holding fixed the static product values, there is a range where state dependence first decrease and then increase product differentiation level.

Table C.4.3: Product Differentiation with Variety Seeking Effect

<table>
<tr><td></td><td>Similar, strong VS</td><td>Dissimilar, weak VS</td></tr>
<tr><td>Seller 1's consumer:</td><td>

Differentiation ↑</td><td>

Differentiation ↑</td></tr>
<tr><td>Seller 2's consumer:</td><td>

Differentiation ↑</td><td>

Differentiation ↓</td></tr>
</table>

Note: The table shows product differentiation level with variety seeking effect. The columns corresponds two cases when product are similar and variety seeking is relatively strong, and when product are dissimilar and variety seeking is relatively weak. The rows are products' values from consumers who purchased product 1, and 2 in the previous period, respectively. Arrows shows the differentiation level change in each cases.

C.5 Industry Practice

In this section, we discuss the current targeted marketing practice in the industry, and the potential improvement by taking into account consumer variety-seeking preferences. We find that consumers' variety-seeking preferences are not accounted for in current industry targeted pricing and are likely not emphasized enough in ranking and recommendation systems on several common food delivery platforms.

C.5.1 Targeted Pricing

Current industry practice in pricing is more observable than ranking and recommendation systems on food delivery platforms. We find several common platforms' promotion activities facing merchants on their websites. Uber Eats allows merchants to targeted promotions to either all customers in their delivery zone or new customers that have not ordered from their restaurant. DoorDash's 'Order Again and Save' promotion is for customer retention and allows merchants to pay \$0.99 fee per order to target their existing DoorDash consumers who haven't ordered from them in over 45 days. Similarly, Grubhub's promotion for past customers is also focused on past customer retention, and it only allows merchants to target consumers who have not ordered in the last 90 days to order again. No popular food delivery platforms allow targeted pricing to old customers with recent orders. Table C.5.1 provides a summary of the current merchant-level promotions allowed on popular food delivery platforms.

C.5.2 Personalized Recommendation

Besides ranking and pricing, our findings also have important implications for the design of personalized recommendation. We do not have data on personalized recommendation for our platform, so we can not do counterfactual analysis about it using our data. However, Figure C.5.1 shows the current common practice of personalized recommendation on several popular food delivery platforms. The recommendation sections such

Table C.5.1: Current Targeted Pricing Available for Merchants on Popular Food Delivery Platforms

Platform	Current Consumer Types for Merchant Targeted Promotions
Uber Eats	New consumers All consumers in the restaurant's delivery zone
DoorDash	New consumers Consumers who haven't ordered from the restaurant in over 45 days All consumers
Grubhub	New consumers Consumers who haven't ordered in the last 90 days All consumers

Note: The table summarizes the current promotions allowed on popular food delivery platforms. Information comes from the platforms' websites for merchants. None of the popular platforms allow for promotions targeting existing consumers who ordered recently.

as UberEats's 'Order Again', Grubhub's 'Reorder your favorites', and DoorDash's 'Your Favorites' all understand and emphasize the importance of consumer's persistent heterogeneous preferences, but it does not seem like they consider consumer variety-seeking preferences by avoiding recent orders.

Koren et al. (2009) discuss several techniques for recommendation systems in the context of Netflix movie recommendation. Such systems are particularly useful for entertainment products such as movies, music, and TV shows. Among the models discussed, the only model that captures dynamics in demand is the temporal dynamics model, which account for the time-drifting nature of user preference by allowing preference parameters as a function of time. But the specific model captures the changes in persistent preference over time,[14] which is different from the high-frequency dynamics driven by variety-seeking preferences.

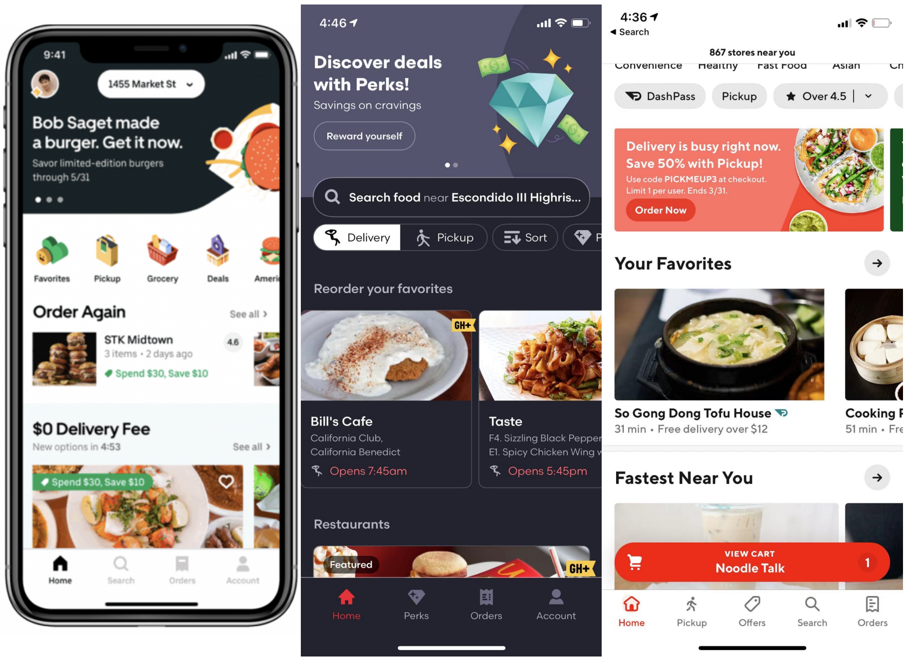
Uber Eats posts research results on Uber Engineering blogs, and based on the recent reports, variety-seeking dynamics are not captured in their recommendation or ranking algorithm. Jain et al. (2019) [15] discuss that Uber Eats use the graph learning method GraphSAGE developed by Hamilton et al. (2017) to power their recommendation system. In this method the aggregators are symmetric(permutation invariant),[16] and a random permutation of the users' orders does not affect the graph and thus recommendation, thus the high frequency dynamics induced by variety-seeking preferences are not included

[14] Koren et al. (2009) describe this with an example: "..., a fan of the psychological thrillers genre might become a fan of crime dramas a year later. Similarly, humans change their perception of certain actors and directors."

[15] Food Discovery with Uber Eats: Using Graph Learning to Power Recommendations `https://eng.uber.com/uber-eats-graph-learning/`

[16] The method is based on a graph where nodes are users and sellers, and edges are weighted by historical number of orders. For example, consumer 1 who ordered CBAAA and consumer 2 who ordered ABACA will generate the same graph and the method will predict that they will have similar order patterns.

Figure C.5.1: Recommendations on Popular Food Delivery Platforms



Note: The graph shows screenshots of the homepage recommendation from popular platforms including Uber Eats, Grubhub, and DoorDash. The frequently ordered kitchens are recommended as 'Order Again' and 'Your Favorites' without punishment on the recency of orders from the recommended kitchens.

in this method or widely used matrix factorization and collaborative filtering methods for recommendation systems. Meanwhile, methods that process inputs in a sequential manner such as LSTM have limitations in scalability.

C.6 Upper Confidence Bound Algorithm

Currently one of the most commonly used optimal solutions for multi-armed bandit problems is the upper confidence bound (UCB) algorithm.[17] One general form of the method is to choose the option with the optimal UCB index, UCB_i, which is defined as follows:

$$UCB_i = \hat{\mu}_i + f(T_i) \tag{C.7}$$

where $\hat{\mu}_i$ is the empirical mean of former experience. $f(T_i)$ is a decreasing function of the sample times the learner has observed for arm i. This index illustrates the explore vs exploit tradeoff, where the first term $\hat{\mu}_i$ captures the exploit part and the second term $f(T_i)$ represents potential reward from exploration. The explore vs exploit tradeoff

[17]This method is currently used by Uber Eats. "..., we applied the upper confidence bound (UCB, one of the methodologies of MAB) approach to facilitate the exploration of new restaurants or restaurants with low impressions. We calculate a UCB score for each restaurants based on metrics such as its historic impression number, total click number, and boosting factor. The UCB score, along with other objectives discussed above, decides the ranking order among restaurants. A new restaurant will have a relatively high UCB score initially and hence rank highly, increasing its exposure. As the new restaurant gathers other impressions, the UCB score will smoothly decrease and gradually traster ranking weight back to other objective scores such as relevance." Wang et al. (2018): Food Discovery with Uber Eats: Recommending for the Marketplace `https://eng.uber.com/uber-eats-recommending-marketplace/`

implied by this solution suggests less switching among familiar arms and more switching towards new arms, which corresponds to the two tests we perform in Section 3.6.1 and Section 3.6.2.

Note that the explore part is different from variety seeking in the sense that it is based on cumulative experience with the arm, and the recency of experience does not matter, whereas in variety seeking the recency of an order is crucial.