

Online Appendix to “Promotional Reviews: An Empirical Investigation of Online Review Manipulation”

By DINA MAYZLIN, YANIV DOVER, JUDITH CHEVALIER

I. Robustness Check

Until late 2010, Orbitz, like Expedia, only accepted reviews from individuals who had booked their stay at Orbitz.com. Starting in late 2010, Orbitz allowed others to submit hotel reviews, but reviews from verified customers are identified as “Verified” and are given higher weight in calculating the Orbitz Reviewer Score for each property. Here we use only verified reviews from Orbitz. In our sample, only 1.2% of Orbitz reviews were unverified. Hence, while the Orbitz data is not a perfect substitute for Expedia data due to Orbitz’s hybrid model, its smaller size, and the fact that there may sharper differences between TripAdvisor and Orbitz reviewer preferences, we expect that Orbitz’s verified reviews are costlier to manipulate than TripAdvisor’s reviews. If our results are driven by important (and subtle) differences between the customer pools at Expedia and TripAdvisor, robustness of our results for Orbitz may be valuable.

In particular, in January 2013, a little more than a year after the original sample was collected, we obtained TripAdvisor and Orbitz reviews for the hotels in our original sample.¹ For our hotels in the 2013 sample, we have a total of 105,232 verified Orbitz reviews; in contrast the number of TripAdvisor reviews is 486,769. Summary statistics are shown in Table 1. One characteristic of Orbitz reviews that stands out is the relatively low share of 5- star reviews.

We estimate the following equation:

¹The reason behind the later data collection effort is that Orbitz had relatively few reviews in October of 2011: 1) there were 104 hotels that had reviews at TripAdvisor and Expedia but no reviews on Orbitz, and 2) for the hotels with reviews at both Orbitz and TripAdvisor, the number of reviews at Orbitz was about three-quarters of the number of Expedia reviews. Unfortunately, during our later data collection effort we also discovered that Expedia had expanded its reviews in a way that compromised our identification effort. In March of 2012, they merged their original database of verified reviews with reviews from other sites. Since we could not ascertain that all the new reviews had been verified in the same manner as the earlier reviews, we could not re-do the analysis with the more recent Expedia data. Note, however, that the change took place after our original sample was collected.

TABLE 1—USER REVIEWS AT TRIPADVISOR AND ORBITZ, JANUARY 2013

	Mean	Standard deviation	Minimum	Maximum
Number of TripAdvisor reviews	193.78	261.66	1	2468
Number of Orbitz reviews	41.89	62.68	1	1077
Average TripAdvisor star rating	3.61	0.70	1	5
Average Orbitz star rating	3.61	0.68	1	5
Share of TripAdvisor 1-star reviews	0.12			
Share of TripAdvisor 2-star reviews	0.10			
Share of Orbitz 1-star reviews	0.10			
Share of Orbitz 2-star reviews	0.09			
Share of TripAdvisor 5-star reviews	0.33			
Share of Orbitz 5-star reviews	0.23			
Total number of hotels	2512			

Note: The sample consists of the subset of hotels in our base sample that 1) were also available on TripAdvisor and Orbitz in January 2013, and 2) had at least one review on TripAdvisor and one verified review on Orbitz at that time.

$$(1) \quad \frac{NStarReviews_{ij}^{TA}}{TotalReviews_{ij}^{TA}} - \frac{NStarReviews_{ij}^{Orb}}{TotalReviews_{ij}^{Orb}} = X_{ij}B_1 + OwnAf_{ij}B_2 + Nei_{ij}B_3 + \\ NeiOwnAf_{ij}B_4 + \sum \gamma_j + \varepsilon_{ij}$$

These results are reported in Table 2. This analysis repeats the regression specifications of Table 3 in the main body of the paper, replacing Orbitz verified reviews with Expedia reviews. Regressions results are qualitatively similar to the results found in Table 3 of the paper. Turning to the 1- and 2-star reviews, we find, as in Table 3, the own hotel owner characteristics are small and insignificant. However, we also find small and insignificant hotel neighbor effects. Note that the neighbor effects are all in the hypothesized direction. As in Table 3, we find that independent hotels have more 5-star reviews on TripAdvisor versus Orbitz, and hotels from large ownership entities have fewer 5-star reviews. In the Orbitz specification, the magnitude of the independence effect is somewhat larger than in our Expedia specifications, while the magnitude and significance of the multi-unit owner effect is smaller (albeit the effect is still significant at 10% significance). Overall, we take these results as suggestive that our findings are robust when

examining alternative sites.

TABLE 2—TRIPADVISOR VERSUS ORBITZ RESULTS

		Difference in share of 1- and 2-star reviews	Difference in share of 1- and 2-star reviews	Difference in share of 5- star reviews
X_{ij}	Site rating	0.0035 (0.0069)	0.0043 (0.0070)	0.0041 (0.0081)
	Hotel age	0.0003** (0.00015)	0.0003* (0.0002)	-0.0009*** (0.0002)
	All Suites	0.0027 (0.0088)	0.0030 (0.0088)	0.0147 (.0104)
	Convention Center	0.0190** (0.0095)	0.0206** (0.0101)	-0.0405*** (0.0124)
	Restaurant	0.0078 (0.0073)	0.0073 (0.0074)	-0.0081 (0.0098)
	Hotel tier controls?	Yes	Yes	Yes
	Hotel location controls?	Yes	Yes	Yes
$OwnAf_{ij}$	Hotel is Independent		0.0080 (0.0103)	0.045*** (0.0107)
	Multi-unit owner		-0.0018 (0.0060)	-0.0146* (0.008)
Nei_{ij}	Has a neighbor	0.0127 (0.0082)	0.0151 (0.0102)	-0.0044 (0.0113)
$NeiOwnAf_{ij}$	Has independent neighbor		0.0028 (0.0085)	0.006 (0.0096)
	Has multi-unit owner neighbor		-0.0044 (0.0079)	0.0124 (0.0095)
γ_j	City-level fixed effects?	YES	YES	YES
	Num. of observations	2512	2512	2281
	R-squared	0.03	0.03	0.07

Note: *** p<0.01, ** p<0.05, * p<0.10
Heteroskedasticity robust standard errors in parentheses. All neighbor effects calculated for 0.5km radius. Regression estimates of Equation (1). The dependent variable in all specifications is the share of reviews that are N star for a given hotel at TripAdvisor minus the share of reviews for that hotel that are N star at Orbitz. Heteroskedasticity robust standard errors in parentheses.

II. Comparative Statics under Endogenous Prices

Here we continue the proof in the Appendix of the main body of the paper. We next derive the equilibrium prices, which are set in stage II. The firm does not expect manipulation to change its market share in expectation, given the optimal discounting by the consumer. Hence, the maximization problem in stage II is the following

$$(2) \quad \Pi_{A,Stage 2}^* = \max_{p_A} p_A \left[\frac{1}{2} + \frac{p_B - p_A}{2t} \right] - \delta_A \frac{(e_{A,A}^*)^2}{2} - \gamma_A \frac{(e_{A,B}^*)^2}{2}$$

$$(3) \quad \Pi_{B,Stage 2}^* = \max_{p_B} p_B \left[\frac{1}{2} + \frac{p_A - p_B}{2t} \right] - \delta_B \frac{(e_{B,B}^*)^2}{2} - \gamma_B \frac{(e_{B,A}^*)^2}{2}$$

After the appropriate substitutions (Proposition 1 in the Appendix of the paper provides $e_{A,A}^*$, etc.), taking the first order conditions, and some algebra, we have the following expressions for the equilibrium prices:

$$(4) \quad p_A = \frac{3t + 2tw_B}{4(1 + w_A)(1 + w_B) - 1}$$

$$(5) \quad p_B = \frac{3t + 2tw_A}{4(1 + w_A)(1 + w_B) - 1}$$

where

$$w_A = \frac{\delta_A + \gamma_A}{4t(\delta_A\gamma_A)}; \quad w_B = \frac{\delta_B + \gamma_B}{4t(\delta_B\gamma_B)}$$

We can show that

$$\frac{\partial e_{A,A}^*}{\partial \delta_A} = \frac{\mu_s}{2t} \left[\frac{\frac{\partial p_A}{\partial \delta_A} \delta_A - p_A}{\rho^2} \right] = (1 + w_B) - t\delta_A [4(1 + w_A)(1 + w_B) - 1]$$

In particular, this derivative is guaranteed to be negative if $t\delta_A > 1$, the same result as in the case of exogenous prices (see Corollary 1 in the Appendix of the paper). If $t\delta_A > 1$, we also obtain that

$$\frac{\partial e_{A,B}^*}{\partial \gamma_A} < 0, \quad \frac{\partial e_{B,B}^*}{\partial \delta_B} < 0, \quad \text{and} \quad \frac{\partial e_{B,B}^*}{\partial \gamma_B} < 0$$