Online Appendix for "Consumer Search and Purchase: An Empirical Investigation of Retargeting Based on Consumer Online Behaviors"

Zhenling Jiang, Tat Chan, Hai Che, Youwei Wang

1 Robustness with Correlation between Pre- and Post- Click Error Terms

In the main model, we assume that the pre- and post- click error terms e_{ij} (equation 2) and ϵ_{ij} (equation 3) are independent. It is conceivable that the two sets of error terms may be either positively or negatively correlated. For example, a consumer who likes an option on the search research page more than other consumers (represented by a high e_{ij}) will likely continue to appreciate it after click (represented by a high ϵ_{ij}). If so, the two error terms may be positively correlated. On the other hand, if an option attracts more clicks than usual (e.g. when it moves up in ranking) with a higher e_{ij}, consumers clicking into it will likely realize a lower than expected actual value from a lower ϵ_{ij}, leading to a negative correlation. We assume that e_{ij} and ϵ_{ij} are jointly distributed as bivariate normal distributions. To operationalize the correlation, we let

$$\epsilon_{ij} = \frac{\rho}{\sigma_e} \cdot e_{ij} + \sqrt{\sigma_\epsilon^2 - \rho^2} \cdot \eta_{ij}$$

where $\eta_{ij} \sim N(0,1)$. Under this specification, ρ represents the correlation coefficient. Consumers have rational expectation of ρ and takes it into account in their search decisions.

While it is difficult to pin down the correlation, especially when the majority of consumers' searches are short, we show that the estimation and counterfactual results are robust to possible correlations between the two error terms. To evaluate the impact, we re-estimate the model while fixing the correlation ρ to be -0.1 and 0.1 respectively to represent scenarios with possible negative and positive correlations. The estimation results are shown in Table 1. The estimates of key model parameters are very close to the main model.

We further conduct counterfactuals based on the estimation results. Tables 2 - 4 are analogous to Tables $6 - 8$ in the main results. Importantly, all of the findings from counterfactuals remain unchanged. We still find that providing sellers the information on consumers' search history will significantly improve the effectiveness of retargeting, and that seller recommendation is more effec-

Table 1: Estimation Results with Correlation ρ

Utility parameters:	*Fix* $\rho = 0.1$		*Fix* $\rho = -0.1$	
Constant	-3.7791 (0.0064)		-3.4058 (0.0063)	
Price	-0.0921 (0.0028)		-0.0935 (0.0029)	
High star	0.2231 (0.0047)		0.2221 (0.0046)	
Same location	0.2173 (0.0041)		0.2224 (0.0043)	
Age (log)	0.0219 (0.0018)		0.0211 (0.0020)	
Male	0.0111 (0.0047)		0.0140 (0.0048)	
Missing demographics	0.0612 (0.0049)		0.0612 (0.0046)	
Number of past transactions (log)	0.0061 (0.0005)		0.0060 (0.0007)	
Working hours	0.0230 (0.0006)		0.0229 (0.0008)	
Weekend	-0.0046 (0.0016)		-0.0024 (0.0017)	
Search cost parameters:				
Mean of log search cost	-5.4349 (0.0016)		-4.3958 (0.0021)	
Std. dev. of log search cost	0.5716 (0.0010)		0.5596 (0.0011)	
Estimated $\xi's$ *and* $\omega's$	*Pre-click* ξ_j	*Post-click* ω_j	*Pre-click* ξ_j	*Post-click* ω_j
Seller 1	0.7755 (0.0031)	-0.4540 (0.0106)	0.7800 (0.0032)	-0.5139 (0.0149)
Seller 2	0.5154 (0.0036)	-0.1953 (0.0083)	0.5202 (0.0035)	-0.2127 (0.0086)
Seller 3	0.2045 (0.0039)	0.0799 (0.0076)	0.2069 (0.0039)	0.0714 (0.0075)
Seller 4	0.0697 (0.0042)	0.2604 (0.0074)	0.0719 (0.0042)	0.2585 (0.0077)
Seller 5	0.3721 (0.0047)	-0.3208 (0.0085)	0.3710 (0.0048)	-0.3113 (0.0091)
Seller 6	-0.0265 (0.0054)	-0.5533 (0.0083)	-0.0269 (0.0053)	-0.5685 (0.0090)
Seller 7	0.0409 (0.0055)	-0.0486 (0.0081)	0.0368 (0.0055)	-0.0297 (0.0077)
Seller 8	0.1247 (0.0055)	0.1939 (0.0073)	0.1240 (0.0051)	0.1949 (0.0074)
Seller 9	-0.1979 (0.0061)	-0.0105 (0.0079)	-0.2011 (0.0060)	0.0173 (0.0076)
Seller 10	-0.2130 (0.0060)	0.1921 (0.0078)	-0.2169 (0.0063)	0.1962 (0.0078)
Seller 11	-0.1926 (0.0062)	-0.2479 (0.0076)	-0.1874 (0.0061)	-0.2397 (0.0072)
Seller 12	-0.0886 (0.0068)	0.2544 (0.0075)	-0.0914 (0.0071)	0.2549 (0.0074)
Seller 13	0.0023 (0.0071)	0.2689 (0.0075)	0.0074 (0.0071)	0.2656 (0.0083)
Seller 14	-0.2482 (0.0063)	0.4817 (0.0080)	-0.2489 (0.0062)	0.4821 (0.0076)
Seller 15	-0.1220 (0.0073)	0.0774 (0.0082)	-0.1226 (0.0074)	0.0237 (0.0087)
Seller 16	-0.1231 (0.0074)	0.0870 (0.0082)	-0.1256 (0.0072)	0.1537 (0.0078)
Seller 17	-0.3459 (0.0171)	0.1400 (0.0089)	-0.3380 (0.0175)	0.1504 (0.0081)
Seller 18	-0.1057 (0.0102)	0.1983 (0.0091)	-0.1041 (0.0140)	0.1989 (0.0081)
Seller 19	-0.1005 (0.0197)	0.1103 (0.0082)	-0.0956 (0.0220)	0.0974 (0.0079)
Seller 20	-0.3410 (0.0174)	-0.5143 (0.0369)	-0.3597 (0.0176)	-0.4892 (0.0373)
Observations	104,819		104,819	
LL	-288,006.6		-287960.9	
AIC	576,109.2		576017.8	
BIC	576,568.1		576476.7	

tive than offering coupons as a retargeting strategy. These results suggest that our main findings are robust to possible correlation between the pre- and post-click error terms.

Table 2: Results from Random Retargeting with Coupons and the Heterogeneity

	Fix $\rho = 0.1$		*Fix* $\rho = -0.1$	
	Conversion rate of retargeting	Expected value (¥)	Conversion rate of retargeting	Expected value (¥)
Random retargeting for all consumers w/o purchase	0.0077%	0.0083	0.0071%	0.0077
Heterogeneity Based on Observed Consumer Characteristics:				
Past number of transactions				
0 – 30	0.0069%	0.0074	0.0063%	0.0067
31 – 150	0.0075%	0.0081	0.0070%	0.0074
151 or above	0.0084%	0.0091	0.0079%	0.0084
Gender				
Male	0.0083%	0.0090	0.0078%	0.0083
Female	0.0080%	0.0086	0.0074%	0.0080
Age				
18 – 25	0.0078%	0.0085	0.0073%	0.0078
26 – 30	0.0083%	0.0090	0.0078%	0.0084
31 or above	0.0085%	0.0092	0.0080%	0.0086
Time of Visit				
Working hours	0.0080%	0.0086	0.0074%	0.0080
Non-working hours	0.0072%	0.0078	0.0067%	0.0072
Day of Visit				
Weekday	0.0078%	0.0084	0.0072%	0.0077
Weekend	0.0075%	0.0081	0.0070%	0.0076
Heterogeneity Based on Consumer Search Behaviors:				
Own or Competitors' Customers				
Send to own customers	0.0965%	0.1068	0.0090%	0.0958
Send to competitors' customers	0.0021%	0.0023	0.0022%	0.0023
Search Intensity				
Searched one seller	0.0055%	0.0059	0.0051%	0.0054
Searched two sellers	0.0234%	0.0254	0.0222%	0.0240
Searched three or more sellers	0.0318%	0.0347	0.0300%	0.0325
Order of Search (among own customers)				
Searched two sellers:				
First searched seller	0.2277%	0.2446	0.1986%	0.2122
Second searched seller	0.1496%	0.1632	0.1470%	0.1599
Searched three sellers:				
First searched seller	0.2412%	0.2590	0.2009%	0.2149
Second searched seller	0.1611%	0.1756	0.1494%	0.1619
Last searched seller	0.1235%	0.1352	0.1198%	0.1307
Heterogeneity Based on Seller Characteristics:				
Sales Volume				
High	0.0134%	0.0144	0.0129%	0.0138
Low	0.0020%	0.0022	0.0014%	0.0015
Click Rate				
High	0.0131%	0.0139	0.0127%	0.0134
Low	0.0023%	0.0027	0.0016%	0.0019
Conversion Rate				
High	0.0095%	0.0104	0.0086%	0.0094
Low	0.0059%	0.0064	0.0057%	0.0059

Table 3: Counterfactual – Retargeting with Coupons with Correlation ρ

	Random Retargeting	Retargeting by the Last Click	Retargeting through Second-Priced Bids: No Consumer Information	Retargeting through Second-Priced Bids: with Consumer Search History
Fix $\rho = 0.1$				
Conversion rate (%)	0.008	0.082	0.038	0.096
Consumer welfare (¥)	550	4,841	2,266	5,697
Seller profit (¥)	-	-	799	9,618
Platform profit (¥)	-	-	15,593	33,230
Total profit (¥)	3,507	36,630	16,392	42,848
Fix $\rho = -0.1$				
Conversion rate (%)	0.007	0.074	0.039	0.084
Consumer welfare (¥)	535	4,395	2,327	4,955
Seller profit (¥)	-	-	811	8,120
Platform profit (¥)	-	-	16,060	29,269
Total profit (¥)	3,247	33,166	16,871	37,389

Table 4: Counterfactual – Retargeting by Seller Recommendation with Correlation ρ

	Random Retargeting	Retargeting through Second-Priced Bids: No Consumer Information	Retargeting through Second-Priced Bids: with Consumer Search History
Fix $\rho = 0.1$			
Conversion rate (%)	0.107	0.254	0.258
Consumer welfare (¥)	29,153	72,864	73,562
Seller profit (¥)	-	9,432	8,112
Platform profit (¥)	-	154,668	157,328
Total profit (¥)	60,854	164,100	165,440
Fix $\rho = -0.1$			
Conversion rate (%)	0.125	0.372	0.389
Consumer welfare (¥)	34,259	103,557	108,192
Seller profit (¥)	-	9,212	14,166
Platform profit (¥)	-	191,278	197,487
Total profit (¥)	71,916	200,490	211,653

2 Estimation with Three Price and Star Level Groups

In the main estimation results, we estimate the price coefficient by using the variation in clicking and purchase patterns between high price (average price ¥2,334) group and low price group (average price ¥2,102). Similarly, the coefficient for star level is also estimated using the variation between high star (average 12.91) group and low star group (average 9.12). In this appendix, we present

results when breaking the sellers into three price and seller star groups. Table 5 shows the average price and star level for each seller, as well as the categorization of the price and star level groups.

We run the estimation with the three price level groups and three star level groups as a robustness check. With three price groups, the average price is ¥2,412 for the high price group, ¥2,206 for the medium price group and ¥2,092 for the low price group. We directly estimate the price coefficient $\beta_p$for ¥100. Using the low price group as a baseline, the perceived utility difference is $1.14 \cdot \beta_p$ for sellers in the medium price group, and $3.2 \cdot \beta_p$ for sellers in the high price group. With three star level groups, the average star level is 13.67 for the high star group, 11.52 for the medium star group and 7.68 for the low star group. We estimate the star level coefficient $\beta_s$for 1 star level difference. Using the low star level group as a baseline, the perceived utility difference is $3.84 \cdot \beta_s$ for sellers in the medium star group, and $5.99 \cdot \beta_s$ for sellers in the high star group.

The estimated price coefficient for ¥100 is -0.051. This is close to the results in the main analysis with -0.040 for ¥100 (-0.0932 with ¥232 average price difference between high and low price groups). The estimated star coefficient for 1 star level difference is 0.045, which is close to the main analysis with 0.059 for 1 star level difference (0.2234 for 3.79 average star level difference between high and low star groups). These results show the robustness of inferring the coefficients for price and star level using the variation across sellers.

Table 5: Price and Star Level for Each Seller

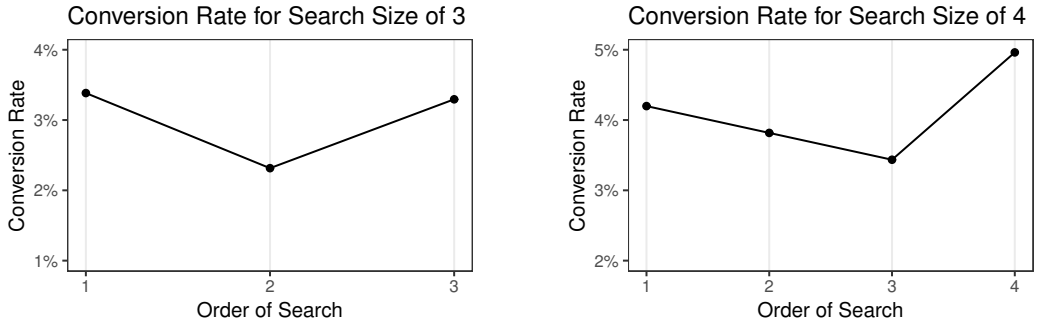
Seller:	Price	Two Price Groups	Three Price Groups	*Seller:*	Star Level	Two Star Level Groups	Three Star Level Groups
9	¥2,082	L	L	19	6.0	L	L
15	¥2,082	L	L	16	7.0	L	L
4	¥2,082	L	L	7	7.4	L	L
10	¥2,089	L	L	13	9.0	L	L
16	¥2,102	L	L	5	9.0	L	L
1	¥2,103	L	L	12	10.7	L	M
12	¥2,105	L	L	15	11.0	L	M
7	¥2,123	L	M	8	11.0	L	M
17	¥2,126	L	M	18	11.0	L	M
3	¥2,127	L	M	20	12.0	H	M
20	¥2,183	H	M	17	12.0	H	M
11	¥2,202	H	M	9	12.0	H	M
19	¥2,290	H	M	14	12.0	H	M
2	¥2,292	H	M	3	12.0	H	M
8	¥2,308	H	M	1	13.0	H	H
5	¥2,339	H	H	6	13.0	H	H
6	¥2,359	H	H	11	13.0	H	H
13	¥2,376	H	H	10	13.0	H	H
18	¥2,489	H	H	2	15.0	H	H
14	¥2,499	H	H	4	15.0	H	H

3 Robustness Analysis with Another Product

The main analysis is based on one product with the largest number of searches and purchases. In this appendix, we show robustness of the analysis using the iPad 4 16G WiFi product, which has the second largest number of searches and purchases in our data.

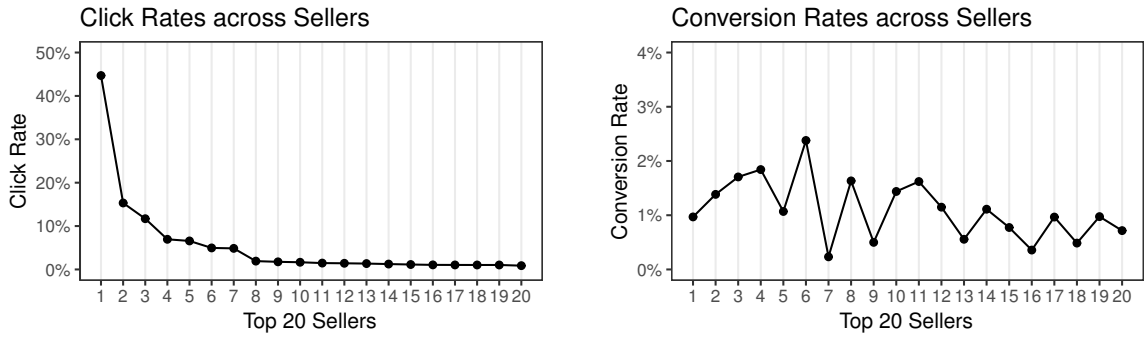
We show that the reduced-form data patterns in the main analysis continue to hold. We observe a U-shaped relationship between the order of search and conversion rate for consumers with search size 3 or 4 (Figure 1), which is consistent with an optimal sequential search model.

Figure 1: Order of Search and Conversion Rate



For the top 20 sellers with the iPad 4 16G WiFi product, similarly, we see a large variation in click rates and conversion rates across the sellers (Figure 2).

Figure 2: Click Rates and Conversion Rates across Sellers



In addition, we show estimation results with the iPad 4 16G WiFi. Table 6 is analogous to the estimation results using the main product category (iPad Mini 16G WiFi) reported in Table 5. The estimates are very close between the two products.

Table 6: Estimation Results for iPad 4

Utility parameters:	Estimates (std. err.)	
Constant	-3.5378 (0.0075)	
Price	-0.1638 (0.0050)	
High star	0.2911 (0.0064)	
Same location	0.2413 (0.0051)	
Age (log)	0.0123 (0.0026)	
Male	0.0263 (0.0058)	
Missing demographics	0.0395 (0.0057)	
Number of past transactions (log)	0.0024 (0.0009)	
Working hours	0.0033 (0.0011)	
Weekend	-0.0126 (0.0019)	
Search cost parameters:		
Mean of log search cost	-4.7622 (0.0023)	
Std. dev. of log search cost	0.6527 (0.0013)	
Estimated $\xi's$ and $\omega's$		
Seller 1	0.6550 (0.0030)	-0.2744 (0.0082)
Seller 2	0.2673 (0.0038)	0.0003 (0.0073)
Seller 3	0.1568 (0.0042)	0.1586 (0.0069)
Seller 4	-0.0014 (0.0044)	0.2418 (0.0072)
Seller 5	0.0368 (0.0048)	0.0230 (0.0073)
Seller 6	-0.0387 (0.0057)	0.3394 (0.0071)
Seller 7	0.3613 (0.0059)	-0.4544 (0.0073)
Seller 8	0.2253 (0.0065)	0.2409 (0.0075)
Seller 9	-0.2249 (0.0079)	-0.0535 (0.0077)
Seller 10	0.0450 (0.0081)	0.2915 (0.0077)
Seller 11	-0.3133 (0.0071)	0.2405 (0.0078)
Seller 12	-0.1128 (0.0081)	0.1500 (0.0076)
Seller 13	0.1631 (0.0079)	-0.2057 (0.0080)
Seller 14	-0.1761 (0.0074)	0.1374 (0.0081)
Seller 15	-0.2880 (0.0103)	0.0699 (0.0074)
Seller 16	-0.1806 (0.0079)	-0.2017 (0.0080)
Seller 17	-0.4104 (0.0211)	0.1665 (0.0092)
Seller 18	-0.0275 (0.0094)	-0.1260 (0.0084)
Seller 19	-0.1776 (0.0225)	0.0721 (0.0082)
Seller 20	0.0406 (0.0213)	-0.8164 (0.0341)
Observations	79,689	
LL	-225,569.9	
AIC	451,235.8	
BIC	451,681.5	