Shirking with Good Reputation? Evidence from Hotel Industry*

Fang-Chang Kuo†

April 21, 2021

Preliminary and Incomplete. Comments are Greatly Appreciated.

Abstract

This paper empirically examines the impact of online reputation on investment in hotel industry. Recent theory suggests that reputation could have ambiguous effects on investments. Using detailed firm-level data on investment expenditures, and online consumer ratings from Taiwanese hotels, I adopt a regression discontinuity design based on TripAdvisor's rating display system and identify treatment effects. The regression discontinuity estimates show that that higher ratings negatively impact investment expenditures while lower ratings tend to encourage investment. The findings are consistent with Board and Meyer-ter-Vehn (2013), in which their good news model predicts that firms shirk when they have good reputation.

Keywords: reputation; investment; regression discontinuity design; hotel industry

*This paper is revised from a chapter of my Ph.D. dissertation at University of Texas at Austin. I am indebted to my advisors Eugenio Miravete, Daniel Ackerberg, Robert Town, and Jason Duan. I also thank seminar participants at the UT-Austin, TER 2019, and Fu Jen Catholic University for their valuable comments. Any remaining errors are my own.

†National Chung Cheng University, Department of Economics, No.168, Sec.1, University Rd., Minhsiung, Chiayi 62102, Taiwan; Email: *fckuo@ccu.edu.tw*; https://fangchangkuo.weebly.com/

1 Introduction

In the past two decades, reputation mechanisms have been widely used in many markets. Consumers regularly consult online ratings on Yelp, Amazon, or TripAdvisor before making purchase decisions. In turn, firms, or products with better reputation are rewarded by consumers with more sales, higher prices, and lower exit probabilities. While the impact of reputation on consumer demand is well-studied in the literature (Chevalier and Mayzlin 2006, Cabral and Hortacsu 2010, Anderson and Magruder 2012, Luca 2016), little do we know about supply side responses to reputation. In particular, how does reputation affect firm's incentives to invest in product quality?

The relationship between reputation and investment is ambiguous and remains unexplored empirically. The only theoretical prediction comes from Board and Meyer-ter Vehn (2013). On the one hand, firms may view their reputation as a valuable asset and try to maintain it by keep investing in quality. In this situation, reputation will have positive impact on investment. On the other hand, firms with good reputation may run down its reputation by delaying investment because consumers believe that product quality is still good. Firms are essentially leveraging asymmetric information as consumers can't perfectly observe product quality. In this case, better reputation tends to give firms weaker incentives to invest.

In this paper, I study the relationship of reputation and investment using a novel dataset with information on hotel's investment expenditures and online consumer ratings from TripAdvisor. To identify the casual impact of reputation, I use a regression discontinuity design (RDD) exploiting the rounding thresholds in TripAdvisor Bubble Rating System.

The specific setting I consider is that of the hotel industry in Taiwan. The hotel industry is ideal when it comes to answering the research question in this paper. First, hotels provide an experience good which consumers are unable to observe the true quality before actually staying at the hotel. Consumers need to make choice based on reputation, which is commonly defined as consumer belief about product quality.

Online rating is one of the important pieces of information in a hotel's reputation as it shows experiences from past consumers, and reveals perceived product quality in a straightforward manner. This makes hotel industry one of the earliest industries to adopt online ratings. Second, product quality is not fixed in this market as the investment hotels make are directly linked to product quality. Unlike other industries, the capacity, number of rooms for a hotel, is largely fixed based on initial construction (Kalnins 2006). Besides, the marginal costs are very low for hotel industry. Therefore, the investment is unlikely to increase capacity or decrease marginal cost of production. These features result in a relatively straightforward channel for investment to improve product quality with few confounding factors. Tangible quality of a hotel may also deteriorates due to wear and tear as age grows. Hotel can maintain quality through regular maintenance or major renovation. The setting is fairly close to the one considered in Board and Meyer-ter Vehn (2013).

Empirical strategy in this paper leverages the fact that consumer ratings on TripAdvisor are rounded off to nearest half-bubble in a five point scale. Using a regression discontinuity (RD) design, I compare investments close to cutoffs. The empirical results show heterogeneous responses at different cutoffs. Hotels tend to invest less when their TripAdvisor ratings jump from 4 points to 4.5 points while hotels with lower ratings, 3 to 3.5, increases their investment expenditures. Both intensive and extensive margins exhibit similar pattern. This results are consistent with the work-shirk equilibrium in Board and Meyer-ter Vehn (2013). Hotels with better ratings may run down its reputation by lowering investment since consumers believe that product quality is still good, and marginal return to investment is capped. Meanwhile, hotels with lower ratings keep investing in product quality as long as marginal cost is lower than marginal benefit of investment.

To the best of my knowledge, this paper is the first to empirically examine the relationship of reputation and investment. Understanding the relationship is of interest for several reasons. First, while Board and Meyer-ter Vehn (2013) provide a nice model to analyze the relationship, different type of signals lead to distinct work-shirk or shirk-work equilibrium. Therefore, there is no universal prediction for all kinds of markets or industries. It becomes an empirical question to analyze this relationship case by case. Second, the question is related to the design of reputation mechanism. With the increasing importance of reputation mechanisms, e-commerce platforms such as Taobao, eBay, and Amazon, or online platforms like TripAdvisor, Yelp, or Google Map may design different reputation mechanisms to improve market outcomes and compete with other platforms. It would be necessary to understand potential impacts, not only on prices, or sales, but also on endogenous product characteristics from sellers since dynamic investment in product quality can directly affect consumers welfare. Third, the relationship helps understand the investment dynamics when reputation concerns are important in a market. Why do some firms keep investing, while some firms do not? This paper sheds some light on investment dynamics from a reputational perspective.

Empirical findings in this paper adds to the board literature on the impact of user ratings with the focus of impact of ratings on endogenous product qualities through investment decisions. In this stand of literature, Chevalier and Mayzlin (2006) studies the impact of online reviews on book sales on Amazon.com and Barnesandoble.com, and find positive impact. Anderson and Magruder (2012) and Luca (2016) use restaurants ratings on Yelp.com, and adopt regression discontinuity framework. Anderson and Magruder (2012) find that higher ratings causes restaurants to sell out more frequently. Luca (2016) shows positive revenue effect of ratings. This paper is closely related to Hollenbeck et al. (2019) since both are interested in hotel's strategic reaction to online ratings and exploiting TripAdvisor display rules in a RD design. Hollenbeck et al. (2019) examines advertising strategy and identifies a negative effect of online rating on ad expenditures while this paper focuses on investment strategy. This paper is also related to Jin and Leslie (2003) as both papers look into the effects of information on endogenous quality decisions. Jin and Leslie (2003) takes advantage of a mandatory disclosure of hygiene grade, and finds that the mandatory disclosure policy, which exogenously increases information provision, improved product qualities, measured by health inspection scores and the number of foodborne illness hospitalizations.

This paper is built on and is motivated by theoretical predictions from Board and Meyer-ter Vehn (2013). They model quality as an endogenous type instead of a exogenous one, and allow firm to make investment to dynamically affect product quality. Consumers receive signs of qualities in different information structures, good news and bad news. They find that in good news model there exists a work-shirk equilibrium in which better reputation gives weaker incentive to invest in quality. Consistent with work-shirk equilibrium, empirical results indicate that better-rated hotels have weaker investment incentives both o the intensive and extensive margin.

There are growing literature on the welfare effects from online ratings or reputation mechanisms. For example, Saeedi (2019) uses dynamic model to quantify the welfare effects from eBay registered-store and powerseller status, and finds that reputation mechanism increases market size by 61%. Lewis and Zervas (2016) quantifies the welfare effects of online reviews in the hotel industry. Removing review information will lower consumers surplus by around $ 124 million. Fang (2020) studies how online review platforms, Yelp, help consumers learn about product qualities using restaurant revenue data in Texas, and finds that Yelp speeds up the learning process. Compared to these recent papers, this paper does not try to quantify welfare effects through a structural model. Instead, it uses a reduced-form approach and highlights that reputation could adversely affect endogenous product quality and consumers' welfare might therefore be lower.

The remainder of this paper is organized as follows. Section 2 introduces the data and presents descriptive evidence. Section 3 conducts a regression discontinuity analysis. In Section 4, I provides some robustness checks. Section 5 concludes this paper.

2 Data and Descriptive Evidence

To identify the effect of online ratings on hotel's investment incentives, I construct a monthly panel of TripAdvisor ratings and hotel investment expenditures. Hotel investment information from 2009 to 2016 is acquired from the Bureau of Tourism in Taiwan. The Bureau provides a panel hotel financial performances at hotel-year-month level for over 95% of legitimate hotels in Taiwan.[1] The panel contains information of room revenues, average daily rates, sales, employment, number of rooms, customer information, and expenses in fixed assets and durable goods. Monthly expenses spent on durable goods and fixed assets are used to define investment variable.[2]

Panel of hotel investment expenditure is supplemented with with a panel of consumer ratings from TripAdvisor.[3] TripAdvisor is one of the most popular travel review platforms. As of September 2020, TripAdvisor provides over close to 1 billion reviews over 9 million accommodations, restaurants, experiences, and airlines.[4] Consumer ratings on hotels are used to measure hotels' online reputation. Individual consumer reviews are scraped from TripAdvisor. Each review contains hotel information, stay date, review date, and consumer ratings from 1 to 5. The TripAdvisor dataset contains 569,858 consumer ratings in 1,324 hotels. I then calculate monthly cumulative average rating for each hotel in every month.

Average ratings are converted into TripAdvisor Bubble Ratings according to the TripAdvisor rounding rules. Figure 1 shows a sample search result on TripAdvisor. Instead of showing numeric values of average ratings, average rating for each hotel is rounded

[1] The data was collected originally for administrative purposes.

[2] Under this definition, this study implicitly restricts investment in quality to the tangible dimension. Intangible quality like service quality could also be improved through on-job training or other human capital investments. However, these investments in other quality dimensions are not observed in this data.

[3] Focusing only on TripAdvisor has its own limitations. Apparently, I do not incorporate information from other rating platforms. Ratings on other platforms may not be irrelevant. TripAdvisor ratings may not be representative measure of reputation for some hotels. In other words, hotels may not respond to the discrete ratings on TripAdvisor since they may have significantly better or worse ratings on other platforms.

[4] See: https://tripadvisor.mediaroom.com/US-about-us

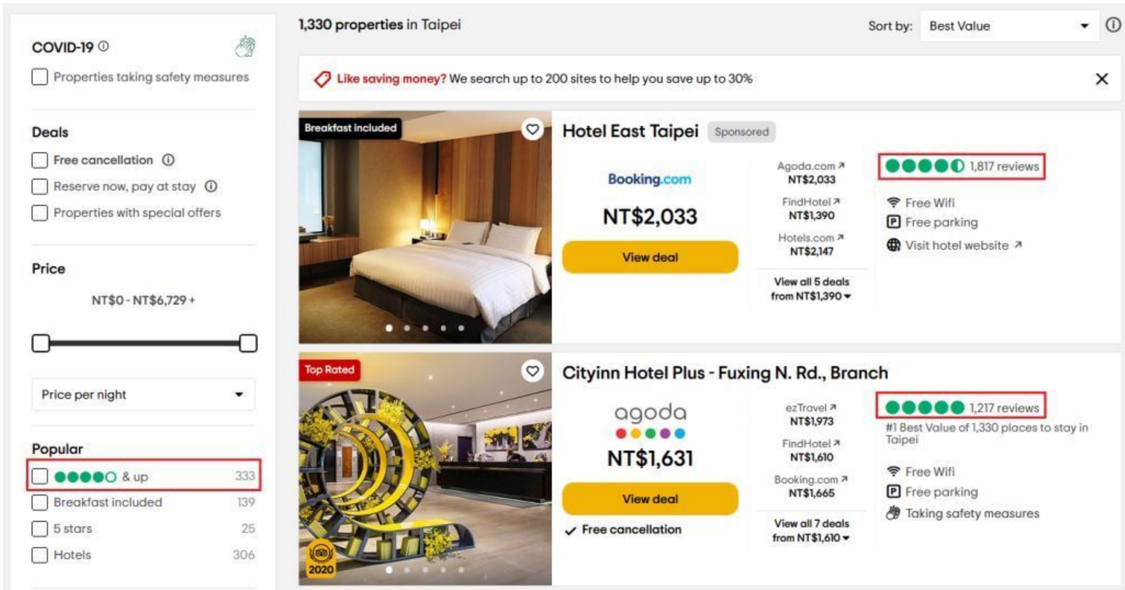


Figure 1: Tripadvisor's Search Results and Filter

to the nearest half-bubble, and is prominently displayed in light green color. Consumer sees these whole and half bubbles. The rounding rules can be described by a step function which includes jumps for every 0.5 gap. Figure 2 summarizes the step function from average rating of 2.5 and above. This step function features discontinuities when the average rating moves around thresholds. For example, an average rating of 4.23 will be displayed as 4 in bubble rating while average 4.25 gets additional half bubble with 4.5 bubble rating.

Two data sets are matched using hotel information, and year-month. The final data contains 38,516 hotel-year-month observation from 2009 January to 2016 June. For each observation, main variables are investment expenditure, average TripAdvisor rating, and TripAdvisor Bubble Ratings. For the rest of my analyses, I limit observations to hotels with 25 or more reviews to reduce impacts of extreme ratings. Figure 3 shows the distribution of hotels over discrete Bubble Ratings. Most of the hotels have Bubble Ratings in 3.5, 4.0, and 4.5 while lower or higher ratings only account for small fractions. This pattern is similar to those on other platforms.[5]

[5] Fang (2020) presents restaurants ratings on Yelp, Google, and TripAdvisor. The distributions exhibits similar patterns. Unconditional distribution is presented in Figure A1.

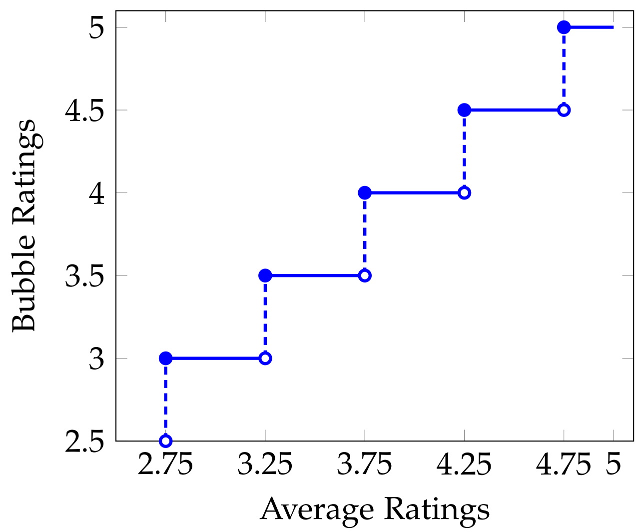


Figure 2: TripAdvisor Bubble Rating System

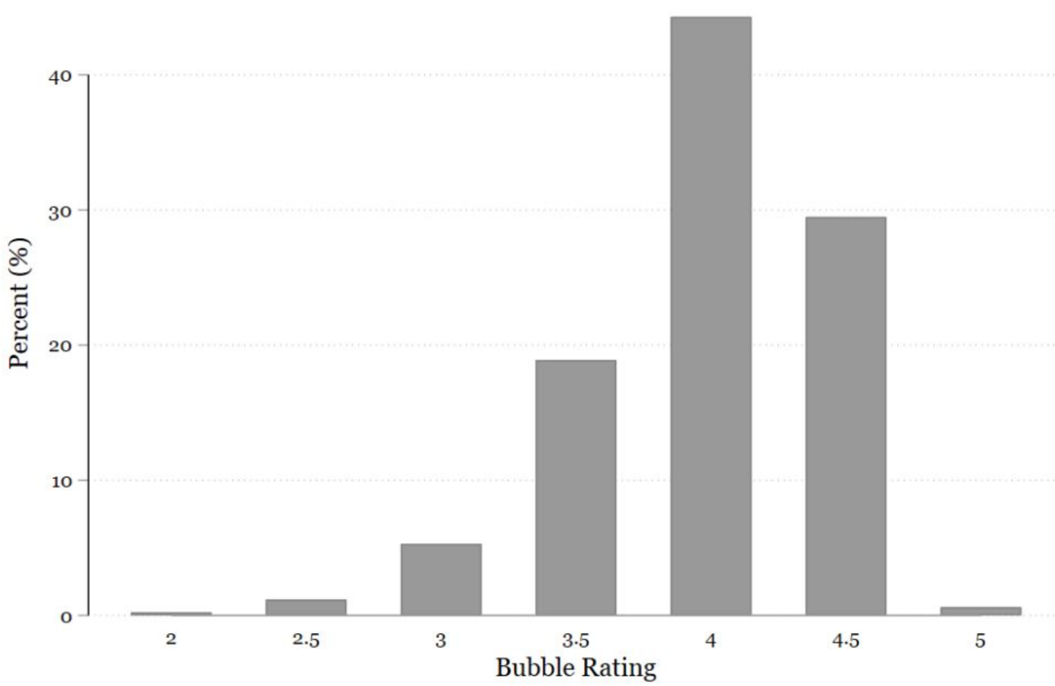
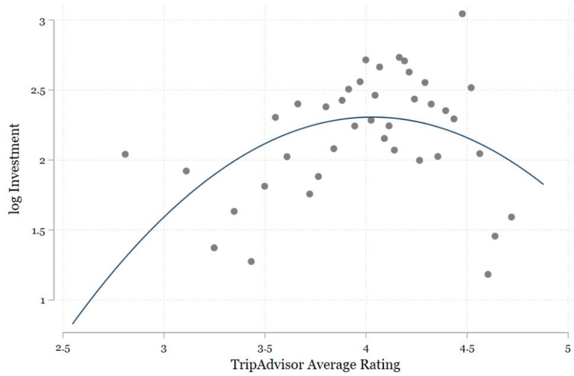
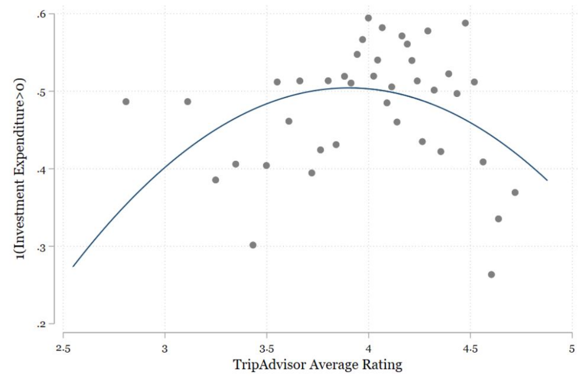


Figure 3: Distribution of Hotels over Tripadvisor Bubble Ratings

Notes: Only hotels with 25 or more reviews are included.



(a) Binned Scatter Plot: Intensive Margin



(b) Binned Scatter Plot: Extensive Margin

Figure 4: Relationship between investment and TripAdvisor average rating

Notes: The above two figures present binscatter plots and associated quadratic fits in two different measures of investment. The x-axis is the cumulative average TripAdvisor rating in the previous month. Only hotels with 25 or more reviews are included.

With the dataset, Figure 4 presents binned scatter plots for the unconditional relationship between average TripAdvisor rating and investment. Figure 4a uses logarithm of investment expenditures to measure investment intensity, while Figure 4b considers binary investment decisions as positive expenditures are treated as 1 and 0 otherwise. Both figures show inverted-U relationships. Investment incentives are the strongest when average ratings lie between 4.0 and 4.5. As rating gets better, investment decreases. This descriptive pattern is consistent with the work-shirk equilibrium in Board and Meyer-ter Vehn (2013).

3 Empirical Analysis

My empirical strategy exploit rounding thresholds on TripAdvisor. Instead of showing raw average ratings, TripAdvisor displays ratings in a discrete fashion. Average ratings above cutoffs will receive an additional half point more than those just below cutoffs. I apply a sharp regression discontinuity design comparing investment decisions of hotels around these thresholds.

Following notations in Lee and Lemieux (2010), RD treatment effect parameter τ is defined by

$$\tau = \lim_{x \downarrow c} E[Y_{it}|X_{it} = x] - \lim_{x \uparrow c} E[Y_{it}|X_{it} = x] \tag{1}$$

Empirical model can be written as:

$$\begin{aligned} &Y_{it} = \alpha + \tau\, \mathbf{1}(X_{it} > c) + \beta_1 (X_{it} - c) + \beta_2\, \mathbf{1}(X_{it} > c) \times (X_{it} - c) + \epsilon_{it}, \\ &\forall X_{it} \in (c - h, c + h) \end{aligned} \tag{2}$$

where Y_i is the outcome variable, investment; X_i is the running variable, TripAdvisor average rating in previous period; $c \in \{3.25, 3.75, 4.25, 4.75\}$ is one of the thresholds in TripAdvisor Bubble Rating System; h is the bandwidth around thresholds; τ is the coefficient of interest as it captures the discrete jump at cutoffs; β_1 and β_2 are separate slopes below and above cutoffs which allow for flexible linear relationships.

Estimation follows nonparametric approach proposed by Hahn et al. (2001) by running local linear regression to estimate the limits around discontinuities. Bandwidth is chosen based on a data-driven algorithm proposed by Imbens and Kalyanaraman (2012). Alternative bandwidths and order of polynomials will also be used in the empirical estimation.

3.1 Validity Checks

In this section, I first show two validity check before presenting RD results. First, Figure 5 show a histogram of running variable, average rating. The cutoffs are highlighted by vertical red dash lines. The bin width is set to 0.02 in order to observed number of observations above and below cutoffs. For cutoffs at 3.25, 4.25, and 4.75, the number of observations around thresholds. However, at 3.75 cutoff, the density above the cutoff is significantly higher than that right below the cutoff, indicating potential manipulation to increase the Bubble Rating. I further plot densities at various cutoffs following McCrary (2008) in Figure 6. It seems Figure 6b exhibits potential discontinuity at cutoff 3.75 while

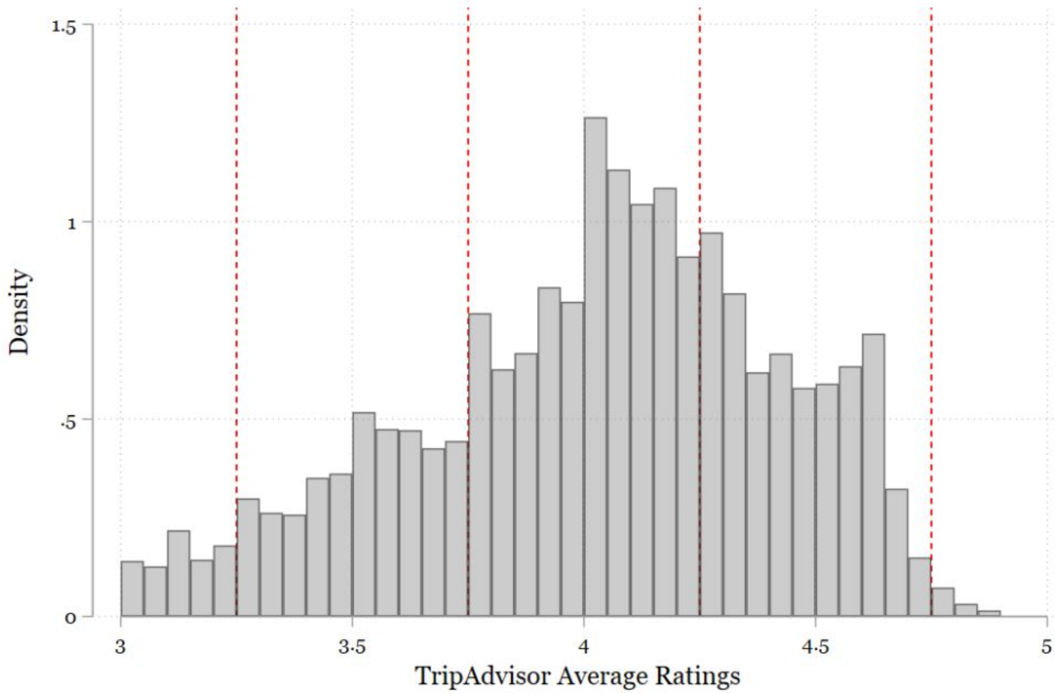


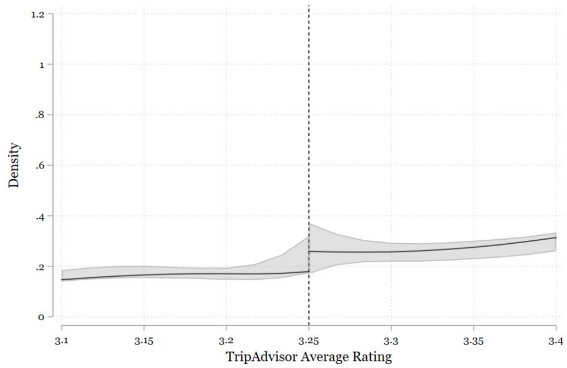
Figure 5: Histogram of TripAdvisor Average Ratings

Notes: The above figure shows histogram of TripAdvisor average ratings above 3 with bin width of 0.05. Red dash lines indicate various cutoffs. Alternative bin width 0.025 is used in Figure **??**. This figure only includes observations with more than 20 reviews.

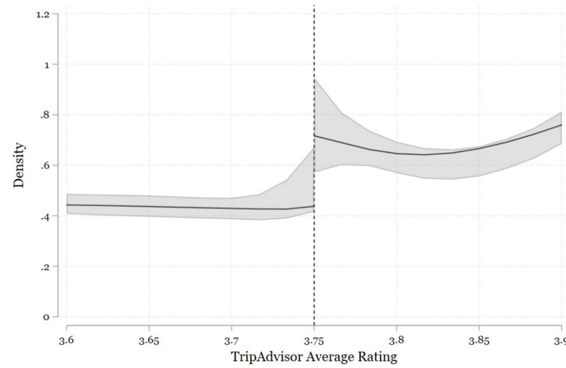
others look relatively continuous around cutoffs.[6]

Following Cattaneo et al. (2020), I conduct formal manipulation tests at four cutoffs. Test results in Table 1 confirm graphical intuitions from Figure 5 and Figure 6. At 3.75 cutoff, the T-statistics is approximately 3, which is significant at any reasonable significant level. Other cutoffs do not exhibit noticeable discontinuities. One potential explanation of such result can be drawn from one of the search filter options in Figure 1. The search filter options allow consumers to refine their search for hotels according to multiple characteristics. One option is to limit hotels to four bubbles or more. In this case, the filter immediately reduces the number of hotels from 1330 to 333, excluding

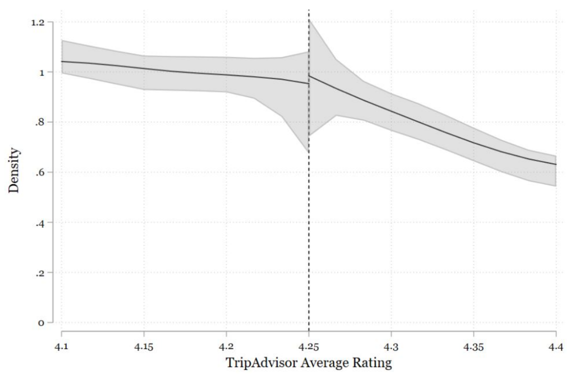
[6] As Hollenbeck et al. (2019) also focuses on TripAdvisor cutoffs for hotels, they do not conduct separate density tests for each threshold. Instead, they aggregate all observations around thresholds by calculating distance from thresholds. Anderson and Magruder (2012) also aggregate Yelp's ratings.



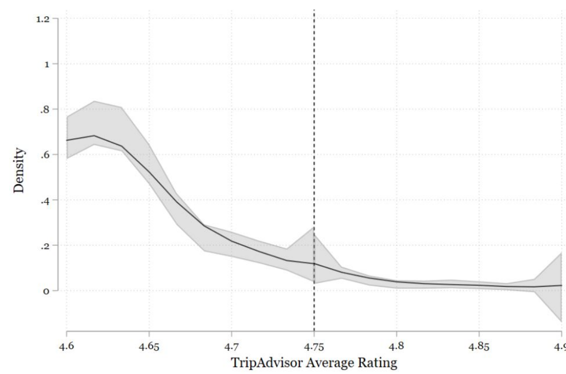
(a) 3.25



(b) 3.75



(c) 4.25



(d) 4.75

Figure 6: Density Tests at Four Cutoffs

75% of hotels. It provides additional strong incentives for hotels to acquire 4 bubbles if their average ratings are just below the cutoff 3.75. For other cutoffs, without such filter option, the densities are relatively continuous and are valid for regression discontinuity design.

Another validity test consider predetermined characteristics. The idea is that if the assignments around cutoffs are random, then the predetermined characteristics should be very similar for both sides at cutoffs. In Table 2, I estimate RD parameters at all four cutoffs for five baseline covariates.[7] None of the estimate are statistically significant. Besides, the signs are mixed for the same covariates, indicating that observation above

[7] Perhaps some covariates are not determined completely before the assignment of bubble ratings, I still include them in the analysis.

Table 1: Manipulation Tests

	Cut-off points:			
	3.25	3.75	4.25	4.75
T-statistic	0.503	3.012	1.207	−0.322
P-value	0.615	0.003	0.228	0.747

Notes: This table reports results from RD manipulation tests using local polynomial density estimation. I use triangular kernel density function for the tests. For detailed implementations, please see Cattaneo et al. (2018). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: RD Estimates for Baseline Covariates

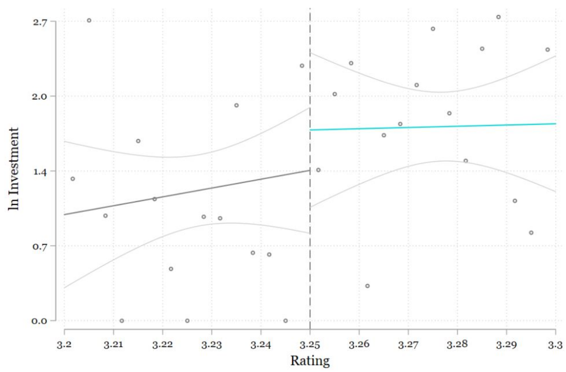
Outcome	Baseline Covariates				
	(1) Age	(2) Brand	(3) No. of rooms	(4) No. of reviews	(5) Empl. per room
3.25 RD	3.580	0.000	17.925	−10.558	0.087
	(4.325)	(0.000)	(37.789)	(11.193)	(0.073)
3.75 RD	−1.644	0.096	21.639	27.553	0.040
	(6.353)	(0.161)	(47.433)	(31.101)	(0.133)
4.25 RD	−1.218	0.022	−37.826	−36.637	0.031
	(2.453)	(0.094)	(39.475)	(49.796)	(0.233)
4.75 RD	1.566	−0.017	33.729	−66.151	0.113
	(3.036)	(0.240)	(84.876)	(84.924)	(0.249)

Notes: The estimation results are based on local-linear non-parametric RD estimation. Only hotels with more 25 reviews are included. Bandwidths are fixed at 0.05. All specifications use triangular kernel function. Standard errors are robust and clustered at firm level.

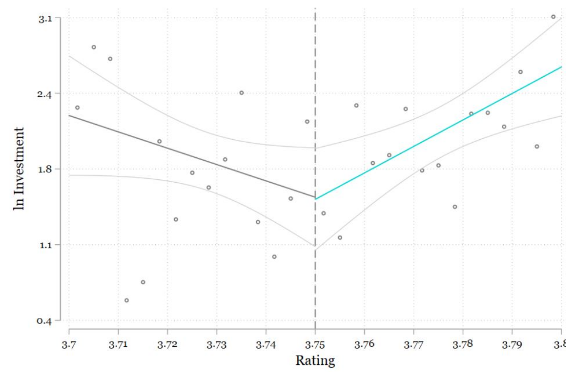
and below cutoffs do not differ in a systematic way.

3.2 RD Results

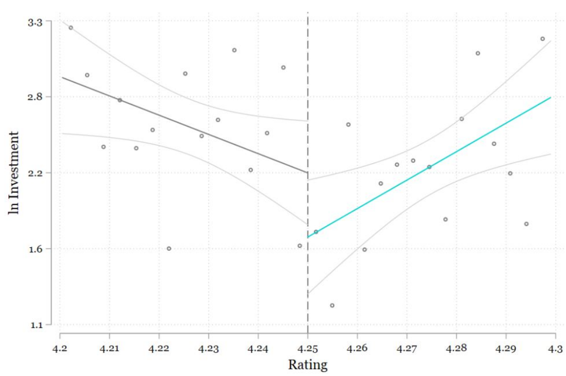
I first present visual evidence that hotels strategically change investment behaviors in response to different jumps of their online rating. Figure 7 show the relationship between online rating in last period and logarithm of investment expenditure in this period at four different thresholds. There are positive jumps at cutoffs of 3.25, and 4.75. But these jumps are not clear. At 4.25 cutoff, the gap is negative and obvious, indicating the effect of moving from 4 Bubble to 4.5 Bubble on hotel investment is negative.



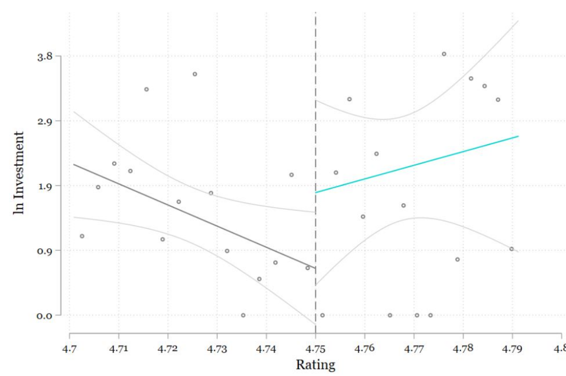
(a) 3.25



(b) 3.75



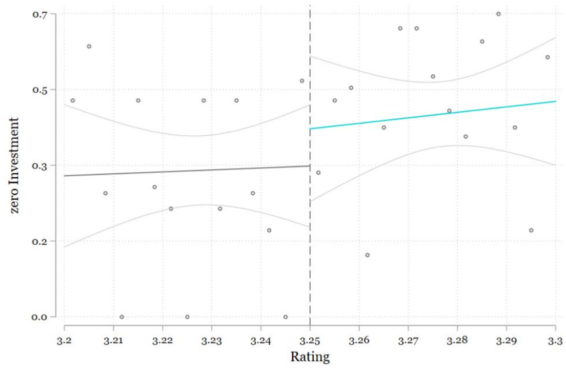
(c) 4.25



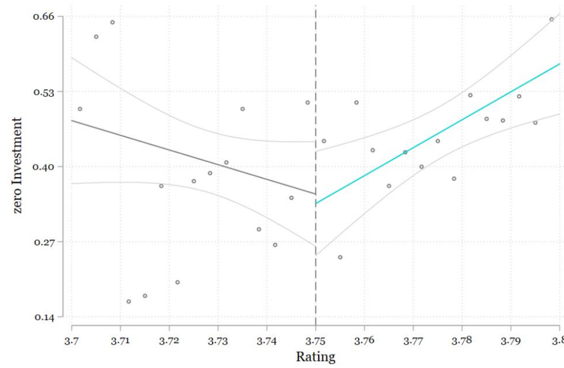
(d) 4.75

Figure 7: Log Investment versus Average Ratings at Various Cutoffs

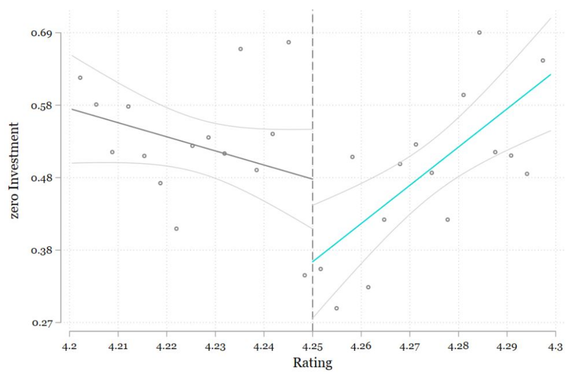
As Figure 8 focuses on the intensive margin using logarithm of investment expenditure as outcome variable, Figure 7 show the relationship in the extensive margin using a dummy variable for whether hotel make any investment or not. The overall patterns



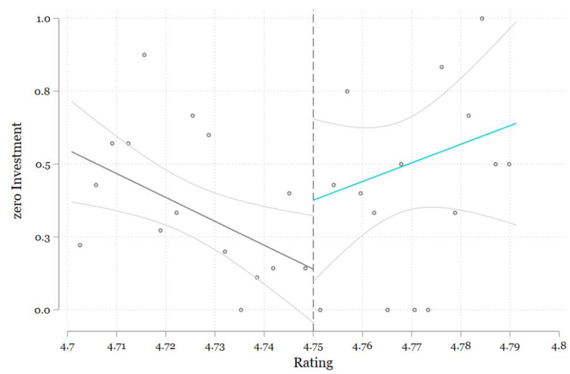
(a) 3.25



(b) 3.75



(c) 4.25



(d) 4.75

Figure 8: Investment Probability versus Average Ratings at Various Cutoffs

is similar to those in Figure 8. Only cutoff of 4.25 shows a clear negative gap at threshold. Similar to the intensive margin, clear positive jumps at 3.25 and 4.75 are observed. It is worth mentioning that while cutoff 3.75 does not pass manipulation test, the both graphical illustrations indicate that there are no visible discontinuities at 3.75.

Previous figures implies a negative RD treatment effect at cutoff 4.25. Here I conduct non-parametric local polynomial estimation. Table 3 and Table 4 presents the estimation results. Different panels correspond to various cutoffs. In each panel, the first row provides RD estimated treatment effects with different specifications. There are several important choices. First, I include a set covariates in the estimation to reduce sampling errors and control for potential factors which could impact investment decisions (Lee

Table 3: Investment and Online Reputation: Intensive Margin

Outcome	Log of Investment					
Polynomial	Linear				Quadratic	Cubic
Bandwidth	$\tilde{h}_1$	$\hat{h}_1$	$\hat{h}_1/2$	$2\hat{h}_1$	$\hat{h}_2$	$\hat{h}_3$
Covariates	No	Yes	Yes	Yes	Yes	Yes
Panel A: 3.25 Cutoff						
RD_Estimate	1.361***	0.348*	0.356	0.138	0.426**	0.440**
	(0.521)	(0.187)	(0.237)	(0.164)	(0.201)	(0.220)
Bandwidth	0.234	0.236	0.118	0.472	0.365	0.445
Obs	1,424	1,437	682	3,019	2,296	2,862
Panel B: 3.75 Cutoff						
RD_Estimate	−0.546	−0.315**	−0.346*	−0.270**	−0.319*	−0.324
	(0.499)	(0.147)	(0.199)	(0.112)	(0.191)	(0.224)
Bandwidth	0.151	0.151	0.076	0.302	0.179	0.215
Obs	4,756	4,322	2,023	9,099	7,108	9,387
Panel C: 4.25 Cutoff						
RD_Estimate	−0.500	−0.257*	−0.254	−0.253**	−0.249	−0.231
	(0.470)	(0.142)	(0.186)	(0.109)	(0.180)	(0.213)
Bandwidth	0.152	0.158	0.079	0.317	0.194	0.221
Obs	3,929	4,064	2,124	7,752	4,951	5,666
Panel D: 4.75 Cutoff						
RD_Estimate	0.190	0.098	0.069	0.120	0.106	0.187
	(0.732)	(0.396)	(0.481)	(0.286)	(0.490)	(0.523)
Bandwidth	0.101	0.087	0.044	0.175	0.090	0.105
Obs	433	332	148	1,125	342	455

Notes: Only hotels with more 25 reviews are included. Bandwidths are computed for different order of polynomial at various cutoffs. One common MSE-optimal bandwidth is used for both sides around cutoffs. All specifications use triangular kernel function.

and Lemieux 2010). I include lagged investment variable, logarithm of average price, occupancy rate, hotel age, international brand dummy, hotel capacity, average number of employees per room, number of reviews. Optimal bandwidth $\hat{h}_p$ is chosen based on the algorithm proposed by Imbens and Kalyanaraman (2012). The subscript p indicates the order of local polynomials.[8] Bandwidth choices are indicated on the third rows in each panel.

Estimates in Panel C in Table 3 confirm graphical patterns in previous figures. With covariates, the RD estimates are smaller but more precise. Negative and significant estimates at threshold are also economically significant as the magnitudes are between 27% to 35% decreases in investment expenditure. This result implies that hotel with better rating may shirk by lowering investment expenditures. Panel A shows large positive treatment effect estimates at cutoff 3.25 without covariates. But estimates become weakly significant with covariates. Estimated average increase in investment is approximately 14% to 44% as hotel jump from 3 bubbles to 3.5 bubbles. The huge jump in investment could be induced by more intensive competition since hotels with discrete ratings of 3.5 or more account for more than 85% of the population. Close or similar ratings could encourage hotels to compete in quality dimension by making more investment. Moreover, the investment expenditures could also protect hotel's ratings from falling. For cutoff at 4.75, the estimates can be resulted from the fact that there are not many observations to identify the casual effect from obtaining additional half bubble at 4.0. Cutoff 3.75 shows negative estimates. However, due to the fact the density is not continuous in previous validity checks. It cannot be interpreted as causal.

Table 4 provides RD estimation results in terms of extensive margins, which is measured by a dummy variable for positive investment expenditures. The empirical results are qualitatively similar to those in Table 3. Cutoffs at 4.75 and 3.75 have mostly positive but insignificant RD estimates. Discontinuity at 4.25 is negative and significant, and discontinuity at 3.25 is positive and significant with larger magnitude.

Comparing to previous literature studying outcomes using regression discontinuity design on half point rounding thresholds (Anderson and Magruder 2012, Hollenbeck

[8] For local polynomial approach, it usually uses lower-order of polynomials. For different order of polynomials, I recompute the optimal bandwidth $\hat{h}_p$.

Table 4: **Investment and Online Reputation: Intensive Margin**

Outcome	$\mathbf{1}$(Investment > 0)					
Polynomial	Linear				Quadratic	Cubic
Bandwidth	$\tilde{h}_1$	$\hat{h}_1$	$\hat{h}_1/2$	$2\hat{h}_1$	$\hat{h}_2$	$\hat{h}_3$
Covariates	No	Yes	Yes	Yes	Yes	Yes
Panel A: 3.25 Cutoff						
RD_Estimate	0.271^{**}	0.059^{*}	0.071^{*}	0.040	0.073^{*}	0.073
	(0.126)	(0.034)	(0.043)	(0.028)	(0.040)	(0.045)
Bandwidth	0.259	0.304	0.152	0.609	0.376	0.451
Obs	1,948	2,305	1,079	5,072	2,853	3,466
Panel B: 3.75 Cutoff						
RD_Estimate	0.021	0.011	0.038	-0.004	0.022	0.033
	(0.107)	(0.030)	(0.036)	(0.025)	(0.033)	(0.035)
Bandwidth	0.282	0.248	0.124	0.496	0.352	0.412
Obs	5,230	4,395	2,091	9,416	6,762	7,872
Panel C: 4.25 Cutoff						
RD_Estimate	-0.084	-0.054^{**}	-0.067^{*}	-0.040^{**}	-0.067^{*}	-0.078^{*}
	(0.093)	(0.025)	(0.037)	(0.019)	(0.037)	(0.043)
Bandwidth	0.156	0.188	0.094	0.376	0.176	0.215
Obs	4,010	4,793	2,489	8,976	4,515	5,493
Panel D: 4.75 Cutoff						
RD_Estimate	0.056	0.037	0.075	0.051	0.078	0.094
	(0.168)	(0.081)	(0.082)	(0.060)	(0.088)	(0.095)
Bandwidth	0.108	0.058	0.029	0.117	0.063	0.090
Obs	486	203	104	589	220	342

Notes: Only hotels with more 25 reviews are included. Bandwidths are computed for different order of polynomial at various cutoffs. One common MSE-optimal bandwidth is used for both sides around cutoffs. All specifications use triangular kernel function.

et al. 2019), my RD results feature non-monotonic effects over various thresholds. Anderson and Magruder (2012) finds that better ratings lead to lower availability for restaurants on Yelp.com. Hollenbeck et al. (2019) identifies substitution between advertising spending and online ratings. In this paper, the relationship is inverted-U shape where lower ratings induce strong investment incentives and then at higher ratings the investment incentives diminish.

4 Robustness of RD Results

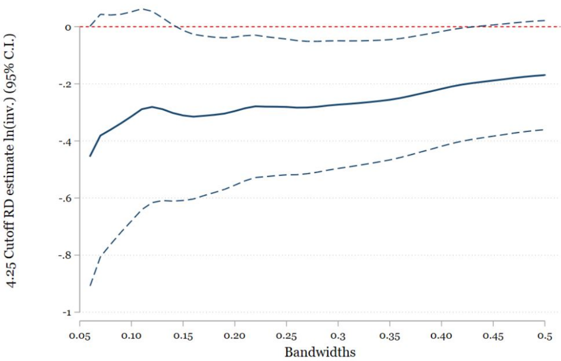
This section examines the robustness of RD results under alternative specifications.

4.1 Bandwidth Sensitivity

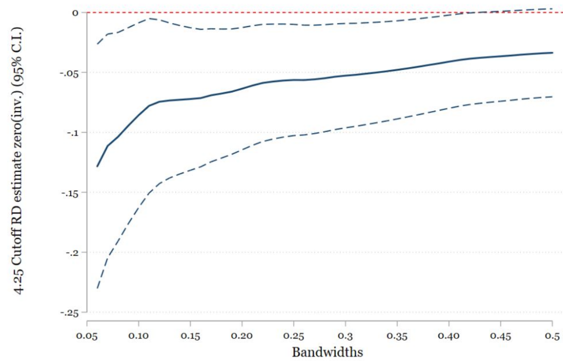
In this section, I use various alternative bandwidths for RD estimation to check if the estimates are sensitive to changes in bandwidth. Figure 9 summarizes the results for cutoff 3.25 and 4.25. The estimates remain stable for wide range of bandwidths ranging from 0.05 to 0.50, and are close to those in Table 3 and Table 4. At 4.25 cutoff, investments are lowered by around 40 to 20%. For tighter bandwidths, the standard error are relatively large. On the other hand, estimates get closer to zero when larger bandwidths are used in estimation since observations away from cutoffs are used in the estimation. For the other two cutoffs, results are presented in Figure A5 in Online Appendix. Similar to previous tables, the estimates are insignificant and close to zero.

4.2 Placebo Cutoffs

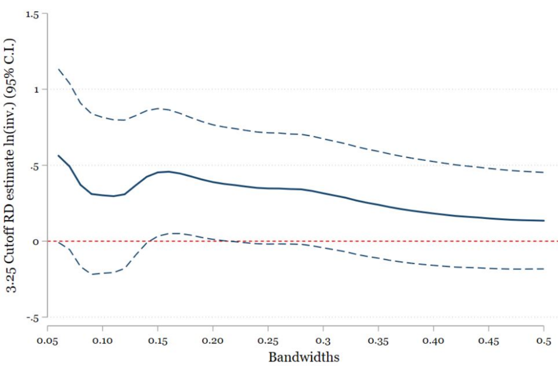
In this section, I examine treatment effects at placebo cutoffs. Specifically, I replace the true cutoffs, 3.25 and 4.25, with fake cutoff values, at which the change in bubble ratings does not happen. At those artificial cutoffs, RD estimates are expected to capture no effect. Figure 10 presents the results. For 4.25 cutoff, moving away from true cutoff value results in insignificant estimates close to zero. The pattern holds for both intensive margin and extensive margin. However, for cutoff 3.25, there is not noticeable shift for RD estimates at artificial cutoffs. For intensive margin, the estimate gets closer to zero



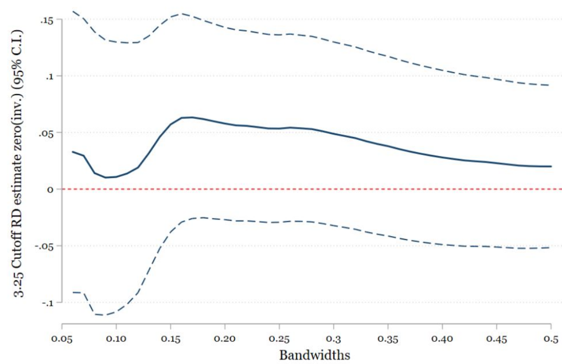
(a) 4.25 Intensive Margin



(b) 3.25 Extensive Margin



(c) 3.25 Intensive Margin



(d) 3.25 Extensive Margin

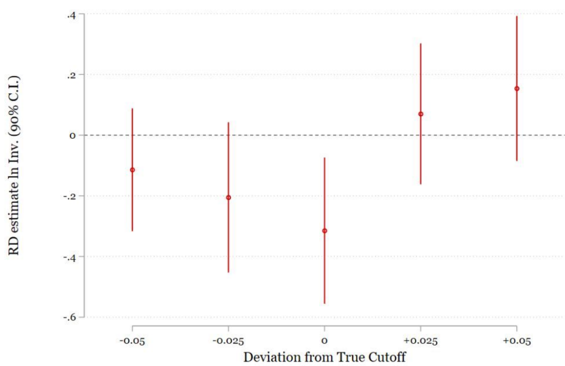
Figure 9: RD Estimates with Alternative Bandwidths: 3.25 and 4.25 Cutoffs

Notes: The above figures are based on RD estimates using local-linear non-parametric estimation. Only hotels with more 25 reviews are included. Covariates include lagged dependent variable, and other controls. Incremental change of bandwidth is 0.01. All specifications use triangular kernel function. Standard errors are robust and clustered at firm level.

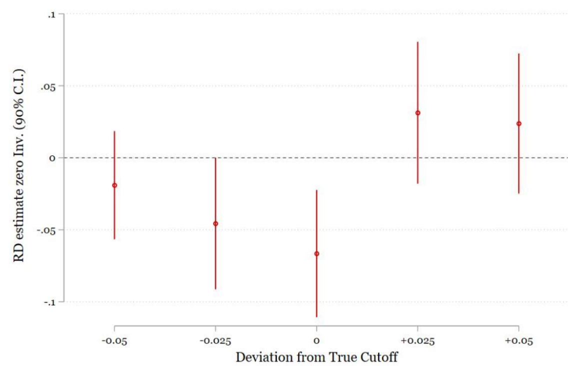
at smaller fake cutoffs. For extensive margin, the estimates do not really move as the estimate at true cutoff is already noisy.

4.3 Fixed Effects Specification

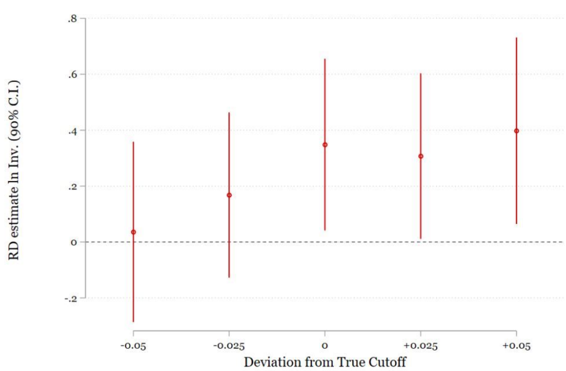
The fact that my data is a monthly panel of hotel investment and TripAdvisor rating make it tempting to add firm fixed effects, and time fixed effects. The econometric



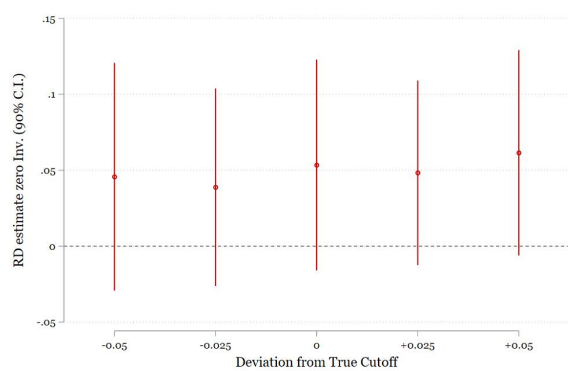
(a) 4.25 Intensive Margin



(b) 4.25 Extensive Margin



(c) 3.25 Intensive Margin



(d) 3.75 Extensive Margin

Figure 10: Placebo Tests at Artificial Cutoffs

Notes: The above figures are based on RD estimates using local-linear non-parametric estimation. Only hotels with more 25 reviews are included. Covariates include lagged dependent variable, and other controls. Artificial cutoffs of are used. All specifications use triangular kernel function. Standard errors are robust and clustered at firm level.

model then becomes:

$$
\begin{aligned}
&Y_{it} = \alpha + \tau\,\mathbf{1}(X_{it} > c) + \beta_1(X_{it} - c) + \beta_2\,\mathbf{1}(X_{it} > c) \times (X_{it} - c) + \mu_i + \delta_t + \epsilon_{it}, \\
&\forall X_{it} \in (c - h, c + h)
\end{aligned} \tag{3}
$$

where μ_i and δ_t denote firm fixed effects and time fixed effects. Estimation of τ can be done by estimating a fixed effect regression. However, it is important to note that inclusion of fixed effects is not required for identification in RD design (Lee and Lemieux

Table 5: Fixed Effects Specification RD Estimates for Intensive Margin

	3.25	3.75	4.25	4.75
Above Cutoff	0.114	0.102	-0.299**	0.237
	(0.174)	(0.179)	(0.129)	(0.536)
Average Rating	-2.310	-3.048	1.474	20.016*
	(2.582)	(2.120)	(2.381)	(11.305)
Above Cutoff X Average Rating	6.622*	1.512	0.761	-28.383
	(3.369)	(2.839)	(3.418)	(18.100)
Lagged log Inv.	0.344***	0.324***	0.422***	0.303
	(0.066)	(0.051)	(0.059)	(0.257)
Covariates	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Hotel FE	Yes	Yes	Yes	Yes
Bandwidth	.168	.152	.130	.070
Observations	1001	2273	3125	203
R-square	0.659	0.690	0.655	0.732

Notes: Only hotels with more 25 reviews are included. Covariates include lagged dependent variable, and other controls. One common MSE-optimal bandwidth is used for both sides around cutoffs. Standard errors are robust and clustered at firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2010). Previous studies exploiting rating thresholds also have some discrepancy about adding fixed effects.[9] Therefore, here I conduct RD estimation with two-way fixed effects similar to Hollenbeck et al. (2019). However, I include lagged dependent variables and other baseline covariates to reduce sampling variance as suggested by Lee and Lemieux (2010). Fixed effect regression results are reported in Table 5 and Table 6. One noticeable difference from the previous RD results is that RD estimates at 3.25 cutoff become smaller and insignificant. Potential cause could be that observations around 3.25 cutoff have little within variations as firm fixed effects absorb cross-section variations. On the other hand, RD estimate for 4.25 cutoff remain significant and have similar magnitude.

[9] While Hollenbeck et al. (2019) includes both time fixed effects and firm fixed effects, Anderson and Magruder (2012) conducts RD estimation by running pooled-OLS estimation.

Table 6: Fixed Effects Specification RD Estimates Extensive Margin

	3.25	3.75	4.25	4.75
Above Cutoff	-0.003	0.035	-0.058**	0.121
	(0.054)	(0.040)	(0.025)	(0.125)
Average Rating	-0.112	-0.696	0.063	2.076
	(0.737)	(0.443)	(0.419)	(2.278)
Above Cutoff X Average Rating	1.108	0.490	0.227	-1.612
	(0.861)	(0.665)	(0.616)	(5.139)
Lagged 1(Inv.> 0)	0.066***	0.071***	0.066***	0.036
	(0.014)	(0.012)	(0.008)	(0.054)
Covariates	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Hotel FE	Yes	Yes	Yes	Yes
Bandwidth	.173	.150	.125	.065
Observations	1027	2255	3033	176
R-square	0.726	0.687	0.679	0.740

Notes: Only hotels with more 25 reviews are included. Covariates include lagged dependent variable, and other controls. One common MSE-optimal bandwidth is used for both sides around cutoffs. Standard errors are robust and clustered at firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4 Alternative Measures of Investment

work in progress...

5 Conclusion

This paper provides empirical evidence supporting work-shirk equilibrium in Board and Meyer-ter Vehn (2013). In particular, I identify the impact of online rating on investment expenditures by exploiting TripAdvisor Rounding rules in a regression discontinuity design framework. My estimates show that the relationship between online rating and hotel investment is not linear. With higher rating, hotels tend to decrease their investment intensity or frequency, suggesting a substitution relationship between quality and online reputation. However, at lower ratings, the incentives for investment are stronger if hotels jump from 3 bubbles to 3.5 bubbles.

Reference

Anderson, M. and Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, 122(563):957–989.

Board, S. and Meyer-ter Vehn, M. (2013). Reputation for quality. *Econometrica*, 81(6):2381–2462.

Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354.

Fang, L. (2020). Evaluating the effects of online review platforms on restaurant revenues, consumer learning and welfare. *Working Paper*.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Hollenbeck, B. (2017). The economic advantages of chain organization. *The RAND Journal of Economics*, 48(4):1103–1135.

Hollenbeck, B., Moorthy, S., and Proserpio, D. (2019). Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science*, 38(5):793–811.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3):933–959.

Jin, G. Z. and Leslie, P. (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, 118(2):409–451.

Kalnins, A. (2006). Markets: The us lodging industry. *The Journal of Economic Perspectives*, 20(4):203–203.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355.

Lewis, G. and Zervas, G. (2016). The welfare impact of consumer reviews: A case study of the hotel industry. Technical report, Working Paper.

Luca, M. (2016). Reviews, reputation, and revenue: The case of yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016).

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–55.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714. The regression discontinuity design: Theory and applications.

Saeedi, M. (2019). Reputation and adverse selection: theory and evidence from ebay. *The RAND Journal of Economics*, 50(4):822–853.

Online Appendix

A Additional Figures and Tables

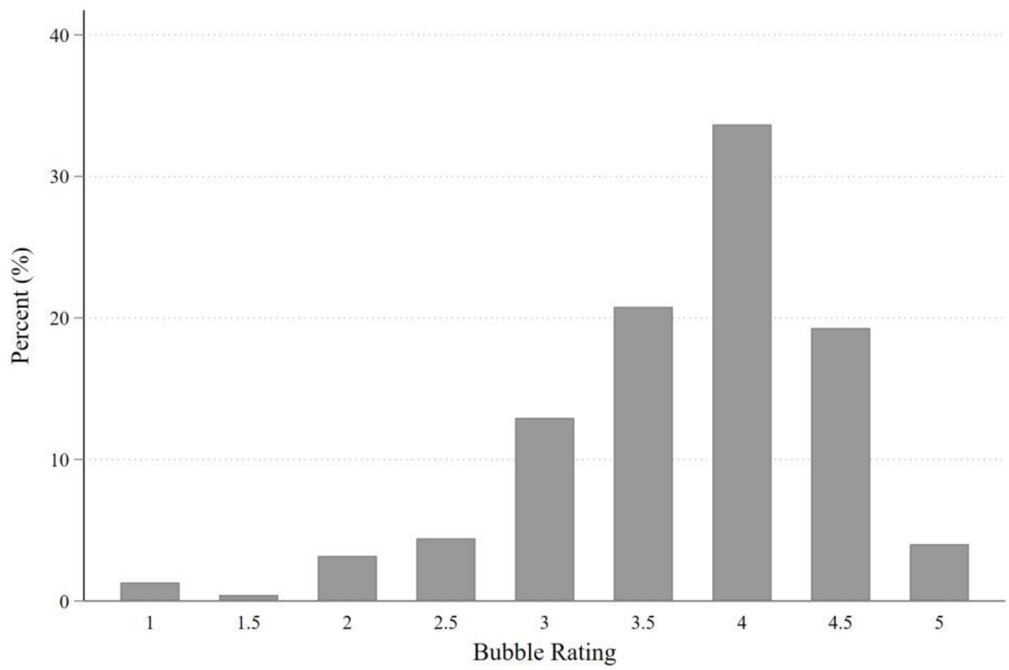


Figure A1: Distribution of Bubble Ratings

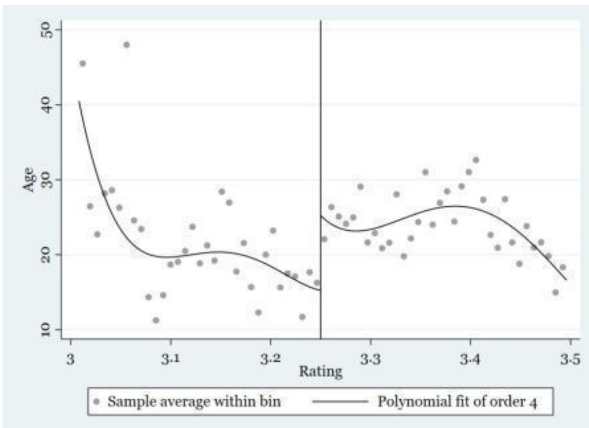
	Below	Above	Diff. (S.E.)
Age	15.634	17.001	-1.368 (0.812)
No. of Rooms	169.301	181.984	-12.68 (8.292)
International Chain	0.118	0.154	-0.0357 (0.0201)
Star Rating	4.228	4.179	0.0488 (0.0828)
No. of Reviews	125.642	120.355	5.287 (9.378)
No. of Empl. per Room	0.782	0.791	-0.00976 (0.0805)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

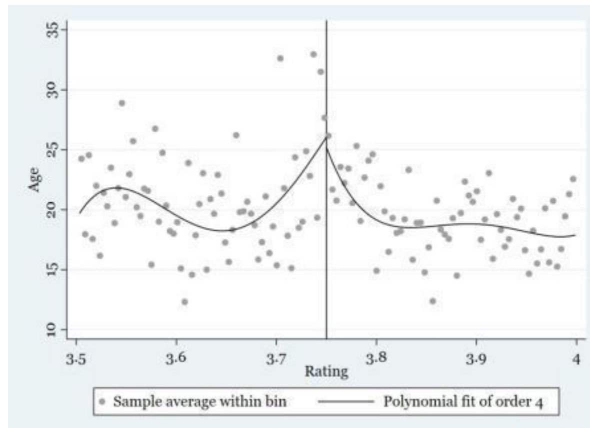
Table A1: Percentage Frequency Table for TripAdvisor Bubble Ratings

	Pct. (%)	Cum. Pct. (%)
2.0 Bubbles	0.32	0.32
2.5 Bubbles	1.55	1.87
3.0 Bubbles	6.24	8.10
3.5 Bubbles	19.76	27.86
4.0 Bubbles	43.37	71.24
4.5 Bubbles	28.07	99.30
5.0 Bubbles	0.70	100.00
Total	100.00	

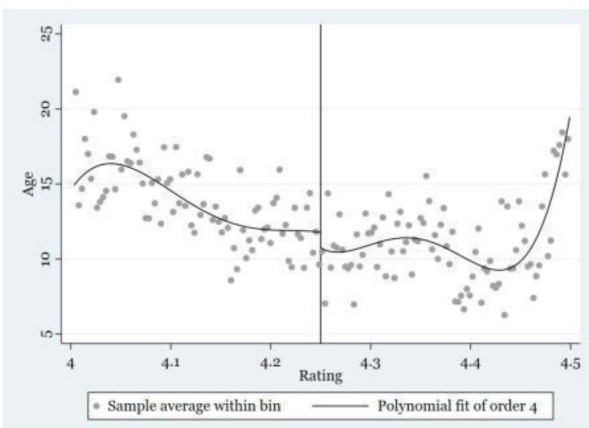
Note: Bubble ratings are restricted to observations with at least 20 reviews.



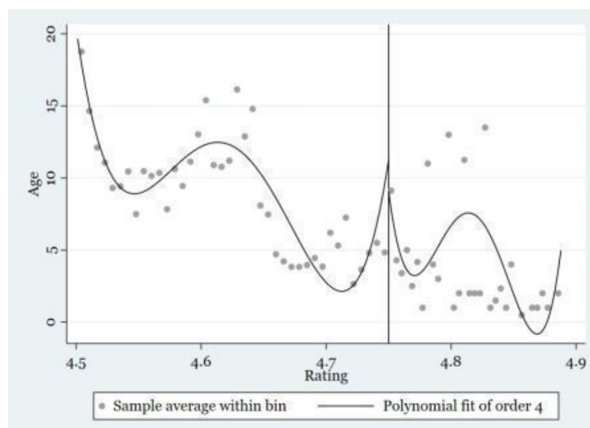
(a) 3.25



(b) 3.75

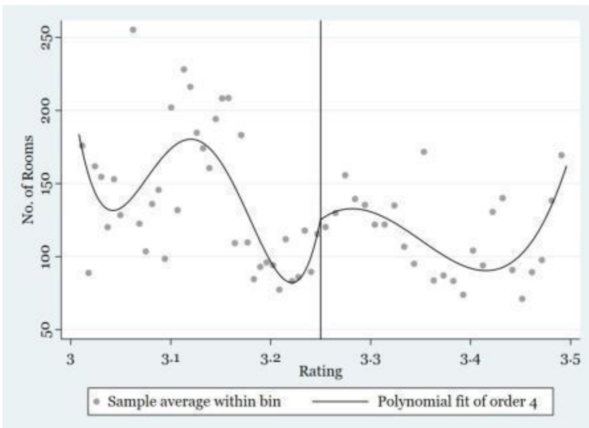


(c) 4.25

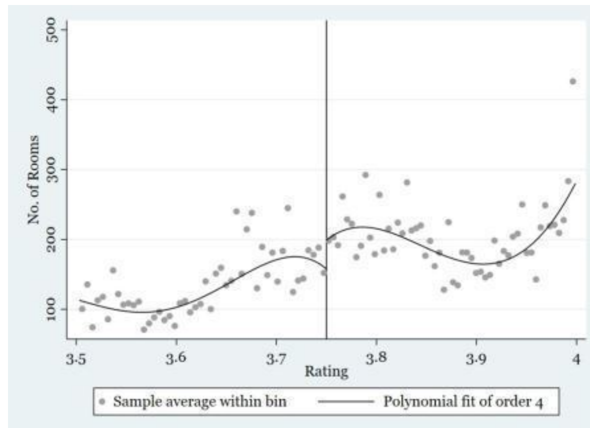


(d) 4.75

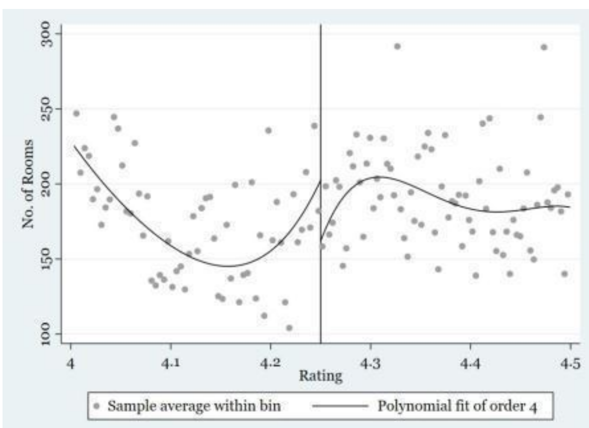
Figure A2: RD Plots for Predetermined Covariate: Age



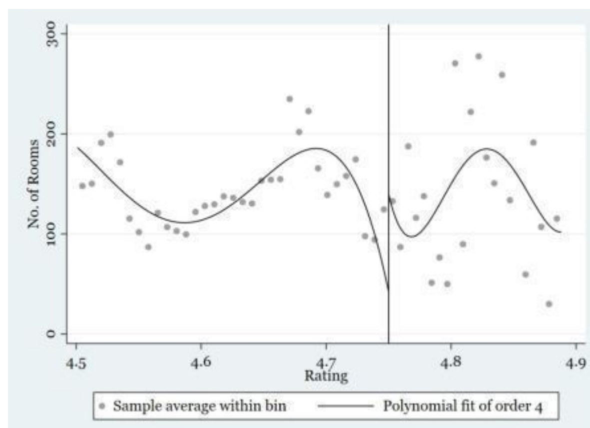
(a) 3.25



(b) 3.75

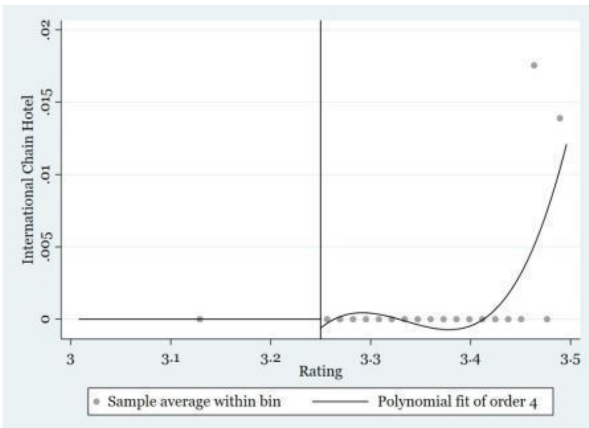


(c) 4.25

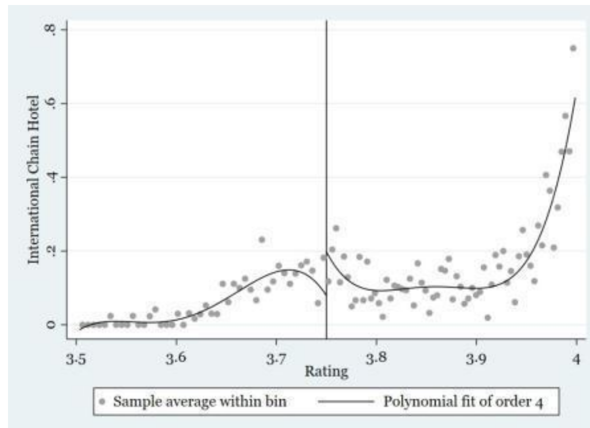


(d) 4.75

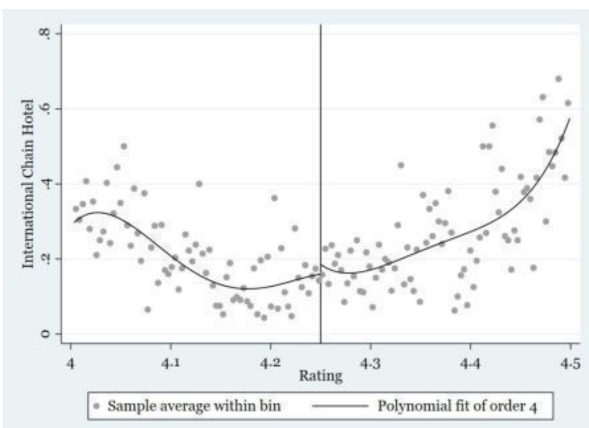
Figure A3: RD Plots for Predetermined Covariate: Number of Rooms



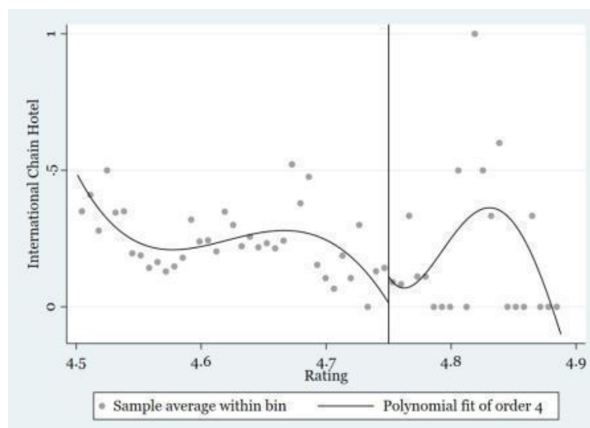
(a) 3.25



(b) 3.75



(c) 4.25



(d) 4.75

Figure A4: RD Plots for Predetermined Covariate: International Chain Hotels

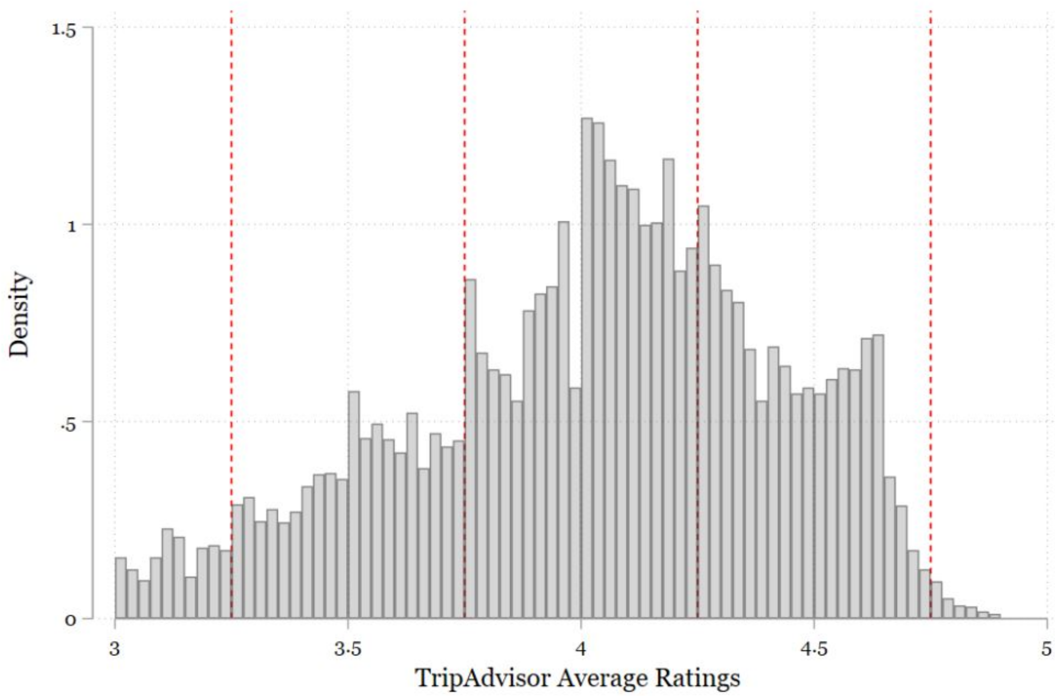
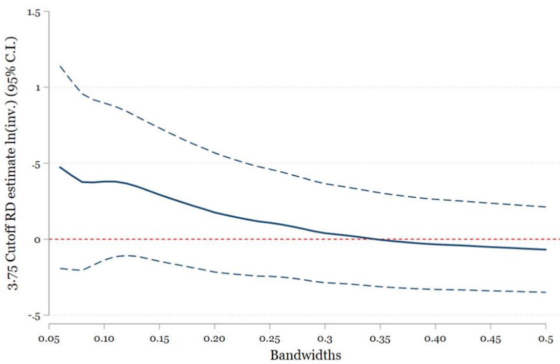
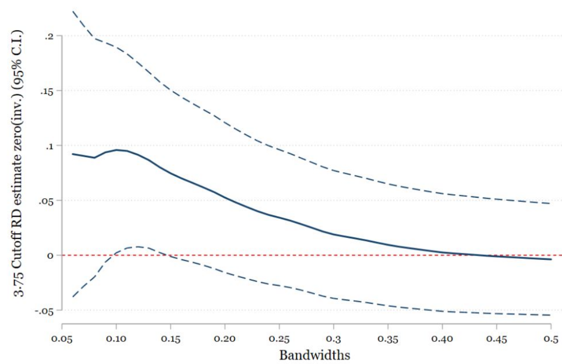


Figure A5: Histogram of TripAdvisor Average Ratings

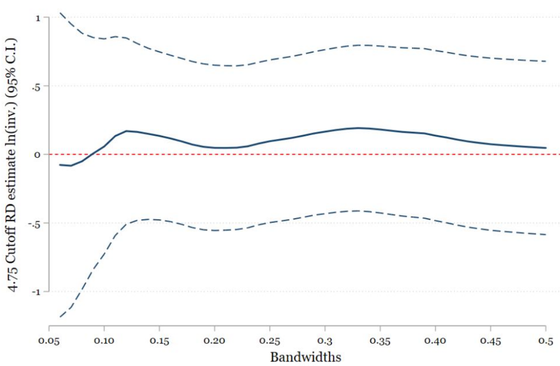
Notes: The above figure shows histogram of TripAdvisor average ratings above 3 with bin width of 0.025. Cutoffs are indicated by red dash lines. Only observations with more than 20 reviews are included.



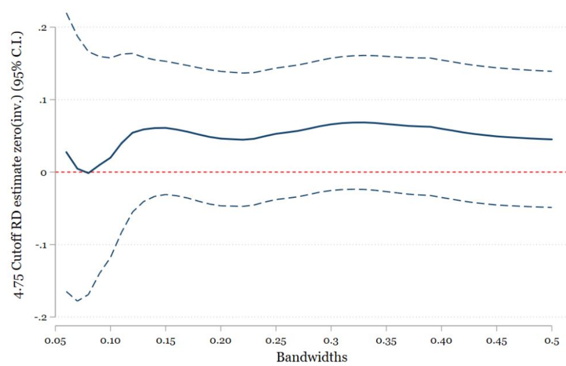
(a) 3.75 Intensive Margin



(b) 3.75 Extensive Margin

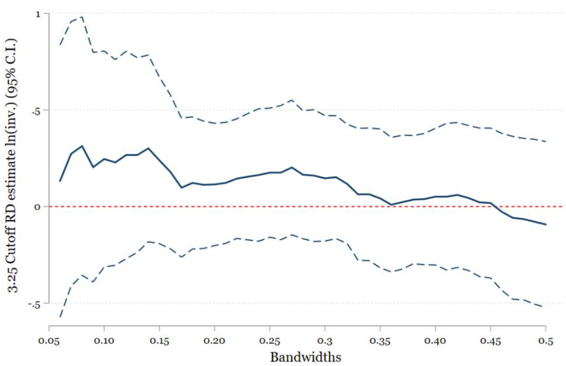


(c) 4.75 Intensive Margin

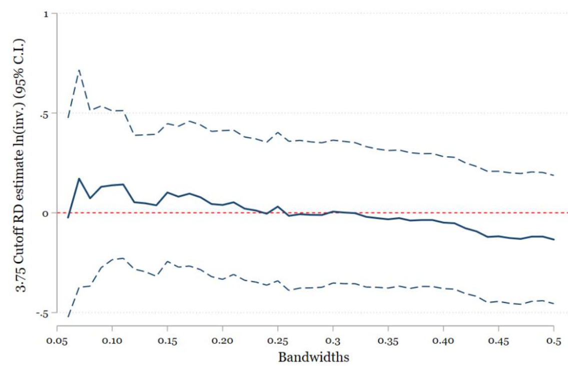


(d) 4.75 Extensive Margin

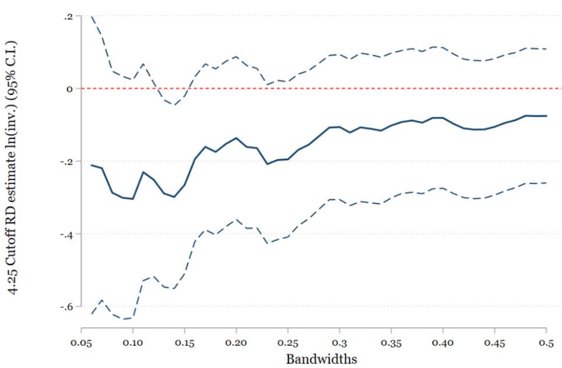
Figure A6: RD Estimates with Alternative Bandwidths: 3.75 and 4.75 Cutoffs



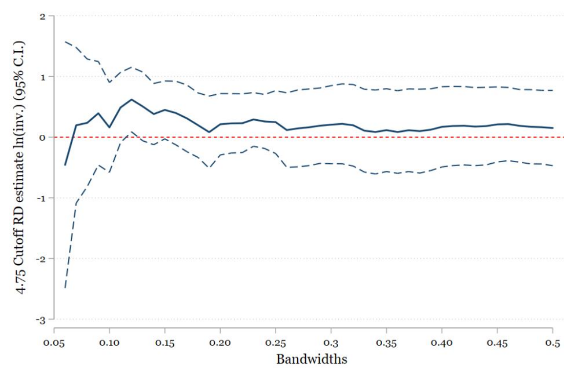
(a) 3.25 Intensive Margin



(b) 3.75 Extensive Margin

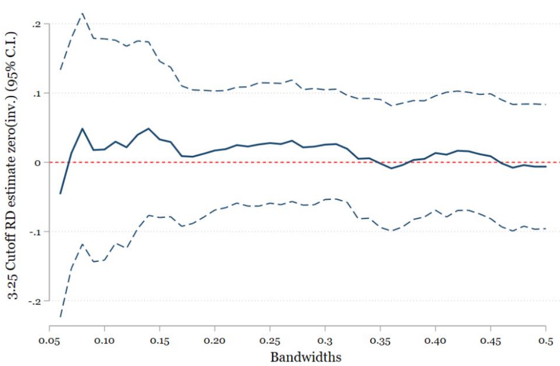


(c) 4.25 Intensive Margin

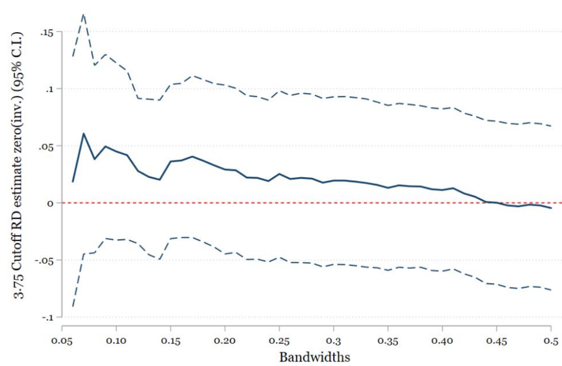


(d) 4.75 Extensive Margin

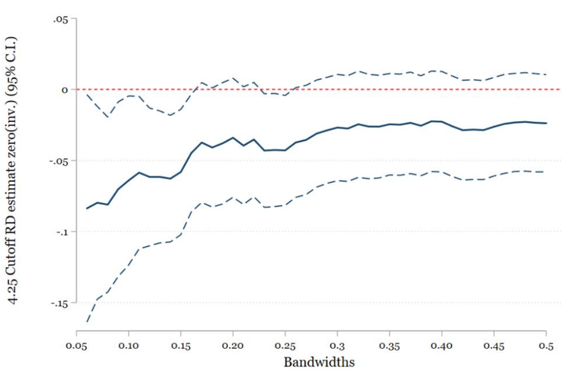
Figure A7: Fixed Effects RD Estimates with Different Bandwidths (Intensive Margin)



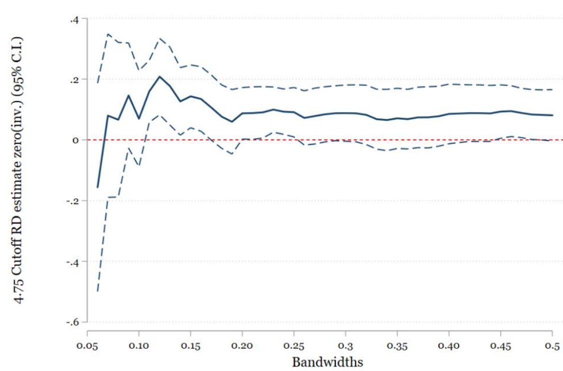
(a) 3.25 Intensive Margin



(b) 3.75 Extensive Margin



(c) 4.25 Intensive Margin



(d) 4.75 Extensive Margin

Figure A8: Fixed Effects RD Estimates with Different Bandwidths (Extensive Margin)