

**Mobile Location-Based Recommendation Systems and Local Businesses:
Evidence from the Restaurant Industry
Online Supplementary Appendix**

Appendix I: Prior Research Streams and Contribution of This Study

In Appendix I, we position this study relative to prior work across three dimensions: research context (online vs. offline/local), consumer engagement outcomes (exposure, selection, conversion), and contextual factors (non-spatial vs. spatial). This paper is distinct in examining the role of recommendation systems in offline, location-based environments while tracing consumer engagement across the full purchase funnel and incorporating multiple spatial contextual factors.

Table A1. Prior Research Streams and Contribution of This Study

Study	Research Context		Consumer Engagement			Contextual Factors	
	Online Consumer Behavior	Offline Consumer Behavior	Exposure	Selection (View)	Conversion (Purchase)	Non-spatial Factors	Spatial Factors
Adamopoulos et al. (2022)		O			O	O	O
Chang and Chin (2010)	O				O	O	
Cooke et al. (2002)	O				O	O	
De et al. (2010)	O				O	O	
Lee and Hosanagar (2019)	O			O	O	O	
Lee and Hosanagar (2021)	O			O	O	O	
Lee et al. (2020)	O			O	O	O	
Li et al. (2022)	O		O		O	O	
Oestreicher-Singer and Sundararajan (2012)	O				O	O	
Peng and Liang (2023)	O		O	O	O	O	
Senecal and Nantel (2004)	O			O		O	
Wan et al. (2024)	O			O	O	O	
Yuan et al. (2025)	O			O	O	O	
Zheng et al. (2025)	O				O	O	
Zhou et al. (2024)	O				O	O	
This Study		O	O	O	O	O	O

Appendix II: Recommendation Algorithm of *SmartAround*

SmartAround is a location-based recommendation service that curates nearby places to visit, eat, and shop based on the user's current location, time, age, gender, and other contextual information. Our partnering search engine company indicates that *SmartAround* utilizes a variety of factors and algorithms to generate its recommendations, including:

1. **Proximity:** *SmartAround* prioritizes places that are close to the user's current location, utilizing the search engine's proprietary *location-based* recommendation algorithm.
2. **Popularity:** *SmartAround* recommends places that are popular among users with similar attributes (e.g., age and gender) in the area and at the current time (i.e., *collaborative filtering* algorithm).
3. **Relevance:** *SmartAround* considers the user's preferences, interests, and the type of place they are looking for, based on the user's past search history regarding restaurant choices (i.e., *content-based filtering* algorithm).

Although the recommendation algorithm of *SmartAround* remained unchanged throughout the study period, the broader system is periodically updated to improve accuracy and relevance. Here is an illustrative example of how *SmartAround* works: “A user is looking for a place to eat lunch in Seoul. S/he opens the company's mobile application and taps on the *SmartAround* icon. *SmartAround* uses the user's current location and time of day to recommend a list of nearby restaurants. The list includes a variety of restaurants, such as Korean, Japanese, Chinese, and Western. *SmartAround* also includes information about each restaurant, such as its rating, price range, and cuisine. The user can then browse the list of restaurants and choose one that they are interested in. If the user is unsure which restaurant to choose, they can read reviews from other users or tap on the restaurant's website to learn more (e.g., store and food photos, menus, and so on).”

Appendix III: Sample Composition

Table A2. Excluded Restaurants due to Entry and Exit

	Included Restaurants (%)	Excluded Restaurants (%)
Seoul		
Nonhyeon-dong	75.92	24.08
Bongcheon-dong	71.28	28.72
Seogyo-dong	77.28	22.72
Seocho-dong	77.05	22.95
Yeoksam-dong	74.65	25.35
Yeonnam-dong	71.28	28.72
Busan		
Bujeon-dong	72.70	27.30
U-dong	74.36	25.6
Incheon		
Guwol-dong	70.72	29.28
Songdo-dong	70.23	29.77
Overall Average	73.93	26.07

Notes: The rate of dropped observations (entries/exits) is stable across the different districts in our data. Although Incheon shows a slightly higher exclusion rate (approximately 29.5%), we find no systematic differences between districts within Incheon.

Appendix IV: Illustration of the Consumer Funnel

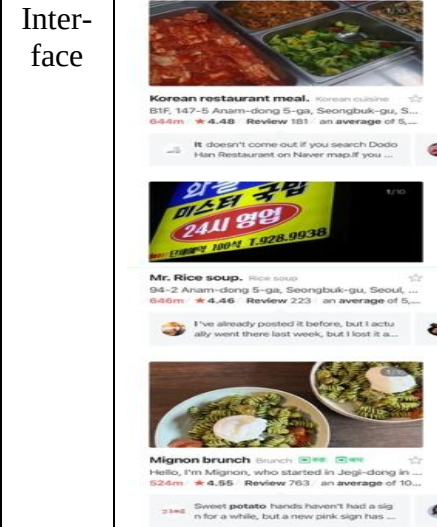
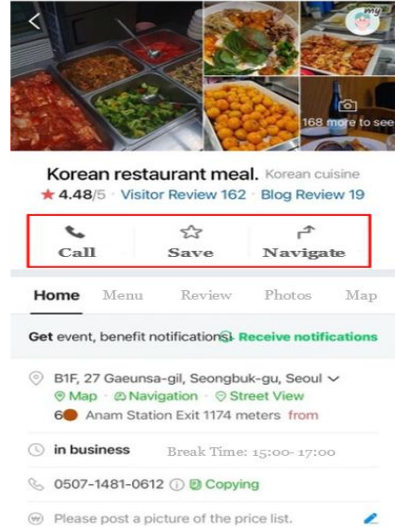
Inter- face			
Stage	Stage 1: <i>Exposure</i>	Stage 2: <i>Click</i>	Stage 3: <i>Visit</i>

Figure A1. Consumer Purchasing Funnel on SmartAround

Appendix V: Instrument Validity for the Relevance Condition

This appendix reports the analyses supporting the relevance of our instrument. Relevance requires that Competitors $\times$ Post shifts algorithmic exposure, and that this relationship is activated by the official launch rather than reflecting a pre-existing association between competitor density and exposure. Unlike the exclusion restriction, relevance is empirically testable.

We first assess whether the first-stage relationship is present before the official launch. Figure A2 plots the daily-level leads-and-lags estimates by competitor-density quartile; the weekly-level counterpart appears as Figure 4 in the main manuscript. Exposure levels do not differ systematically across quartiles before April 3, 2019, and diverge sharply thereafter, with the lowest-density quartile gaining the most.

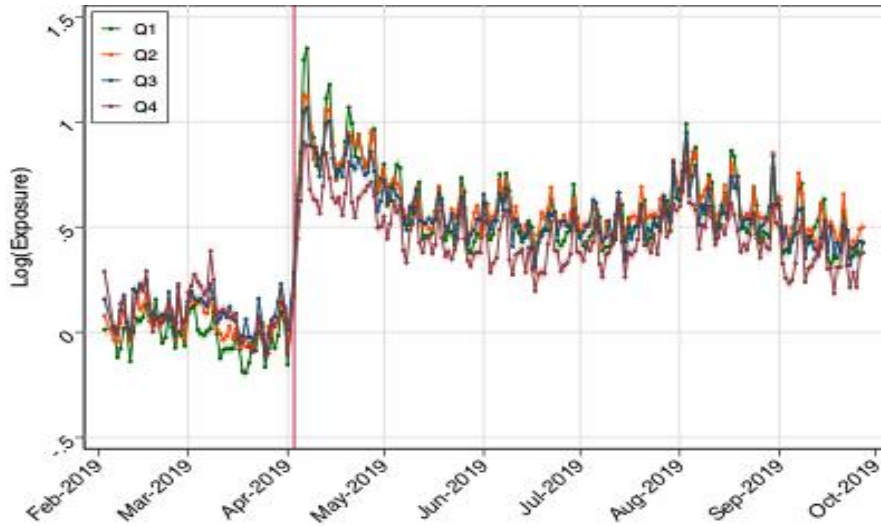


Figure A2. Trends in Exposures by the Quartile of the Number of Nearby Restaurants

Notes: The figure plots the coefficients β_{jk} 's from estimating the following leads-and-lags analysis: $Exposure_{it} = \sum_{j=\{1,2,3,4\}} \sum_{s=k-1, s \neq 1}^{k+T} \beta_{jk} \cdot 1(j^{th} \text{ Quartile}_i) 1(t = k - s) + \delta_i + \epsilon_{it}$ where $1(j^{th} \text{ Quartile}_i)$ is an indicator for whether the restaurant i is in the j -th quartile of the number of nearby restaurants, $1(t = k - s)$ is an indicator variable equal to 1 if the time period t is s days before the treatment date k (i.e., the official launch was April 3rd, 2019), and δ_i denotes restaurant fixed effects. We omit the day before the treatment (i.e., time period $k - 1$) as the baseline. The standard errors are clustered at the restaurant level and plotted as the shaded region. The weekly-level counterpart is reported as Figure 4 in the main manuscript.

Table A3 reports the underlying coefficients for all three funnel outcomes. For each of the three funnel outcomes ($\text{Log}(Exposure)$, $\text{Log}(Click)$, and $\text{Log}(Visit)$), we estimate weekly interaction coefficients between competitor-density quartile indicators (Q2, Q3, Q4 relative to

Q1) and weekly time dummies covering the eight weeks preceding the official launch of SmartAround (Weeks 5–12 of 2019; January 28 – March 24, 2019). More precisely, the

estimating equation is: $\text{Log}(\text{Exposure})_{it} = \sum_{j=\{2,3,4\}} \sum_{s=k-1, s \neq 1}^{k+T} \beta_{jk} \cdot$

$1(j^{\text{th}} \text{ Quartile}_i)1(t = k - s) + \delta_i + \delta_t + \epsilon_{it}$ where $1(j^{\text{th}} \text{ Quartile}_i)$ is an indicator for whether the restaurant i is in the j -th quartile based on the number of nearby restaurants within a 400-meter radius, $1(t = k - s)$ is an indicator variable equal to 1 if the time period t is s weeks before the treatment date k (i.e., the official launch was April 3rd, 2019), and δ_i denotes restaurant fixed effects. Q1 (lowest competitor density) is the omitted reference group.

In Table A3, Panel A reports $\text{Log}(\text{Exposure})$, Panel B $\text{Log}(\text{Click})$ and Panel C $\text{Log}(\text{Visit})$. For $\text{Log}(\text{Visit})$, pre-launch coefficients are economically negligible (all estimates fall within the range of -0.001 to $+0.001$), and none are statistically significant, providing direct empirical support for the parallel-trends assumption at the outcome level. For $\text{Log}(\text{Exposure})$ and $\text{Log}(\text{Click})$, pre-launch coefficients are predominantly small and statistically insignificant; the individual estimates that reach conventional significance are inconsistent in sign across quartiles and time, a pattern consistent with multiple testing rather than a systematic pre-trend. Table A4 supplements this with a formal test for linear differential pre-trends by competitor density.

Figure A3 reports the pseudo-launch placebo test. Restricting the sample to the pre-launch period (February 3 – April 2, 2019), we iteratively assign each date from February 11 to March 25 as a pseudo-launch date and re-estimate the first stage, yielding a distribution of 43 placebo coefficients. These cluster tightly around zero: competitor density does not systematically predict changes in exposure at arbitrary pre-launch dates. The actual first-stage coefficient using the true launch date falls well outside this distribution, in the far-left tail, confirming that the instrument's predictive power is activated by the launch event rather than reflecting a pre-existing relationship.

Table A3. Estimates of Leads-and-Lags Analysis

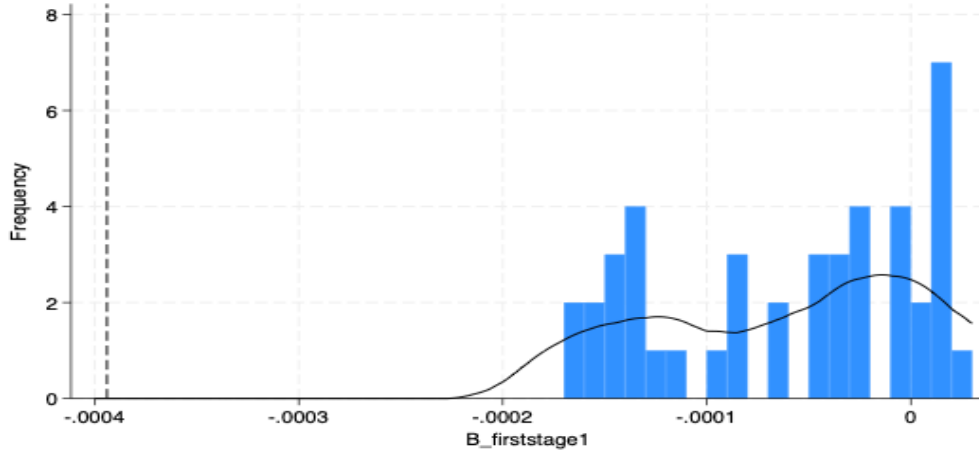
	Q2 relative to Q1	Q3 relative to Q1	Q4 relative to Q1
Panel A.		Log(<i>Exposure</i>)	
2019 Week 5 (Jan. 28 to Feb. 3)	-0.01793 (0.02715)	0.03025 (0.04228)	0.24289*** (0.06347)
2019 Week 6 (Feb. 4 to Feb. 10)	-0.03009 (0.02196)	-0.00658 (0.01541)	0.04456* (0.02103)
2019 Week 7 (Feb. 11 to Feb. 17)	-0.03402 (0.02049)	0.01419 (0.01780)	0.05143** (0.02194)
2019 Week 8 (Feb. 18 to Feb. 24)	-0.04136 (0.02526)	-0.06690*** (0.01942)	0.00007 (0.02363)
2019 Week 9 (Feb. 25 to Mar. 3)	-0.00123 (0.01883)	0.00640 (0.01902)	0.11675*** (0.02775)
2019 Week 10 (Mar. 4 to Mar. 10)	0.01522 (0.02425)	0.03200 (0.02143)	0.17846*** (0.02599)
2019 Week 11 (Mar. 11 to Mar. 17)	-0.02285 (0.01413)	0.02175* (0.01176)	0.09557*** (0.01395)
2019 Week 12 (Mar. 18 to Mar. 24)	-0.04287*** (0.01231)	-0.00614 (0.02279)	-0.00982 (0.03193)
Panel B.		Log(<i>Click</i>)	
2019 Week 5 (Jan. 28 to Feb. 3)	0.00923 (0.00658)	0.00032 (0.00275)	0.06758** (0.02820)
2019 Week 6 (Feb. 4 to Feb. 10)	-0.00708*** (0.00195)	0.02038* (0.01099)	-0.00208 (0.00410)
2019 Week 7 (Feb. 11 to Feb. 17)	-0.00320 (0.00373)	-0.01177*** (0.00309)	0.00936 (0.00531)
2019 Week 8 (Feb. 18 to Feb. 24)	-0.00264 (0.00382)	-0.00395 (0.00398)	0.00411 (0.00331)
2019 Week 9 (Feb. 25 to Mar. 3)	0.00201 (0.00690)	-0.00737 (0.00505)	0.03380** (0.01450)
2019 Week 10 (Mar. 4 to Mar. 10)	0.00910 (0.00692)	0.01071* (0.00518)	0.04484** (0.01526)
2019 Week 11 (Mar. 11 to Mar. 17)	-0.00203 (0.00362)	0.02095** (0.00914)	0.01067** (0.00461)
2019 Week 12 (Mar. 18 to Mar. 24)	-0.00196 (0.00196)	0.00127 (0.00294)	-0.00491 (0.00325)
Panel C.		Log(<i>Visit</i>)	
2019 Week 5 (Jan. 28 to Feb. 3)	0.00047 (0.00045)	0.00041 (0.00053)	0.00105 (0.00102)
2019 Week 6 (Feb. 4 to Feb. 10)	-0.00031 (0.00056)	-0.00007 (0.00055)	0.00014 (0.00087)
2019 Week 7 (Feb. 11 to Feb. 17)	0.00014 (0.00053)	-0.00043 (0.00040)	-0.00079 (0.00081)
2019 Week 8 (Feb. 18 to Feb. 24)	0.00000 (0.00046)	0.00004 (0.00042)	0.00011 (0.00068)
2019 Week 9 (Feb. 25 to Mar. 3)	-0.00043 (0.00039)	-0.00004 (0.00026)	0.00024 (0.00086)
2019 Week 10 (Mar. 4 to Mar. 10)	-0.00012 (0.00026)	0.00013 (0.00039)	0.00018 (0.00053)
2019 Week 11 (Mar. 11 to Mar. 17)	0.00036 (0.00083)	0.00030 (0.00041)	0.00084 (0.00051)
2019 Week 12 (Mar. 18 to Mar. 24)	0.00022 (0.00043)	-0.00019 (0.00032)	-0.00006 (0.00047)
Coefficients on the lags		Omitted	
Store FE		Yes	
Time FE		Yes	
Adj. R^2	Panel A. 0.8473	Panel B. 0.5412	Panel C. 0.1892
Observations		581,095	

Notes: Dependent variable is Log(*Exposure*) in Panel A, Log(*Click*) in Panel B, and Log(*Visit*) in Panel C. Each cell reports the coefficients β_{jk} 's from a leads-and-lags analysis. The estimating equation is: $Dependent\ Variable_{it} = \sum_{j=\{2,3,4\}} \sum_{s=k-1, s \neq 1}^{k+T} \beta_{jk} \cdot 1(j^{th}\ Quartile_i)1(t = k - s) + \delta_i + \delta_t + \epsilon_{it}$ where $1(j^{th}\ Quartile_i)$ is an indicator for whether the restaurant i is in the j -th quartile based on the number of nearby restaurants within a 400-meter radius, $1(t = k - s)$ is an indicator variable equal to 1 if the time period t is s weeks before the treatment date k (i.e., the official launch was April 3rd, 2019), and δ_i denotes restaurant fixed effects. Q1 (lowest competitor density) is the omitted reference group. The sample is collapsed onto the weekly level. Only the coefficients on the leads are reported and those on the lags are estimated but suppressed for brevity. All specifications include store and time fixed effects. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4. Testing Linear Pre-trends by Competitor Density

	(1) $\text{Log}(\text{Exposure})$	(2) Exposure
<i>Time</i>	-0.0013*** (0.0002)	-0.1130*** (0.0374)
<i>Time</i> $\times$ 1(2 nd Quartile)	0.0001 (0.0002)	0.0411 (0.0413)
<i>Time</i> $\times$ 1(3 rd Quartile)	-0.0001 (0.0002)	0.0363 (0.0414)
<i>Time</i> $\times$ 1(4 th Quartile)	-0.0008*** (0.0002)	0.0145 (0.0409)
<i>Control variables</i>		
<i>DAU</i>	0.0011*** (0.0000)	0.0110*** (0.0020)
<i>PM10</i>	0.0004*** (0.0000)	0.0093*** (0.0034)
<i>Temperature</i>	0.0010*** (0.0004)	0.3516*** (0.0504)
<i>Rainfall</i>	-0.0002 (0.0001)	0.0216*** (0.0071)
Store FE	Yes	Yes
Adj. R^2	0.7729	0.6495
Observations	927,607	927,607

Notes: The dependent variable is $\text{Log}(\text{Exposure})$ in Column (1) and Exposure in Column (2). The table reports the coefficients from estimating the following regression: $Y_{it} = \beta_1 \text{Time} + \beta_2 (\text{Time} \times 1(2^{\text{nd}} \text{ Quartile})) + \beta_3 (\text{Time} \times 1(3^{\text{rd}} \text{ Quartile})) + \beta_4 (\text{Time} \times 1(4^{\text{th}} \text{ Quartile})) + \delta_i + e_{it}$ where $1(j^{\text{th}} \text{ Quartile}_i)$ is an indicator for whether the restaurant i is in the j -th quartile of the number of nearby restaurants, and δ_i denotes restaurant fixed effects. The sample is restricted to the pre-period. The standard errors are clustered at the restaurant level. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Figure A3. Distribution of First-Stage Coefficients Across Pseudo-Launch Dates**

Notes: The histogram displays the distribution of 43 placebo first-stage coefficients from the pseudo-launch placebo test. The sample is restricted to the pre-launch period (February 3 – April 2, 2019). Each coefficient is obtained from a separate first-stage regression of $\text{Log}(\text{Exposure})$ on a pseudo-instrument ($\text{Competitors} \times \text{Pseudo-Post}$), where the pseudo-launch date is iteratively assigned to each date from February 11 to March 25, 2019. Dates within the first and last seven days of the pre-period are excluded to ensure sufficient observations on both sides of each pseudo-treatment date. The dashed vertical line indicates the actual first-stage coefficient (-0.000394) estimated using the true launch date of April 3, 2019 with the full sample. The solid curve represents a kernel density estimate. All regressions include store, region-by-time, and category-by-time fixed effects, and control for DAU, PM10, temperature, and rainfall. Standard errors are clustered at the restaurant level.

Appendix VI: Instrument Validity for the Exclusion Restriction

The exclusion restriction requires that $Competitors \times Post$ affects clicks and visit intentions only through algorithmic exposure. As discussed in Section 4.3, this is an identifying assumption rather than a testable property because the instrument and the most plausible confounders both vary at the neighborhood level. In our setting, the relevant neighborhood unit is the dong, a neighborhood-level administrative district in South Korea. This appendix reports the diagnostics that probe specific alternative channels.

One way to examine empirically whether the instrument provides variation independent of neighborhood confounders is to control for competitor density outside the algorithm's reach. Table A5 reports the circular-equivalent radius implied by the administrative area of each dong in our sample. Excluding Songdo-dong, an outlier at 53.4 km², the median implied radius is approximately 930 meters, indicating that restaurants within roughly one kilometer of a focal restaurant are typically located in the same dong. Our main outer-ring controls are therefore the counts of competitors in the 400–1,000 meter and 400–1,200 meter rings, each interacted with the Post indicator. These rings sit closest to the focal restaurant and thus proxy its neighborhood characteristics most directly. One caveat is that the algorithm's 400-meter cutoff is centered on the user rather than the focal restaurant, so a competitor 400–800 meters from the focal restaurant can still enter the same user's candidate set when the user stands between them. Controlling for these rings therefore absorbs both neighborhood variation and part of the algorithmic mechanism, which makes the resulting attenuation a conservative upper bound on neighborhood confounding. As a complementary check, we also construct rings beginning at 800 meters (800–1,000 and 800–1,200), beyond which no competitor can fall within 400 meters of any user who is also within 400 meters of the focal restaurant. If the first stage merely reflected a generic post-launch neighborhood shock, outer-ring density should absorb the instrument's explanatory power.

Table A5. Administrative Area and Implied Radius of Sample Neighborhoods

Dong	City/District	Area (km²)	Approx. radius (m)
Yeonnam-dong	Seoul, Mapo-gu	0.65	455
Bujeon-dong	Busan, Busanjin-gu	1.28	638
Seogyo-dong	Seoul, Mapo-gu	1.65	725
Yeoksam-dong	Seoul, Gangnam-gu	2.30	856
Nonhyeon-dong	Seoul, Gangnam-gu	2.72	930
Guwol-dong	Incheon	5.40	1,311
Seocho-dong	Seoul, Seocho-gu	6.51	1,439
Bongcheon-dong	Seoul, Gwanak-gu	8.16	1,611
U-dong	Busan, Haeundae-gu	11.03	1,874
Songdo-dong	Incheon	53.40	4,123

Notes: Dong is a neighborhood-level administrative district in South Korea. Area is obtained from official administrative records. Approximate radius is computed as $\sqrt{(\text{Area} / \pi)}$, assuming a circular shape. The sample covers 10 dongs across Seoul, Busan, and Incheon.

Table A6 shows that it does not. Column (1) reproduces the baseline IV estimate of 0.078 for reference. Controlling for the 400–1,000 meter ring (Column 3), the exposure coefficient is 0.062 ($p < 0.01$, first-stage $F = 193.92$); controlling for the 400–1,200 meter ring (Column 5), it is 0.071 ($p < 0.01$, $F = 258.82$). The attenuation of roughly 10 to 20 percent is consistent with the two density measures being spatially correlated by construction and, as noted above, with partial absorption of the algorithmic mechanism; even under the conservative reading that all of it reflects neighborhood confounding, roughly 80 percent of the baseline effect survives. The complementary rings beginning at 800 meters (Columns 7 and 9) leave the coefficient essentially unchanged at 0.077 ($F = 315.07$) and 0.078 ($F = 317.71$). The click-to-visit coefficient is unchanged at 0.053 across all specifications. We present this analysis as a targeted diagnostic consistent with, rather than proof of, the exclusion restriction.

Table A6. 2SLS IV Estimates Controlling for Outer-Ring Competitor Density

	Log(Click)	Log(Visit)	Log(Click)	Log(Visit)	Log(Click)	Log(Visit)	Log(Click)	Log(Visit)	Log(Click)	Log(Visit)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log(Exposure)	0.07756*** (0.00866)		0.06158*** (0.01067)		0.07101*** (0.00924)		0.07733*** (0.00866)		0.07797*** (0.00870)	
Log(Click)		0.05334*** (0.00197)		0.05339*** (0.00197)		0.05335*** (0.00197)		0.05333*** (0.00197)		0.05334*** (0.00197)
Post × Competitors b/t 400m and 1000m			-0.00001*** (0.00000)	0.00000 (0.00000)						
Post × Competitors b/t 400m and 1200m					-0.00000*** (0.00000)	0.00000 (0.00000)				
Post × Competitors b/t 800m and 1000m							-0.00001** (0.00000)	-0.00000 (0.00000)		
Post × Competitors b/t 800m and 1200m									-0.00000 (0.00000)	-0.00000 (0.00000)
Other Ctrl Variables: PM10, Temperature, Rainfall	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Category-by-Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistics	316.40	6586.89	193.92	6562.02	258.82	6572.12	315.07	6583.13	317.71	6586.10
Adj. R ²	0.0556	0.0282	0.0578	0.0282	0.0573	0.0282	0.0557	0.0282	0.0555	0.0282
Observations	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050

Notes: Dependent variable is Log(Click) for odd-numbered columns and Log(Visit) for even-numbered columns. The instrument is Post × Number of Nearby Restaurants, where the number of nearby restaurants is the count of other restaurants within a 400-meter radius measured during the pre-launch period. In even-numbered columns, the right-hand-side variable Log(Click) is the fitted value from the first-stage funnel regression. Columns (1) and (2) report the baseline specification. Columns (3)–(10) add outer-ring competitor counts interacted with the Post indicator as control variables, using rings of 400–1,000m, 400–1,200m, 800–1,000m, and 800–1,200m respectively. Because the algorithm's 400-meter radius is centered on the user rather than the restaurant, two restaurants can enter the same recommendation candidate set whenever they lie within 800 meters of each other. Rings beginning at 400 meters therefore retain some algorithmic overlap with the focal restaurant, whereas rings beginning at 800 meters or beyond are algorithmically disconnected by construction and absorb neighborhood-level density alone. Control variables include DAU, PM10, temperature, and rainfall; coefficients on these controls are suppressed for brevity. All specifications include store, region-by-time, and category-by-time fixed effects. *F*-statistic reports the Kleibergen-Paap rk Wald *F*-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A distinct concern for the exclusion restriction is that denser commercial areas simply attracted more platform users after the launch, generating higher clicks through a demand channel unrelated to per-restaurant exposure. If this channel were operating, competitor density should predict differential post-launch changes in local platform adoption. Table A7 regresses dong-level daily active users on the interaction of within-400-meter competitor density and the *Post* indicator, with dong and date fixed effects. Column (1) reports the

outcome in levels and Column (2) in logs. The interaction coefficient is statistically insignificant in both specifications (-0.033 , $p = 0.380$ in levels; -0.00006 , $p = 0.828$ in logs), and the log estimate implies that a one-unit increase in competitor density is associated with a differential post-launch change in adoption of well under one hundredth of a percent. We find no evidence that platform adoption trajectories diverged by competitor density after the launch.

Table A7. Testing Differential Post-Launch Changes in Dong-Level Platform Adoption

	(1) <i>DAU</i>	(2) <i>Log(DAU)</i>
<i>Instrument (Competitor density $\times$ Post)</i>	-0.03345 (0.03574)	-0.00006 (0.00028)
Dong FE	Yes	Yes
Time FE	Yes	Yes
R^2	0.7608	0.8144
Observations	1,864	1,864

Notes: The dependent variable is dong-level daily active users (*DAU*) in Column (1) and *Log(DAU)* in Column (2). The unit of observation is the dong-day. The table reports coefficients from regressing the outcome on the interaction of the number of competitors within a 400-meter radius, measured during the pre-launch period, and an indicator for the post-launch period (i.e., from April 3, 2019). The main effects of competitor density and the post-launch indicator are absorbed by the dong and date fixed effects, respectively. Standard errors are clustered at the dong level and reported in parentheses. The regression includes eight dong clusters (*DAU* is available for eight of the ten dongs); given the small number of clusters, the standard errors should be interpreted with caution, though the point estimates are economically negligible in both specifications. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As a further control for spatial heterogeneity in platform adoption, we augment the main specification with a $DAU \times Post$ interaction, allowing dong-level daily active users to have a differential effect in the post-launch period. Table A8 reports the results. The exposure coefficient is 0.072 ($p < 0.01$, $F = 311$) with the baseline specification and 0.072 ($p < 0.01$, $F = 306$) with region-category-time fixed effects, both cases essentially unchanged. These results indicate that our findings are not driven by differential post-launch demand-side adoption across neighborhoods.

While *DAU* captures platform-level activity, offline neighborhood conditions may evolve independently of app usage. We obtain dong-level daily foot-traffic data from a telecommunications provider, which measures pedestrian movement irrespective of platform engagement. These data cover only the Seoul metropolitan area, so this analysis uses the Seoul subsample ($N = 589,932$). Column (4) of Table A8 reports the results. The exposure coefficient is 0.070 ($p < 0.05$, $F = 18.27$), consistent with the full-sample baseline. Foot

traffic itself enters significantly, and its estimated coefficient is negative, indicating that, conditional on the fixed effects and controls, higher dong-level pedestrian activity is associated with fewer clicks rather than more. Crucially, its inclusion does not meaningfully alter the coefficient of interest. We note that this test covers a subset of the sample and therefore speaks to one channel in one region rather than to the exclusion restriction in general.

Table A8. 2SLS IV Estimates with DAU and Foot Traffic Controls

	(1) Baseline	(2) DAU $\times$ Post	(3) DAU $\times$ Post + Region-Category Time FE	(4) Foot Traffic
$\text{Log}(\text{Exposure})$	0.07756*** (0.00866)	0.07168*** (0.00870)	0.07167*** (0.00875)	0.06952** (0.03255)
<i>Control Variables</i>				
DAU	0.00010*** (0.00001)	0.00020*** (0.00001)	0.00020*** (0.00001)	0.00011*** (0.00001)
DAU $\times$ Post		-0.00010*** (0.00001)	-0.00010*** (0.00001)	
Foot Traffic (in 100,000's)				-0.00283*** (0.00000)
Other Control Var: PM10, Temp., Rainfall	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes
Region-Time FE	Yes	Yes		Yes
Category-Time FE	Yes	Yes		Yes
Region-Category-Time FE			Yes	
F-statistics	316.40	311.15	305.61	18.27
R ²	0.0556	0.06	0.06	0.08
Observations	3,725,050	3,725,050	3,725,050	589,932

Notes: Dependent variable is $\text{Log}(\text{Click})$. The instrument is $\text{Post} \times \text{Number of Nearby Restaurants}$, where the number of nearby restaurants is the count of other restaurants within a 400-meter radius measured during the pre-launch period. Column (1) reports the baseline specification with region-by-time and category-by-time fixed effects. Column (2) adds a $\text{DAU} \times \text{Post}$ interaction term, which allows dong-level daily active users to have a differential effect in the post-launch period. Column (3) adds the $\text{DAU} \times \text{Post}$ interaction together with region-category-time fixed effects. Column (4) retains the baseline fixed-effects structure and adds dong-level daily foot-traffic data obtained from a telecommunications service provider as an additional control; this analysis is restricted to the Seoul metropolitan area ($N = 589,932$). The coefficients on the additional controls, *DAU*, *PM10*, *temperature*, and *rainfall*, are suppressed for brevity. All specifications include store fixed effects. *F*-statistic reports the Kleibergen-Paap rk Wald *F*-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To account for other general neighborhood-level time-varying factors that may be correlated with competitor density and independently affect outcomes, we report specifications that absorb time trends at progressively finer geographic levels. Table A9 presents the results. Column (1) reports the baseline with region-by-time and category-by-time fixed effects

(0.078, $F = 316$). Column (2) replaces these with region-category-time fixed effects, absorbing cuisine-specific variation within each region; the coefficient is unchanged at 0.078 ($F = 311$). Columns (3) and (4) add dong-time and dong-category-time fixed effects. In both, the first stage weakens substantially ($F = 15.5$ and 17.7) and the exposure coefficient becomes insignificant.

Table A9. 2SLS IV Estimates with Granular Spatial-Time Fixed Effects

	(1) Baseline	(2) Region-Category Time FE	(3) Dong Time FE	(4) Dong-Category Time FE
$\text{Log}(\text{Exposure})$	0.07756*** (0.00866)	0.07762*** (0.00870)	-0.10221 (0.06779)	-0.08514 (0.06021)
Other Control Variables: DAU, PM10, Temp., Rainfall	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes
Region Time FE	Yes			
Category Time FE	Yes		Yes	
Region-Category Time FE		Yes		
Dong Time FE			Yes	
Dong-Category Time FE				Yes
F -statistics	316.40	310.87	15.50	17.68
Adj. R^2	0.0556	0.0543	-0.3176	-0.2483
Observations	3,725,050	3,725,050	3,725,048	3,725,048

Notes: Dependent variable is $\text{Log}(\text{Click})$. The instrument is $\text{Post} \times \text{Number of Nearby Restaurants}$, where the number of nearby restaurants is the count of other restaurants within a 400-meter radius measured during the pre-launch period. Column (1) reports the baseline specification with region-by-time and category-by-time fixed effects. Column (2) replaces these with region-category-time fixed effects, which absorb cuisine-specific variation within each region. Column (3) includes dong-time and category-time fixed effects, where dong refers to a neighborhood-level administrative district in South Korea. Column (4) includes dong-category-time fixed effects. In Columns (3) and (4), the first stage weakens substantially because the number of competitors within 400 meters varies primarily across dongs rather than within them, so dong-level time fixed effects absorb the cross-sectional variation that identifies the instrument. Additional control variables include DAU, PM10, temperature, and rainfall; coefficients are suppressed for brevity. All specifications include store fixed effects. F -statistic reports the Kleibergen-Paap rk Wald F -statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We report these specifications for transparency, but they do not constitute evidence against the instrument. Because the average area of a dong district in Korea is about 1.4 km², the number of competitors within 400 meters varies primarily across dongs rather than within them. Consequently, dong-level time fixed effects absorb the very cross-sectional variation that identifies the instrument. The collapse of the first stage is a mechanical consequence of this collinearity rather than a sign that neighborhood-level confounders were driving the

baseline result. It does, however, illustrate the limit of what fixed effects can establish here: the instrument and the confounders of concern operate at the same geographic level, which is why we turn to the outer-ring negative control, which absorbs neighborhood density while preserving within-400-meter variation.

Taken together, these diagnostics probe the neighborhood-level channels most likely to threaten the exclusion restriction, and none materially alters the estimated effect. They do not, however, establish exclusion. What remains assumed is that no unobserved confounder operates exclusively within the algorithm's 400-meter radius.

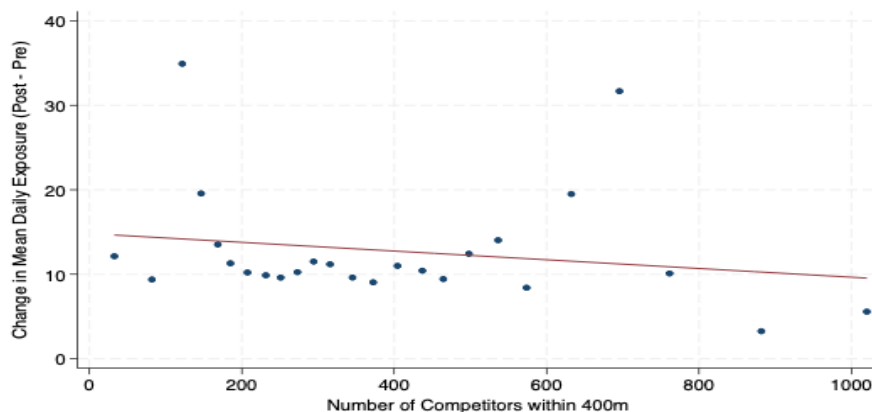
Appendix VII: Instrument Validity for Monotonicity

The interpretation of the IV estimate requires that the instrument shift exposure in the same direction for all restaurants. In our setting, this means that a larger number of competitors within 400 meters should never increase a restaurant's own exposure. The assumption is plausible in our setting. Because the algorithm draws its recommendation candidate set from a fixed 400-meter radius, additional competitors mechanically dilute each restaurant's probability of being surfaced. Monotonicity is not directly testable, but the pattern in the data is consistent with it.

Figure A4 plots the mean post-launch change in daily exposure against the number of competitors within 400 meters, using 25 equally sized bins. The relationship is clearly downward-sloping: restaurants with more competitors receive a smaller exposure boost after the launch. The decline is most pronounced above roughly 800 competitors, where the average change approaches zero. Importantly, every bin is positive, indicating that all restaurants gained exposure after the launch. What varies is the magnitude of that gain, which declines broadly in competitor density. A small number of bins at intermediate competitor counts, around 100 and 700, lie above the fitted line; we read these as idiosyncratic variation in local user traffic rather than as evidence that the instrument shifts exposure in opposite directions for different restaurants.

We emphasize that this evidence is suggestive rather than dispositive. The binned means describe the average relationship between competitor density and the exposure change, and cannot rule out that individual restaurants within a bin experienced offsetting movements. What the figure establishes is that the aggregate pattern runs in one direction across the full support of the instrument, which is what the monotonicity assumption requires at the level our design can observe.

Figure A4. Post-Launch Change in Mean Exposure by Competitor Density



Notes: The figure plots the mean post-launch change in daily exposure (the post-period average minus the pre-period average) against the number of competitors within a 400-meter radius, using 25 equally sized bins. The fitted line is from a linear regression of the binned means on competitor count. The downward slope indicates that restaurants with more nearby competitors receive a smaller exposure boost following the official launch of SmartAround on April 3, 2019, consistent with the monotonicity assumption required for the IV estimate. All 25 bins have positive mean exposure changes, and the binned mean change is broadly decreasing in competitor density, though the magnitude declines in competitor density.

Appendix VIII: Robustness to Outlier Restaurants

As documented in Figure 4 and Appendix V, restaurants in the fourth quartile of competitor density exhibit a slight differential pre-trend relative to other quartiles. We therefore re-estimate the model omitting this group. We also address the possibility that our findings are driven by outlier restaurants more generally, re-estimating on a trimmed sample that excludes restaurants in the top or bottom 5% of the distribution for age, cumulative reviews, distance from public transportation, and cumulative exposure during the beta trial. The trimming removes approximately 28.6% of the sample, leaving 11,945 restaurants.

Table A10 reports the estimation results. Column (1) reports the baseline estimation result using the full sample. Columns (2) and (3) report the IV estimates using the trimmed subsample and the subsample excluding the 4th quartile restaurants, respectively. The stability of the coefficients across these restrictive subsamples suggests that our main IV estimates are robust and not driven by observable outliers or by the pre-trend among high-density restaurants.

Table A10. 2SLS IV Estimates of the Trimmed Sample

	(1)	(2)	(3)
<i>Log(Exposure)</i>	0.07756*** (0.00866)	0.07636*** (0.00618)	0.07616*** (0.02418)
<i>Control Variables</i>			
<i>DAU</i>	0.00010*** (0.00001)	0.00007*** (0.00000)	0.00008*** (0.00001)
<i>PM10</i>	0.00008 (0.00007)	-0.00012** (0.00005)	0.00012 (0.00014)
<i>Temperature</i>	-0.00063 (0.00046)	-0.00010 (0.00040)	-0.00132** (0.00063)
<i>Rainfall</i>	-0.00009** (0.00004)	-0.00006 (0.00004)	-0.00014*** (0.00005)
Store FE	Yes	Yes	Yes
Region-Time FE	Yes	Yes	Yes
Category-Time FE	Yes	Yes	Yes
5% Trimmed		Yes	
4 th Quartile Dropped			Yes
<i>F</i> -statistics	316.40	296.54	44.96
<i>Adj. R</i> ²	0.0556	0.0365	0.0502
No. of Unique Restaurants	16,720	11,945	12,552
Observations	3,725,050	2,676,165	2,795,153

Notes: Dependent variable is *Log(Click)*. The instrument is *Post × Number of Nearby Restaurants*, where the number of nearby restaurants is the count of other restaurants within a 400-meter radius measured during the pre-launch period. Column (1) reports the baseline estimate using the full sample. Column (2) reports the estimate using a trimmed sample that excludes restaurants in the top or bottom 5% of the distribution for age, cumulative reviews, distance from public transportation, and cumulative exposure during the beta trial, reducing the sample to 11,945 unique restaurants. Column (3) excludes restaurants in the fourth quartile of competitor density (i.e., the highest number of nearby competitors), reducing the sample to 12,552 unique restaurants. Control variables include *DAU*, *PM10*, temperature, and rainfall. All specifications include store, region-by-time, and category-by-time fixed effects. *F*-statistic reports the Kleibergen-Paap rk Wald *F*-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix IX: Matching Estimates for Observable Imbalance

A potential concern with our IV strategy is that restaurants in high- and low-competitor-density areas may differ systematically in observable characteristics, and that these differences may confound the estimated effect of algorithmic exposure. To address this, we implement two matching strategies designed to improve the comparability between high- and low-density restaurants.

To implement matching, we first modify the instrument from its original continuous form ($Post \times \text{Number of Nearby Restaurants}$) to a binary form ($Post \times 1[\text{Number of Nearby Restaurants} > \text{Median}]$). This modification is necessary because matching requires a binary treatment indicator to define comparable groups of high- and low-competitor-density restaurants on which covariate balance can be enforced.

Using the binary instrument, restaurants are matched on a comprehensive set of covariates that capture both consumer-observable attributes and time-invariant restaurant characteristics: log pre-period exposure, log number of reviews, log number of photos, franchise status, and indicators for online menu availability and advertisement (URL to an external website and whether the restaurant has been featured on TV). Because these covariates are measured before the launch or are structurally time-invariant, they are not post-treatment controls and do not introduce “bad-controls” bias. Coarsened Exact Matching (CEM) reduces the sample to 14,825 unique restaurants (3,303,498 restaurant-day observations), and Propensity-Score Matching with a 1:3 ratio (PSM 1:3) reduces the sample to 14,374 restaurants (3,203,292 observations).

Table A11 reports variable-by-variable standardized mean differences (SMD) between above-median and below-median competitor-density restaurants before and after matching. Panel A and B report diagnostics for CEM and PSM, respectively. Prior to matching, several covariates exhibit imbalance above the conventional 0.1 threshold: number of reviews (SMD = 0.257 for CEM, 0.447 for PSM), number of photos (0.162, 0.241), online menu availability (0.162), and TV feature indicator (0.108). After CEM, most covariates achieve adequate balance below 0.1, though number of reviews (0.162), number of photos (0.106), and the menu indicator (0.119) remain slightly above or at the boundary. PSM achieves stronger

balance: all covariates fall below the 0.1 threshold after matching, with the largest remaining SMD being pre-launch exposure (0.063) and franchise status (0.062). The difference in SMD values before matching across the two panels reflects the different samples entering each procedure, but both methods substantially reduce imbalance across all covariates.

Table A11. Matching Diagnostics: Standardized Mean Differences

Variable	SMD Before	SMD After	Balance status
Panel A. CEM			
Pre-launch Exposure	-0.018	-0.027	Balanced (both below 0.1)
Pre-launch Click	0.078	0.037	Improved; balanced
Pre-launch Visit	0.026	-0.028	Balanced (both below 0.1)
Number of Reviews	0.257	0.162	Improved but remains above 0.1
Number of Photos	0.162	0.106	Improved but borderline (≈ 0.1)
Franchise	-0.027	-0.041	Balanced (both below 0.1)
Menu Indicator	0.162	0.119	Improved but remains above 0.1
URL Indicator	0.100	0.087	Now below 0.1
Premium Advertising Indicator	0.067	0.009	Substantially improved; balanced
TV Feature Indicator	0.108	0.038	Substantially improved; balanced
Panel B. PSM			
Pre-launch Exposure	-0.072	0.063	Balanced (both below 0.1)
Pre-launch Click	0.034	-0.029	Balanced (both below 0.1)
Pre-launch Visit	-0.027	-0.015	Balanced (both below 0.1)
Number of Reviews	0.447	0.054	Substantially improved; balanced
Number of Photos	0.241	0.014	Substantially improved; balanced
Franchise	-0.027	0.062	Balanced (both below 0.1)
Menu Indicator	0.162	0.030	Substantially improved; balanced
URL Indicator	0.100	0.014	Substantially improved; balanced
Premium Advertising Indicator	0.067	-0.027	Balanced (both below 0.1)
TV Feature Indicator	0.108	0.013	Substantially improved; balanced

Notes: Standardized mean differences (SMD) for each matching covariate, computed before and after matching. The conventional balance threshold is $|SMD| < 0.1$. Covariates that remain above 0.1 after matching are flagged in the rightmost column; these covariates exhibit strongly right-skewed distributions and are common targets of residual imbalance in restaurant samples. CEM and PSM matched sample size are 14,825 and 14,374 unique restaurants, respectively.

In Table A12, we report the IV estimates across the unmatched and matched samples. Column (1) reports the baseline estimate using the continuous instrument, and Column (2) reports the estimate using the binary instrument; the two coefficients are virtually identical (0.078 vs. 0.078), confirming that the binary reclassification does not alter the main estimate. The IV coefficient on Log(Exposure) is remarkably stable across all four specifications: 0.078 in the unmatched continuous-IV baseline (Column (1)), 0.078 with the binary IV (Column (2)), 0.078 in the CEM sample (Column (3), $F = 301.48$), and 0.089 in the PSM 1:3 sample (Column (4), $F = 267.09$). This stability provides strong evidence that the IV results

are not driven by observable imbalance between high- and low-competitor-density restaurants. The matching procedure also ensures covariate balance on the specific information that consumers observe in restaurant listings such as reviews, photos, menu availability, URLs, and advertising indicators, mitigating the concern that these consumer-observable attributes confound the relationship between algorithmic exposure and restaurant outcomes.

Table A12. 2SLS IV Estimates After Matching

	(1)	(2)	(3)	(4)
Log(<i>Exposure</i>)	0.07756*** (0.00866)	0.07750*** (0.00994)	0.07798*** (0.00848)	0.08860*** (0.00696)
Instrument Variable	Post × No. of Nearby Restaurants	Post × 1(No. of Nearby Restaurants > Median)	Post × 1(No. of Nearby Restaurants > Median)	Post × 1(No. of Nearby Restaurants > Median)
Control Variables: DAU, PM10, Temp., Rainfall	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes
Region-Time FE	Yes	Yes	Yes	Yes
Category-Time FE	Yes	Yes	Yes	Yes
Matching			CEM	PSM(1:3)
F-statistics	316.40	246.21	301.48	267.09
Adj. R ²	0.0556	0.0556	0.0472	0.0520
No. of Unique Restaurants	16,720	16,720	14,825	14,374
Observations	3,725,050	3,725,050	3,303,498	3,203,292

Notes: Dependent variable is Log(*Click*). Column (1) reports the baseline 2SLS estimate using the continuous instrument, *Post* × *Number of Nearby Restaurants*, where the number of nearby restaurants is the count of other restaurants within a 400-meter radius measured during the pre-launch period. Column (2) replaces the continuous instrument with a binary instrument, *Post* × 1(*Number of Nearby Restaurants* > *Median*), where the median is computed across all restaurants in the pre-launch period. Columns (3) and (4) apply Coarsened Exact Matching (CEM) and Propensity-Score Matching with a 1:3 ratio (PSM 1:3), respectively, using the binary instrument. Matching covariates include log pre-period exposure, log number of reviews, log number of photos, franchise status, and indicators for online menu availability and advertisement (URL to an external website and whether the restaurant is broadcast on TV). Control variables include DAU, PM10, temperature, and rainfall. The coefficients are suppressed for brevity. All specifications include store, region-by-time, and category-by-time fixed effects. *F*-statistic reports the Kleibergen-Paap rk Wald *F*-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix X: Robustness to Pre-Launch (Beta Trial) Spillover

Prior to the official launch of SmartAround on April 3, 2019, a beta version of the service was available to a limited set of early-adopter users. A potential concern is that beta-period usage may have generated spillover effects that complicate the interpretation of the post-launch estimates. If early usage was concentrated in particular neighborhoods, it could have increased baseline exposure, shaped consumer awareness, or induced restaurant-side responses, such that the pre-launch window may not represent a clean no-treatment baseline. We address this concern through descriptive evidence and three targeted robustness checks.

First, we document the exposure to potential spillover effects. In Table A13, we report the distribution of average daily exposure during the pre-launch period (February 3 – April 2, 2019) across all 16,720 restaurants. The distribution is highly right-skewed (skewness = 45.6). The median restaurant received only 1.6 daily exposures, and approximately 78% of restaurants received 10 or fewer. A small set of outlier restaurants in high-traffic locations received several thousand daily exposures. Figure A5 plots the empirical cumulative distribution function (CDF) of average daily pre-launch exposure across all restaurants. The distribution rises sharply near zero, with over 90% of restaurants receiving fewer than 50 daily exposures, and then flattens into a long right tail. This pattern confirms that beta-period exposure was minimal for the vast majority of restaurants and concentrated among a small subset of establishments in high-traffic locations.

Table A13. Summary Statistics of Pre-Period Exposure

Statistics	Value
Percentile	
1%	0
5%	0.018
10%	0.070
25%	0.317
50% (median)	1.614
75%	8.351
90%	23.860
95%	37.570
99%	83.0
Maximum	4,693.088
Mean	9.800
Std. dev.	75.777
Variance	5,742.13
Skewness	45.647
Kurtosis	2,297.638
No. of Restaurants	16,720

Notes: Unit of observation is the restaurant. *Exposure* is measured as the average daily number of algorithmic recommendations received by each restaurant during the pre-launch period (February 3 – April 2, 2019).

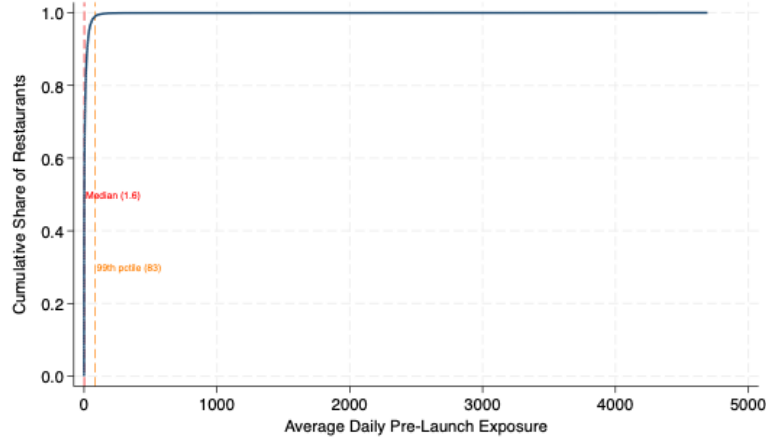


Figure A5. Cumulative Distribution of Pre-Period Exposure

Notes: The figure plots the empirical cumulative distribution function (CDF) of average daily exposure across all 16,720 restaurants during the pre-launch period (February 3 – April 2, 2019). The dashed red line indicates the median (1.6 daily exposures) and the dashed orange line indicates the 99th percentile (83 daily exposures). *Exposure* is measured as the average daily number of algorithmic recommendations received by each restaurant during the pre-launch period.

In Table A14, we further explore whether restaurant characteristics differ by the level of pre-launch exposure. Time-varying local conditions (DAU, temperature, rainfall, and PM10) are virtually identical across quartiles, as expected given that these variables vary at the day-level and are orthogonal to restaurant-level exposure assignment. However, observable differences are indeed present. Restaurants with higher beta-period exposure tend to be older, have more reviews and photos, and are more likely to have registered menu information and external URLs on the platform. These patterns indicate that more established and informationally richer restaurants attracted greater user attention even during the low-activity beta period. Although all of these cross-sectional differences are time-invariant and are fully absorbed by our store fixed effects, we next turn to examine whether these pre-period differences affect our main findings.

Table A14. Characteristics by Pre-Period Exposure Quartile

Variable	Q1 (Lowest)	Q2	Q3	Q4 (Highest)	Overall
<i>Avg. daily exposure</i>	0.115	0.823	4.060	34.276	9.800
(sd)	(0.410)	(1.627)	(4.407)	(25.774)	(8.082)
<i>Avg. daily clicks</i>	0.001	0.007	0.023	0.278	0.077
(sd)	(0.009)	(0.044)	(0.114)	(0.563)	(0.182)
<i>Avg. daily visits</i>	0.000	0.000	0.001	0.006	0.002
(sd)	(0.000)	(0.001)	(0.005)	(0.038)	(0.011)
<i>Avg. daily sales</i> (1,000 (KRW))	500.088	625.374	824.846	1,348.537	849.628

(sd)	(523.193)	(599.317)	(693.547)	(863.989)	(681.035)
DAU	78.810	81.282	78.184	78.673	79.242
Temperature	5.598	5.315	5.144	5.144	5.300
Rainfall	1.251	1.258	1.246	1.241	1.249
PM10	62.110	62.457	61.958	62.119	62.162
Restaurant age (days)	944.812	1,072.144	1,250.209	1,311.048	1,144.208
Number of reviews	10.908	30.256	68.339	218.467	81.861
Franchise (share)	0.129	0.225	0.271	0.234	0.215
Walking distance (m)	441.667	462.800	493.154	472.828	467.566
Public cafeteria (share)	0.499	0.584	0.530	0.427	0.510
Region: Seoul (share)	0.700	0.705	0.645	0.651	0.675
Region: Busan (share)	0.190	0.146	0.146	0.146	0.157
Region: Incheon (share)	0.111	0.149	0.209	0.202	0.168
Number of photos	4.443	5.534	7.309	10.045	7.202
Menu registered (Menu Indicator)	0.222	0.354	0.457	0.602	0.409
External URL (URL Indicator)	0.159	0.265	0.352	0.434	0.302
Premium Advertising	0.008	0.018	0.032	0.057	0.029
Featured on TV	0.005	0.019	0.036	0.120	0.045

Notes: Summary statistics for restaurant characteristics during the pre-launch period (February 3 – April 2, 2019), stratified by quartile of average daily beta-period exposure. Means are reported; standard deviations are reported in parentheses for the four time-varying outcomes (*exposure*, *clicks*, *visits*, *sales*). Restaurants with higher beta-period exposure tend to be older, have substantially more reviews, more photos, and are more likely to have registered menu information and URLs on the platform. All cross-sectional differences are time-invariant and are fully absorbed by store fixed effects in our main IV regressions. Time-varying local conditions (*DAU*, *temperature*, *rainfall*, *PM10*) are virtually identical across quartiles.

Second, we re-estimate our main specification after progressively excluding restaurants with the highest pre-treatment exposure. Table A15 reports results for both the exposure-to-click and click-to-visit stages. Columns (1)–(2) exclude restaurants in the top 1% of the pre-treatment exposure distribution (average daily exposure above 83.0), Columns (3)–(4) exclude the top 5% (above 37.57), and Columns (5)–(6) exclude the top 10% (above 23.86). The exposure-to-click coefficient ranges from 0.078 when excluding the top 1% to 0.053 when excluding the top 10%, while the click-to-visit coefficient remains remarkably stable between 0.050 and 0.052 across all thresholds. The attenuation in the exposure coefficient at higher exclusion thresholds is consistent with reduced instrument variation in the remaining sample. However, the persistence of statistical significance across all thresholds indicates that the main findings are not driven by heavily exposed restaurants.

Table A15. 2SLS IV Estimates after Excluding High-Exposure Restaurants

	(1) Log(<i>Click</i>)	(2) Log(<i>Visit</i>)	(3) Log(<i>Click</i>)	(4) Log(<i>Visit</i>)	(5) Log(<i>Click</i>)	(6) Log(<i>Visit</i>)
Log(<i>Exposure</i>)	0.07822*** (0.00748)		0.07380*** (0.00617)		0.05268*** (0.00549)	
Log(<i>Click</i>)		0.05140*** (0.00199)		0.05172*** (0.00229)		0.05025*** (0.00271)
Other Control Var: <i>DAU</i> , <i>PM10</i> , <i>Temp.</i> , <i>Rainfall</i>	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Category-Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Drop Criteria	Top 1%	Top 1%	Top 5%	Top 5%	Top 10%	Top 10%
F-statistic	323.08	8678.83	309.68	7865.35	287.39	6473.80
Adj. R ²	0.05	0.03	0.04	0.03	0.05	0.03
Observations	3,687,638	3,687,638	3,537,898	3,537,898	3,349,896	3,349,896

Notes: Dependent variable is Log(*Click*) for odd-numbered columns and Log(*Visit*) for even-numbered columns. Columns (1)–(2) exclude restaurants with average daily pre-launch exposure in the top 1% (above 83.0); Columns (3)–(4) in the top 5% (above 37.57); and Columns (5)–(6) in the top 10% (above 23.86). For exposure-to-click stages (odd columns), the instrument is *Competitors* $\times$ *Post*. For click-to-visit stages (even columns), the right-hand-side variable is fitted log clicks from the exposure-to-click stage. Additional control variables include *DAU*, *PM10*, *temperature*, and *rainfall*; coefficients are suppressed for brevity. All specifications include store, region-by-time, and category-by-time fixed effects. F-statistic reports the Kleibergen-Paap rk Wald F-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Third, we address the concern that pre-launch spillover may contaminate the baseline by dropping the weeks immediately preceding the official launch. Figure A6 plots total daily exposure across all restaurants over the sample period. The pre-launch period exhibits a stable, low level of aggregate exposure with no discernible upward trend in the weeks leading up to the official launch. A sharp increase in exposure occurs only after April 3, 2019 (indicated by the vertical line), consistent with the launch activating the recommendation system at scale. Table A16 reports the regression results after dropping the pre-launch weeks. Columns (1)–(2) drop the two weeks prior to launch (March 20 – April 2, 2019) and Columns (3)–(4) drop the four weeks prior to launch (March 6 – April 2, 2019). The exposure-to-click coefficient is 0.080 ($F = 332.22$) when dropping two weeks and 0.082 ($F = 313.47$) when dropping four weeks; the click-to-visit coefficient is 0.053 and 0.056, respectively. All estimates are essentially identical to the full-sample baseline, confirming that our results are not sensitive to the inclusion of the weeks closest to the official launch.



Figure A6. Trend in Total Daily Exposure

Notes: The vertical line indicates the official launch date (April 3, 2019). *Exposure* is measured as the total number of algorithmic recommendations across all 16,720 restaurants per day.

Table A16. 2SLS IV Estimates after Dropping Pre-Launch Weeks

	(1)	(2)	(3)	(4)
	Log(<i>Click</i>)	Log(<i>Visit</i>)	Log(<i>Click</i>)	Log(<i>Visit</i>)
Log(<i>Exposure</i>)	0.08049*** (0.00865)		0.08249*** (0.00912)	
Log(<i>Click</i>)		0.05324*** (0.00199)		0.05564*** (0.00216)
Other Control Var.: DAU, PM10, Temp., Rainfall	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes
Region-Time FE	Yes	Yes	Yes	Yes
Category-Time FE	Yes	Yes	Yes	Yes
Instrument	Yes	Yes	Yes	Yes
Drop <i>N</i> Weeks Before Official Launch	2 weeks	2 weeks	4 weeks	4 weeks
<i>F</i> -statistic	332.22	6,507.39	313.47	6,360.75
Adj. <i>R</i> ²	0.05	0.03	0.05	0.03
Observations	3,608,765	3,608,765	3,264,312	3,264,312

Notes: Dependent variable is Log(*Click*) for Columns (1) and (3) and Log(*Visit*) for Columns (2) and (4). Columns (1)–(2) drop the two weeks (March 20 – April 2, 2019) immediately preceding the official launch. Columns (3)–(4) drop the four weeks (March 6 – April 2, 2019) immediately preceding the official launch. The instrument is *Competitors* × *Post* for exposure-to-click stages; the click-to-visit stages use fitted log clicks from the exposure-to-click stage. Additional control variables include DAU, PM10, temperature, and rainfall; coefficients are suppressed for brevity. All specifications include store, region-by-time, and category-by-time fixed effects. *F*-statistic reports the Kleibergen-Paap rk Wald *F*-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, we include pre-exposure-by-time fixed effects to absorb any group-specific temporal shocks that may differ between high- and low-exposure restaurants. We classify restaurants by whether their average pre-treatment exposure falls above or below the sample median and include separate time fixed effects for each group. Table A17 reports the results. Columns (1)–(2) report the baseline specification for reference. With the pre-exposure-by-time fixed effects (Columns (3)–(4)), the exposure-to-click coefficient is 0.068 ($p < 0.01$, $F = 281.04$) and the click-to-visit coefficient is 0.051 ($p < 0.01$, $F = 6,237.56$), both closely aligned with the baseline. The stability of these estimates confirms that differential time trends between high- and low-beta-exposure restaurants do not drive our results.

Table A17. 2SLS IV Estimates with Pre-Exposure Specific Time Fixed Effects

	(1) Log(<i>Click</i>) Baseline	(2) Log(<i>Visit</i>) Baseline	(3) Log(<i>Click</i>) Pre-Exp $\times$ Time FE	(4) Log(<i>Visit</i>) Pre-Exp $\times$ Time FE
Log(<i>Exposure</i>)	0.07756*** (0.00866)		0.06812*** (0.00925)	
Log(<i>Click</i>)		0.05334*** (0.00197)		0.05118*** (0.00204)
Other Control Var.:				
DAU, PM10, Temp., Rainfall	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes
Region-Time FE	Yes	Yes	Yes	Yes
Category-Time FE	Yes	Yes	Yes	Yes
Pre-Exposure Time FE			Yes	Yes
Instrument	Yes	Yes	Yes	Yes
<i>F</i> -statistic	316.40	6,586.89	281.04	6,237.56
<i>Adj. R</i> ²	0.06	0.03	0.05	0.03
Observations	3,725,050	3,725,050	3,725,050	3,725,050

Notes: Dependent variable is Log(*Click*) for Columns (1) and (3) and Log(*Visit*) for Columns (2) and (4). Columns (1)–(2) report the baseline specification. Columns (3)–(4) include separate time fixed effects for restaurants above and below the median of average pre-treatment exposure. The instrument is *Competitors* $\times$ *Post* for exposure-to-click stages; the click-to-visit stages use fitted log clicks from the exposure-to-click stage. Additional control variables include DAU, PM10, temperature, and rainfall; coefficients are suppressed for brevity. All specifications include store, region-by-time, and category-by-time fixed effects. *F*-statistic reports the Kleibergen-Paap rk Wald F-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In summary, we do not claim that the beta period was perfectly clean. However, our additional analyses consistently indicate that beta-period exposure was not the main driver of our original finding.

Appendix XI: Robustness to Alternative Functional Forms

Because clicks and visits are sparse at the daily restaurant level, we verify that our findings are not sensitive to the functional form treatment of zero-heavy outcome variables. Table A18 reports estimates from three alternative specifications alongside the baseline log-log results. Columns (1) and (2) reproduce the baseline log-log IV estimates for reference. Columns (3) and (4) apply the inverse hyperbolic sine (asinh) transformation to both the dependent and independent variables. The asinh transformation behaves like the natural logarithm for large values but is defined at zero, providing a natural robustness check for sparse outcomes. The exposure-to-click coefficient is 0.087 ($p < 0.01$, $F = 317.65$) and the click-to-visit coefficient is 0.047 ($p < 0.01$, $F = 6,084.26$), both closely aligned with the baseline estimates. Columns (5) and (6) report Poisson pseudo-maximum likelihood (PPML) estimates using a control function approach. Because PPML is a nonlinear estimator, we adopt the control function adaptation for count-data settings: the first-stage residuals from the exposure equation are included as additional regressors, with the instrument entering only the first stage. The exposure-to-click coefficient is 1.780 ($p < 0.01$) and the click-to-visit coefficient is 5.782 ($p < 0.01$). These coefficients are not directly comparable in magnitude to the log-log and asinh estimates because PPML models $\log(E[y])$ rather than $E[\log(y)]$; the relevant comparison is the sign, statistical significance, and directional consistency, all of which are preserved. The significance of the first-stage residuals in both equations (-0.914 , $p < 0.05$ and -2.868 , $p < 0.01$) confirms the presence of endogeneity and validates the control function correction. Columns (7) and (8) aggregate outcomes to the weekly restaurant level, which smooths daily zeros and substantially reduces sparsity. The exposure-to-click coefficient is 0.105 ($p < 0.01$, $F = 319.58$) and the click-to-visit coefficient is 0.051 ($p < 0.01$, $F = 4,389.80$), again consistent with the baseline.

Across all alternative specifications, the pattern of statistical significance and directional consistency is preserved, confirming that our main findings are not an artifact of daily-level zeros in the outcome variables.

Table A18. Regression Estimates Using Various Functional Forms

	Log(<i>Click</i>)	Log(<i>Visit</i>)	Asinh(<i>Click</i>)	Asinh(<i>Visit</i>)	<i>Click</i> ; PPML	<i>Visit</i> ; PPML	Log(<i>Click</i>) ; Weekly	Log(<i>Visit</i>) Weekly
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(<i>Exposure</i>)	0.07756*** (0.00866)				1.78003*** (0.37600)		0.10535*** (0.01023)	
Log(<i>Click</i>)		0.05334*** (0.00197)				5.78236*** (0.28960)		0.05112*** (0.00323)
Asinh(<i>Exposure</i>)			0.08651*** (0.00980)					
Asinh(<i>Click</i>)				0.04686*** (0.00191)				
First-stage Residuals					-0.9144** (0.37946)	-2.8675*** (0.30148)		
Other Control Var.: DAU, PM10, Temp., Rainfall	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Category-by-Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Control Function	No	No	No	No	Yes	Yes	No	No
<i>F</i> -statistic	316.40	6586.89	317.65	6084.26			319.58	4389.80
Adj. <i>R</i> ²	0.0556	0.0282	0.0433	0.0415			0.1617	0.0645
Observations	3,725,050	3,725,050	3,725,050	3,235,836	3,087,236	819,109	569,053	497,495

Notes: Columns (1) and (2) report baseline 2SLS IV estimates using the log-log specification. The dependent variable is Log(*Click*) in Column (1) and Log(*Visit*) in Column (2). The instrument is *Post* $\times$ *Number of Nearby Restaurants*, where the number of nearby restaurants is the count of other restaurants within a 400-meter radius measured during the pre-launch period. In Column (2), the right-hand-side variable Log(*Click*) is the fitted value from the first-stage funnel regression. Columns (3) and (4) apply the inverse hyperbolic sine (asinh) transformation to both the dependent and independent variables. Columns (5) and (6) report Poisson pseudo-maximum likelihood (PPML) estimates using a control function approach, with first-stage residuals included as additional regressors; the dependent variables are clicks and visits in levels. PPML coefficients are not directly comparable in magnitude to the log-log and asinh estimates because PPML models log(*E*[*y*]) rather than *E*[log(*y*)]. Columns (7) and (8) aggregate outcomes to the weekly level. Control variables include DAU, PM10, temperature, and rainfall. All specifications include store, region-by-time, and category-by-time fixed effects. *F*-statistic reports the Kleibergen-Paap *rk* Wald *F*-statistic. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix XII: Heterogeneous Treatment Effects by Accessibility over Time

We investigate whether the impact of the MLBRS evolves over time by splitting our post-launch sample into two periods: a short-term window (Period 1: April 3, 2019 – June 30, 2019) and a long-term window (Period 2: July 1, 2019 – September 27, 2019). We re-estimate our heterogeneous IV model using interactions with these time period indicators.

Table A19 presents the estimation results. In the exposure-to-click stage (Columns (1)–(3)), we find that although the main effect of exposure on clicks attenuates slightly over time (dropping from 0.062 in Period 1 to 0.044 in Period 2 in Column (3)), the heterogeneous benefit for low-accessibility restaurants remains remarkably stable. The interaction coefficient ($\text{Log}(\text{Exposure}) \times 1(\text{LowAccess})$) is 0.012 in both the short and long term. This suggests that although the baseline exposure-to-click effect attenuates over time for accessible restaurants, the *structural* benefit of the MLBRS (its ability to redistribute attention to geographically disadvantaged businesses) is persistent and not a transient effect.

In the click-to-visit stage (Columns (4)–(6)), the conversion elasticity from clicks to visits is stronger in the short term (0.070) compared to the long term (0.035). However, the interaction term for low-accessibility restaurants remains statistically insignificant across both periods. This persistence confirms that the incremental gain in clicks generated by the MLBRS for geographically disadvantaged businesses does not translate into a significant incremental increase in visit intentions, reinforcing our finding in Subsection 5.3 that algorithmic exposure alone is insufficient to drive conversion in the absence of digital information and reputation signals.

Table A19. 2SLS IV Estimates by Accessibility over Time

	Log(Click)			Log(Visit)		
	(1)	(2)	(3)	(4)	(5)	(6)
Period 1:						
<i>April 3 to June 30, 2019</i>						
Log(Exposure)	0.04320***	0.04784***	0.06172***			
× Period 1	(0.00411)	(0.00694)	(0.00771)			
Log(Exposure)	0.00971***	0.01033***	0.01232***			
× Period 1 × 1(LowAccess)	(0.00127)	(0.00190)	(0.00202)			
Period 2:						
<i>July 1 to September 27, 2019</i>						
Log(Exposure)	0.02776***	0.02749***	0.04394***			
× Period 2	(0.00444)	(0.00767)	(0.00808)			
Log(Exposure)	0.01017***	0.00955***	0.01183***			
× Period 2 × 1(LowAccess)	(0.00142)	(0.00212)	(0.00216)			
Period 1:						
<i>April 3 to June 30, 2019</i>						
Log(Click)				0.07027***	0.06974***	0.06964***
× Period 1				(0.00228)	(0.00226)	(0.00228)
Log(Click)				-0.00157	0.00052	0.00096
× Period 1 × 1(LowAccess)				(0.00440)	(0.00485)	(0.00485)
Period 2:						
<i>July 1 to September 27, 2019</i>						
Log(Click)				0.03468***	0.03467***	0.03452***
× Period 2				(0.00255)	(0.00254)	(0.00251)
Log(Click)				0.00219	0.00250	0.00283
× Period 2 × 1(LowAccess)				(0.00354)	(0.00372)	(0.00371)
Control Variables						
DAU	0.00012***	0.00011***	0.00010***	0.00000**	0.00000**	0.00000**
	(0.00000)	(0.00001)	(0.00001)	(0.00000)	(0.00000)	(0.00000)
PM10	0.00006**	0.00023**	0.00006	-0.00000	-0.00000	-0.00000
	(0.00002)	(0.00011)	(0.00010)	(0.00000)	(0.00001)	(0.00001)
Temperature	-0.00090***	0.00010	0.00064	-0.00009***	-0.00026***	-0.00024**
	(0.00020)	(0.00057)	(0.00060)	(0.00003)	(0.00010)	(0.00010)
Rainfall	0.00001	-0.00006	-0.00006*	-0.00001	0.00001	0.00001
	(0.00001)	(0.00004)	(0.00004)	(0.00000)	(0.00001)	(0.00001)
Store FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes			Yes		
Region-by-Time FE		Yes	Yes		Yes	Yes
Category-by-Time FE			Yes			Yes
Instrument	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic	77.20	27.30	27.79	1588.59	1286.26	1344.61
Adj. R ²	0.0413	0.0393	0.0382	0.0316	0.0315	0.0308
Observations	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050	3,725,050

Notes: Dependent variable is *Log(Click)* for Columns (1)–(3) and *Log(Visit)* for Columns (4)–(6). *Log(Click)* and *Log(Visit)* denote the logged values of *Click* and *Visit*. *1(LowAccess)* denotes restaurants with low accessibility. A low-accessibility restaurant is defined as one located more than 30 meters away from the nearest main road and more than 400 meters away from the nearest subway station or bus stop. Control variables include *DAU*, *PM10*, *Temperature*, and *Rainfall*. For Columns (1)–(3), we use $Competitors_j \times 1(Post_t)$ as an instrument for $Log(Exposure)$ where $Competitors_j$ denotes the count of other restaurants within a 400-meter radius of the restaurant j , and $1(Post_t)$ is an indicator variable for the time periods after the official launch of *SmartAround* (i.e., from April 3 to September 27, 2019). This IV is interacted with relevant binary indicators for the other interaction terms. For Columns (4)–(6), fitted log clicks derived from the prior exposure-to-click stage are used as the right-hand-side variable in the conversion-stage equation. These fitted log clicks are interacted with the relevant binary indicators for the interaction terms. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix XIII: Heterogeneous Treatment Effects by Digital Information and Reputation Signals: High-Accessibility Subsample

In Subsection 5.3, we present evidence consistent with our proposed mechanism that digital information and reputation signals serve to mitigate high travel frictions. We replicate our heterogeneity analysis in Subsection 5.3 using the subsample of high-accessibility restaurants (Subsection 5.3 used the subsample of low-accessibility restaurants). If our proposed mechanism holds, we should observe that the marginal benefit of these signals is attenuated for restaurants that are already geographically accessible by car or public transit. Because these establishments face significantly lower physical barriers to visit, consumers should be less reliant on these signals to justify a visit compared to their low-accessibility counterparts. Consistent with our expectations, although stronger signals still yield a positive effect for accessible restaurants, the magnitude of this effect is consistently smaller than what we observed for low-accessibility restaurants.

Table A20. 2SLS IV Estimates for High-Accessibility Restaurants by Digital Information and Reputation Signals

	Log(Click)				Log(Visit)			
	Digital Information and Reputation Signals				Digital Information and Reputation Signals			
	Number of Reviews	Number of Photos	Availability of Menu	Ad. (TV/SNS)	Number of Reviews	Number of Photos	Availability of Menu	Ad. (TV/SNS)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(Exposure)	0.07300*** (0.01128)	0.07068*** (0.01152)	0.06496*** (0.01255)	0.06664*** (0.01209)				
Log(Exposure)	0.02865*** (0.00691)	0.02798*** (0.00552)	0.03011*** (0.00605)	0.03016*** (0.00587)				
× Stronger Signal								
Log(Click)					0.02626*** (0.00580)	0.04875*** (0.00245)	0.04248*** (0.00287)	0.04221*** (0.00178)
Log(Click)					0.02940*** (0.00599)	0.00670* (0.00346)	0.01530*** (0.00356)	0.01893*** (0.00331)
× Stronger Signal								
<i>Control Variables</i>								
DAU	0.00008*** (0.00001)	0.00008*** (0.00001)	0.00008*** (0.00001)	0.00008*** (0.00001)	0.00000 (0.00000)	0.00000 (0.00000)	0.00000 (0.00000)	0.00000 (0.00000)
PM10	0.00007 (0.00007)	0.00013 (0.00008)	0.00013* (0.00008)	0.00013 (0.00008)	-0.00003** (0.00001)	-0.00003* (0.00001)	-0.00003* (0.00001)	-0.00003* (0.00001)
Temperature	-0.00017 (0.00054)	-0.00034 (0.00055)	-0.00039 (0.00056)	-0.00041 (0.00056)	-0.00004 (0.00013)	-0.00003 (0.00013)	-0.00003 (0.00013)	-0.00003 (0.00013)
Rainfall	-0.00015*** (0.00005)	-0.00014*** (0.00005)	-0.00013*** (0.00005)	-0.00014*** (0.00005)	0.00002 (0.00001)	0.00002 (0.00001)	0.00002 (0.00001)	0.00002 (0.00001)
Store FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE								
Region-by-Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Category-by-Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic	257.38	120.31	120.84	113.96	978.22	1668.63	1110.83	2100.09
Adj. R ²	0.0559	0.0571	0.0578	0.0581	0.0274	0.0271	0.0273	0.0274
Observations	2530069	2530069	2530069	2530069	2530069	2530069	2530069	2530069

Notes: Dependent variable is $\text{Log}(\text{Click})$ for Columns (1)–(4) and $\text{Log}(\text{Visit})$ for Columns (5)–(8). $\text{Log}(\text{Click})$ and $\text{Log}(\text{Visit})$ denote the logged values of *Click* and *Visit*. The sample is restricted to high-accessibility restaurants, defined as restaurants located within 30 meters of the nearest main road or within 400 meters of the nearest subway station or bus stop. Stronger Signal is a binary indicator defined from the signal measure specific to each column. As noted in Section 4.2 of the paper, review and photo counts are recorded as single values as of the last day of data collection; the indicators therefore reflect signal strength over the observation window rather than a strictly pre-launch snapshot. For the number of reviews, the high-quality signal indicator equals 1 if the restaurant has an accumulated number of reviews above the sample median. For the number of photos, the quality indicator equals 1 if the restaurant has a number of store photos above the sample median. For the availability of the menu, the indicator equals 1 if a digital menu is available. For the presence of advertisement, the indicator equals 1 if the restaurant features a verified store description, TV appearance, or linked social media account. Detailed definitions of *Exposure*, *Click*, *Visit* and control variables (*DAU*, *PM10*, *Temperature*, and *Rainfall*) can be found in the data description section of the paper. For Columns (1)–(4), we use $\text{Competitors}_j \times 1(\text{Post}_t)$ as an instrument where Competitors_j denotes the count of other restaurants within a 400-meter radius of the restaurant j , and $1(\text{Post}_t)$ is an indicator variable for the time periods after the official launch of *SmartAround* (i.e., from April 3, 2019 to September 27, 2019). For Columns (5)–(8), fitted log clicks derived from the prior exposure-to-click stage are used as the right-hand-side variable in the conversion-stage equation. These fitted log clicks are interacted with the corresponding high-quality signal indicators. Standard errors are clustered at the restaurant level and reported in parentheses. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the exposure-to-click stage (Columns (1)–(4)), the interaction coefficients for high accessibility restaurants range from 0.028 to 0.030. Comparing this to the low accessibility results (where coefficients ranged from 0.035 to 0.044), we find that the attention boost from these signals is approximately 20–30% weaker for accessible restaurants. For instance, the interaction effect of *Advertisements* on clicks drops from 0.044 (*Low Accessibility*) to 0.030 (*High Accessibility*). This suggests that when a restaurant is accessible, consumers are less reliant on the digital information and reputation signals before making a click.

In the click-to-visit stage (Columns (5)–(8)), the pattern persists but is more nuanced. The interaction effect of *Menus* on visits drops from 0.020 (*Low Accessibility*) to 0.015 (*High Accessibility*), supporting the idea that detailed product information is less critical when the cost of walking over to check the menu in person is low. Interestingly, the effect of *Reviews* remains robust across both groups, consistent with user-generated social proof being associated with conversion regardless of location.

Appendix XIV: Robustness to Stage Dependency

The sequential nature of consumer decisions along the purchase funnel, from exposure to click to visit, implies that unobserved factors at earlier stages may propagate to later stages. To account for this stage dependence, we implement a control function specification that explicitly allows for correlated unobservables across funnel stages.

Formally, the estimator is a recursive triangular system. Letting ε_1 , ε_2 , and ε_3 denote the error terms in the exposure, click, and visit equations:

$$\begin{aligned} \text{Exposure} &= f(Z, X, FE) + \varepsilon_1 \\ \text{Click} &= g(\text{Exposure}, X, FE) + \gamma_1 \varepsilon_1 + \varepsilon_2 \\ \text{Visit} &= h(\text{Click}, X, FE) + \delta_1 \varepsilon_1 + \delta_2 \varepsilon_2 + \varepsilon_3 \end{aligned}$$

where Z denotes the instrument, X the controls, and FE the fixed effects. The control function residuals proxy for ε_1 and ε_2 , and the coefficients γ_1 , δ_1 , and δ_2 capture the cross-stage dependence.

In the first stage, we regress $\text{Log}(\text{Exposure})$ on our instrument ($\text{Competitors} \times \text{Post}$) and obtain the residual $\hat{v}_1$, which captures unobserved factors that influence exposure but are orthogonal to the instrument. We include $\hat{v}_1$ as a control in the click equation, where we regress $\text{Log}(\text{Click})$ on $\text{Log}(\text{Exposure})$. From this equation, we obtain the predicted $\text{Log}(\text{Click})$, representing the component of clicking behavior explained by exogenous variation in exposure generated via the official launch of the MLBRS. To obtain $\hat{v}_2$, we regress the observed $\text{Log}(\text{Click})$ on this predicted $\text{Log}(\text{Click})$ and take the residual. The second residual should then capture the component of clicking behavior unexplained by the instrumented exposure pathway. This could reflect unobserved factors such as idiosyncratic browsing behavior, listing salience, or time-varying restaurant appeal that independently influence both clicks and visits (i.e., endogeneity in the click-to-visit equation). In the click-to-visit equation, we include both $\hat{v}_1$ and $\hat{v}_2$ as controls. The inclusion of $\hat{v}_1$ allows unobserved factors originating at the exposure stage to propagate through to visit decisions, capturing the cross-stage (or stage-dependent) correlation in unobservables. The inclusion of $\hat{v}_2$ corrects for the endogeneity of clicks in the visit equation by absorbing unobserved click-stage factors that would otherwise bias the estimated effect of clicks on visits.

Identification of the click-to-visit effect requires that the instrument's variation reach visits only through the preceding funnel stages. The reduced-form estimate supports this: the instrument does not directly predict visit intentions (coefficient = $3.05\text{e-}07$, $p = 0.548$), indicating no detectable reduced-form association between competitor density and visits, consistent with the maintained sequential exclusion assumption.

Table A21 presents the results. Columns (1) and (2) report the baseline IV estimates for reference, whereas Columns (3)–(5) report the control function specification. In the click equation (Column (3)), the coefficient on $\text{Log}(Exposure)$ is 0.078 ($p < 0.01$), identical to the baseline, while $\hat{v}_1$ is statistically insignificant, suggesting that residual endogeneity in exposure is not a major concern for click outcomes. In the visit equation (Column (4)), the coefficient on $\text{Log}(Click)$ is 0.057 ($p < 0.01$), closely aligned with the baseline estimate of 0.053. Crucially, $\hat{v}_1$ is statistically significant (0.004, $p < 0.01$), providing direct evidence of cross-stage correlated unobservables. The control function absorbs these correlated unobservables, yet the stability of the $\text{Log}(Click)$ coefficient relative to the baseline suggests that stage dependence does not materially alter the main finding. The second residual $\hat{v}_2$ is also statistically significant (-0.068 , $p < 0.01$), confirming the presence of endogeneity at the click stage and validating the need for instrumentation in the click-to-visit equation.

To further address the selection concern (that consumers can only visit if they first click), Column (5) re-estimates the visit equation restricting the sample to restaurant-day observations with positive clicks. The sample reduces from 3,235,836 to 256,266 observations. The coefficient on $\text{Log}(Click)$ increases to 0.127 ($p < 0.01$), consistent with a stronger intensive-margin effect among restaurants that have entered the click stage. Both control function residuals remain significant ($\hat{v}_1 = 0.041$, $p < 0.05$; $\hat{v}_2 = -0.619$, $p < 0.05$), confirming that cross-stage dependence and click-level endogeneity are present and accounted for even conditional on selection into the click stage.

Because the procedure generates regressors at each stage, conventional standard errors do not account for the estimation uncertainty introduced by generated regressors. We therefore report bootstrap standard errors, resampling restaurant clusters and re-estimating the full sequential procedure: first-stage residual, click equation, second residual, and visit equation in each of 100 replications. The bootstrap standard errors are close to the conventional clustered errors: 0.0016 versus 0.0017 for $\text{Log}(Click)$, 0.0012 versus 0.0011 for $\hat{v}_1$, and

0.0185 versus 0.0167 for $\hat{v}_2$. All coefficients remain highly significant, indicating that generated-regressor uncertainty does not affect inference.

Table A21. 2SLS IV and Control Function Regression Estimates

	(1) Log(<i>Click</i>)	(2) Log(<i>Visit</i>)	(3) Log(<i>Click</i>)	(4) Log(<i>Visit</i>)	(5) Log(<i>Visit</i> <i>Clicks</i> >0)
Log(<i>Exposure</i>)	0.07756*** (0.00866)		0.07756*** (0.00881)		
Log(<i>Click</i>)		0.05334*** (0.00197) [0.00188]		0.05694*** (0.00170) [0.00159]	0.12721*** (0.00994)
$\hat{v}_1$ (1 st stage residuals)			-0.01291 (0.00887)	0.00417*** (0.00108) [0.00119]	0.04077** (0.01736)
$\hat{v}_2$ (2 nd stage residuals)				-0.06792*** (0.01665) [0.01845]	-0.61947** (0.26623)
Control Variables:					
DAU, PM10, Temp., Rainfall	Yes	Yes	Yes	Yes	Yes
Store FE	Yes	Yes	Yes	Yes	Yes
Region-by-Time FE	Yes	Yes	Yes	Yes	Yes
Category-by-Time FE	Yes	Yes	Yes	Yes	Yes
Instrument	Yes	Yes	No	No	No
Control Function	No	No	Yes	Yes	Yes
F-statistic	316.40	6586.89			
Adj. R ²	0.0556	0.0282	0.3123	0.0561	0.0562
Observations	3,725,050	3,725,050	3,725,050	3,235,836	256,266

Notes: Columns (1) and (2) report baseline 2SLS IV estimates for the exposure-to-click and click-to-visit stages, respectively. The instrument is *Post* $\times$ *Number of Nearby Restaurants*, where the number of nearby restaurants is the count of other restaurants within a 400-meter radius measured during the pre-launch period. In Column (2), the right-hand-side variable Log(*Click*) is the fitted value from the first-stage funnel regression. Columns (3)–(5) report control function estimates. In Column (3), we regress Log(*Click*) on Log(*Exposure*) and the first-stage residual $\hat{v}_1$, obtained from regressing Log(*Exposure*) on the instrument. In Column (4), we regress Log(*Visit*) on Log(*Click*), $\hat{v}_1$, and the second-stage residual $\hat{v}_2$, obtained from regressing observed Log(*Click*) on predicted Log(*Click*). Column (5) replicates the specification in Column (4) but restricts the sample to restaurant-day observations with positive clicks, directly modeling the intensive margin of the conversion stage. The significance of $\hat{v}_1$ in Columns (4) and (5) indicates the presence of cross-stage correlated unobservables; the significance of $\hat{v}_2$ confirms the endogeneity of clicks in the visit equation. Control variables include DAU, PM10, temperature, and rainfall; coefficients are suppressed for brevity. All specifications include store, region-by-time, and category-by-time fixed effects. F-statistic reports the Kleibergen-Paap rk Wald F-statistic for the IV specifications in Columns (1) and (2). Standard errors clustered at the restaurant level are reported in parentheses. Because the sequential and control function procedures generate regressors at each stage, we also report bootstrap standard errors in brackets for the coefficients in Columns (2) and (4), computed by resampling restaurant clusters and re-estimating the full procedure in each of 100 replications. The bootstrap and clustered standard errors are closely comparable, and inference is unchanged. Asterisks denote statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

- Adamopoulos P, Ghose A, Tuzhilin A (2022) Heterogeneous demand effects of recommendation strategies in a mobile application: Evidence from econometric models and machine-learning instruments. *MIS Q.* 46(1):101–150.
- Chang CC, Chin YC (2010) The impact of recommendation sources on online purchase intentions: The moderating effects of gender and perceived risk. *World Acad. Sci. Eng. Technol.* 66(6):111–114.
- Cooke AD, Sujan H, Sujan M, Weitz BA (2002) Marketing the unfamiliar: The role of context and item-specific information in electronic agent recommendations. *J. Mark. Res.* 39(4):488–497.
- De P, Hu Y, Rahman MS (2010) Technology usage and online sales: An empirical study. *Manag. Sci.* 56(11):1930–1945.
- Lee D, Gopal A, Park SH (2020) Different but equal? A field experiment on the impact of recommendation systems on mobile and personal computer channels in retail. *Inf. Syst. Res.* 31(3):892–912.
- Lee D, Hosanagar K (2019) How do recommender systems affect sales diversity? a cross-category investigation via a randomized field experiment. *Inf. Syst. Res.* 30(1):239–259.
- Lee D, Hosanagar K (2021) How do product attributes and reviews moderate the impact of recommender systems through purchase stages? *Manag. Sci.* 67(1):524–546.
- Li X, Grahl J, Hinz O (2022) How do recommender systems lead to consumer purchases? a causal mediation analysis of a field experiment. *Inf. Syst. Res.* 33(2):620–637.
- Oestreicher-Singer G, Sundararajan A (2012) Recommendation networks and the long tail of electronic commerce. *MIS Q.* 36(1):65–83.
- Peng J, Liang C (2023) On the differences between view-based and purchase-based recommender systems. *MIS Q.* 47(2):875–900.
- Senecal S, Nantel J (2004) The influence of online product recommendations on consumers' online choices. *J. Retail.* 80(2):159–169.
- Wan X, Kumar A, Li X (2024) Retargeted vs. generic product recommendations: when is it valuable to present retargeted recommendations? *Inf. Syst. Res.* 35(3):1403–1421.
- Yuan Z, Chen AJY, Wang Y, Sun T (2025) How recommendation affects customer search: A field experiment. *Inf. Syst. Res.* 36(1):84–106.
- Zheng S, Tong S, Kwon HE, Burtch G, Li X (2025) Frontiers: Recommending what to search: sales volume and consumption diversity effects of a query recommender system. *Mark. Sci.* 44(3):516–524.
- Zhou M, Zhang J, Adomavicius G (2024) Longitudinal impact of preference biases on recommender systems' performance. *Inf. Syst. Res.* 35(4):1634–1656.