Advertising strategy in the presence of reviews: An empirical analysis [1]

Brett Hollenbeck
Anderson School of Management, University of California Los Angeles

Sridhar Moorthy
Rotman School of Management, University of Toronto

Davide Proserpio
Marshall School of Business, University of Southern California

November 12, 2018

[1]The authors would like to thank Judith Chevalier, Pedro Gardete, Tobias Klein, Jason Roos, Navdeep Sahni, Ken Wilbur, and audiences at the 2017 QME Conference, 2017 NYU Conference on Digital, Mobile Marketing, and Social Media Analytics, 2018 ISMS Marketing Science Conference, 2018 NBER Economics of Digitization Conference, 2018 Bass FORMS Conference, 2018 ACM Conference on Economics and Computation, and 2018 Economics of Advertising Workshop, for their helpful comments. Sridhar Moorthy thanks the Social Sciences and Humanities Research Council of Canada, and Brett Hollenbeck thanks the Morrison Center for Marketing and Data Analytics, for their generous support of this research.

Abstract

We study the relationship between online reviews and advertising spending in the hotel industry. Combining a dataset of TripAdvisor reviews with other datasets describing these hotels' advertising expenditures, we show, first, that the quality information embedded in reviews has a causal demand-side effect on ad spending. Second, this effect is negative: hotels with higher ratings spend less on advertising than hotels with lower ratings. This suggests that hotels treat TripAdvisor ratings and advertising spending as substitutes, not complements. Third, the relationship is stronger for independent hotels than for chains and stronger in less differentiated markets than in more differentiated markets. The former suggests that a strong brand name continues to provide some immunity to reviews and the latter that the advertising response is stronger when ratings are more likely to be pivotal. Finally, we show that the relationship between TripAdvisor ratings and advertising has strengthened over time just as TripAdvisor has become more popular among users. This provides further corroboration that the effect we are observing operates through the demand side and not the cost side.

1 Introduction

Over the last fifteen years, one of the major developments in the consumer's shopping environment has been the growth and proliferation of online review platforms such as TripAdvisor and Yelp, providing independent quality information on experience goods. According to the Pew Research Center, in 2016, 82% of U.S. adults read online reviews occasionally or regularly before purchasing a product for the first time; 40% did so almost always.[1] The ready availability of experiential information from past users and professional reviewers is a potentially significant demand shock affecting firms in these industries. Effectively, experience goods have become search goods in many categories. This raises many questions, among them: How does this affect the advertising strategy of firms? What is the relationship, if any, between the information on quality revealed in online reviews and firms' advertising spending decisions, and how has this changed over time? In this paper, we report on these questions in the context of the hotel industry.

The only theoretical prediction we have about the relationship between advertising spending and product quality is the one from Nelson's (1970, 1974) signaling theory. He argue that advertising spending for experience goods should be positively related to product quality because only firms with high quality products would be willing to invest in advertising to signal quality. However, the many analyses that have followed in his wake—Kihlstrom and Riordan (1984), Milgrom and Roberts (1986), Hertzendorf (1993), Zhao (2000), and Linnemer (2002), among others—have mainly served to document the difficulties of consummating the argument. The reason is, in most realistic contexts, there are multiple forces pulling in different directions, among them: (i) the cost-side effects of quality on advertising spending, which arise as soon as advertising has some role other than pure signaling—such as raising awareness—and these tend to be negative (Bagwell 2007, p. 1777); (ii) the demand-side effects of quality on advertising spending, which arise whenever at least some consumers are informed about quality before purchase, and these can be positive (Archibald et al. 1983, Lei

[1] See: `http://www.pewinternet.org/2016/12/19/online-reviews/`

2015) or negative (Chen and Xie 2005); (iii) interactions between the two, which can lead to a net positive or negative effect (Schmalensee 1978); and finally, (iv) strategic interactions in prices and advertising among firms, which can lead to strategic complementarity or strategic substitutability effects (Chen and Xie 2005, Lei 2015).

Perhaps reflecting these difficulties, past empirical studies have also failed to show a consistent advertising spending-quality relationship. For instance, Rotfeld and Rotzoll (1976), looking at 12 convenience-goods categories, find a positive correlation between advertising and quality (as reported in *Consumer Reports* and *Consumers Bulletin*) among all brands—advertising and non-advertising—but not within the subset of brands that advertise. Caves and Greene's (1996) more comprehensive study of nearly 200 categories reports median correlations around zero.

Online data from review platforms such as TripAdvisor present a unique opportunity to get a clearer picture of the multiple forces at work. They allow us to identify a causal effect of quality on advertising spending and to label it as a demand-side effect. Both accomplishments stem from the same feature of the data: TripAdvisor rounds up or down the average ratings of reviewers to the nearest whole or half ratings (on a 5-star scale) and displays only those rounded ratings to users. What the consumer sees are these whole and half ratings—$1, 1.5, \cdots, 4.5, 5$ stars—not the average ratings underlying them.[2] This has two effects: (i) it creates a dissociation between perceived quality (displayed ratings) and actual quality (average ratings) (see Figure 1), and (ii) it provides us with a ready-made regression-discontinuity design (RDD) to identify a causal effect. By focusing on the random, discrete variation in perceived quality around the discontinuities—when the average rating changes from, say, 3.24 (displayed rating 3) to 3.25 (displayed rating 3.5)—and measuring the effect of this variation on advertising spending, we identify a causal effect of quality on advertising spending.[3] Furthermore, this is self-evidently a demand-side effect: while "large variations"

[2]A particularly motivated consumer could go to the raw data—reviewers' actual ratings—and calculate the averages for herself, but we conjecture that most consumers would not do that.

[3]For other applications of regression-discontinuity designs using user reviews see Anderson and Magruder (2012), Luca (2016), and Lei (2017).

in average quality can plausibly have both demand-side and cost-side effects,[4] the essentially random variation in average quality in the neighborhood of the discontinuity thresholds can only be plausibly characterized as demand-side effects.

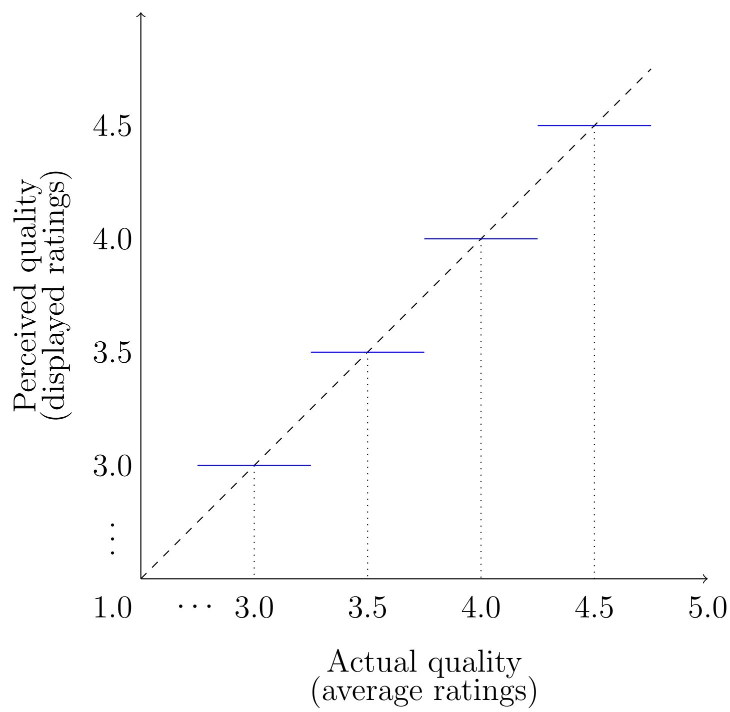


Figure 1: How perceived quality varies with actual quality at TripAdvisor

The hotel industry is an ideal setting to study the relationship between online reviews and advertising for several reasons. First, it was one of the earliest adopters of online reviews. A large corpus of reviews has accumulated, showing good variation both in the cross-section and in the time-series. Second, because hotels are experience goods and serve people from many widely-dispersed locations, online word-of-mouth is especially important relative to offline word-of-mouth. Finally, the industry is large and important in its own right. Hotels generated $196 billion in sales and employed 2 million individuals in 2012 according to the U.S. Economic Census. The industry also spent $2.1 billion on advertising in 2015. Separately from the broader question of the relationship between advertising and

[4] In the hotel industry, arguably, most costs are fixed. So even large-scale average ratings variation probably isn't accompanied by significant marginal cost variation.

quality, it is also of direct interest how the increasing importance of ratings platforms has effected advertising decisions in this industry.

Our empirical analysis is based on four datasets, one comprising all TripAdvisor hotel reviews from 2002 onwards, two others containing detailed information on the advertising strategies of the hotels featured in those reviews, and the fourth describing various characteristics of the hotels. The review dataset contains all U.S. hotels listed on TripAdvisor, and the advertising dataset includes information about each hotel's monthly advertising spend, disaggregated by media—TV, newspapers, magazines, radio, outdoor, and internet display and search advertising. One contribution of this paper is simply the collection and matching of these datasets, allowing for the first time-series and cross-sectional study of the empirical relationship between online reviews and firms' advertising strategies. This is noteworthy because our data are not just a sample of a particular market in a particular time period but rather the entire experience of an industry over virtually the entire period online reviews have existed.

Our main results are the following. First, total ad spending by hotels has fallen slightly from 2002 to 2015. The fall is quite sharp in traditional advertising—print and television, mainly—but this is largely offset by the growth in internet search and display advertising. Second, we find a distinct inverse-U shaped relationship between hotels' *average* ratings and ad spending, both in traditional media and in online advertising. This is consistent with previous studies in the literature that have also documented a cross-sectional inverse-U relationship between quality and advertising spending (Horstmann and Moorthy 2003). Third, we find a significant causal relationship between *displayed* TripAdvisor ratings and advertising spending: higher ratings consistently lead to lower advertising. We observe this effect in all the traditional media as well as in the online media. Moreover, this effect obtains both at the extensive margin (decision to advertise) and at the intensive margin (advertising level given advertising). These results collectively suggest that hotels are treating their TripAdvisor displayed ratings as a substitute for advertising spending. This is as if hotels with higher ratings are targeting the informed consumer while hotels with lower ratings are targeting the uninformed consumer. Fourth, this relationship is stronger for independent hotels than for chains, consistent with prior research showing that online reviews have larger effects on independent hotels' sales than on chain hotels' sales (Hollenbeck 2018). A strong brand name continues to provide some immunity to reviews—as has been found in other contexts as well, such as movies (Eliashberg and Shugan 1997). Fifth, when we examine how the ratings-advertising relationship operates in markets differing in vertical differentiation—less differentiated markets being those with a small standard deviation of ratings—we find that less differentiated markets show a weaker relationship, suggesting that ratings have a bigger effect on ad spending when they are more likely to be pivotal. Finally, comparing the relationship between online ratings and advertising spending in the early years of TripAdvisor (2002-2005) with the relationship in more recent years (2012-2015), we find that the relationship has strengthened over the years. This makes sense. As we have argued, the demand-side effect of online ratings operates through more consumers getting informed about quality before purchase. In the early years, the number of informed consumers was small, and firms didn't have to pay much attention to their ratings; but over time, as more consumers have started paying attention to ratings, firms have been forced to respond to them.

2 Background

As noted earlier, survey evidence shows that a large number of people consult online reviews before making purchase decisions. It is only logical, then, that online reviews should affect sales. Indeed, this is what the literature finds: Chevalier and Mayzlin (2006) and Sun (2012) for books, Anderson and Magruder (2012) and Luca (2016) for restaurants, Jin and Kato (2006) and Cabral and Hortaçsu (2010) for eBay auctions, and Lewis and Zervas (2016) for hotels. Luca (2016) shows, in addition, that the effect is particularly strong for independent restaurants (compared to chain restaurants). In a similar vein, Hollenbeck (2018) shows that online reviews have significantly reduced the revenue premium enjoyed by hotel chains over independent hotels.

In contrast to this large literature showing the effect of online reviews on demand, there is relatively little work on how online reviews affect firms' actions. This is an important lacuna: not only is the effect of online reputation on firm behavior interesting in its own right, it also has implications for the literature on consumer response. If firms change their pricing, advertising, or other actions when their ratings change, prior work documenting an effect of ratings on sales might actually be documenting the combined effect of ratings and firm response.[5] One setting where firms' responses to ratings have been studied is eBay, but even here, the focus has been on price responses only (Melnik and Alm 2002, Houser and Wooders 2006, Resnick et al. 2006). Arguably, eBay auctions are unique in several respects, not the least being that sellers on the platform tend to be small and online-only. One non-eBay paper is by Lewis and Zervas (2016) who use TripAdvisor data to show that hotel prices increase in response to online ratings. Lei (2017) analyzes Yelp restaurant ratings data from 2014 using methodology similar to ours and finds, like us, that displayed ratings have a negative effect on advertising spending in the cross-section.

On the theoretical side, we have Chen and Xie (2005, 2008) and Lei (2015). The first is a duopoly model with horizontal and vertical differentiation, the second is a monopoly matching model in two dimensions, and the third is a vertical differentiation duopoly model. In the first, reviews are assumed to provide accurate information about product quality, while advertising can mislead those who don't read reviews. In the second, reviews provide information comprehensively (i.e., on both product dimensions) but may or may not be accurate, while advertising-supplied information is accurate but may or may not be comprehensive. In the third, reviews provide unbiased signals of underlying quality, and advertising creates awareness and directs advertising-sensitive consumers to reviews. The central concern of

[5] We thank an anonymous referee for this insight.

the two Chen and Xie papers is what happens to advertising strategy as reviews become available, i.e., they seek a time-series prediction of advertising strategy. By contrast, Lie's paper focuses on cross-sectional variation in advertising strategy between a "high quality" firm and a "low quality" firm. The main result of the first paper show that when reviews can be incorporated into ads but the horizontal differentiation is so strong that prices don't change when reviews are published, then both the low-quality and the high-quality firm reduce their advertising expenditures. The second paper argues that information provided in ads may increase or decrease as reviews become available. Finally, the third paper shows that many kinds of equilibria are possible—neither firm advertising, only the firm with better (worse) reviews advertising, and both firms advertising (advertising is assumed to be a binary decision). But, in general, the analysis suggests that the firm with better reviews will advertise more.

As noted in the Introduction, this work is also related to the large empirical literature relating advertising spending to product quality. As far as we know, none of this work has used online ratings as a measure of quality—the typical study pre-dates the Internet era (Rotfeld and Rotzoll 1976, Caves and Greene 1996, Moorthy and Zhao 2000, Horstmann and Moorthy 2003). Furthermore, these studies are correlational, not causal. As such, they are unable to separate the cost-side effects of quality from the demand-side effects. Our regression discontinuity design allows us to identify a causal effect, and because the variation we exploit is a variation in perceived quality, we can be confident that what we are finding is a demand-side effect of quality on advertising spending, not a cost-side effect.

3 Data

To study the effect of online ratings on hotels' advertising spending empirically, we examine data from four sources: Kantar Media, TripAdvisor, SpyFu, and STR.

Kantar Media. Advertising spending data from Kantar Media (and its previous incarnations, TNS Media Intelligence and LNA) have been the basis of a number of studies in advertising (e.g., Caves and Greene 1996, Shum 2004, Kim and McAlister 2011, Honka et al. 2017). Kantar's data cover all major media: TV, radio, magazines, newspapers, internet display, and outdoor.[6] Their methodology is essentially a bottom-up approach, combining direct monitoring of ads and information supplied by media outlets.[7]

Monthly advertising expenditures, by media, are available for each hotel brand and "product," brand being a higher-level aggregation than product. Generally, product refers to a specific hotel property, but more generally it refers to a specific "advertised product." For example, for Best Western Hotels, brand is "Best Western Hotels," and there are over 100 advertised products, including "Best Western Hotels: Bethlehem PA," "Best Western Hotels: Online," and "Best Western Hotels & Minnesota State Tourism: Combo."[8] Brand ad expenditure is the sum of all product ad expenditures. We obtained monthly advertising expenditures for 16,852 brand-product pairs for a period covering 14 years, from 2002 to 2015. From this dataset, we remove products containing the string "Combo" because such products include advertising spending on 2+ brand-product pairs. This leaves us with 15,973 distinct brand-product pairs.

TripAdvisor. Our hotel user ratings data come from TripAdvisor. Launched in 2000, TripAdvisor is now one of the most popular review platforms on the internet.[9] In an average month, it has about 350 million unique visitors worldwide.[10] In addition, TripAdvisor's ratings are widely displayed on other travel search platforms such as Hotels.com, Orbitz,

[6]To be more specific, 18 media categories are identified: Network TV, Spot TV, Spanish Language Network TV, Cable TV, Syndication, Magazines, Sunday Magazines, Local Magazines, Hispanic Magazines, B-to-B Magazines, National Newspapers, Newspapers, Hispanic Newspapers, Network Radio, National Spot Radio, Local Radio, US Internet Display, and Outdoor.

[7]For more details, see `http://stradegy.kantarmediana.com/Stradegy/Help/Methodology.aspx?.pl=Methodology`.

[8]As this examples indicate, "product" may indicate a specific property, but it may also indicate "type of ad" and advertising partnerships.

[9]Indeed, it is one of the most visited websites on the internet (`http://www.comscore.com/Insights/Rankings/Revised-Top-50-Digital-Media-Properties-for-October-and-November-2016`).

[10]See: `https://www.tripadvisor.nl/pages/factsheet.html`.

Travelocity, and Expedia.com. TripAdvisor ratings therefore have a very large audience, extending beyond visitors to TripAdvisor.com itself. As of May 2016, TripAdvisor had over 500 million customer reviews on over 6 million accommodations, restaurants, and attractions.

We created a script to search and scrape all data on TripAdvisor for the accommodation properties (hotels, B&Bs, inns, etc.) located in the U.S. This yielded reviews on 91,783 properties; 82,589 of these had at least one review in the period January 2002-December 2015. The total number of reviews is 13,947,126.

SpyFu. We use SpyFu for the one advertising medium missing from Kantar: online search advertising. SpyFu is a company that tracks online search advertising. Their methodology is to search millions of keywords on Google, Bing, and Yahoo and record the URLs these searches return, with their positions in paid (and organic) listings. From this raw material they obtain estimates of monthly search advertising spending by each URL, using Google's Keyword Planner tool.[11] Their reach is extensive and includes even very specific keywords such as "Dockside Inn Fort Pierce" or "hotel new brunswick."

To use the SpyFu data, we proceeded as follows. First, from TripAdvisor, we obtained each hotel's homepage URL.[12] This yielded about 31,000 URLs, out of which 10,398 were for independent hotels. Then, using the SpyFu API, we obtained search advertising spending information for all the independent hotels for which SpyFu provided this information. This procedure yielded monthly search advertising spending data on 9,718 independent hotels. Chain hotels' search ad spending cannot be obtained from SpyFu because SpyFu doesn't provide ad spending estimates at the sub-URL level.[13]

[11] This tool provides the average traffic, cost per thousand impressions (CPM), and the cost per click (CPC) of the average ad, for every keyword.

[12] For most hotels, TripAdvisor provides a link to the hotel homepage, if one exists.

[13] For instance, the Hyatt Regency in Princeton, NJ uses the URL https://www.hyatt.com/en-US/hotel/new-jersey/hyatt-regency-princeton/princ and the Hyatt Regency in Buffalo, NY uses the URL https://www.hyatt.com/en-US/hotel/new-york/hyatt-regency-buffalo-hotel-and-conference-center/buffa. SpyFu does not report separate spending numbers for these URLs; instead, it aggregates "all Hyatt spending" into www.hyatt.com.

STR. Finally, from STR, a company that tracks the hotel industry, we obtain: (1) a list of all the hotel chains in the U.S, along with the number of properties in each chain; (2) basic census data for a large fraction of U.S hotel properties, including hotel name, location, price caategory, class, ownership, and capacity, among others;[14] adn (3) a panel of hotel prices (average daily rates) at the hotel-year-month level of a subset of hotels in the STR census—the ones that chose to report such information to STR.

3.1 Matching the datasets

The first challenge is matching up the hotels in the four datasets: TripAdvisor, Kantar, SpyFu, and STR.

Matching Kantar and TripAdvisor. This match poses the biggest challenge because, while TripAdvisor provides detailed information—exact name and address—for each hotel, Kantar's ad spending numbers are organized by brand-product, a much sketchier hotel designation. To perform this match at scale, we implemented the following algorithm:

1. Perform a Google search of all the Kantar brand-products and examine the URLs of the top-10 search results.

2. Identify the subset of URLs that correspond to a TripAdvisor hotel URL.

3. Extract the TripAdvisor hotel ID(s) for those URL(s).

4. If the TripAdvisor hotel ID extracted is unique, the algorithm returns the pair (TripAdvisor ID, Kantar ID) as a possible match. We obtained 10,470 such unique matches.

5. Compute a similarity score between the Kantar product name and the TripAdvisor hotel name and retain only the results that are "highly similar."

[14] The STR hotel census contains information on about 63,502 properties, which is about 69% of the properties listed on TripAdvisor.

Then we manually checked the output of the algorithm for correctness.

This algorithm matches 6,312 Kantar brand-products with 5,666 TripAdvisor hotels. This is about 60% of the matches obtained in Step 4 of the algorithm or about 40% of the total Kantar sample. The attrition is due to the fact that our algorithm returns only Kantar hotels that are likely to be uniquely identifiable, while the full Kantar sample contains all hotels, uniquely identifiable or not.[15] Out of the 5,666 TripAdvisor hotels correctly matched, 5,563 received a review before the year 2016; in total, these 5,563 hotels had 3,308,450 reviews or about 594 reviews per hotel.

Using these matches, we proceeded to construct a monthly panel of hotel ratings and advertising spending. The final dataset contains 762,233 hotel-year-month observations for 5,563 TripAdvisor hotels (4,020 independent and 1,543 chains) that were reviewed between January 2002 and December 2015. Most of the analysis we present below is based on this dataset.

Adding STR information. We augment the above dataset with information from the STR census. Matching hotels between TripAdvisor and STR is a much easier task because both datasets contain hotel name and exact location (complete address). This matching yields 3,996 hotels (about 72% of the 5,563 hotels we started with). Out of these 3,996 hotels, STR provided us with financial information (prices) for 2,810 hotels.[16]

Adding SpyFu advertising spending. Out of 9,718 independent hotels for which we have SpyFu data, 3,520 can be linked through the TripAdvisor hotel ID to the 4,020 independent hotels in the Kantar-TripAdvisor matched dataset. For these 3,520 independent

[15]To further elaborate, our algorithm returns a unique TripAdvisor ID, but it cannot guarantee that the result is the hotel we are looking for; this is why in step 5 we only retain matches with similar names. Google's search results may be "wrong" because: (i) the hotel we searched for did not have a TripAdvisor page, and Google returned a result for a hotel with a similar name, (ii) the hotel we searched for is closed, and Google returned the TripAdvisor page of the hotel currently open at the same location, and (iii) the name of the hotel on the TripAdvisor page is different enough from the Kantar name that the match is discarded even though it is correct.

[16]Only a fraction of hotels that are in the STR census decide to report financial information to STR.

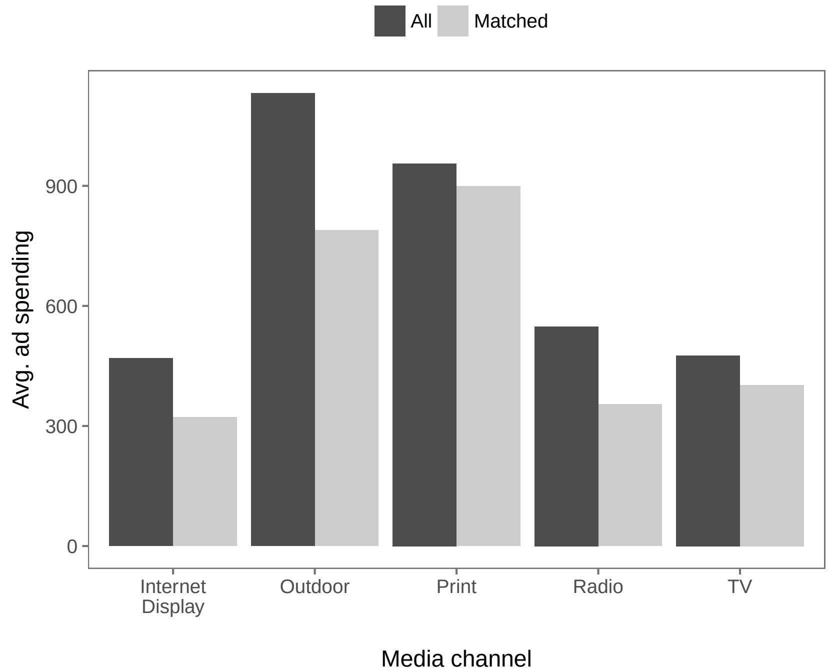


Figure 2: Comparing average advertising spending by media for all hotels in the full Kantar sample and the matched sample.

hotels, then, we have essentially a complete advertising profile: advertising spending in all the traditional media as well as advertising spending in internet display and internet search.

Besides this dataset with a complete advertising profile, we created a supplementary dataset to focus on search advertising specifically. This dataset contains TripAdvisor ratings and search ad spending information for 9,008 independent hotels over 10 years (from 2006 to 2015), a total of 439,506 hotel-year-month observations. We use this second dataset to check whether the search advertising results obtained on the smaller Kantar-matched dataset hold true in the larger dataset.

3.2 Checks for matching bias and selection

Here, we compare the full Kantar dataset with the matched sample to see how the two samples might be different, if at all. First, we compare the advertising spending levels in the different media (internet display, outdoor, print, radio, and TV) for the matched and full

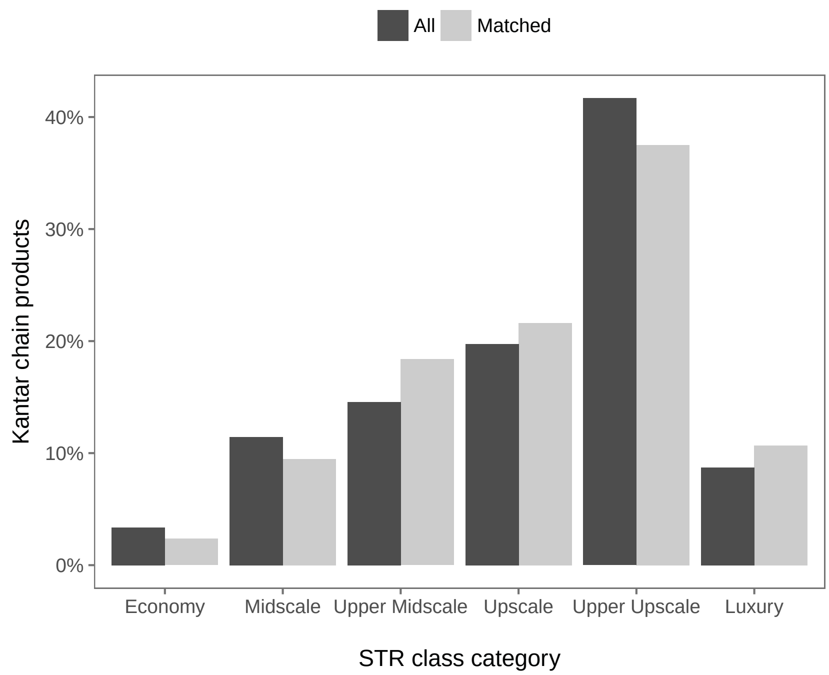


Figure 3: Comparing STR class distribution for hotel chains in the full Kantar sample and the matched sample.

Kantar sample in Figure 2. In general, while ad spending in each medium is smaller in the matched sample than in the full Kantar sample, the distribution is quite similar: outdoor and print are the media where hotels spend the most and Internet dispaly, radio, and TV the media where they spend less. Second, we compare the hotel class distributions for hotel chains in the matched sample and in the full Kantar sample in Figure 3.[17] These distributions are remarkably similar. Finally, we compare the two samples on average fraction of months with positive advertising and fraction of chains. The average fraction of months with positive advertising is 0.14 in both the full and matched sample; the fraction of chain hotels is 0.27 in the full sample and 0.26 in the matched sample. These results suggest that our matched sample is quite similar to the full Kantar sample.

[17]This comparison is only possible for chain hotels because we are able to identify chain name and class for the whole Kantar sample using a list of US chains and their class obtained from STR.

4 Descriptive evidence

In this section we describe the general patterns in the ratings and advertising data. We start with Table 1 where we show the summary statistics of TripAdvisor user ratings and advertising spending for the years 2002 (the start of our sample) and 2015 (the end of our sample). Comparing the second and third columns shows how user ratings and advertising spending have changed over a 14-year period.

	2002	2015
Hotels	2,805	5,446
Fraction of months with advertising	0.18	0.12
Fraction of hotels reviewed	0.25	1.00
Ratings and reviews		
Avg. hotel rating	3.89	4.08
Avg. reviews per hotel	1.71	558.07
Average monthly advertising expenditure($)		
Internet display	73	162
Print	1,576	526
Outdoor	202	94
Radio	30	35
Television	40	33
Total (Kantar media)	1,921	849
Internet search[18]	1,326	1,153

Table 1: Summary statistics

First, both average user ratings and number of reviews have increased over time. Ratings increased by about 0.2 stars while the number of reviews grew exponentially from about 2 reviews per hotel in 2002 to over 550 per hotel in 2015. Second, spending on traditional media decreased over the same period. Average monthly spending per property in the traditional media—print, radio, TV, and outdoor—decreased by about 56% from 2002 to 2015, from about \$2,000/month in 2002 to about \$850/month in 2015.[19] This decrease was particularly

[18] Only for independent hotels. Column 1 now refers to the year 2006, the first year for which search advertising data was available from SpyFu.

[19] These are nominal spending numbers, not inflation-adjusted numbers. If we were to plot the latter, the decrease would be even steeper.

pronounced for print advertising, followed by outdoor and TV advertising spending; see Figure 4. The decrease in ad spending in traditional media is offset, however, by a large

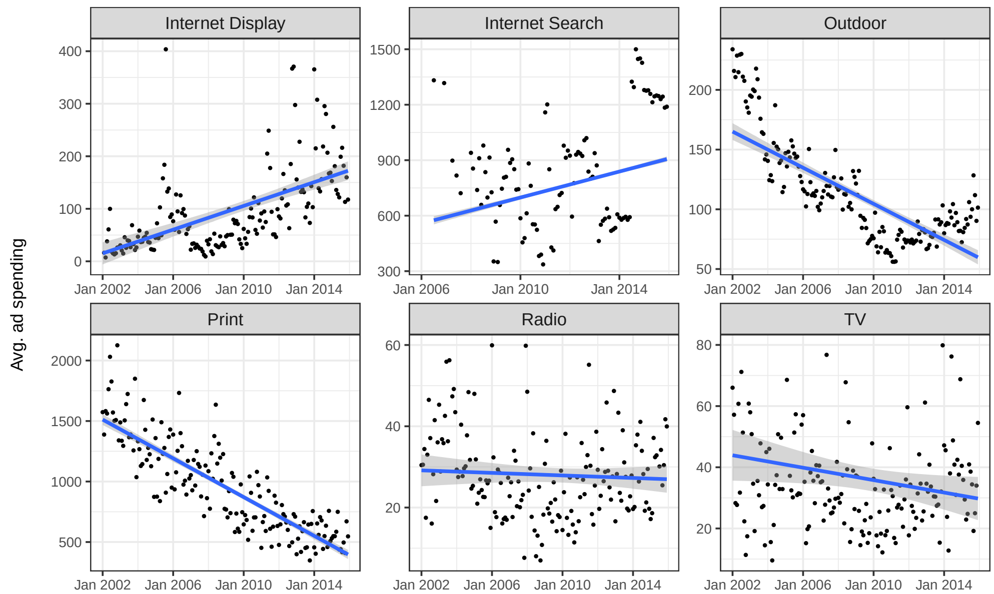


Figure 4: Average year-month ad spending by media channel. The blue line represents a linear fit and the grey shade the 95% confidence interval. Note that internet search advertising covers independent hotels only.

increase in online advertising. Internet display advertising shows a steep increase of about 116%, and search advertising (by independent hotels) shows an even steeper increase, from essentially zero in 2002 to about \$900 per month by 2015.

Figure 5 shows monthly average advertising spending in Kantar media by different types of hotels: different hotel classes and independent versus chain hotels. Luxury hotels advertise the most, spending almost as much as all the other hotel tiers combined; they also decline the most. Comparing chains to independent hotels, ad spending, excluding internet search, decreased more for the former. Figure 10 in the Appendix shows the effect of search advertising for independent hotels. Because search ads partially replace traditional ads, total ad spending does not fall as much, although it declines by 34% from 2002-2015.

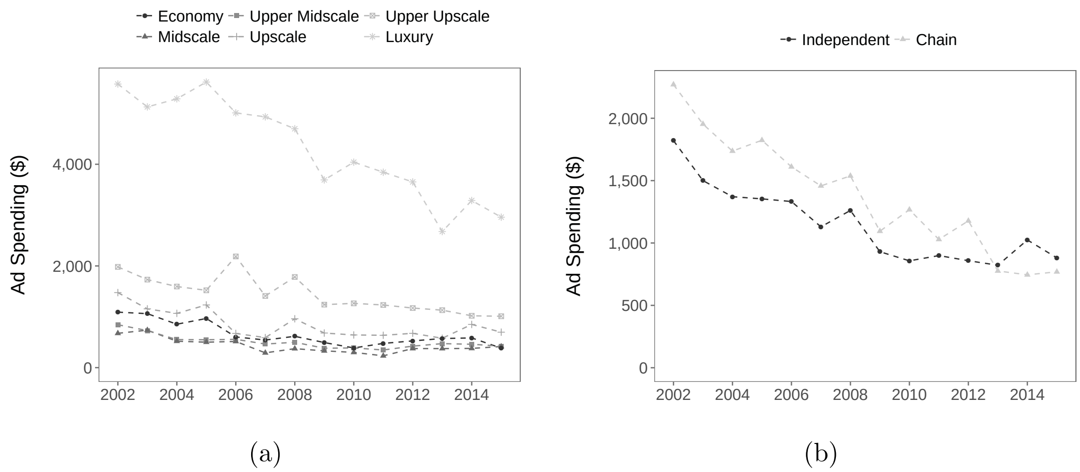


Figure 5: Year-over-year average monthly ad spending in Kantar media by hotel class and independent versus chain hotels. Note: internet search advertising is not included.

Finally, the relationship between advertising spending and TripAdvisor ratings is plotted in Figure 6 for two time periods, the years 2002-2005 and the years 2012-2015. We make two observations. First, comparing the two panels, the relationship between ratings and ad spending is very noisy in the early years, whereas it coalesces into a fairly well-defined inverted-U in the later years. A similar pattern is evident for search advertising by independent hotels in Figure 7. (In both figures, even in the later years, there is a lot of noise at the low ratings end, reflecting the small sample sizes there.) Since review platforms' influence has steadily increased over the years (Lewis and Zervas 2016), this suggests that hotels started reacting to TripAdvisor reviews only after consumers started responding to them.[20] And once they start doing so, the inverted-U relationship between ad spending and average ratings that develops is reminiscent of previous results in the literature such as Horstmann and Moorthy (2003) for restaurants and Dhar and Moorthy (2017) for movies.

[20]Indeed, Lewis and Zervas's (2016) work shows that the relationship between ratings and hotel revenue became significant only in 2006, after which it steadily increases year-over-year.

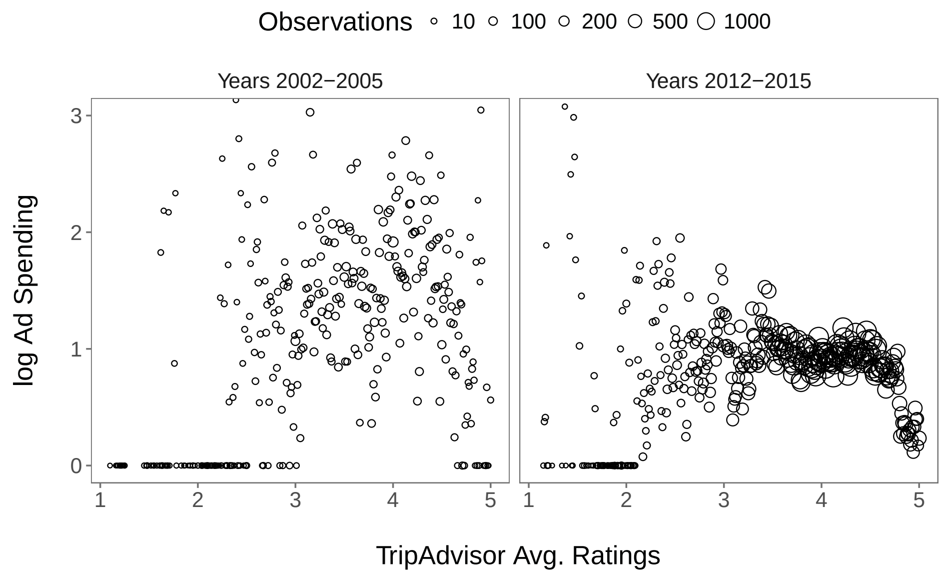


Figure 6: Relationship between advertising spending (excluding search advertising) and hotel ratings: 2002-2005 versus 2012-2015.

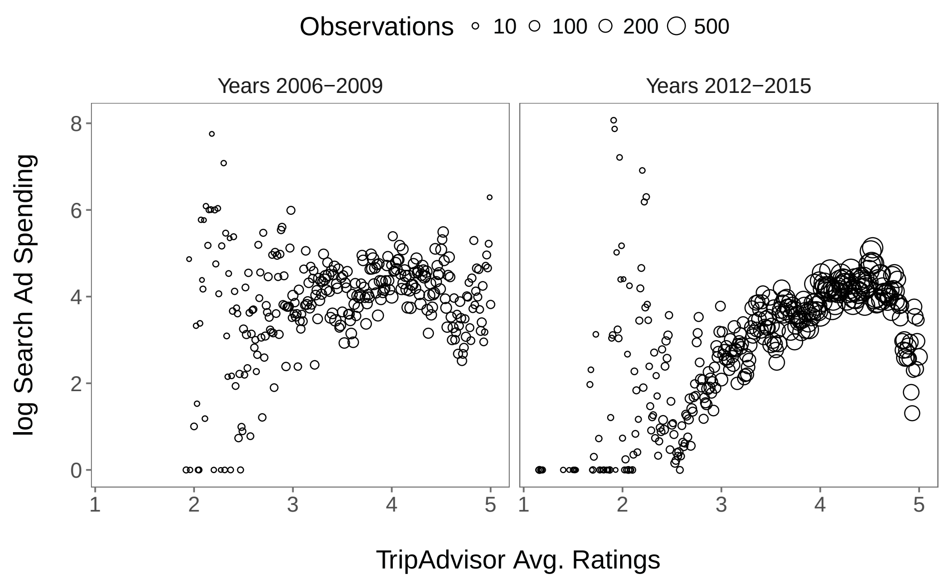


Figure 7: Relationship between search advertising spending and hotel ratings: 2006-2009 versus 2012-2015; independent hotels only, Kantar-matched SpyFu data

5 Empirical analysis

5.1 Econometric framework

We estimate a demand-side causal effect of online ratings on advertising spending using a regression discontinuity design (RDD). The RDD exploits the rounding rule TripAdvisor uses to convert average ratings into displayed ratings. Specifically, TripAdvisor's displayed ratings are the average ratings of reviewers rounded to the nearest half- or full-star. Thus, for example, a hotel with an average rating of 3.74 is shown as a 3.5-star hotel, while a hotel with an average rating of 3.75 stars is shown as a 4-star hotel. If we assume that average ratings (after a sufficient number of reviewers) reflect the actual quality of hotels subject to some random noise, this rounding mechanism creates discrete, random variation in perceived quality that is effectively exogenous to a hotel's true quality around rounding thresholds (see Figure 1). To the extent that hotel marginal costs depend on hotel quality, they depend on true quality, not the discontinuous changes in displayed ratings produced by TripAdvisor's rounding rule. Therefore, any variation in a firm's advertising level that correlates with average ratings crossing those rounding thresholds represents a causal demand-side effect of quality, not a cost-side effect.

Implicit in this identification strategy are two assumptions. First, average ratings are a continuous function of underlying true quality. In practice, this means that we need to focus our estimation on a narrow band of observations around each rounding threshold. Our large dataset allows us to live with this restriction while retaining sufficient statistical power. Second, conditional on the underlying quality, variation in which firms are above a threshold and which are below is essentially random.

The last assumption would be problematic if firms could manipulate ratings (Mayzlin et al. 2014). The concern would be that hotels might write fake "user reviews" awarding high ratings to themselves and giving low ratings to their competitors. If firms could successfully pull off this maneuver around the rounding thresholds and simultaneously adjust their

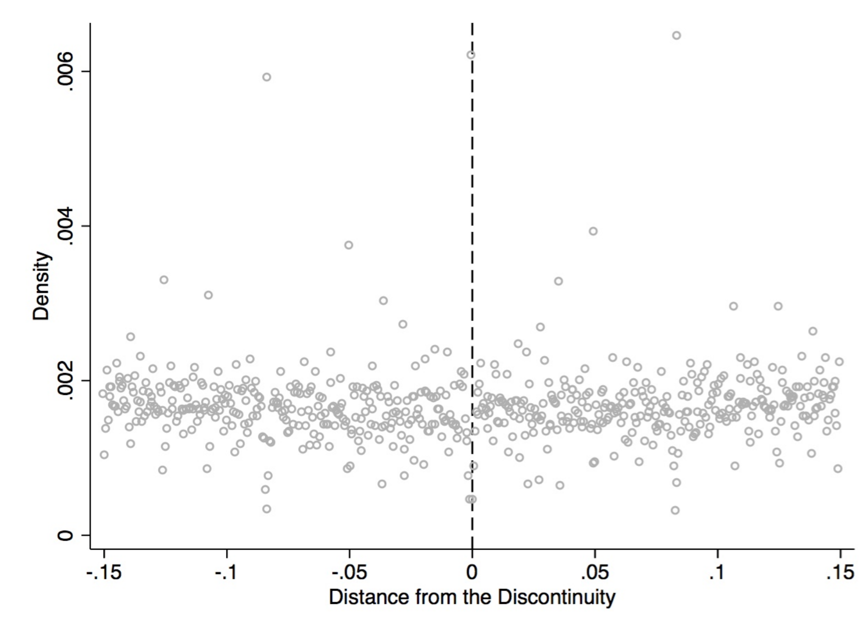


Figure 8: McCrary test: distribution of average ratings near rounding thresholds

advertising levels, then any correlation between advertising spending and displayed ratings, while still technically a demand-side effect, is probably best characterized as a supplier-driven demand-side effect.

We perform a series of tests to alleviate the concern that review manipulation is driving our results. These are discussed in detail in Section 6.1, but here we provide only a brief synopsis. The primary check is an RDD to test whether the running variable—in this case, average rating—is continuous in a neighborhood of the rounding thresholds. If there is upward manipulation of ratings, we would expect to see relatively few firms with ratings just below the threshold and a clump of firms with ratings just above the threshold; if there is both upward and downward manipulation of ratings, we should see clumps both below and above the threshold. Figure 8 shows that the distribution of average ratings is essentially uniform with neither dip nor bump below the rounding cutoffs and no bulge above the rounding cutoffs.[21] The visual evidence is confirmed with the standard McCrary test of continuity for RD designs (McCrary 2008) in Table 2. Second, we show that hotels just above the threshold

[21]While round numbers such as the rounding cutoffs are more common, the rounding cutoffs themselves are no more likely than other non-rounding round numbers such as 4.15 and 4.35.

do not differ systematically form those just below the threshold (see Table 3). Finally, we follow Anderson and Magruder (2012) and present several additional tests based on reviews, reviewers, and hotels characteristics to provide evidence that potentially fake reviewers are continuous around the rounding thresholds. These are discussed in Section 6.1.

	BW=0.05	BW=0.075
Above Threshold	0.00117	−0.00017
	(0.00118)	(0.00039)
Avg. Ratings	−0.03048	−0.00621
	(0.02947)	(0.00697)
Above Threshold × Avg. Ratings	0.03805	0.01192
	(0.04667)	(0.00998)
Year-month FE	Yes	Yes
Brand FE	Yes	Yes
N	251	375
R^2	0.54	0.59

Note: The dependent variable is density of average ratings, computed on a bin size of 0.0004 stars. Pooled RDDs; only hotels with 20 or more reviews included. Robust standard errors in parentheses.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 2: McCrary test

5.2 Specification

Having assured ourselves that the RDD identification assumptions are likely to hold in our sample, we proceed to estimating our RD regression. The specification is as follows:

$$\begin{aligned} \log \text{Ad Spending}_{it} &= \beta_1 \text{Above Threshold}_{it} + \beta_2 \text{Avg Ratings}_{it} \\ &+ \beta_3 \text{Above Threshold}_{it} \times \text{Avg Ratings}_{it} + \alpha_i + \tau_t + \epsilon_{it}. \end{aligned} \tag{1}$$

Here, t refers to the beginning of each month. Thus, Avg Ratings$_{it}$ is average rating of hotel i at the beginning of month t. The dependent variable, however, aggregates ad spending over the subsequent six months, i.e., log Ad Spending$_{it}$ is the logarithm of total advertising

	Below	Above	Difference (SE)
Hotel is Chain	0.38	0.38	0.0017
			(0.004 0)
Hotel Rooms	197.70	195.40	2.34
			(1.96)
Number of Reviews	259.50	258.10	1.34
			(3.43)
Hotel Class	4.01	4.01	−0.0043
			(0.013)
Hotel Location	4.42	4.41	0.012
			(0.011)
Hotel has Meeting Space	0.79	0.79	0.000 89
			(0.003 3)
Hotel Brand	18.60	18.50	0.063
			(0.23)
Hotel Price	164.60	165.70	−1.16
			(0.96)

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 3: Randomization check: Comparison of hotel characteristics above and below the threshold

spending of hotel i in the period $[t, t + 6 \text{ months}]$. We sum ad spending over six months because (i) ads are purchased ahead of when they are delivered and (ii) aggregating over six months reduces noise in the variable even at the cost of making our analysis overly conservative in finding an effect of ratings on advertising.[22]

Above Threshold$_{it}$, whose coefficient is the main object of interest, is an indicator of whether the average rating of hotel i at time t falls above the rounding threshold (for it to be rounded up to the nearest half- or full-star). The inclusion of separate slopes for average ratings above and below the threshold allows for a more flexible specification; it increases our confidence that β_1 actually represents the difference in advertising spending between hotels

[22]Ideally, we would observe when ad spending decisions are made and relate them directly to average and displayed ratings at that time. Since we only observe actual spending and not planned spending, we use this time-lagging strategy. Because this weakens the relationship between the dependent variable and our treatment of interest, it only makes it harder to detect a ratings-ad spending effect. Under-measuring the magnitude of the effect can be thought of as a conservative approach; it enhances our ability to prevent type-1 errors.

that differ by a half-star in their displayed ratings. All specifications include year-month fixed effects, τ_t, and brand fixed effects, α_i.[23]

Finally, we made two important design choices. First, we take advantage of the large dataset available to us by limiting attention to hotels that are within 0.05 stars of each rounding threshold. Second, to have confidence that average ratings actually represent underlying true quality, we limit all our analyses to observations in which hotels have 20 or more reviews. In Section 6.2, we carry out a number of tests to check the robustness of our specification to different functional forms, bandwidths, and different aggregation windows of the dependent variable.

5.3 How online ratings affect advertising spending

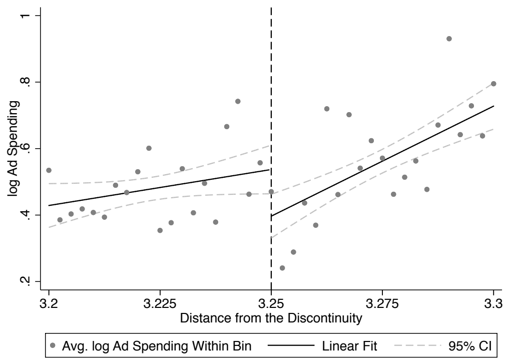
We start our analysis by presenting visual evidence that hotels' ad spending levels are sensitive to their TripAdvisor ratings. In Figure 9, we show the relationship between average ratings at time t and logarithm of total advertising spending in the period $[t, t + 6 \text{ months}]$ at different rounding thresholds: 3.25, 3.75, 4.25, and 4.75.[24] There are clear jumps at thresholds 3.25, 4.25, and 4.75, and a less clear one at threshold 3.75. Recall that crossing a threshold increases the displayed rating by half a star. Thus, Figure 9 shows that for all but the 3.75 threshold, the discrete increase in displayed ratings around the rounding thresholds results in a reduction in the amount of advertising spending.

Table 4 presents detailed RDD estimates for the visual evidence. The coefficient of interest, Above Threshold$_{it}$, is negative for all thresholds and statistically significant for all except the 3.75 threshold. The results suggest that hotels above the threshold spend on average between 6% and 15% less on advertising than hotels below the thresholds.

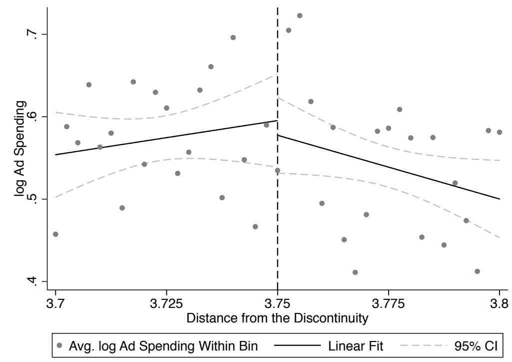
Because of the consistent negative effects of displayed ratings on advertising spending across thresholds, in all subsequent analyses, we pool the thresholds together. Column 1

[23] We thank the anonymous reviewers for this suggestion.

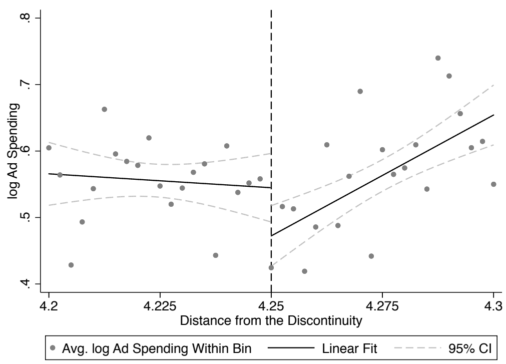
[24] At levels below 3 stars, there are not enough observations to estimate an effect of being above or below the threshold. Fewer than 5% of hotels had a rating below 3 stars in 2015.



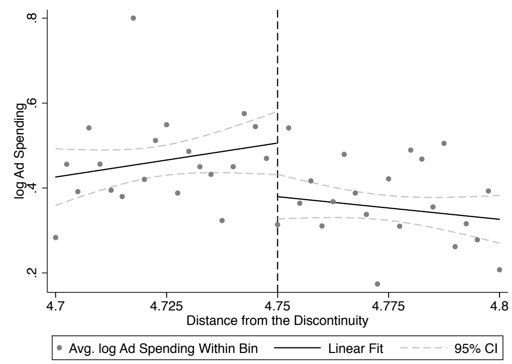
(a)



(b)



(c)



(d)

Figure 9: Relationship between advertising spending and TripAdvisor average ratings at various rounding thresholds

of Table 7 repeats Table 4 analysis on the pooled sample. The results suggest that an extra half-star in TripAdvisor's displayed ratings causes hotels to spend about 7% less on advertising.

5.4 Effects by media

Next, we look at how the effect of ratings on ad spending varies across advertising media. We use the same specification as before, but the dependent variable now is log of ad spending within a particular ad medium. Results are presented in Table 5. As the table shows, our earlier finding is replicated within each media as well. Search advertising is by far the most responsive, followed by print, Internet display, outdoor, and TV; radio is the least responsive.

	3.25	3.75	4.25	4.75
Above Threshold	−0.156***	−0.028	−0.062*	−0.108**
	(0.044)	(0.032)	(0.029)	(0.042)
Avg. Ratings	3.409**	1.269	−0.704	2.657*
	(1.090)	(0.806)	(0.736)	(1.072)
Above Threshold × Avg. Ratings	0.896	−1.814	3.463***	−3.448*
	(1.479)	(1.060)	(0.978)	(1.382)
Year-month FE	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes
N	10825	22113	24824	10215
R^2	0.085	0.11	0.12	0.10

Note: The dependent variable in each column is log of ad spending in the following 6 months. All columns use a bandwidth of 0.05 stars around the rounding cutoff in the column header. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 4: How TripAdvisor ratings affect ad spending: RDD estimates

	Internet display	Internet search	Outdoor	Print	Radio	TV
Above Threshold	−0.027***	−0.206*	−0.017*	−0.039*	−0.009*	−0.011*
	(0.008)	(0.102)	(0.007)	(0.015)	(0.004)	(0.005)
Avg. Ratings	0.313	2.335	0.437*	0.703	0.071	0.168
	(0.195)	(2.509)	(0.175)	(0.389)	(0.097)	(0.120)
Above Threshold × Avg. Ratings	0.156	1.707	−0.074	0.262	0.230	−0.161
	(0.256)	(3.421)	(0.239)	(0.517)	(0.130)	(0.155)
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes	Yes	Yes
N	67977	22455	67977	67977	67977	67977
R^2	0.029	0.19	0.026	0.094	0.055	0.023

Note: The dependent variable is log of ad spending in the following 6 months in particular media. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only hotels with 20 or more reviews are considered. Robust standard errors in parentheses. Internet search contains only observations for independent hotels.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 5: Ad spending effects by media type

5.5 Extensive versus intensive margins

Because many firms do not advertise in many periods, we examine whether the effect of ratings on advertising operates on the extensive margin or the intensive margin. In other words, are firms responding to their ratings by varying whether they advertise or do they just adjust the quantity of advertising? To answer this, we repeat the RD analysis, first on the subset of firms that had a positive ad spending level in the past year, and second, using the full sample. But now the dependent variable is a dummy variable, equaling one if the firm advertises in the following six months, zero otherwise. Results of these analyses are presented in Table 6. We see a significant response in both margins for overall spending. Firms are less likely to advertise if they are above the rounding threshold and, among the firms who advertise, the amount they spend is significantly lower for firms above the threshold.

	Extensive margin	Intensive margin
Above Threshold	−0.021***	−0.16***
	(0.006)	(0.042)
Avg. Ratings	0.27	1.87
	(0.152)	(1.047)
Above Threshold × Avg. Ratings	0.016	1.04
	(0.203)	(1.405)
Year-month FE	Yes	Yes
Brand FE	Yes	Yes
N	23030	73777
R^2	0.095	0.218

Note: The dependent variable in column 1 is a dummy variable for whether or not the firm advertised in the following 6 months; in column 2 it is the log of ad spending in the following 6 months for firms that advertised in the preceding year. Both columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 6: Ad spending effects on the intensive and extensive margins

5.6 Hotel type effects

In this section, we explore whether the ratings-ad spending relationship varies by whether the hotel is independent or part of a chain.

	All hotels	Chains	Independent	Independent (including search ads)
Above Threshold	−0.070***	−0.024	−0.088***	−0.229*
	(0.018)	(0.029)	(0.022)	(0.099)
Avg. Ratings	1.035*	0.572	1.173*	1.839
	(0.445)	(0.737)	(0.556)	(2.428)
Above Threshold × Avg. Ratings	0.655	0.517	0.789	2.621
	(0.591)	(0.982)	(0.738)	(3.315)
Year-month FE	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes
N	67977	22546	45431	22455
R^2	0.079	0.21	0.012	0.18

Note: The dependent variable is log of ad spending in the following 6 months for the hotels covered by the column heading. The first 3 columns do not include search advertising. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parenthesis.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 7: Ad spending effects by hotel type

Table 7, columns 2-4, presents the results. In columns 2 and 3, we report the results for chain hotels and independent hotels, respectively, excluding spending on search ads. In column 4, we report the results for independent hotels including search ads in the total ad spending.[25] We observe that the relationship for chains is negative, but not statistically significant, whereas for independent hotels (column 3), it is ($p < .001$). The latter is also economically meaningful: a half-star increase in TripAdvisor ratings reduces independent hotels' ad spending by about 9%. Although the coefficients for "chain hotels" and "independent hotels" are not statistically different from one another, this is suggestive that the ad responses of independent hotels drives much of the overall result. Finally, in column 4,

[25]As we explained in Section 3, we are able to obtain Internet search data only for independent hotels starting in the year 2006.

when we repeat the analysis for independent hotels including search ad spending, we find an even stronger result.

Why might independent hotels be more responsive to their ratings than chain hotels? We conjecture that this is because the chains' national branding insulates them from reviews—both on the positive side and on the negative side. When demand is insensitive to reviews, these hotels have less reason to respond.

	Fewer than 100 properties		More than 100 properties	
	Non-luxury	Luxury	Non-luxury	Luxury
Above Threshold	−0.350***	−0.072	−0.025	−0.064
	(0.096)	(0.135)	(0.034)	(0.078)
Avg. Ratings	6.880**	−7.000*	0.260	7.040***
	(2.446)	(3.423)	(0.847)	(1.964)
Above Threshold X Avg. Ratings	−10.100**	16.600***	0.530	−7.170**
	(3.284)	(4.690)	(1.125)	(2.628)
Year-month FE	Yes	Yes	Yes	Yes
N	2465	3115	12588	4386
R^2	0.063	0.043	0.014	0.079

Note: The dependent variable in each column is log of search ad spending in the following 6 months. Only hotels with 20 or more reviews are included. Chain size and luxury class come from STR. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 8: How ad spending effects vary by chain size and quality

We next test how the effect varies within chains, based on size and quality-tier differences. We separately estimate for small (less than 100 properties) and large (more than 100 properties) chains, and distinguish between luxury and non-luxury chains (as classified by STR). (100 properties is approximately the 25th percentile of chain size in our data.) Table 8 shows a significant relationship only for small non-luxury chains, suggesting that brand strength may be a function not only of size, but also prominence. Luxury chains such as Ritz-Carlton or W, though small, might still be prominent enough to insulate them from reviews.

5.7 Market effects

Now, we examine how the degree of differentiation in a market affects the relationship between ratings and ad spending. We operationalize extent of differentiation by using the STR definition of a market as a Metropolitan Statistical Area (MSA) and calculating the standard deviation of average ratings in each market each year. Our hypothesis is that in markets with relatively low differentiation in TripAdvisor hotel ratings, the boost in displayed ratings around the rounding thresholds might be more pivotal. In such markets, then, we should see a bigger ad response than in markets where average ratings are already well-differentiated to begin with.

	Kantar media spending		Search spending	
	High ratings std dev	Low ratings std dev	High ratings std dev	Low Ratings std dev
Above Threshold	−0.062	−0.075***	−0.097	−0.23*
	(0.033)	(0.021)	(0.133)	(0.097)
Avg. Ratings	0.012	1.51**	−3.78	2.49
	(0.846)	(0.525)	(3.394)	(2.420)
Above Threshold × Avg. Ratings	3.01**	−0.45	11.8**	−3.52
	(1.108)	(0.708)	(4.425)	(3.265)
Year-month FE	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	No	No
N	21788	46342	10723	24764
R^2	0.098	0.085	0.019	0.014

Note: The left (right) column refers to markets where the standard deviation of average ratings is higher (lower) than the median standard deviation. The dependent variable in each column is log of ad spending in the following 6 months. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 9: Ad spending effects by market competitiveness

Table 9 shows the results of this analysis. Comparing markets with above-the-median differentiation and markets with below-the-median differentiation, we find that ratings have a stronger negative effect on ad spending in less-differentiated markets than in more-differentiated markets, and that this difference is especially pronounced for search advertising.

5.8 Early versus late-period effects

In Figure 3, we saw that the ratings-advertising relationship in 2002-2005 was mostly noise, whereas in 2012-2015 a well-defined inverted-U relationship was present. We hypothesize that this difference is driven by the growing popularity of TripAdvisor among users. In the early years of our dataset, it is likely that few people visited TripAdvisor to read reviews and make buying decisions based on them. However, by 2015, reviews and review platforms such as TripAdvisor had become extremely popular and many consumers were using them to make buying decisions. (As we noted in footnote 9, by November 2016, TripAdvisor was

	2002-2005	2012-2015
Above Threshold	−0.121	−0.106***
	(0.141)	(0.024)
Avg. Ratings	4.705	1.124
	(3.283)	(0.602)
Above Threshold × Avg. Ratings	−0.703	0.522
	(4.572)	(0.804)
Year-month FE	Yes	Yes
Brand FE	Yes	Yes
N	2191	34496
R^2	0.18	0.066

Note: The dependent variable is log of ad spending in the following 6 months during the period of the column heading. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 10: Effect of TripAdvisor ratings on ad spending, early versus late

among the top 50 sites in the entire internet.) We test this in Table 10 in which we report the results of two RD regressions, one based on 2002-2005 data and another based on 2012-2015 data. We find that in the 2002-2005 period, although the point estimate of the ratings effect is large, the coefficient is not statistically significant because of the large amount of noise in the data. During the 2012-2015 period, however, the relationship between ratings and advertising becomes highly significant. In fact, as expected, the negative relationship in the later period is even stronger than in our earlier regressions where we aggregated across all periods (Table 7, column 1). This analysis suggests that it is not the mere presence of reviews that induces firms to alter their advertising level, but rather, the realization that large numbers of consumers are making purchase decisions based on those reviews.

6 Robustness checks

In this section, we describe two types of robustness checks. First, we report tests that show that ratings manipulation on TripAdvisor cannot explain our results. Second, we evaluate the robustness of our results to alternative formulations of the RDD: alternative functional forms, alternative bandwidths, placebo thresholds, and inclusion of hotel prices as a control.

6.1 Ratings manipulation

Our interpretation of the observed RDD effects as a causal demand-side effect of quality on advertising spending relies on the assumption that random exogenous factors, rather than the motivated actions of the actors themselves, are responsible for the movement of average ratings around the rounding thresholds. If, instead, hotels were responsible for those movements, then our results would be better characterized as supplier-driven-demand-side effects.

However, it is not easy to explain our results as ratings manipulation driven. For instance, it is not enough to say that hotels write positive fake reviews to push their average ratings just above TripAdvisor's rounding thresholds (Luca 2016, Anderson and Magruder 2012, Lei 2017). To invalidate our RDD results, ratings manipulation must occur differently above and below the rounding thresholds. For example, if all hotels manipulated their reviews upward, this would simply shift the average ratings distribution to the right, without impacting our results. Alternatively, if some hotels manipulated their reviews upward, while their competitors did just the opposite, the average ratings distribution would still be the same, and our results wouldn't be affected.

We provide several pieces of evidence to argue that ratings manipulation does not occur differently above and below the rounding thresholds. First, hotels just above and just below the rounding thresholds are almost identical in their observable characteristics (Table 3). Second, as Figure 8 shows, the density of average ratings, is constant around the rounding thresholds. There are no jumps, up or down, below or above the thresholds. Third, a formal McCrary (2008) test confirms what we see visually. RDDs using density of average ratings as the running variable, with two different bandwidths (0.05 and 0.075 stars) show no evidence of discontinuities in the average ratings density around the rounding thresholds (Table 2).

Another approach to testing for ratings manipulations involves checking the continuity of variables associated with fake reviews and reviewers (Anderson and Magruder 2012). Following Mayzlin et al. (2014) and Luca and Zervas (2016) who observe that positive fake reviews tend to be 5-star, we check whether the fraction of 5-star reviews is significantly different above and below the rounding thresholds. Further, because fake reviewers have, on average, fewer reviews than genuine reviewers (Luca and Zervas 2016), we check whether there are significant differences between reviewers who post ratings above and below the discontinuities. The results of these analyses are in Table 11, columns 1-4. The coefficient of interest, Above Threshold, is statically insignificant in all cases, suggesting that review manipulation does not occur differently above and below the threshold.

Next, we turn to analyzing the text of the reviews. Using LIWC (Pennebaker et al. 2015), a software for automated text analysis, we compute a measure of "authenticity" for every review.[26] LIWC's authenticity algorithm is based on a series of studies that tested whether computer-based text analysis could differentiate between honest and deceptive linguistic styles (Newman et al. 2003). While Pennebaker's formula is proprietary, he discusses the factors that go into determining authenticity in his book (Pennebaker 2011).[27] Using LIWC, we check whether review authenticity scores differ above and below the rounding thresholds.

[26]For a recent application of LIWC's authenticity metric, to test for fake financial news, see Kogan et al. (2017).

[27]In Appendix B we provide some evidence that LIWC authenticity scores capture, at least in part, the authenticity of a review.

	Fraction of 5-star reviews	Average reviews per reviewer	Reviewers with 0 prior reviews	Reviewers with < 6 prior reviews	Fraction of reviews with response
Above Threshold	0.0048	−0.0116	−0.0009	−0.0001	0.0009
	(0.0037)	(0.0177)	(0.0024)	(0.0020)	(0.0029)
Avg. Ratings	0.0774	0.9076*	−0.1047	−0.0144	0.1088
	(0.0930)	(0.4407)	(0.0605)	(0.0506)	(0.0751)
Above Threshold	0.1583	−0.7688	0.1321	0.0075	−0.2584**
× Avg. Ratings	(0.1252)	(0.5978)	(0.0825)	(0.0687)	(0.0977)
Year-month FE	Yes	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes	Yes
N	73595	46118	46869	46869	73595
R^2	0.14	0.36	0.25	0.37	0.21

Note: The dependent variable in column 1 is fraction of 5-stars reviews; in column 2 it is average reviews per reviewer prior to the current review; in columns 3 and 4 it is, respectively, fraction of reviewers with no prior reviews and fraction of reviewers with less than 6 prior reviews; in colum 5 the fraction of reviews with a manager response. All columns use pooled RDDs with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 11: Test for review manipulation: reviews and reviewers characteristics

The results of this analysis, reported in Table 12, show no difference in text authenticity above and below the thresholds.

Yet another test leverages Mayzlin et al. (2014) findings that certain types of hotels are less likely to post fake positive reviews. Specifically, one of the findings of Mayzlin et al. (2014) is that independent hotels owned by a small owner will generate more fake positive reviews than independents hotels owned by a big owner. Similar to Mayzlin et al. (2014), we use the STR census dataset to identify hotels managed by small (the owner manages one hotel) and big owners (the owner manages two or more hotels). Then, we re-estimate our main specification on a subset of independent hotels whose owner is big (and thus less likely to post positive fake reviews). We report these results in Table 13, in column 1 for owners with two or more hotels and in column 2 (for robustness) for owners with five or more hotels. In both cases, our results continue to hold.

Finally, we check for the possibility that hotels increasing their ratings by providing responses to reviews might explain our results. This is based on the work of Proserpio and Zervas (2017) who show that hotel managers who respond to reviews get higher ratings. In column 5 of Table 11, we check whether the fraction of reviews with responses shows dis-

	Authenticity	log(Authenticity)
Above Threshold	0.0207	−0.0008
	(0.0901)	(0.0018)
Avg. Ratings	−2.2240	−0.0171
	(2.2519)	(0.0441)
Above Threshold × Avg. Ratings	1.2701	0.0075
	(3.0644)	(0.0602)
Year-month FE	Yes	Yes
Brand FE	Yes	Yes
N	73595	73595
R^2	0.15	0.14

Note: The dependent variable in column 1 is the average LIWC authenticity measure of reviews written by hotel i at year-month t, while in column 2 is the log of average LIWC authenticity measure of hotel i at year-month t. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 12: Test for review authenticity

	2+ owners	5+ owners
Above Threshold	−0.114**	−0.074*
	(0.036)	(0.034)
Avg. Ratings	0.596	−0.151
	(0.962)	(0.912)
Above Threshold × Avg. Ratings	2.059	2.688*
	(1.231)	(1.173)
Year-month FE	Yes	Yes
Brand FE	Yes	Yes
N	17393	15585
R^2	0.020	0.027

Note: The dependent variable in each column is the log of ad spending of hotel i in the following 6 months . All columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 13: Test for review manipulation: independent hotels managed by large versus small owners

continuities around the rounding thresholds. Again, the coefficient of interest is statistically insignificant.

In summary, this section provides evidence that review manipulation is not driving our results.

6.2 Alternative RDD specifications

Quadratic polynomial estimator. All our analyses so far have been based on a linear estimator. In this section, we test the sensitivity of our results to a different functional form, a quadratic polynomial estimator. We do so by adding to Equation 1 a quadratic term for the cumulative average rating variable, and allow for a separate effect of the quadratic term above and below the discontinuity. The specification we estimate is as follows:

$$\begin{aligned} \log \text{Ad Spending}_{it} &= \beta_1 \text{Above Threshold}_{it} + \beta_2 \text{Avg Ratings}_{it} + \beta_3 \text{Avg Ratings}_{it}^2 \\ &+ \beta_4 \text{Above Threshold}_{it} \times \text{Avg Ratings}_{it} \\ &+ \beta_5 \text{Above Threshold}_{it} \times \text{Avg Ratings}_{it}^2 + \alpha_i + \tau_t + \epsilon_{it}. \end{aligned} \tag{2}$$

We report the results of this model in Table 14. The estimates are qualitatively similar to those using a local linear estimator, suggesting that our results are not sensitive to the functional form used for the estimation.

Different bandwidths. We test the sensitivity of our RDD to alternative bandwidths. We test three different bandwidths – 0.025, 0.5, and 0.075 – in Table 15. The coefficient of interest, Above Threshold, is statistically significant for every bandwidth tested; however, the effect of ratings t is stronger for smaller bandwidths.

Placebo test. In order to test whether medium-term trends in quality and advertising both rising or falling together may be causing the RDD to falsely pick up a discontinuous change in advertising around the rounding thresholds, we estimate the regression discontinuity on

	All Firms
Above Threshold	−0.081**
	(0.027)
Avg. Ratings	0.670
	(1.916)
Above Threshold × Avg. Ratings	2.941
	(2.395)
Avg. Ratings^2	−6.882
	(34.556)
Above Threshold × Avg. Ratings^2	−32.068
	(44.864)
Year-month FE	Yes
Brand FE	Yes
N	67977
R^2	0.079

Note: The dependent variable is log of ad spending over the next 6 months. All columns use pooled RDDs with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 14: Quadratic specification

	BW=0.025	Bw=0.05	BW=0.075
Above Threshold	−0.107***	−0.070***	−0.025+
	(0.026)	(0.018)	(0.015)
Avg. Ratings	3.953**	1.035*	1.086*
	(1.367)	(0.445)	(0.445)
Above Threshold × Avg. Ratings	−1.604	0.655	−1.093*
	(1.723)	(0.591)	(0.445)
Year-month FE	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes
N	33523	67977	100292
R^2	0.074	0.079	0.085

Note: The dependent variable is log of ad spending over the next 6 months. All columns use pooled RDDs with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.

Significance levels: + p<0.1, * p<0.05, ** p<0.01, *** p<0.001.

Table 15: Sensitivity analysis: different bandwidths

placebo thresholds, i.e., thresholds where there is no jump in the displayed ratings. The results are in Table 16. The placebo thresholds we chose were 3.1, 3.6, 4.1, and 4.6. The coefficient of interest, Above Threshold, is quite close to zero and statistically insignificant.

	All Firms
Above Threshold	−0.015
	(0.018)
Avg. Ratings	−0.308
	(0.448)
Above Threshold × Avg. Ratings	0.439
	(0.617)
Year-month FE	Yes
Brand FE	Yes
N	69971
R^2	0.097

Note: The dependent variable in each column is the log of ad spending over the next 6 months. All columns use a pooled RDD around placebo cutoffs (3.1, 3.6, 4.1, 4.6 stars), with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parentheses.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 16: Sensitivity analysis: placebo test

Controlling for hotel prices. Luca (2016), using a RDD strategy similar to ours, demonstrates the existence of a causal relationship between ratings and revenue for restaurants. Because of his results, one may argue that hotels price is an omitted variable in our model, and that it can potentially bias our results. To reduce this concern, we reestimate equation 1 on a subset of hotels for which we have Average Daily Rates (ADR) (see Section 3). We report the results of this specification in Table 17. In column 1, hotel prices are not included; in column 2, they are. First, we notice that the results on the subset of hotels for which we have hotel prices are similar to those obtained with the full sample (see Table 7, column 1). Second, when we insert the logarithm of ADR as a control, the coefficient remain negative, statistically significant, and similar in magnitude to that reported in column 1, suggesting that our results are not affected by the omission of hotel prices.

	(1)	(2)
Above Threshold	-0.076^{**}	-0.072^{**}
	(0.027)	(0.026)
Avg. Ratings	1.841^{**}	1.572^{*}
	(0.671)	(0.667)
Above Threshold × Avg. Ratings	0.068	0.281
	(0.893)	(0.889)
log Hotel Price		0.368^{***}
		(0.020)
Year-month FE	Yes	Yes
Brand FE	Yes	Yes
N	36412	36412
R^2	0.12	0.13

Note: The dependent variable is log of ad spending in the following 6 months for the hotels covered by the column heading. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms for which we obtained ADR and with 20 or more reviews are included. Robust standard errors in parenthesis.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 17: How TripAdvisor average ratings affect ad spending, controlling for hotel prices

Different aggregation windows for advertising spending. Recall that we aggregate advertising spending over the six months following our ratings measurement. As we explained in Section 5.1, the motivation for this choice is twofold: (i) ads are often purchased far ahead of when they are delivered, and (ii) we wanted to reduce noise in the variable.

Table 18 shows that our results are robust to different aggregation windows. Column 1 shows our baseline specification using a time window of [t,t+6]; column 2 shows the results using a shorter window of [t,t+3]; column 3 shows "next month" spending and, finally, column 4 shows the results for a forward-looking window of [t+3,t+6]. In each case, we find a negative and significant effect.

	(1) log(Next 6 months)	(2) log(Next 3 months)	(3) log(Next 1 month)	(4) log(Next 3-6 months)
Above Threshold	−0.070***	−0.062***	−0.036***	−0.067***
	(0.018)	(0.014)	(0.009)	(0.016)
Avg. Rating	1.035*	0.90*	0.41	1.11**
	(0.445)	(0.352)	(0.233)	(0.395)
Avg. Rating × Above Threshold	0.655	0.64	0.48	0.15
	(0.59)	(0.470)	(0.311)	(0.528)
Year-month FE	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes
Observations	67977	70966	72832	67977
R^2	0.079	0.066	0.046	0.070

Note: The dependent variable is log of ad spending summed over the aggregation window in the column heading. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only firms with 20 or more reviews are included. Robust standard errors in parenthesis.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 18: Results with different time windows for advertising spending

7 Conclusion

This paper has examined the cross-sectional relationship between online quality ratings and advertising spending in the hotel industry, using a 14-year panel of TripAdvisor hotel reviews matched to advertising data from Kantar Media and SpyFu. Our results suggest that hotels' TripAdvisor ratings have a causal, demand-side effect on their advertising spending decisions. Hotels with higher ratings spend less on advertising than hotels with lower ratings. The effect is robust, seen both in aggregate advertising spending as well as in individual media spending, at the intensive margin as well as at the extensive margin. In short, the evidence is strong that hotels with high TripAdvisor ratings treat their ratings as a substitute for advertising and hotels with low TripAdvisor ratings treat advertising as a substitute for their ratings.

Beneath this broad substitution relationship, there are several interesting nuances. Independent hotels respond to their ratings, but chains generally do not—except small non-luxury chains. This suggests to us that having a strong well-known brand provides some immunity to reviews, a result previously seen in other contexts, such as movies (Eliashberg and Shugan 1997). We also find that hotels in less differentiated markets are more responsive to their online ratings than hotels in more differentiated markets, suggesting that firms are more motivated to respond when ratings are more likely to be competitively pivotal. Finally, in the time-series, we see that hotels have become more responsive over the years, just as TripAdvisor's influence has grown. This tells us that it is not the presence of reviews *per se* that triggers a reaction, but rather the recognition that consumers are using those reviews. It further corroborates our conclusion that what we are observing is a demand-side effect, and not a cost-side effect.

Our empirical analysis, based on regression-discontinuity designs, wouldn't be possible without the exogenous discontinuities in TripAdvisor's displayed ratings. However, this also contributes a limitation. While we can say with confidence that online ratings substitute for advertising ratings in the neighborhood of the discontinuities, we cannot say what happens causally, far from those discontinuities. Nor can we say whether large-scale changes in average ratings have only demand-side effects, or cost-side effects also. Finally, our results are likely sensitive to the particular institutional context of TripAdvisor ratings, in so far as TripAdvisor does not allow hotels to use its ratings in their advertising copy. In other contexts, this might be different. For example, in the movie industry, critics' ratings are routinely featured in advertising copy. Perhaps for this reason, Dhar and Moorthy (2017) report a positive relationship between ratings and ad spending for small, independent movies.

References

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

Archibald, Robert B, Clyde A Haulman, Carlisle E Moody Jr. 1983. Quality, price, advertising, and published quality ratings. *Journal of Consumer Research* **9**(4) 347–356.

Bagwell, Kyle. 2007. The economic analysis of advertising. *Handbook of industrial organization* **3** 1701–1844.

Cabral, Luís, Ali Hortaçsu. 2010. The dynamics of seller reputation: Evidence from ebay*. *The Journal of Industrial Economics* **58**(1) 54–78.

Caves, Richard E, David P Greene. 1996. Brands' quality levels, prices, and advertising outlays: empirical evidence on signals and information costs. *International Journal of Industrial Organization* **14**(1) 29–52.

Chen, Yubo, Jinhong Xie. 2005. Third-party product review and firm marketing strategy. *Marketing Science* **24**(2) 218–240.

Chen, Yubo, Jinhong Xie. 2008. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science* **54**(3) 477–491.

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

Dhar, Tirtha, Sridhar Moorthy. 2017. On the marketing of experience goods: the case of movies. Working Paper, University of Toronto.

Eliashberg, Jehoshua, Steven M. Shugan. 1997. Film critics: Influencers or predictors? *The Journal of Marketing* 68–78.

Hertzendorf, Mark N. 1993. I'm Not a High-Quality Firm-But I Play One on TV. *The RAND Journal of Economics* **24**(2) 236–247. doi:10.2307/2555760.

Hollenbeck, Brett. 2018. Online Reputation Mechanisms and the Decreasing Value of Chain Affiliation. *Journal of Marketing Research* **55**(5).

Honka, Elisabeth, Ali Hortaçsu, Maria Ana Vitorino. 2017. Advertising, consumer awareness, and choice: Evidence from the US banking industry. *The RAND Journal of Economics* **48**(3) 611–646.

Horstmann, Ignatius J., Sridhar Moorthy. 2003. Advertising spending and quality for services: The role of capacity. *Quantitative Marketing and Economics* **1**(3) 337–365.

Houser, Daniel, John Wooders. 2006. Reputation in auctions: Theory, and evidence from eBay. *Journal of Economics & Management Strategy* **15**(2) 353–369.

Jin, Ginger Zhe, Andrew Kato. 2006. Price, quality, and reputation: Evidence from an online field experiment. *The RAND Journal of Economics* **37**(4) 983–1004.

Kihlstrom, Richard E, Michael H Riordan. 1984. Advertising as a signal. *journal of Political Economy* **92**(3) 427–450.

Kim, MinChung, Leigh M. McAlister. 2011. Stock market reaction to unexpected growth in marketing expenditure: Negative for sales force, contingent on spending level for advertising. *Journal of Marketing* **75**(4) 68–85.

Kogan, Shimon, Tobias J Moskowitz, Marina Niessner. 2017. Fake News in Financial Markets. Working Paper, Yale University.

Lei, Ying. 2015. How do firms advertise when customer reviews are available? Working Paper, Boston University.

Lei, Ying. 2017. Local restaurants advertising response to a better online rating. Working Paper, Peking University.

Lewis, Gregory, Georgios Zervas. 2016. The welfare impact of consumer reviews: A case study of the hotel industry. Working Paper, Boston University.

Linnemer, Laurent. 2002. Price and advertising as signals of quality when some consumers are informed. *International Journal of Industrial Organization* **20**(7) 931–947. doi:10.1016/S0167-7187(01)00081-9.

Luca, Michael. 2016. Reviews, reputation, and revenue: The case of yelp. com. Working Paper, Harvard Business School.

Luca, Michael, Georgios Zervas. 2016. Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science* **62**(12) 3412–3427.

Mayzlin, Dina, Yaniv Dover, Judith Chevalier. 2014. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* **104**(8) 2421–55.

McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics* **142**(2) 698–714.

Melnik, Mikhail I, James Alm. 2002. Does a seller's eCommerce reputation matter? evidence from eBay auctions. *The Journal of Industrial Economics* **50**(3) 337–349.

Milgrom, Paul, John Roberts. 1986. Price and advertising signals of product quality. *Journal of political economy* **94**(4) 796–821.

Moorthy, Sridhar, Hao Zhao. 2000. Advertising Spending and Perceived Quality. *Marketing Letters* **11**(3) 221–233. doi:10.1023/A:1008135126025.

Nelson, Phillip. 1970. Information and consumer behavior. *Journal of political economy* **78**(2) 311–329.

Nelson, Phillip. 1974. Advertising as information. *Journal of political economy* **82**(4) 729–754.

Newman, Matthew L, James W Pennebaker, Diane S Berry, Jane M Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin* **29**(5) 665–675.

Pennebaker, James W. 2011. The secret life of pronouns: What our words say about us. *New Scientist* **211**(2828) 42–45.

Pennebaker, JW, RJ Booth, RL Boyd, ME Francis. 2015. Linguistic inquiry and word count: Liwc2015. austin, tx: Pennebaker conglomerates.

Proserpio, Davide, Georgios Zervas. 2017. Online reputation management: Estimating the impact of management responses on consumer reviews. *Forthcoming Marketing Science* .

Resnick, Paul, Richard Zeckhauser, John Swanson, Kate Lockwood. 2006. The value of reputation on eBay: A controlled experiment. *Experimental Economics* **9**(2) 79–101.

Rotfeld, Herbert J, Kim B Rotzoll. 1976. Advertising and product quality: are heavily advertised products better? *Journal of Consumer Affairs* **10**(1) 33–47.

Schmalensee, Richard. 1978. A Model of Advertising and Product Quality. *Journal of Political Economy* **86**(3) 485–503. doi:10.1086/260683.

Shum, Matthew. 2004. Does advertising overcome brand loyalty? Evidence from the breakfast-cereals market. *Journal of Economics & Management Strategy* **13**(2) 241–272.

Sun, Monic. 2012. How does the variance of product ratings matter? *Management Science* **58**(4) 696–707.

Zhao, H. 2000. Raising awareness and signaling quality to uninformed consumers: A price-advertising model. *Marketing Science* 390–396.

A Additional results using the full sample of hotels with search advertising data

In Figure 10, we replicate panel (b) of Figure 5; in Figure 11, we replicate Figure 7; and in Table 19 we report the effect of ratings on search advertising in this sample. These results are consistent with those reported earlier in the paper.

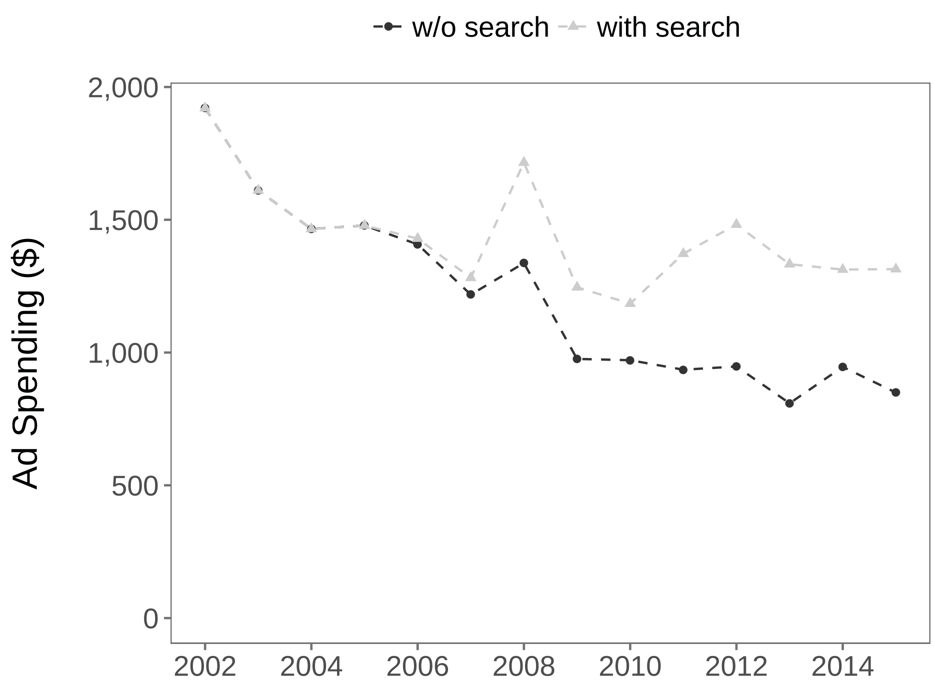


Figure 10: Year-over-year average monthly ad spending for independent hotels with and without search advertising

B Testing LIWC authenticity scores

In 2012, TripAdvisor introduced a program that allows hotels to post solicited user reviews on its site.[28] Hotels that join the program send emails soliciting reviews to their customers at the end of each stay. A review that is collected with the help of the TripAvisor program

[28]See: `https://www.tripadvisor.com/ReviewExpress`

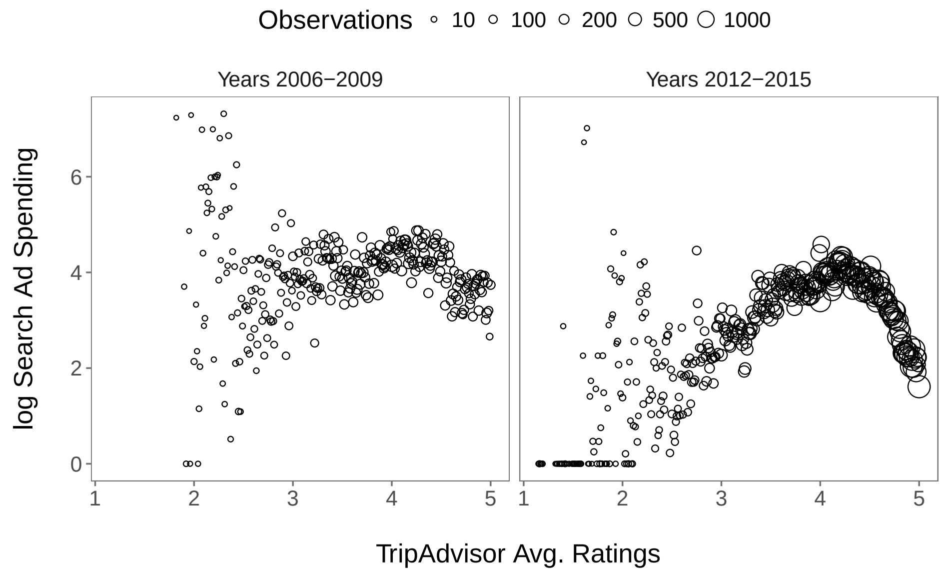


Figure 11: Relationship between search advertising spending and hotel ratings: 2006-2009 versus 2012-2015; all independent hotels in the SpyFu data set

is then published on the TripAdvisor website, and it is publicly identified as a solicited (by the hotel) review. Because solicited reviews are provided by real customers who have stayed at the hotel, they are less likely to be fake. We compare the authenticity of solicited and non-solicited reviews using LIWC, and find that solicited reviews are, on average, deemed more authentic (the average LIWC authenticity scores are 50.48 and 49.74, respectively, which is statistically significant). This test suggests that the LIWC authenticity scores are picking up at least a part of what makes a review authentic.

Table 19: How online user ratings affect search ad spending: RDD estimates

	3.25	3.75	4.25	4.75	Pooled
Above Threshold	0.272	−0.324*	−0.271*	−0.495***	−0.303***
	(0.206)	(0.143)	(0.110)	(0.105)	(0.064)
Avg. Ratings	−11.694*	6.277	4.898	2.771	2.939
	(5.238)	(3.555)	(2.755)	(2.744)	(1.625)
Above Threshold × Avg. Ratings	21.486**	3.489	−6.888	−5.741	−1.676
	(6.989)	(4.767)	(3.694)	(3.576)	(2.157)
Year-month FE	Yes	Yes	Yes	Yes	Yes
N	5765	12365	19972	19331	57433
R^2	0.21	0.18	0.18	0.22	0.20

Note: The dependent variable is log of search ad spending in the following 6 months. All columns use a pooled RDD with a bandwidth of 0.05 stars. Only independent firms with 20 or more reviews are included. Robust standard errors in parenthesis.

Significance levels: * p<0.05, ** p<0.01, *** p<0.001.