Online Appendix for Delivery in the City: Differentiated Products Competition among New York Restaurants

Jacob Cosman* Tianran Dai† Nathan Schiff‡

September 2022

A Empirical Methods and Match Quality Testing

A.1 Selection model and identification strategy

We start with the following reduced form model of restaurant outcomes, analogous to Section 3.2:

$$Y_{rt} = \beta_r * D_{rt} + u_r + u_{L_r} + \xi_{rt} + \xi_{L_r t} + \varepsilon_{rt} \tag{A1}$$

Following the potential outcomes framework, let Y_{rt}^1 be the outcome of a restaurant at time t when there is entry (treatment) and Y_{rt}^0 represent the outcome when there is not entry (control). From Equation A1, these terms and the switching equation may be expressed as follows:

$$\begin{aligned} Y_{rt}^0 &= u_r + u_{L_r} + \xi_{rt} + \xi_{L_r t} + \varepsilon_{rt} \\ Y_{rt}^1 &= \beta_r \mathbb{I}\{t \geq k_r\} + Y_{rt}^0 \\ Y_{rt} &= D_{rt} * Y_{rt}^1 + (1 - D_{rt}) * Y_{rt}^0 \end{aligned} \tag{A2}$$

We want to estimate the effect of new competition on incumbent restaurants, the average treatment effect on the treated (ATT), β:

$$ATT = E[Y_{rt}^1 - Y_{rt}^0 | D_{rt} = 1] = E[\beta_r | D_{rt} = 1] = \beta \tag{A3}$$

We do not observe what restaurants that faced new competition would have counterfactually done in the absence of this competition ($Y_{rt}^0 | D_{rt} = 1$). Further, it is highly likely that factors determining restaurant outcomes also affect entry. To model entry we assume that a new competitor enters near restaurant r at time t if expected profit (modeled as a latent variable) is positive.[1]

$$D_{rt} = \mathbb{I}\{\theta_r + \theta_{L_r} + \psi_{rt} + \psi_{L_r t} \geq 0\} \tag{A4}$$

Equation A4 shows that the entry process may also be a function of characteristics of incumbent restaurant r and location L_r, both time-varying ($\psi_{rt}, \psi_{L_r t}$) and invariant (θ_r, θ_{L_r}). As discussed in Section 3.2, we address

*Zillow

†School of Economics, Shanghai University of Finance and Economics

‡School of Economics, Shanghai University of Finance and Economics. Corresponding author: nschiff@gmail.com. Schiff acknowledges support from National Science Foundation of China grant 71950410628, "Competition and Integration: The Economic Geography of Markets."

[1] In equation A4 we are treating entry as a process independent of the characteristics of the entrant. We address entrant characteristics with the analysis in Table 5 and the entry location analysis in Section D.1.

the endogeneity of entry with a difference-in-difference matching strategy. Given potential entry in period k, define the difference in an outcome d periods before entry and d periods after as $\Delta Y_{rk} = Y_{r,k+d} - Y_{r,k-d}$. Then we can estimate β from this difference:

$$ATT = E[\Delta Y^1_{rk} - \Delta Y^0_{rk} | \Delta D_{rk} = 1] = E[\beta_r | \Delta D_{rk} = 1] = \beta \tag{A5}$$

This differencing removes any correlation between the time-invariant terms in the outcome equation and the selection equation.[2] Entry and outcomes could still both be influenced by the time-varying terms ξ and ψ and therefore we use matching to mitigate this form of selection bias. We use a two-stage process to match treated restaurants with control restaurants using both characteristics of the incumbent restaurant's location $X(L_r)$ and the restaurant's menu text M_r. Letting $\hat{P}(X(L))$ denote the predicted intensity of entrants at location L_r, our identifying assumption is conditional mean independence:

$$E[\Delta Y^0_{rk} | \hat{P}(X(L)), M_r, \Delta D_{rk} = 1] = E[\Delta Y^0_{rk} | \hat{P}(X(L)), M_r, \Delta D_{rk} = 0]$$

A.2 Cosine distance: details and implementation

We can compare menu M to menu M' by comparing their ngram weights on the set of J ngrams, where J is the superset of ngrams from both menus for some pre-chosen ngram size (we use a size of 3). If a menu has count m_i occurrences of ngram i then the weight x_i of this ngram is:

$$x_i = \frac{m_i}{\sum_{j=1}^{J} m_j} \tag{A6}$$

Damashek defines the "cosine similarity" $S_{M,M'}$ of two documents (menus) M and M' as the cosine of the angle between their ngram vectors (with elements denoted by x_{Mj} and $x_{M'j}$):

$$S(M, M') = \frac{\sum_{j=1}^{J} x_{Mj} x_{M'j}}{\left(\sum_{j=1}^{J} x^2_{Mj} \sum_{j=1}^{J} x^2_{M'j} \right)^{1/2}} \tag{A7}$$

In Damashek (1995) the author uses his method to assign documents to languages (e.g. "French") and topic areas for news articles in a given language (e.g. "mining"). He finds that Equation A7 performs well for language assignment but has worse performance for topic assignment. He suggests that this is because the ngram vectors of two articles written in the same language will have a great deal of similarity simply due to common and uninformative ngrams in the language or general group to which the documents belong. For example, in English the 3-gram "the" is common but uninformative about topic. To deal with this issue he suggests centering all ngram vectors by subtracting a common vector that captures the ngram distribution of some specific language or subject group. Letting μ represent this common vector of weights the "centered cosine similarity" is:

$$S^c(M, M') = \frac{\sum_{j=1}^{J} (x_{Mj} - \mu_j)(x_{M'j} - \mu_j)}{\left(\sum_{j=1}^{J} (x_{Mj} - \mu_j)^2 \sum_{j=1}^{J} (x_{M'j} - \mu_j)^2 \right)^{1/2}} \tag{A8}$$

[2] In Equation A5 note that $\Delta Y^1_{rk} = Y^1_{r,k+d} - Y^1_{r,k-d} = \beta_r + Y^0_{r,k+d} - Y^0_{r,k-d} = \beta_r + \Delta Y^0_{kt}$

Table A1: Most common n-grams in sample with frequency of occurrence.

_sa	_ch	chi	ed_	and
206624	197278	183113	176519	160072
ick	cke	en_	hic	ken
153950	148003	147005	145687	143927
_wi	th_	ith	wit	sal
123583	113200	111242	111117	105591
ala	nd_	_an	san	lad
96385	88437	83429	79267	78750
ich	_ro	che	_co	ice
76252	75512	73962	73711	73369

In our context, we wish to subtract out the common distribution of restaurant menu ngrams and so we define the vector μ as simply the vector of ngram centroids across all restaurants $r \in R$. As described in Section 3.2, we want to capture a pre-treatment measure of the menu distance between two restaurants. Therefore we use the first observed menu for every restaurant. For the majority of restaurants this is the first period of our data but varies for later entrants.[3] If we weight each menu equally then the centroid for ngram j is:

$$\mu_j = \frac{1}{|R|} \sum_{r \in R} x_{M_r j} \tag{A9}$$

Note that when a menu M has no occurrences of ngram i that ngram receives zero weight, $x_{Mi} = 0$, but this weight of zero still enters the calculation of S^c. Finally, as mentioned earlier, we convert this measure to a distance by subtracting it from 1, yielding our formula for menu distance:

$$\omega\left(M, M'\right) = 1 - \frac{\sum_{j=1}^{J} (x_{Mj} - \mu_j)(x_{M'j} - \mu_j)}{\left(\sum_{j=1}^{J} (x_{Mj} - \mu_j)^2 \sum_{j=1}^{J} (x_{M'j} - \mu_j)^2\right)^{1/2}} = 1 - S^c\left(M, M'\right) \tag{A10}$$

In calculating this measure we use only the names of menu items and exclude the item descriptions (which are often missing). We calculate the menu distance between the initial menu of every restaurant in our sample, yielding a symmetric matrix of pairwise distances between all restaurants.

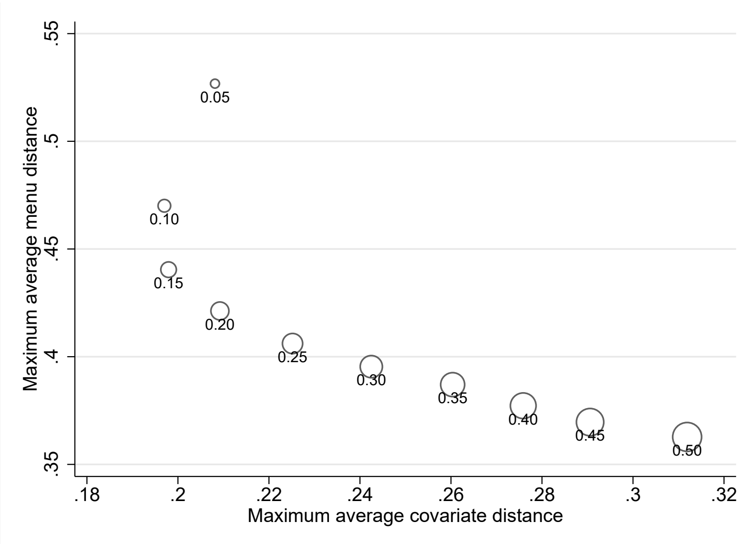
Our sample includes 23620 n-grams. Of these, 10454 appear in the sample at least ten times. Table A1 shows the most common n-grams; as shown, these include the n-grams comprising the words “chicken”, “salad”, and “sandwich”.

A.3 Choice of predicted entrant bandwidth

The two-stage calliper matching process described in the text requires us to choose a bandwidth for the callipers. This bandwidth determines the range of predicted entrant counts in which we search for the closest control observation match by menu distance. Bandwidth selection involves a tradeoff: a small bandwidth ensures a closer match on predicted entrant count in the first stage whereas a wider bandwidth improves the

[3] As discussed below, this is a very large set of n-grams. Therefore, choosing different periods or combining periods is unlikely to have any qualitative effect on our measure. There are a few ngrams that show up in later menus which are missing from our μ vector. We assign these ngrams a μ value of zero.

Figure A1: Menu distance vs covariate distance



Comparison of menu distance (cosine distance) between treated and control pairs with standardized distance between Poisson regression variables. Diameter of circle is proportional to count of matches within callipers (first stage); calliper sizes α are listed under each circle.

prospects of finding a close menu match in the second stage. Crucially, a wider bandwidth also increases the final sample size of matched treated and control pairs.

We explore possible bandwidths through a process that allows us to investigate this tradeoff:

1. We divide observations into quintiles of predicted entrant count $q \in \{1,2,3,4,5\}$.

2. For each observation i in quintile q we find the observation in quintile $-q \neq q$ with the smallest menu distance to observation i. Then, we take the average across each quintile q. We denote the maximum of this average across all quitiles as the "maximum average menu distance".

3. For each observation i in quintile q we select a random observation j from a quintile $-q \neq q$. For each covariate in the Poisson regressions we take the average of the standardized distance between the covariate value for observations i and j. We average this measure across all covariates within a quintile and then take the maximum of this average across all quintiles as the "maximum average covariate distance".

Figure A1 shows the resulting menu distances and propensity covariate distances for a bandwidth of α standard deviations in the log of the predicted entrant count for $\alpha \in \{0.05, 0.10, .., 0.50\}$. Based on these results, we select a bandwidth of 0.25 standard deviations of predicted entrant count for the two-stage calliper matching procedure.

A.4 Trimming the entrant count

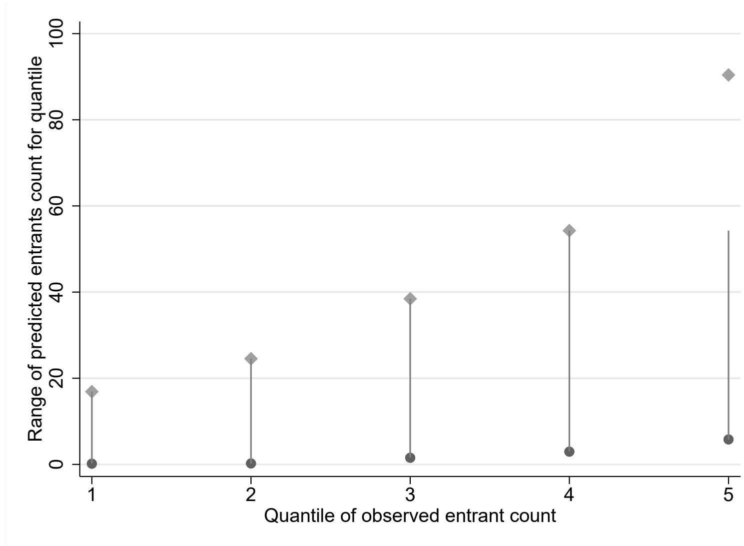
When matching observations with similar predicted entrant counts, we trim observations with very high or very low predicted entrant counts. In a simpler model with a binary treatment variable Crump, Hotz, Imbens and Mitnik (2009) demonstrate that this approach improves the precision of the estimate by ensuring overlap in propensity covariate distributions. Specifically, we only include observations with a predicted entrant

Table A2: Poisson regression coefficients

	Coefficient	Std. err		Coefficient	Std. err
Competitors within 25 m	-0.0081	0.003	Spanish and English	-0.0624	0.013
Competitors within 50 m	-0.0188	0.004	Other IE, limited English	0.002	0.006
Competitors within 100 m	-0.0203	0.005	Other IE, English	-0.0132	0.004
Competitors within 250 m	0.0239	0.009	AP, limited English	0.069	0.006
Competitors within 500 m	0.8017	0.015	AP, English	0.0281	0.006
Competitors within 1 km	0.3726	0.017	Poverty	0.0219	0.006
Competitors within 2.5 km	0.0103	0.014	Income <10k	-0.0146	0.005
Efficiency rent	1.6579	0.371	Income 10k-20k	-0.0207	0.006
One-bedroom rent	-1.7832	0.574	Income 20k-30k	-0.0091	0.005
Two-bedroom rent	6.7964	1.295	Income 30k-40k	-0.0186	0.005
Three-bedroom rent	8.2158	0.685	Income 40k-50k	-0.0369	0.004
Four-bedroom rent	-14.8412	0.82	Income 50k-60k	-0.013	0.004
Age <10	-0.0119	0.012	Income 60k-75k	-0.0105	0.004
10 ≤ Age ≤	0.0618	0.013	Income 75k-100k	0.0125	0.004
18 ≤ Age ≤ 24	0.0454	0.011	Income 100k-150k	-0.0433	0.004
25 ≤ Age ≤ 29	0.1123	0.011	Income 150k-200k	-0.0388	0.005
30 ≤ Age ≤ 39	0.067	0.011	Owner-occupied	0.0877	0.006
40 ≤ Age ≤ 49	0.0288	0.007	Detached house	0.0498	0.011
50 ≤ Age ≤ 59	-0.0119	0.007	3-9 unit structure	0.1651	0.011
60 ≤ Age ≤ 64	0.0538	0.005	10-49 unit structure	0.1533	0.01
65 ≤ Age ≤ 69	0.039	0.006	>50 unit structure	0.0456	0.014
70 ≤ Age ≤ 79	-0.0111	0.007	Built post-2010	-0.0239	0.003
White	-0.0374	0.016	Built 2000-2009	0.0075	0.003
Black	-0.0072	0.013	Built 1990-1999	0.046	0.004
Asian	0.0495	0.012	Rent 1250-1499	0.0237	0.006
Latino	-0.1	0.017	Rent 1500-1999	-0.1682	0.004
Commute out of county	0.0699	0.006	Rent 2000+	0.0176	0.004
Commute out of state	-0.0173	0.003	Rent-to-income 35-40%	-0.0193	0.003
Family household	0.1173	0.017	Rent-to-income 40-50%	0.0209	0.003
Married household	-0.1389	0.012	Rent-to-income 50%+	-0.002	0.005
Roommate household	-0.0328	0.006	House value 500k-750k	0.0111	0.003
Enrolled in college	0.0273	0.014	House value 750k-1m	0.043	0.003
Not enrolled in school	-0.0082	0.018	House value 1m+	-0.0041	0.003
College graduate	-0.0453	0.013	Distance to subway	0.0116	0.011
Spanish, limited English	0.0086	0.009	Constant	2.3709	0.004
Observations	11198				
Adj. Pseudo-R2	0.7485				

Coefficients from poisson model for the number of entrants within 500m. Model uses a LASSO penalty estimator with penalty parameter of 2.5 to select regression variables. Unit of observation is a restaurant.

Figure A2: Range of predicted entrant counts for the five quintiles of observed entrant counts.



count in the *common support* of the quintiles of the observed entrant count. We calculate this common support as follows:

1. Divide the sample into five quintiles according to the observed entrant count at each observation.
2. Calculate the common support for each of the five quintile subsamples in a manner analogous to (Flores, Flores-Lagunes, Gonzalez and Neumann 2012). Let q denote the set of observations in a given quintile subsample. Then, the common support CS_q for quintile subsample q is as follows:
$$CS_q = \left[\max \left\{ \min_{i \in q} \left\{ P(X_{j(i)}) \right\}, \min_{i \notin q} \left\{ P(X_{j(i)}) \right\} \right\}, \min \left\{ \max_{i \in q} \left\{ P(X_{j(i)}) \right\}, \max_{i \notin q} \left\{ P(X_{j(i)}) \right\} \right\} \right] \quad \text{(A11)}$$
3. Find the common support for the overall sample as the union of the common supports of the five quintile subsamples CS_q .

Figure A2 shows the range of predicted entrant counts for each quintile of the distribution of observed entrant counts. Qualitatively, the common support of the sample is the range of predicted entrant counts which lie in at least two quintiles of the observed entrant count. Trimming the sample to only include observations within this common support ensures that we only match treated observations which could potentially be matched to a control observation in another quintile of the observed entrant distribution.

A.5 Testing match quality

In our results section we present spatial competition results for three durations ($d = 4,6,8$) using two different dimensions to define space (geographic and product), as well as results on exit likelihood. Rather than showing separate balance tables for all of these analyses (7 tables), we instead present more general results showing the sample balance for matched restaurants across the distribution of the count of nearby entrants during the sample period.[4] These results demonstrate that treated and control restaurants (which by construction have different entrant counts over the defined duration) are balanced on observables for different

[4] All of our tests on balance and our matched sample regression results use an inner radius of $\rho_T = 500$ meters to define entry near a given restaurant as discussed in Section 3.1. Separate post-match balance tables for any particular analysis and duration are available upon request.

durations. To do so, we group restaurants into quintiles of observed entrant count. Then, we compare the covariates for observations in a specific quintile to observations in all other quintiles before and after matching. Since we use a two-stage matching process, we first show the balance improvement from matching on entrant intensity and then show the additional effect of using menu distance relative to matching on entrant intensity alone.

Testing entrant intensity balance: We follow the general procedure of Hirano and Imbens (2004) by grouping restaurants into quintiles of observed entrant counts; for example, the first quintile consists of locations that have two or fewer nearby entrants. We wish to compare the average value of each location covariate for locations with up to two entrants (quintile 1) to locations with more than two entrants (quintiles 2-5). As recommended in Imbens (2015), we compare covariates using normalized differences. Our approach proceeds as follows:

1. Divide restaurants into quintiles according to the number of nearby entrants over the sample period. Let R_q be the set of restaurants in quintile q and let R_{-q} be the set of restaurants not in R_q.

2. For each quintile q for each restaurant $r \in R_q$ define a candidate set $C(r)$. This is the intersection of R_{-q} and the set of observations lying within the propensity calliper of r — i.e., the observations with a log predicted entrant count within 0.25 standard deviations of the log predicted entry count for r.

3. For each quintile q for each observation $r \in R_q$ randomly sample (with replacement) one thousand observations from $C(r)$. Index these bootstrap draws by b. For each r denote the corresponding bootstrap observation by $s^b(r)$.

4. For each bootstrap iteration b for each locational variable $X_j(L) \in X(L)$ calculate the following absolute normalized difference across all restaurants r and their randomly-selected matches $s^b(r)$:

$$v_{qj}^b = \frac{\left|\text{mean}_{r\in R_q}\left(X_j(L_r)\right) - \text{mean}_{r\in R_q}\left(X_j(L_{s^b(r)})\right)\right|}{\frac{1}{2}\sqrt{\text{var}_{r\in R_q}\left(X_j(L_r)\right) + \text{var}_{r\in R_q}\left(X_j(L_{s^b(r)})\right)}} \tag{A12}$$

5. Take the average over values of v_{qj}^b across all bootstrap iterations b.

Table A3 compares the resulting normalized differences to the normalized differences obtained without callipers — that is, by randomly sampling from R_{-q} rather than $C(r)$ in step 3. Imbens (2015) suggests 0.2 as a reasonable threshold for the normalized difference. With the callipers nearly all covariates fall below this level. Although the normalized distances are generally lower than in the pre-callipers sample, some age brackets and housing characteristics still differ across quintiles.

Testing menu distance balance: In the second stage of the matching process, we match each treated restaurant to the within-calliper control restaurant with a menu at the smallest menu distance. As discussed in Section 3.2, this is intended to produce matched pairs of treated and control observations which would have a similar response to competition.

In order to measure the similarity of the matched pairs, we compare the normalized differences between menu attributes of the treated and control matched pairs with the normalized differences between menu attributes of a counterfactual set of treated and control pairs. We generate this counterfactual set by randomly selecting a control restaurant within the propensity callipers for each treated restaurant. The comparison isolates the improvement in menu similarity using the nearest-neighbor menu distance match from that

Table A3: Entrant intensity covariate balance

Variable		Q1	Q2	Q3	Q4	Q5
Competitors within 100 m	Without callipers	1.78	1.15	0.30	1.03	1.48
	With callipers	0.19	0.05	0.09	0.18	0.04
Competitors within 500 m	Without callipers	2.47	1.84	0.74	1.43	3.40
	With callipers	0.51	0.11	0.08	0.06	0.27
Competitors within 1 km	Without callipers	2.38	1.76	0.86	1.45	3.40
	With callipers	0.28	0.02	0.26	0.05	0.07
One-bedroom rent	Without callipers	1.65	1.41	0.19	1.59	1.40
	With callipers	0.12	0.04	0.20	0.02	0.08
Two-bedroom rent	Without callipers	1.65	1.41	0.19	1.59	1.41
	With callipers	0.12	0.04	0.20	0.02	0.08
White	Without callipers	1.08	0.86	0.35	1.31	0.58
	With callipers	0.04	0.02	0.07	0.07	0.16
Black	Without callipers	0.71	0.62	0.04	0.96	0.85
	With callipers	0.05	0.12	0.03	0.29	0.48
Asian	Without callipers	0.29	0.11	0.41	0.13	0.88
	With callipers	0.20	0.05	0.15	0.05	0.07
Latino	Without callipers	0.83	0.82	0.17	0.98	1.34
	With callipers	0.05	0.06	0.11	0.07	0.03
Family household	Without callipers	2.60	1.16	0.10	1.39	2.38
	With callipers	0.28	0.13	0.09	0.13	0.21
Married household	Without callipers	1.28	0.49	0.17	0.56	1.68
	With callipers	0.03	0.08	0.05	0.25	0.35
Enrolled in college	Without callipers	0.47	0.27	0.32	0.21	1.03
	With callipers	0.09	0.05	0.02	0.30	0.46
College graduate	Without callipers	2.39	1.24	0.11	1.72	1.70
	With callipers	0.25	0.09	0.06	0.03	0.03
Poverty	Without callipers	0.78	0.77	0.02	1.21	0.54
	With callipers	0.02	0.11	0.11	0.19	0.31
Income 75k-100k	Without callipers	0.34	0.05	0.03	0.26	0.16
	With callipers	0.04	0.03	0.02	0.08	0.14
Income 100k-150k	Without callipers	0.42	0.66	0.03	0.67	0.47
	With callipers	0.12	0.05	0.03	0.13	0.19
Income 150k-200k	Without callipers	0.93	0.93	0.11	0.74	1.13
	With callipers	0.04	0.02	0.08	0.08	0.03
Detached house	Without callipers	1.25	0.11	0.66	0.77	0.53
	With callipers	0.18	0.06	0.17	0.30	0.54
3-9 unit structure	Without callipers	0.19	0.39	0.84	0.29	1.10
	With callipers	0.02	0.14	0.20	0.13	0.29
$>$ 50 unit structure	Without callipers	1.09	0.52	0.37	0.94	0.88
	With callipers	0.33	0.16	0.27	0.05	0.03
Built post-2010	Without callipers	0.18	0.14	0.14	0.06	0.28
	With callipers	0.03	0.06	0.19	0.14	0.11
Rent 2000+	Without callipers	1.12	0.85	0.18	0.87	0.82
	With callipers	0.19	0.03	0.06	0.02	0.07
Rent-to-income 50%+	Without callipers	1.11	0.77	0.24	1.10	0.74
	With callipers	0.03	0.06	0.18	0.03	0.06
House value 500k-750k	Without callipers	0.43	0.14	0.07	0.11	0.39
	With callipers	0.18	0.09	0.12	0.16	0.03
House value 750k-1m	Without callipers	0.94	0.06	0.39	0.18	0.32
	With callipers	0.15	0.04	0.03	0.26	0.26
House value 1m+	Without callipers	1.46	0.45	0.27	0.69	0.62
	With callipers	0.27	0.06	0.04	0.12	0.21
Distance to subway	Without callipers	0.95	0.05	0.44	0.50	0.59
	With callipers	0.15	0.05	0.03	0.33	0.81

Sample divided by quintile of entrant count. Only selected covariates shown. Additional covariates available upon request.

Table A4: Balance of menu and restaurant characteristics

Variable		Q1	Q2	Q3	Q4	Q5
Median price	Before matching	0.10	0.05	0.07	0.16	0.15
	After matching	0.13	0.03	0.01	0.04	0.02
95th ptile price	Before matching	0.15	0.07	0.07	0.18	0.24
	After matching	0.23	0.03	0.03	0.15	0.13
Item count	Before matching	0.23	0.05	0.04	0.06	0.31
	After matching	0.15	0.32	0.21	0.22	0.37
Food quality	Before matching	0.11	0.04	0.03	0.04	0.09
	After matching	0.04	0.09	0.13	0.08	0.12
Delivery timeliness	Before matching	0.04	0.03	0.07	0.04	0.09
	After matching	0.15	0.07	0.03	0.09	0.03
Order accuracy	Before matching	0.05	0.05	0.04	0.12	0.16
	After matching	0.02	0.12	0.02	0.01	0.05
Cuisines Jaccard	Before matching	0.90	0.91	0.91	0.91	0.91
	After matching	0.71	0.71	0.70	0.66	0.66
Cuisines equal	Before matching	0.00	0.00	0.00	0.00	0.00
	After matching	0.04	0.04	0.05	0.09	0.09
Cuisines subset	Before matching	0.05	0.04	0.03	0.04	0.04
	After matching	0.31	0.22	0.24	0.34	0.33

Normalized differences for randomly-selected within-calliper control matches compared to matched treated and control pairs. Unmatched values are the average over one hundred repetitions of random selections.

already achieved by matching on predicted entrants. Table A4 compares the menu similarity of the set of matched pairs with the menu similarity of the counterfactual set for each quintile of the nearby entrant count. We report the normalized differences for several menu attributes, as well as three measures of similarity in cuisine categories: the Jaccard distance[5], an indicator for whether cuisine sets are identical, and an indicator for whether one cuisine set is a subset of the other.

As shown, menu distance matching yields improved pairs compared to randomly selected restaurants within the propensity callipers of the treated restaurants. For many menu and restaurant attributes, the normalized differences are smaller for the matched set. An exception is item count; for this variable our matching does not decrease differences across quintiles and in some quintiles the differences are slightly larger after matching. However, general menu lengths tend to be a fixed characteristic of a restaurant—for example, delis tend to have very large item counts—and therefore we expect that much of this difference will be absorbed by restaurant fixed effects in our analyses (see the item count event studies in Figure 6 and further discussion in footnote 20). The last three rows of Table A4 show that the cuisines of matched restaurants are much closer; the Jaccard distance is smaller and a greater proportion have identical cuisines or some overlapping cuisines.

[5]The Jaccard distance between two sets A and B is defined as $1 - \frac{|A \cap B|}{|A \cup B|}$.

B Data and Additional Descriptive Statistics

B.1 Sources of noise

There are three sources of noise in our data which we refer to as 1) "outliers" 2) "missing data" and 3) "time-of-day effects." We use outliers to describe menus that show very unusual values, such as extremely high or extremely low prices or item counts. Many of these reflect idiosyncratic situations, such as when a restaurant lists a catering package for 100 people, priced at $2000, as an item on the menu. We classify these cases as outliers using a set of conservative rules and drop them from all of the analysis, decreasing our sample by 2.4% (roughly 13,500 restaurant periods).[6]

The second source of noise comes from data collection difficulties caused by website changes, which resulted in some missing data. For four consecutive periods starting the week of April 23 we are missing the prices for all menu items, and thus we do not use these periods in most of our analysis. Additionally, we are missing item names for five consecutive periods starting the week of September 24th. Item names in every period are not necessary for our estimation work, but we do need them in order to accurately drop duplicate items, affecting our measurement of item counts and prices.[7] Therefore we also drop these periods from most of our analysis. For a couple periods we did not collect review data (count of reviews, stars, measures of quality), but these variables are mostly unused in our analysis.

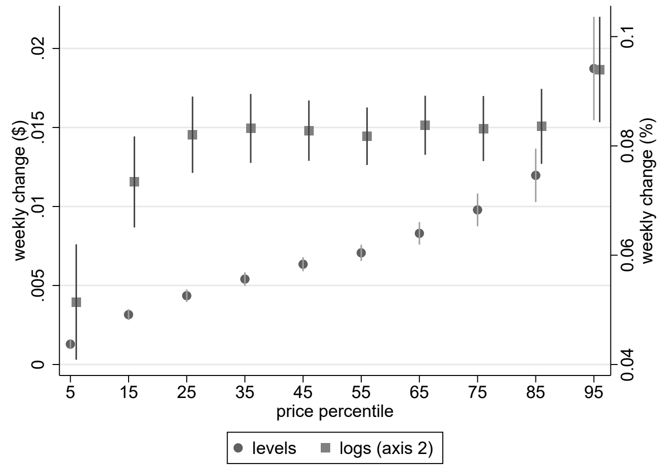
Our third source of noise comes from a unique feature of the website, in which the menus shown to the user can change depending upon the time of day the page is viewed. Some restaurants offer different menus for different meals, such as a breakfast, lunch, dinner, or late night menu. Additionally, when a restaurant is closed users have the option to pre-order, but the items shown may be only those core items that the restaurant always serves (many restaurants still show a full menu). When the restaurant is open the menu may be longer and include daily specials and other items not part of the core set. Since we collect data at different times of day throughout our panel, we sometimes observe just a core menu or short lunch menu, while at other times we see the full menu for that day. This can generate what looks like large period to period changes in the menu, but instead simply reflects the time of day viewed. In these cases the number of items observed in a period may oscillate between two fixed item counts—such as a closed menu and an open menu—providing us a way to identify this situation. We address this source of noise in three ways. First, we define "oscillating periods" as a set of three consecutive periods in which the first to second period absolute change in the log item count is larger than 0.15 log points, and the second to third period change is also larger than 0.15 log points, but the change is in the opposite direction.[8] An absolute change of 0.15 log points is a large change–about the 90th percentile of all period to period changes in log item count–and thus two consecutive large swings in menu length of opposite directions is quite unlikely to be a permanent change to the menu. There are about 50,000 oscillating periods in our data (not already tagged as outliers), about 9% of our sample, and we drop these periods from much of our analysis. Second, for most weeks in our sample we know the exact time the menu was downloaded, as well as the listed hours of the restaurant. Therefore in our main specification we include fixed effects for the hour of day and whether the restaurant was open when the menu was collected. Lastly, we also run our analysis at the restaurant-item level by

[6] Specifically, we drop restaurant periods where the item count is less than 10 or greater than 500, where the median item price is less than $2.5 or greater than $25, and where the mean item price is greater than $50.

[7] Restaurants may list the exact same item, with the same price, multiple times in different sections of the menu, often in a promotional or "popular items" section. For these five periods our item count would be inflated and quantiles of the price distribution would be inaccurate since some items are multiply counted.

[8] In notation, we define oscillating periods as three consecutive periods, $\{t-1, t, t+1\}$, where $abs(ln(itemct_t) - ln(itemct_{t-1})) \geq 0.15$ and $abs(ln(itemct_{t+1}) - ln(itemct_t)) \geq 0.15$ and $(ln(itemct_t) - ln(itemct_{t-1})) \times (ln(itemct_{t+1}) - ln(itemct_t)) < 0$.

Figure A3: Weekly price changes by item price percentile.



examining price changes over time for a constant set of restaurant menu items, which ensures that missing items do not affect our estimates.

It is worth emphasizing that all three sources of noise are completely unrelated to entry and thus our definition of treatment. Further, this noise does not lead to problems of precision in our estimates. Even after dropping observations that could increase measurement error, we still have a large sample and can estimate coefficients with small standard errors.

B.2 Within restaurant menu changes

We look at changes over time within a restaurant by running regressions of the form:

$$Y_{rt} = \beta * weeks_{rt} + \eta_r + \varepsilon_{rt} \tag{A13}$$

The η_r term is a restaurant fixed effect and the "weeks" variable measures the number of weeks (periods) since we first observed the restaurant. In Figure A3 we plot estimates of β using ten different price percentiles as outcomes Y_{rt}, and using the right-hand vertical axis, we plot the coefficients from using the natural logarithm of these ten percentiles (an additional ten regressions). We also show 95% confidence intervals for each estimate, clustering standard errors by restaurant. Figure A3 shows that restaurants increase prices higher in the distribution by a larger amount, a pattern also apparent in Table 2. The logarithm specifications suggest this pattern stems from restaurants increasing prices roughly proportionally across their menus, with most percentiles increasing on average by slightly more than 0.08% per week. At the ends of the menu distribution prices change somewhat differently, with the 5th percentile price decreasing considerably less and the 95th percentile increasing considerably more than the other percentiles. In Table A5 we show results from estimating equation A13 for some additional variables.

B.3 Pre-match differences between treated and control restaurants

As discussed in Section 3.2, a serious concern is that entry intensity may be correlated with location characteristics, making (unmatched) treated and control restaurants systematically different before the treatment period. The left panel of Table A6 uses the $d = 4$ sample to compare restaurant and menu characteristics for unmatched treated and control restaurants, four periods before treatment. The right panel uses the same

Table A5: Regression results for within-restaurant menu changes.

	(1) Itm Ct	(2) p5 Prc	(3) p25 Prc	(4) Med Prc	(5) p75 Prc	(6) p95 Prc	(7) LnMed Prc	(8) Rev Ct
weeks observed	0.088***	0.001***	0.004***	0.007***	0.010***	0.019***	0.001***	5.258***
	(0.005)	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.000)	(0.084)
Observations	456046	456046	456046	456046	456046	456046	456046	404060
Clusters	11274	11274	11274	11274	11274	11274	11274	10369

All specifications include restaurant fixed effects.
Sample excludes outliers, oscillators, missing item/price periods.
Standard errors clustered by restaurant, * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$.

sample and compares demographic characteristics of the restaurants' locations, showing the difference between the percent of the neighborhood with each characteristic and the count of other nearby restaurants. Treated restaurants have fewer items, higher prices at most points of the distribution (although these differences are not significant), more reviews, and higher user ratings. Treated and control restaurants are also in quite different areas. Treated restaurants are located in neighborhoods with younger, less impoverished, and more highly educated residents, whereas control restaurants are found in neighborhoods with a larger black population share, a greater percentage of households married and in families, and a larger share of the single-family detached units in the housing stock. Moreover, a treated restaurant has about 16 more restaurants nearby than a control restaurant. Many of these differences stem from the fact that treated restaurants are in dense, high-income areas with frequent entry and many restaurants; a large percentage are located in lower Manhattan.

These differences highlight an identification challenge likely to be an issue for any study using entry to examine responses to competition. Specifically, locations with high entry intensity have both different demographic characteristics and different types of firms than locations with lower entry intensity. If a researcher only has cross-sectional data on post-entry outcomes then comparing firms near entrants to those further away could yield very misleading results. In our case we would conclude entry leads to shorter menus and higher prices. Further, if firms in areas with high intensity of entry also vary in the frequency with which they make changes, then longitudinal studies (including simple difference-in-difference methods) may also lead to biased conclusions. This motivates the use of the two-stage matching method in this study.

B.4 Pre-match Estimation Results

In this section we estimate the following long differences specification on the pre-matched data:

$$\Delta Y_{r,t} = \beta * (post_{rt} \times D_{rt}) + \eta_t + \eta_r + \varepsilon_{rt} \tag{A14}$$

The dependent variable, $\Delta Y_{r,t}$ is the symmetric difference from $t-d$ to $t+d$. The sample includes all restaurants that were treated or control in period t, thus there are many more control observations than treated. We include restaurant fixed effects, η_r, period fixed effects, η_t, and cluster standard errors by restaurant. For each specification we report the predicted mean change for the control group and the associated SE (calculated using the delta method). This statistic is similar to what the constant would be in a regression without period fixed effects.

Table A6: Statistical tests for difference between treated and control restaurants. All values are measured four periods prior to treatment. Sample excludes outliers and missing price periods.

(a) Menu attributes

	Menu stats	
	t-tests	N
item count	-14.43***	88115
mean item price	0.17**	88115
median item price	0.21***	88115
p25 item price	0.13***	88115
p95 item price	0.00	88115
stars	0.06*	85202
review count	40.87***	79003
order rating	0.71***	86252
food rating	0.44*	86252
delivery rating	1.23***	86252

Tests difference between treated and control.
Calculated using values 4 periods before treatment.
Sample excludes outliers and missing price periods.

(b) Demographic attributes

	Demographics	
	t-tests	N
age2529	0.023***	88476
age3039	0.027***	88476
age7079	-0.003***	88476
racewhite	0.087***	88476
raceblack	-0.053***	88476
hhfamily	-0.086***	88476
hhmarried	-0.038***	88476
educdegree	0.118***	88476
poverty	-0.023***	88476
income100150	0.008***	88476
income150200	0.008***	88476
unitdetached	-0.064***	88476
competitors 500m	16.107***	81877

Tests difference between treated and control.
All demographics calculated as percent of area.
Competitors calculated 4 periods pre-treatment.
Sample excludes outliers and missing price periods.

Table A7: Pre-match regressions using long differences as the dependent variable. All specifications include restaurant fixed effects and period fixed effects; standard errors clustered by restaurant are shown in parentheses. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

	(1) Med Prc	(2) Ln Med Prc	(3) p5 Prc	(4) p95 Prc	(5) Itm Ct	(6) Itm Prc
treated	-0.0160	-0.0021	-0.0044	-0.1044**	0.1033	-0.0027
	(0.0140)	(0.0017)	(0.0076)	(0.0406)	(0.3515)	(0.0047)
Observations	76913	76913	76913	76913	76913	8774924
Unique Restaurants	5521	5521	5521	5521	5521	5267
Treated Count	2129	2129	2129	2129	2129	1686
Pred. Control Mean	0.05	0.01	0.01	0.20	0.89	0.04
SE of Predicted Mean	0.00	0.00	0.00	0.02	0.07	0.00

(a) Four period duration

	(1) Med Prc	(2) Ln Med Prc	(3) p5 Prc	(4) p95 Prc	(5) Itm Ct	(6) Itm Prc
treated	0.0318	0.0038	0.0107	-0.0814	-0.3041	0.0071
	(0.0250)	(0.0029)	(0.0091)	(0.0768)	(0.4450)	(0.0090)
Observations	52822	52822	52822	52822	52822	5730662
Unique Restaurants	4023	4023	4023	4023	4023	3812
Treated Count	1195	1195	1195	1195	1195	804
Pred. Control Mean	0.07	0.01	0.02	0.27	1.06	0.06
SE of Predicted Mean	0.01	0.00	0.00	0.03	0.12	0.00

(b) Six period duration

	(1) Med Prc	(2) Ln Med Prc	(3) p5 Prc	(4) p95 Prc	(5) Itm Ct	(6) Itm Prc
treated	0.0245	0.0034	-0.0221	-0.0117	0.1638	0.0055
	(0.0244)	(0.0032)	(0.0145)	(0.2386)	(0.7183)	(0.0117)
Observations	37941	37941	37941	37941	37941	3998046
Unique Restaurants	3047	3047	3047	3047	3047	2854
Treated Count	634	634	634	634	634	380
Pred. Control Mean	0.09	0.01	0.02	0.30	1.43	0.07
SE of Predicted Mean	0.01	0.00	0.01	0.04	0.18	0.00

(c) Eight period duration

C Robustness and Heterogeneity

C.1 Robustness

In Table A8 we run our main restaurant-level specification on the following set of non-menu variables that Grubhub provides to consumers describing each restaurant: quality ratings, hours of operation, listed cuisines, and count of reviews.[9] We find essentially no effect on any quality ratings, hours of operation, or food quality, and a very small negative coefficient for the number of cuisines, significant at the 10% level. In column 6 we look at the count of reviews, which increases each week and might be interpreted as a very noisy proxy for sales. Interestingly, we find a statistically significant decrease of about 7 reviews for the four period duration. If we just compare the change in review counts from four periods before treatment to four periods after treatment (a "long difference"), then control restaurants have 55 additional reviews and treated restaurants have 45 additional reviews, about a 18% decline. The coefficients for the other two durations are also both negative, but imprecisely estimated, and therefore it is unclear whether this single coefficient indicates a decrease in sales volume resulting from new competition.

A concern with our baseline analysis is that incumbent restaurants may only respond to new competition after longer periods than we have tested. To assess this concern we first run a long difference version of our specification comparing the change in outcomes from $t-d$ to $t+d$ only (just two periods). This range removes the effect of early post-treatment periods, making it more robust to measurement error in entry timing and any anticipatory reactions to new competition, although the pre-treatment coefficients shown in Figure 6 provide no evidence of anticipation. The results from this analysis are similar to those presented in Table 3 and so we omit them for brevity (available upon request). Next, we try re-running our analysis shifting the definition of pre-treatment and post-treatment periods forward by d periods, so that the pre period is $[0,d-1]$ and the post period is $[d+1,2d]$ (actual entry still occurs in the entry window between $t-d/2$ and t). The idea behind this analysis is that if we are not finding any effects in Table 3 because restaurants do not respond in the first d post-entry periods—for example, incumbent restaurants may conduct business as usual while waiting to see how successful is the new entrant—then those first d post-entry periods are actually valid control periods. Further, since our definition of treated and control requires no entry in the $[0,2d]$ periods, we can use the $[d+1,2d]$ range as post-treatment periods without worrying about the effect of additional entrants.

We present the results of this shifted analysis in Table A9. Overall the results are fairly close to those using the original duration in Table 3 and the similarity of the coefficients on "post" suggest that we are capturing consistent changes restaurants make to their menus in the absence of any competitive effects. We do find a statistically significant coefficient for item count in the four period duration, again with a small magnitude (about one half percent of the mean item count). We also now find a statistically significant post-treatment coefficient for the item-level specification in the eight period sample. We think this coefficient probably just represents sampling variation, but even if true, the effect is quite small. A 3.6 cent increase on an average item price of $8.5, which is only a little larger than the general increase in item prices of 3.1 cents.

[9] Other studies find that retail firms in other industries respond to competitive intensity by improving service quality. Auto dealerships carry more inventory (Olivares and Cachon 2009) and supermarkets reduce their inventory shortfalls (Matsa 2011) when competition increases. Longer hours also constitute a form of improved service quality for retail businesses including gas stations (Kügler and Weiss 2016) and outlets of fast-food restaurant chains (Xie 2022) where other forms of differentiation may be infeasible.

Table A8: Fixed effect results for geographic distance treatment. Dependent variables are quality ratings, weekly hours of operation, count of listed cuisines, and count of reviews. All specifications include restaurant fixed effects and period fixed effects. Standard errors clustered by entrant are shown in parentheses. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

	(1) Food Rtng	(2) Delivery Rtng	(3) Order Rtng	(4) Wkly Hrs	(5) Num Cuisines	(6) Review Ct.
treat_post	-0.039	-0.064	-0.022	-0.187	-0.026	-6.881***
	(0.063)	(0.056)	(0.046)	(0.426)	(0.032)	(2.069)
post	0.009	0.036	-0.041	0.246	0.328***	38.034***
	(0.044)	(0.042)	(0.033)	(0.638)	(0.067)	(1.629)
Observations	20233	20233	20233	20162	20652	18800
Clusters	370	370	370	371	371	368
Treated	1941	1941	1941	1942	1944	1934
DepVarMean	86.38	87.12	90.47	66.49	5.11	525.53

(a) Four period duration

	(1) Food Rtng	(2) Delivery Rtng	(3) Order Rtng	(4) Wkly Hrs	(5) Num Cuisines	(6) Review Ct.
treat_post	0.112	0.066	-0.013	-0.459	-0.021	-4.577
	(0.076)	(0.067)	(0.051)	(0.616)	(0.045)	(3.453)
post	-0.145***	-0.089*	-0.118***	3.073**	0.365***	53.652***
	(0.048)	(0.048)	(0.041)	(1.368)	(0.085)	(2.752)
Observations	16629	16629	16629	16704	16944	15431
Clusters	295	295	295	296	296	295
Treated	1244	1244	1244	1246	1246	1240
DepVarMean	86.19	87.02	90.41	66.42	5.00	496.79

(b) Six period duration

	(1) Food Rtng	(2) Delivery Rtng	(3) Order Rtng	(4) Wkly Hrs	(5) Num Cuisines	(6) Review Ct.
treat_post	0.082	0.063	-0.088	-0.343	-0.103*	-1.824
	(0.102)	(0.119)	(0.094)	(0.945)	(0.056)	(5.563)
post	-0.150**	-0.153*	-0.058	6.157**	0.671***	64.180***
	(0.070)	(0.082)	(0.068)	(2.382)	(0.091)	(4.040)
Observations	12016	12016	12016	12074	12180	11128
Clusters	210	210	210	211	211	210
Treated	702	702	702	703	703	701
DepVarMean	86.14	86.83	90.18	64.77	4.95	460.86

(c) Eight period duration

Table A9: Fixed effect results using extended durations. The fourth column shows results from an item-level regression. All specifications include restaurant fixed effects, standard errors clustered by entrant are shown in parentheses. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

	(1) Med Prc	(2) Ln Med Prc	(3) p5 Prc	(4) p95 Prc	(5) Itm Ct	(6) Itm Prc
treated X post	0.009	0.001	-0.000	-0.152	0.807***	0.006
	[-0.022,0.040]	[-0.002,0.005]	[-0.010,0.009]	[-0.346,0.041]	[0.214,1.400]	[-0.006,0.019]
post	0.018*	0.002**	0.006*	0.179**	-0.106	0.024***
	[-0.001,0.037]	[0.000,0.004]	[-0.001,0.012]	[0.006,0.353]	[-0.440,0.228]	[0.018,0.031]
open	-0.003	0.000	0.005*	0.038	2.145***	
	[-0.020,0.014]	[-0.002,0.003]	[-0.000,0.011]	[-0.032,0.108]	[1.653,2.636]	
Observations	19616	19616	19616	19616	19616	2651697
Clusters	367	367	367	367	367	367
Treated	1846	1846	1846	1846	1846	1845
DepVarMean	8.33	2.06	2.41	17.61	148.47	8.67

(a) Four period duration: pre [0,3], post [5,8]

	(1) Med Prc	(2) Ln Med Prc	(3) p5 Prc	(4) p95 Prc	(5) Itm Ct	(6) Itm Prc
treated X post	-0.008	-0.002	-0.012	0.001	0.096	0.017
	[-0.046,0.030]	[-0.006,0.003]	[-0.031,0.006]	[-0.195,0.197]	[-0.490,0.682]	[-0.006,0.039]
post	0.045***	0.006***	0.006	0.141***	0.271	0.025***
	[0.019,0.072]	[0.002,0.009]	[-0.006,0.018]	[0.038,0.243]	[-0.223,0.766]	[0.016,0.033]
open	-0.016	-0.001	0.007*	0.017	1.788***	
	[-0.037,0.005]	[-0.004,0.002]	[-0.001,0.016]	[-0.097,0.131]	[1.271,2.305]	
Observations	17638	17638	17638	17638	17638	2467190
Clusters	293	293	293	293	293	293
Treated	1221	1221	1221	1221	1221	1218
DepVarMean	8.14	2.04	2.31	17.51	154.34	8.50

(b) Six period duration: pre [0,5], post [7,12]

	(1) Med Prc	(2) Ln Med Prc	(3) p5 Prc	(4) p95 Prc	(5) Itm Ct	(6) Itm Prc
treated X post	0.023	0.002	-0.004	0.075	0.590*	0.036**
	[-0.022,0.067]	[-0.003,0.007]	[-0.029,0.021]	[-0.223,0.373]	[-0.092,1.273]	[0.007,0.065]
post	0.054***	0.007***	0.017**	0.174***	0.191	0.031***
	[0.025,0.084]	[0.003,0.011]	[0.002,0.033]	[0.049,0.298]	[-0.292,0.675]	[0.020,0.042]
open	-0.016	-0.001	0.013**	0.012	2.207***	
	[-0.043,0.011]	[-0.005,0.003]	[0.002,0.024]	[-0.077,0.100]	[1.666,2.747]	
Observations	13966	13966	13966	13966	13966	1980004
Clusters	211	211	211	211	211	211
Treated	700	700	700	700	700	699
DepVarMean	8.15	2.05	2.29	17.45	157.95	8.53

(c) Eight period duration: pre [0,7], post [9,16]

C.2 Heterogeneity by Incumbent Characteristics

In this section we examine whether the effect of entry on median price varies by characteristics of the incumbent restaurant or the local area. For characteristics of the restaurant we use the median price, item count, review count (in hundreds), and food rating of each restaurant from the earliest observed period (usually the first period, but always long before treatment). We examine heterogeneity by area characteristics using the number of competitors (in tens) within 500 meters in the first period, the simple population density (population divided by census tract land area), average income, and the distance to the closest subway in 2017.[10] The specification below adds two new interaction terms to our baseline specification, interacting a heterogeneity term *Var* with both the *post* term and the $post \times D_{r,t}$ (treated by post):

$$\begin{aligned} MedPrc_{r,t} =& \beta_1 * post_{r,t} + \beta_2 * (post_{r,t} \times D_{r,t}) + \gamma_1 * (post_{r,t} \times Var_r) + \gamma_2 * (post_{r,t} \times D_{r,t} \times Var_r) \\ &+ \beta_3 * open_{r,t} + \eta_h + \eta_r + \varepsilon_{r,t} \end{aligned} \quad \text{(A15)}$$

We demean each interacted variable *Var* so that β_2 represents the treatment effect at the mean of *Var*, and γ_2 represents the effect of increasing *Var* by one unit from the mean. We show the results in Table A10 for each duration. The second table row in each panel shows the interaction variable and the third table row gives the standard deviation of this variable. For example, a one-standard deviation in population density for the four week sample (panel A) is 14.5 thousand people per square kilometer.

The coefficients on $treated \times post$ are quite similar to those shown earlier in Table 3 and we find that nearly every interaction term ("treated X post X inter.") is small and statistically insignificant. We do find small but statistically significant coefficients for the interaction with the number of local competitors and the distance to the subway for the eight week duration.[11] A one standard deviation increase in the number of competitors is 12 additional restaurants, implying a \$0.06 decrease in median price (competitors is measured in tens of restaurants), which is less than 1% of the average restaurant's median price. Similarly, a one standard deviation increase in the distance to the subway is associated with a \$0.055 increase in median price. This suggests that if there are heterogeneous entry effects across these variables, they are quite small.

C.3 Heterogeneity by Menu Item

In this section we assess whether the response to entry is heterogeneous within a restaurant's menu, meaning that restaurants change prices for some items in different ways than others. Ideally, we could test whether the response differs by item categories—such as appetizers, drinks, or entrees—but unfortunately there is no standard categorization across restaurants (ex: are tapas the same as appetizers?). Instead, we estimate our main specification (eq 6) for the ten different menu item price percentiles from 5% to 95%, in increments of 10%. This strategy allows us to avoid applying our own categorization and does capture broad item categories.[12] From Figure A3 we know that restaurants raises prices approximately proportionally across items and therefore we estimate equation 6 using the natural logarithm of each price percentile. In Figure A4 we plot the $treated \times post$ coefficients for each (ln) percentile using our baseline radius of 500 meters and for each duration. For the four week and eight week durations treatment effects are close to zero with

[10] Population and income data are from the 2015 American Community Survey. Population density is measured in thousands of people per square kilometer, income is measured in thousands of dollars per year, and distance to the nearest subway is measured in kilometers.

[11] These statistically significant coefficients need to be interpreted with caution. With 24 specifications and a significance level of $\alpha = 5\%$, the Bonferroni p-value is $p = 0.05/24 = 0.002$. Using this p-value, none of the interactions would be statistically different from zero, but this correction is likely too conservative since there is correlation across the specifications.

[12] For example, we read through a number of menus and found that the prices of most drinks are in the lowest two percentiles, 5% and 15%.

Table A10: Heterogeneity Analysis. All specifications include restaurant fixed effects, standard errors clustered by entrant are shown in parentheses. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc
treated X post	0.009	0.014	0.007	0.010	0.012	0.008	0.009	0.009
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)
treated X post X inter.	0.010	-0.000	0.008*	0.001	0.007	0.002	0.001	-0.002
	(0.008)	(0.000)	(0.005)	(0.002)	(0.008)	(0.001)	(0.001)	(0.006)
post	0.024**	0.019	0.026**	0.025**	0.022**	0.025**	0.024**	0.025**
	(0.010)	(0.012)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)
post X inter.	-0.011	0.000**	-0.005	-0.001	-0.012**	-0.001	-0.001	0.003
	(0.006)	(0.000)	(0.003)	(0.001)	(0.005)	(0.001)	(0.001)	(0.004)
Observations	20652	20652	19661	20385	20652	20652	20652	20652
Interaction	Med Prc	Itm Ct	Review Ct	Food Rtng	Comp. 500m	Density	Income	Subway Dist
Interaction sd	2.9	86.5	3.7	10.1	2.1	14.5	11.8	1.8
Clusters	371	371	371	370	371	371	371	371

(a) Four period duration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc
treated X post	-0.009	-0.006	-0.008	-0.010	-0.008	-0.008	-0.007	-0.008
	(0.022)	(0.022)	(0.024)	(0.022)	(0.022)	(0.021)	(0.021)	(0.022)
treated X post X inter.	-0.014	0.000	0.001	0.003	-0.006	0.001	0.003	0.009
	(0.018)	(0.000)	(0.005)	(0.003)	(0.012)	(0.002)	(0.002)	(0.009)
post	0.069***	0.067***	0.070***	0.072***	0.069***	0.068***	0.068***	0.069***
	(0.018)	(0.019)	(0.019)	(0.019)	(0.018)	(0.017)	(0.017)	(0.018)
post X inter.	0.004	0.000	-0.001	-0.003	-0.002	0.001	-0.002	-0.006
	(0.016)	(0.000)	(0.003)	(0.002)	(0.010)	(0.001)	(0.002)	(0.004)
Observations	16944	16944	16012	16715	16944	16944	16944	16944
Interaction	Med Prc	Itm Ct	Review Ct	Food Rtng	Comp. 500m	Density	Income	Subway Dist
Interaction sd	2.8	84.2	3.1	10.3	1.7	14.8	12.4	2.1
Clusters	296	296	296	295	296	296	296	296

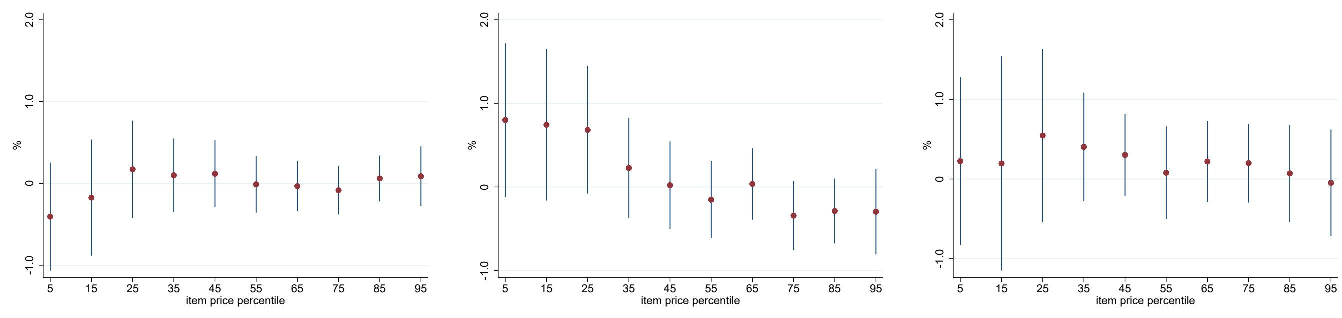
(b) Six period duration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc	Med Prc
treated X post	0.009	0.016	0.007	0.006	0.004	0.010	0.011	0.008
	(0.033)	(0.032)	(0.036)	(0.036)	(0.034)	(0.032)	(0.031)	(0.034)
treated X post X inter.	-0.022	-0.000	0.001	0.007	-0.050***	-0.004	0.007*	0.021**
	(0.034)	(0.000)	(0.008)	(0.005)	(0.017)	(0.003)	(0.004)	(0.011)
post	0.076***	0.068**	0.078**	0.080**	0.078***	0.073***	0.073***	0.077**
	(0.028)	(0.028)	(0.031)	(0.031)	(0.030)	(0.027)	(0.027)	(0.030)
post X inter.	0.015	0.000	-0.005	-0.005	0.015	0.003	-0.005	-0.006
	(0.030)	(0.000)	(0.007)	(0.004)	(0.012)	(0.002)	(0.003)	(0.005)
Observations	12180	12180	11557	12084	12180	12180	12180	12180
Interaction	Med Prc	Itm Ct	Review Ct	Food Rtng	Comp. 500m	Density	Income	Subway Dist
Interaction sd	2.8	83.2	2.8	9.8	1.2	14.9	12.8	2.6
Clusters	211	211	211	210	211	211	211	211

(c) Eight period duration

no obvious pattern across percentiles. In the six week figure there is roughly a monotonic decrease in the estimated coefficients with negative coefficients for the largest price percentiles. These coefficients are still quite small, indicating that treated restaurants decrease prices (relative to control restaurants) by about a half percent over six weeks.

Figure A4: Changes by Price Percentile, %



(a) d=4 (b) d=6 (c) d=8

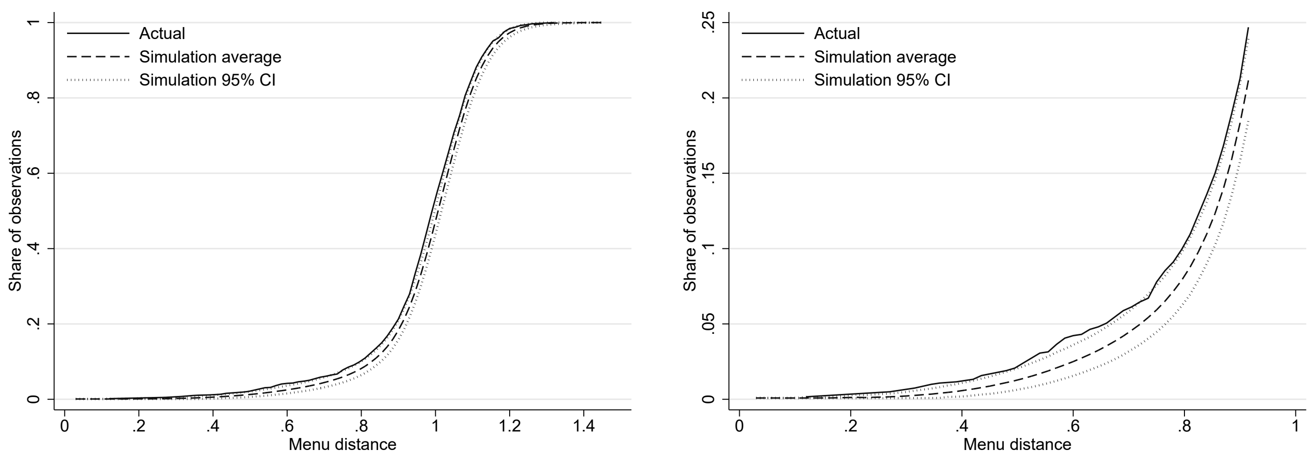
D Additional results

D.1 Location choice analysis

To better understand the entrant location decision, we compare the menu distance between entrants' menus and those of nearby restaurants with the menu distance from a set of simulated counterfactual location choices. Specifically, we compare the observed distribution of menu distance between entrant restaurants and incumbent neighbors (within 500 meters) to a counterfactual distribution generated by repeatedly reshuffling entrants in the $d = 4$ regression sample between observed entrant locations. That is, on each iteration, we randomly reassign entrants among the set of observed entry locations according to a uniform distribution and without replacement. If entrants were strategically locating to soften local competitive intensity, the observed distribution would feature fewer incumbent neighbors at small menu distance than the counterfactual distribution. Restaurant location choices are constrained by many factors (zoning laws, vacancies, availability of suitable space) and therefore limiting the random reassignment to the set of observed entrant locations helps to generate plausible counterfactuals.

Figure A5 shows results generated by randomly reshuffling entrants between the observed entrant locations ten thousand times. As shown, the observed distribution of menu menu distance between entrants and incumbent neighbors is actually concentrated at closer menu distances. For example, about ten percent of entrant-incumbent menu distances are less than 0.8, a share slightly larger than the upper bound of the 95% confidence interval from the bootstrap repetitions. Thus not only is there no evidence of entrants distancing themselves from likely competitors, but rather the distribution shows that similar restaurants are actually more likely to co-locate.

Figure A5: Location Choice Analysis

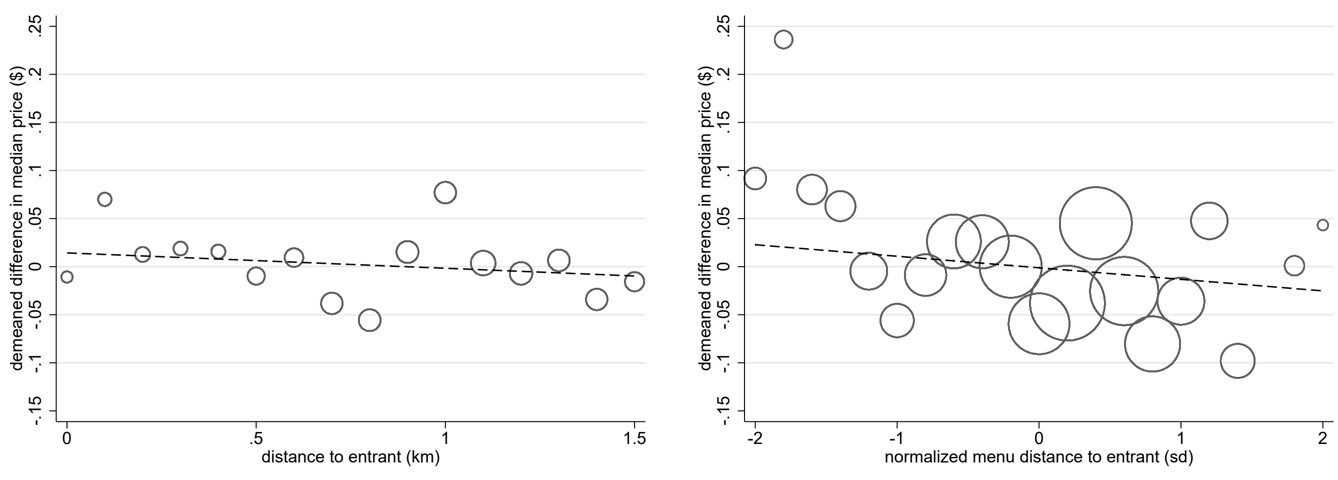


Plots shows cumulative distribution function of menu distance between entrants and incumbent neighbours compared with counterfactual cumulative distribution under random reshuffling. The left panel shows the full distribution. The right panel shows the bottom quintile to emphasize the higher incidence of similar menus in the observed distribution.

D.2 Continuous Measures of Competition: Price Changes by Distance to Entrant

In Figure A6 we plot symmetric differences in median price against distance to the entrant, for both geographic distance and menu distance, using the four week duration sample. We define a symmetric difference within a restaurant as $Y_{r,t+w} - Y_{r,t-w}$—the difference in outcomes with equal time around the treatment period—and then subtract the average of these differences across all restaurants facing the same entrant. Panel A of Figure A6 shows the average change in median price in 0.1km distance bins, with the circle radius representing observation count weights. The demeaned price changes are concentrated within 0.05 of zero with a slight and statistically insignificant negative slope of -0.016. In panel B we plot price changes against standardized menu distance using the mean and standard deviation from the entire distribution of pairwise menu distances—we subtract the mean and divide by the standard deviation—and thus the horizontal axis is measured in standard deviations. We can only match about 40% of entrants to Grubhub restaurants, a requirement in order to calculate menu distance, and thus the sample is less than half the size of the geographic distance sample (see observation counts in Table 4). While panel A shows that geographic distance to the entrant is fairly uniform, menu distance is much more concentrated. In order to focus on the majority of observations, we restrict the plot to two standard deviations from the mean and also drop one outlier bin with few observations and a large positive price change; these exclusions reduce the sample by 5% (we use the full sample in all regression tables). The menu distance plot shows greater variance in price changes but the slope is still small (-0.012) and statistically insignificant.

Figure A6: Median Price Changes by Distance to Entrant, 4 week duration



(a) geographic distance

(b) menu distance

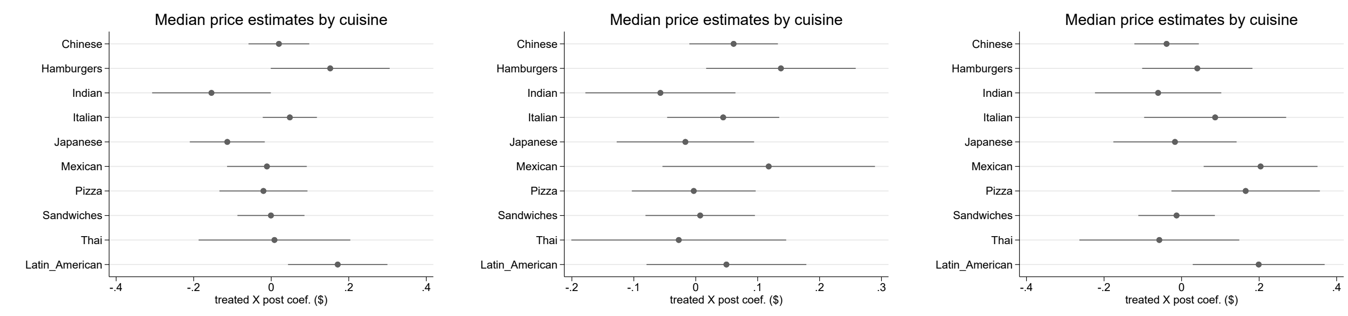
Both graphs plot symmetric median price changes against distance to the entrant, with the price changes demeaned by entrant area. The left-hand plot shows the average change in 0.1 km distance bins while the right-hand plot uses standardized menu distance with a bin size of 0.2 standard deviations. The circle radius indicates observation count weights. In each plot we also show linear fit lines from regressing demeaned price changes on distance. The geographic distance slope is -0.016 (se 0.17) and the menu distance slope is -0.012 (se 0.014). In the right-hand plot we dropped outliers and observations more than 2 sd in menu distance from the mean in order to focus the plot on the remaining 95% of observations.

D.3 Heterogeneity by Cuisine

In this section we run our baseline menu distance specification separately by cuisine. Most restaurants on Grubhub list multiple cuisines and so we define the set of cuisine A restaurants as any restaurant listing A as a cuisine. For consistency, we still define entrants as new competition within the 2nd menu distance percentile of incumbents, but these restaurants nearly always also list the same cuisine. Thus treated restaurants, control restaurants, and entrants all share the cuisine category.

In Figure A7 we plot the coefficients from running the specification on median price for a set of common cuisines with at least 500 restaurant-periods. The observation count in each regression is much smaller, ranging from 500 to 1500 across cuisines and durations, since we have filtered by cuisine. As a result, the confidence intervals (now clustered by treated-control pair) are significantly larger than in Table 5. Most point estimates are concentrated around zero, but a couple cuisines ("Latin American", "Hamburgers") have significant coefficients for some durations.

Figure A7: Median Price Responses by Cuisine, \$



(a) d=4 (b) d=6 (c) d=8

E Implementation of Generalized Propensity Score (GPS) in Exit Analysis

We first re-estimate our Poisson entry model, equation 5, using only entrants from the 54 weeks of the pre-period. We then derive the GPS directly from the predicted number of pre-period entrants using this model. Let λ_r be the predicted number of pre-period entrants within 500 m of restaurant r. This λ_r is an arrival rate (per 54 weeks) for new entrants in the area around restaurant r. We then define the GPS at entrant count n as the Poisson likelihood of n events with rate parameter λ_r:

$$GPS_r(n) = Pr(n|\lambda_r) = \frac{\lambda_r^n e^{-\lambda_r}}{n!} \tag{A16}$$

In Equation A16, $GPS_r(n)$ is a function specific to every restaurant r. It measures the probability that a location with entry rate λ_r receives n entrants over 54 weeks.

We model the hazard of exiting in any one week using a Cox proportional hazard model with a common baseline hazard, $\phi_0(t)$. For restaurant r in a location that received n_r entrants over the 54 periods, the hazard of exiting after t weeks is:

$$\phi_r(t|n_r) = \phi_0(t) * exp(\gamma * n_r) \tag{A17}$$

We then estimate the conditional expectation of the outcome given the treatment and the GPS. Note that in the conditional expectation equation below we evaluate the GPS for restaurant r at the actual number of entrants observed in that location in the pre-period, n_r.

$$\phi_r(t|n_r) = \phi_0(t) * exp(\gamma_1 * n_r + \gamma_2 * GPS_r(n_r)) \tag{A18}$$

Our interest is in the relative hazard (the exponentiated term) which shows how the hazard of exit increases or decreases with entry. Therefore we calculate the dose response function as the relative hazard of exit at a "dose" of n entrants. To do so we take the coefficients from Equation A18, predict the relative hazard at n entrants with the GPS evaluated at n, and then average this predicted relative hazard over all R restaurants:

$$E[\phi_r(t|n)/\phi_0(t)] = \frac{1}{R}\sum_r (exp(\hat{\gamma}_1 * n + \hat{\gamma}_2 * GPS_r(n))) \tag{A19}$$

We use bootstrapping to calculate confidence intervals for Equation A19 using 1000 bootsamples for each dose level.[13]

We run a series of Cox proportional hazard models, as specified by Equation A17, and report the results in Appendix Table A11. We list the p-values from a global test of the proportional hazards assumption in the last table row and find no evidence of non-proportional hazards.[14] When we include observed pre-period entrants (entrant intensity) without any controls (column 3) we find a coefficient of 0.006, indicating that each additional entrant increases the hazard of exiting relative to the baseline by 0.6 percentage points. This implies that a restaurant in a location with an entrant rate of ten entrants in 54 weeks is about 6% more likely to exit in a given week than a restaurant in a location with no entrants; however, this coefficient is imprecisely estimated. In column 4 we add the GPS and find a much larger positive coefficient on entrant intensity. This coefficient is also now statistically significant at the 5% level, but, as emphasized by Hirano and Imbens, has no causal interpretation. Hirano and Imbens suggest using a flexible form for estimating the conditional expectation, and so in columns five and six we add an interaction term and quadratic terms. However, in the most flexible specification (column 6) all of the coefficients are imprecisely estimated and in column 5 the interaction term is insignificant with similar coefficients for entrant intensity and the GPS to those in column 4. Further, a likelihood ratio test comparing the goodness of fit for the simplest specification in column 4 to the more flexible forms in columns 5 and 6 cannot reject that the fit is equal. Therefore we choose the coefficients from the specification in column four to calculate the dose response function. We calculate this dose response at the median value for each entrant count decile and plot the results with bootstrapped 95% confidence intervals in Figure 10.

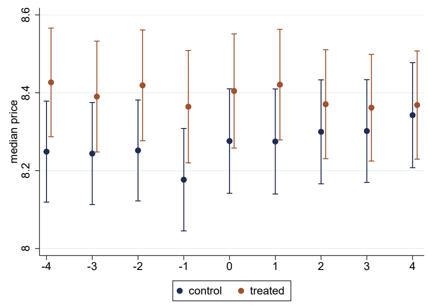
F Supplemental Tables and Figures

[13] We draw with replacement from our estimation sample, re-estimate equation A18, and then calculate equation A19 with the estimates. We repeat this 1000 times and then report the 25th and 975th largest estimates for each dose level as the 95% confidence interval.

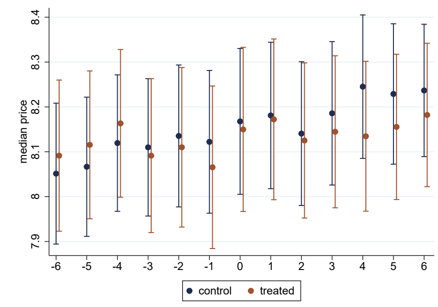
[14] We implemented this test using the "estat phtest" command in Stata, which is based on testing whether the Schoenfeld residuals are correlated with the covariates. We also ran a covariate level test for our main specification (column 4) and found no evidence that the hazards vary non-proportionally over time for either covariate.

Figure A8: Plots show raw (non-regression-adjusted) means and 95% confidence intervals for matched treated and control restaurants by relative treatment period.

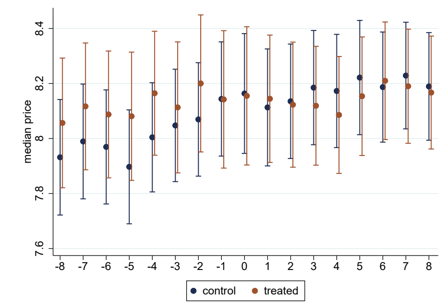
Median Item Price, $



(a) d=4

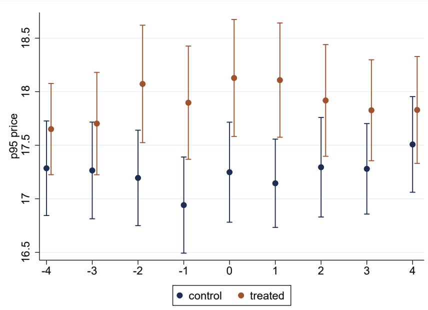


(b) d=6

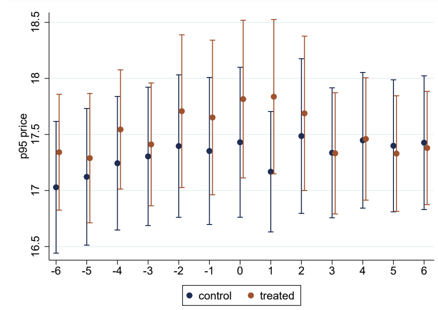


(c) d=8

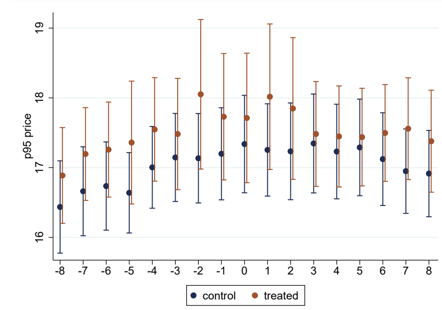
95th Percentile Price, $



(d) d=4

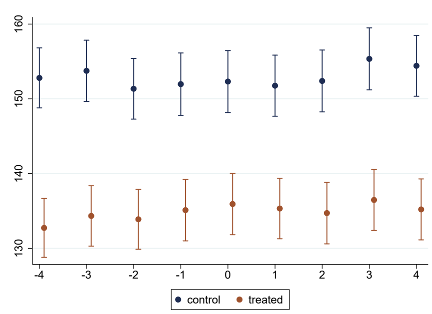


(e) d=6

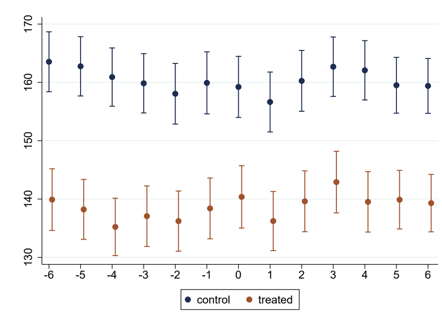


(f) d=8

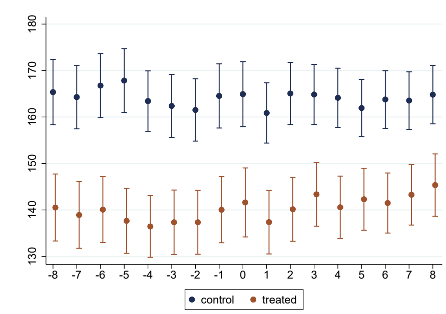
Item Count



(g) d=4



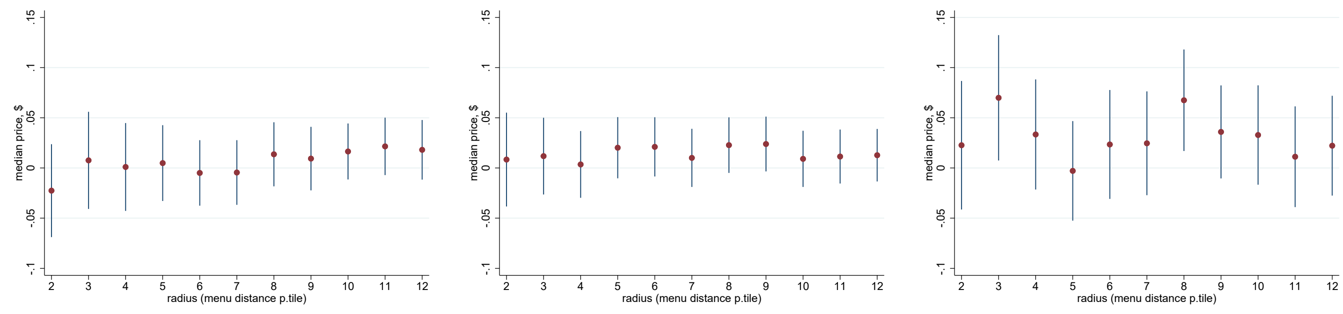
(h) d=6



(i) d=8

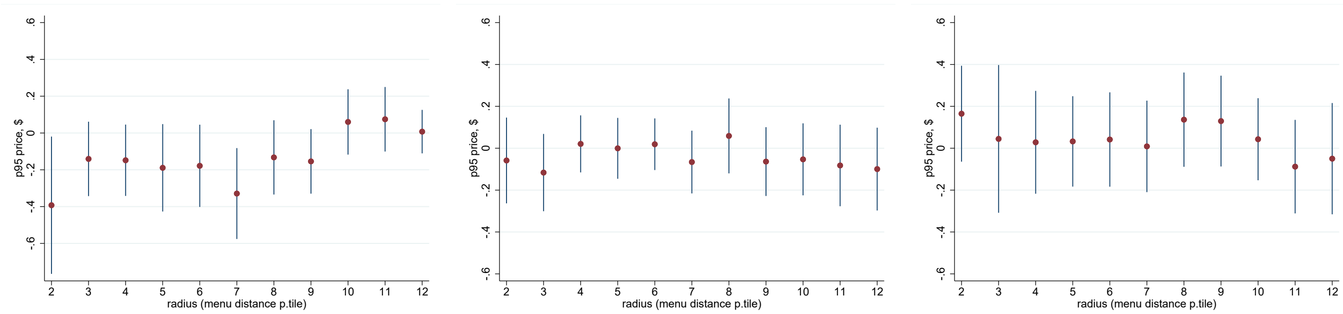
Figure A9: Treatment effects at different menu distance percentiles.
Each point shows the coefficient on $post_{rt} \times D_{rt}$ from a separate regression using menu distance percentile to define treatment.

Median Item Price, $



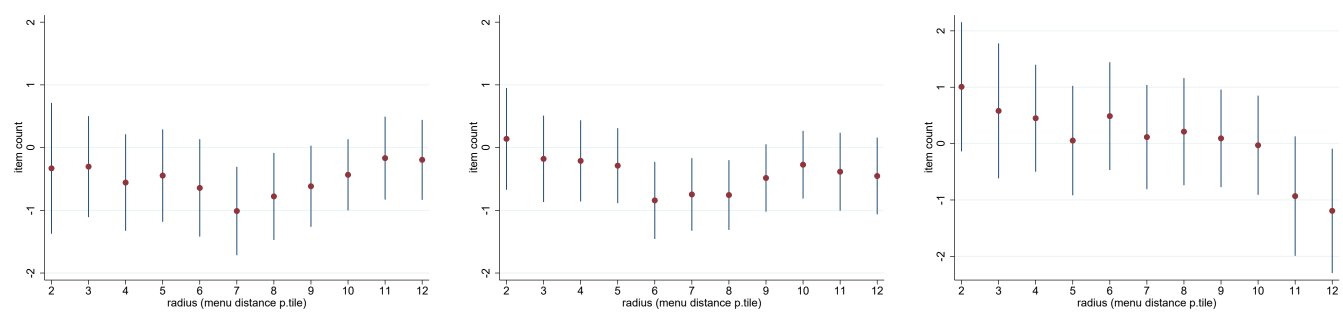
(a) d=4 (b) d=6 (c) d=8

95th Percentile Price, $



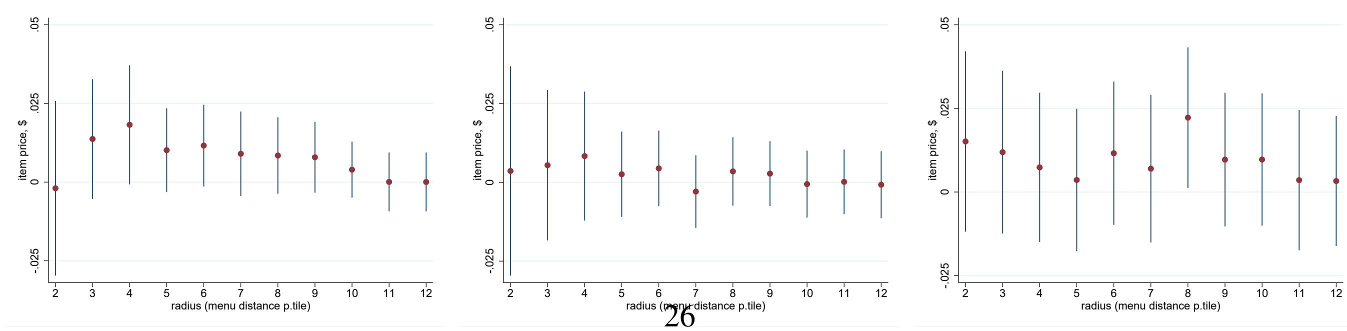
(d) d=4 (e) d=6 (f) d=8

Item Count



(g) d=4 (h) d=6 (i) d=8

Item Price, $



(j) d=4 (k) d=6 (l) d=8

Figure A10: Survival estimates across all entrant quantiles

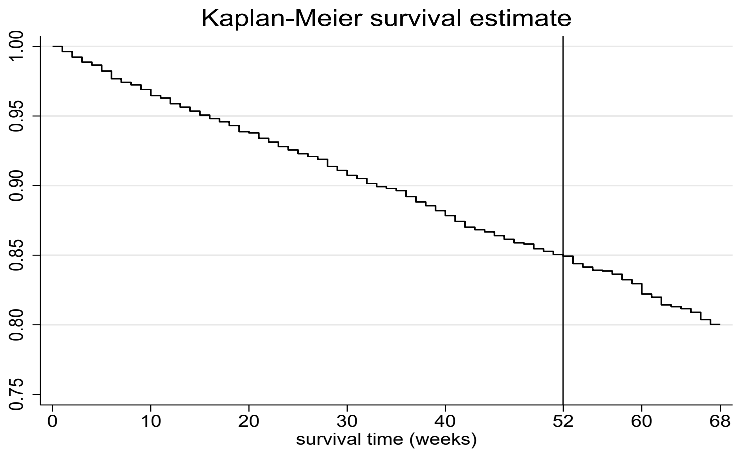


Table A11: Exit Analysis Specifications

	(1) surv. time	(2) surv. time	(3) exit haz.	(4) exit haz.	(5) exit haz.	(6) exit haz.
observed entrants	-0.0356	-0.1446**	0.0060*	0.0118***	0.0155***	0.0181
	(0.0256)	(0.0696)	(0.0034)	(0.0043)	(0.0058)	(0.0236)
predicted entrants		0.1253*				
		(0.0745)				
GPS				0.4777**	0.4693**	0.9988
				(0.2094)	(0.2083)	(0.9063)
obs. ents. X GPS					-0.0703	-0.0982
					(0.0758)	(0.1053)
obs. ents.^{2}						0.0000
						(0.0007)
GPS^2						-0.7450
						(1.1183)
Observations	7016	7016	7016	7016	7016	7016
Likelihood	-29589.5	-29588.1	-12258.2	-12255.7	-12255.3	-12255.0
PH test			0.808	0.858	0.853	0.124

First two specifications show OLS results for survival time (weeks), last four show results from proportional hazards models. For hazard models we show coefficients, not hazard ratios. There are 1401 observed exits in the sample. Significance levels: *** 1 percent, ** 5 percent, * 10 percent.

References

Crump, Richard K, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik, "Dealing with limited overlap in estimation of average treatment effects," *Biometrika*, 2009, *96* (1), 187–199.

Damashek, Marc, "Gauging similarity with n-grams: Language-independent categorization of text," *Science*, 1995, *267* (5199), 843–848.

Flores, Carlos A, Alfonso Flores-Lagunes, Arturo Gonzalez, and Todd C Neumann, "Estimating the effects of length of exposure to instruction in a training program: the case of job corps," *Review of Economics and Statistics*, 2012, *94* (1), 153–171.

Hirano, Keisuke and Guido W Imbens, "The propensity score with continuous treatments," *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, 2004, *226164*, 73–84.

Imbens, Guido W., "Matching Methods in Practice: Three Examples," *Journal of Human Resources*, 2015, *50* (2), 373–419.

Kügler, Agnes and Christoph Weiss, "Time as a strategic variable: business hours in the gasoline market," *Applied Economics Letters*, 2016, *23* (15), 1051–1056.

Matsa, David A, "Competition and product quality in the supermarket industry," *The Quarterly Journal of Economics*, 2011, *126* (3), 1539–1591.

Olivares, Marcelo and Gérard P Cachon, "Competing retailers and inventory: An empirical investigation of General Motors' dealerships in isolated US markets," *Management Science*, 2009, *55* (9), 1586–1604.

Xie, Erhao, "Inference in Games without Equilibrium Restriction: An Application to Restaurants Competition in Opening Hours," *Journal of Business and Economic Statistics*, 2022, *40* (4).