Crowd-Driven Competitive Intelligence: Understanding the Relationship Between Local Market Competition and Online Rating Distributions

Dominik Gutt
Paderborn University
dominik.gutt@uni-paderborn.de

Philipp Herrmann
Consultant
philipp_herrmann@hotmail.de

Mohammad S. Rahman
Purdue University
mrahman@purdue.edu

Appendix

Investigating the Correlation Between Yelp and Zagat Online Ratings

In June and July 2015, we collected data on curated ratings for 28 major U.S. cities with a total of 4,993 restaurants from Zagat.com, a company that looks back on a long-standing history of providing restaurant ratings made by enlisted critics.[1] A series of correlation analyses and regressions on a matched sample of restaurants with Yelp and Zagat ratings reveal that Yelp and Zagat mean ratings exhibit a significantly strong and positive correlation.

Correlations between Yelp ratings and Zagat's food ratings are on average 0.4918 and significant ($p<0.0001$), as depicted in Table A1. The strength of the correlations, when considered on a city-by-city level, ranges from a minimum of $\rho = 0.2948$ to $\rho = 0.7627$ and remains significant throughout. For all 28 cities in our matched sample, the comparatively low value of 0.2948 for San Diego is the only value below 0.3, and most cities—comprising 4,238 out of the total 4,993 restaurants in our matched sample—are above 0.4. We also conducted separate correlation analyses for different cuisine types such as American, Barbeque, Burger, and Chinese and found that correlations ranged from $\rho = 0.7026$ for Barbeque to $\rho = 0.3511$ for Chinese. These values are above conventional thresholds of medium ($\rho = 0.3$) and large ($\rho = 0.5$) correlation effect sizes (Cohen 1992). To account for differences in rating scales (Zagat ratings are given on a scale from 1 to 30 and Yelp ratings on a scale from 1 to 5), we transposed Zagat ratings onto a scale from 1 to 5 and repeated the same correlation analysis. The results remained qualitatively and, by and large, quantitatively unchanged. To verify that the strong correlations were not driven by a higher number of reviews, we calculated separate correlations for restaurants with different numbers of reviews. Correlations ranged from $\rho = 0.3253$ for restaurants with fewer than 25 reviews to $\rho = 0.4881$ for restaurants with fewer than 1,100 reviews. However, the group of restaurants with fewer than 25 reviews comprised only 110 restaurants out of 4,993 total.

[1] We utilized undergraduate research assistance to manually match every restaurant between Yelp and Zagat.

Table A1: City-Level Correlation Between Yelp and Zagat Ratings

	Number of Restaurants	$corr(Yelp, Zagat)$
All Cities	4,993	0.4918
Atlanta, GA	266	0.4066
Austin, TX	256	0.6028
Baltimore, MD	317	0.5125
Boston, MA	197	0.433
Charleston, SC	76	0.5318
Charlotte, NC	62	0.3306
Chicago, IL	169	0.4481
Dallas, TX	94	0.5137
Denver, CO	266	0.4381
Houston, TX	381	0.4591
Las Vegas, NV	242	0.5045
Los Angeles, CA	258	0.4134
Miami, FL	117	0.4728
Minneapolis, MN	201	0.5822
Nashville, TN	48	0.3343
New Orleans, LA	172	0.4367
New York City, NY	399	0.4513
Orlando, FL	55	0.498
Philadelphia, PA	160	0.5765
Scottsdale, AZ	132	0.3962
Portland, OR	276	0.4254
Providence, RI	57	0.5686
San Antonio, TX	97	0.3563
San Diego, CA	352	0.2948
Savannah, GA	24	0.7627
Seattle, WA	80	0.4765
Tampa, FL	64	0.3794
Washington, DC	175	0.5161

Note: All correlations are significant at 1% except for Nashville, Tennessee, which is significant at 5%.

We also conducted a series of regressions of Yelp ratings on Zagat's food ratings to show that Zagat's rating is a strong predictor of Yelp ratings, as depicted in Table A2. These included regressions with neither city-level nor cuisine-level FE (column (1)), with only city-level FE (column (2)), with only cuisine-level FE (column (3)), and with both FEs combined (column (4)). Although Zagat features scores for food, service, and decor, the food score is the strongest predictor for Yelp ratings.

Table A2: Regression Analysis of Yelp and Zagat Ratings

	(1)	(2)	(3)	(4)
VARIABLES	*YELP*	*YELP*	*YELP*	*YELP*
ZAGAT_FOOD	0.30672***	0.30651***	0.33795***	0.33711***
	(0.01269)	(0.01689)	(0.02009)	(0.01388)
ZAGAT_DECOR	–0.03818***	–0.04765***	–0.05363***	–0.06279***
	(0.00899)	(0.01194)	(0.013)	(0.00962)
ZAGAT_SERVICE	0.08336***	0.10904***	0.04960***	0.07680***
	(0.01433)	(0.02326)	(0.01615)	(0.01516)
CITY-LEVEL FE		✓		✓
CUISINE-LEVEL FE			✓	✓
CONSTANT	2.36212***	2.30064***	2.40582***	2.45632***
	(0.04151)	(0.06913)	(0.05763)	(0.05766)
N	4,991	4,991	4,832	4,832
R^2	0.22674	0.2046	0.21414	0.33666

Note: *** p<0.01, ** p<0.05, * p<0.1, Robust standard errors in parentheses. For ease of interpretation we employ Zagat ratings transposed onto a scale of 1 to 5.

Table A3: First-Stage Results *LN_NUMRES*

	All Restaurants	Non-Fast-Food Restaurants	Traditional American Restaurants
VARIABLES	*LN_NUMRES*	*LN_NUMRES*	*LN_NUMRES*
REAL_EST	**0.008*** **	**0.009*** **	**0.009*** **
	(0.003)	(0.004)	(0.004)
FDIC	**0.007*** **	**0.006*** **	**0.004***
	(0.002)	(0.002)	(0.003)
MEDAGE	–0.018***	–0.022***	–0.021***
	(0.005)	(0.006)	(0.006)
MEDINC	–0.004	–0.001	–0.002
	(0.004)	(0.004)	(0.004)
MEDHOUSEVAL	–0.001	–0.002	–0.001
	(0.001)	(0.001)	(0.001)
MEDGROSSRENT	0.051	0.102**	0.087**
	(0.042)	(0.044)	(0.039)
UNEMP	0.010	0.032**	0.026**
	(0.017)	(0.013)	(0.013)
LANDAREA	1.007***	0.56492*	0.692***
	(0.313)	(0.291)	(0.246)
URBAN	0.027***	0.038***	0.029***
	(0.009)	(0.007)	(0.011)
COSTOFLIV	0.002	–0.005	–0.006
	(0.007)	(0.008)	(0.009)
PERCHISP	0.035	–0.119	–0.210
	(0.223)	(0.184)	(0.203)
PECINDIAN	–0.119	–0.705	–0.058
	(0.549)	(0.590)	(0.635)
PERCASIAN	–0.775	–0.346	–1.39463*
	(0.973)	(0.822)	(0.776)
PERCBLACK	–0.026	–0.307*	–0.542**
	(0.166)	(0.184)	(0.218)
RESPOP	0.010***	0.054***	0.102***
	(0.002)	(0.009)	(0.015)
Constant	1.969***	1.017	0.523
	(0.649)	(0.777)	(0.844)
N	348	326	258
F	**61**	**46**	**42**
Adjusted R^2	**0.6509**	**0.6577**	**0.6952**

Note: *** p<0.01, ** p<0.05, * p<0.1, Robust standard errors in parentheses.

Table A4: OLS Results with *HHI* as a Substitute for *LN_NUMRES*

	All Restaurants		Non-Fast-Food Restaurants		Traditional American Restaurants	
VARIABLES	*RANGE*	*AVG*	*RANGE*	*AVG*	*RANGE*	*AVG*
HHI	**–4.09456*** **	**0.37792** **	**–2.98058*** **	**0.33172***	**–2.46366*** **	**0.57596*** **
	(0.33122)	(0.18874)	(0.12863)	(0.19319)	(0.17797)	(0.18137)
MEDAGE	–0.01198*	0.00806**	–0.02379***	0.01351**	–0.01042	0.00954
	(0.00621)	(0.00399)	(0.00834)	(0.00569)	(0.00996)	(0.00796)
MEDINC	0.00706	–0.00524*	0.01554**	–0.00494	0.01147	–0.01192**
	(0.00509)	(0.00316)	(0.00617)	(0.00463)	(0.00781)	(0.00555)
MEDHOUSEVAL	–0.00102	0.00056	–0.00320***	–0.0001	–0.00252**	0.00006
	(0.00085)	(0.00056)	(0.00115)	(0.00081)	(0.00106)	(0.001)
MEDGROSSRENT	–0.03067	0.03923	–0.01619	0.08541*	0.0194	0.08253
	(0.04)	(0.02788)	(0.05356)	(0.04448)	(0.06419)	(0.05273)
UNEMP	–0.00311	0.00878	–0.00823	0.01124	–0.00489	0.0129
	(0.01513)	(0.00897)	(0.01943)	(0.01231)	(0.01963)	(0.01869)
LANDAREA	0.71415*	0.61416**	0.25153	0.28602	0.30657	0.00403
	(0.4211)	(0.25093)	(0.54457)	(0.28018)	(0.58379)	(0.40599)
URBAN	0.01577**	–0.00065	0.03533***	0.0047	–0.01784	0.01518**
	(0.00774)	(0.01024)	(0.01084)	(0.00745)	(0.01277)	(0.00739)
COSTOFLIV	0.01268	–0.00322	0.01569	–0.00709	0.01451	–0.00877
	(0.01014)	(0.00554)	(0.01038)	(0.00873)	(0.00895)	(0.0099)
PERCHISP	–0.09125	–0.05317	0.06725	–0.06877	0.25718	–0.25037
	(0.1928)	(0.11298)	(0.22904)	(0.17576)	(0.29507)	(0.25085)
PECINDIAN	–0.41117	–0.136	–1.89270***	–0.11202	–2.12335*	0.05796
	(0.74308)	(0.42753)	(0.68884)	(0.61128)	(1.25061)	(0.74799)
PERCASIAN	–0.85474	1.53683*	–1.32652	–0.07333	–0.98811	–0.90042
	(0.87113)	(0.79783)	(1.40469)	(0.68741)	(0.89416)	(0.82001)
PERCBLACK	0.06905	–0.17061	0.38287	0.03344	0.41769	–0.01785
	(0.24058)	(0.14322)	(0.36327)	(0.27678)	(0.41772)	(0.36196)
LOC_POP	0.00275*	–0.00081	0.02025***	–0.00066	0.03378**	0.00546
	(0.00144)	(0.00106)	(0.00503)	(0.00432)	(0.01478)	(0.01014)
CONSTANT	2.66901***	2.65958***	2.18665*	2.87798***	1.2073	3.46059***
	(1.00051)	(0.5724)	(1.25547)	(0.88325)	(1.13029)	(1.06617)
N	348	348	326	326	258	258
R^2	0.45496	0.06653	0.58449	0.05126	0.50028	0.09666

Note: *** p<0.01, ** p<0.05, * p<0.1, Robust standard errors in parentheses.

References in the Appendix

Cohen J (1992) A Power Primer. *Psychological Bulletin* 112(1), 155–159.