Advertising strategy in the presence of reviews: An empirical analysis

Hollenbeck, Brett and Moorthy, Sridhar and Proserpio, Davide

UCLA Anderson, University of Toronto Rotman, USC Marshall

14 March 2019

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

March 14, 2019

[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 online ratings have 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 online ratings and advertising has strengthened over time, just as TripAdvisor has become more popular, implying that firms respond to online reviews if and only if consumers respond to them.

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 many industries. Effectively, experience goods have become search goods. This raises many questions, among them, How does this change in the consumer's information environment 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 as online review platforms have become more popular? 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 coming out of Nelson's (1970, 1974) signaling theory. He argued 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 theoretical 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 his argument. The reason is, in most realistic contexts, there are multiple forces pulling in different directions: (i) the cost-side effects of quality on advertising spending, which arise as soon as advertising has a direct demand-enhancing role—such as raising awareness—and these tend to be negative if marginal costs are increasing in quality (Zhao 2000, Bagwell 2007, p. 1777); (ii) the demand-side effects of quality on advertising

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

spending, which arise whenever at least some consumers are informed about quality before purchase, and these can be positive (Archibald et al. 1983, Lei 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 perceived 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 perceived quality on

[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 won't do that.

advertising spending.[3] Furthermore, this is self-evidently a demand-side effect: while "large variations" in average ratings can have both demand- and cost-side effects on advertising spending,[4] small, local variations that produce "big" variations in displayed ratings can only be plausibly characterized as a demand-side effect.[5]

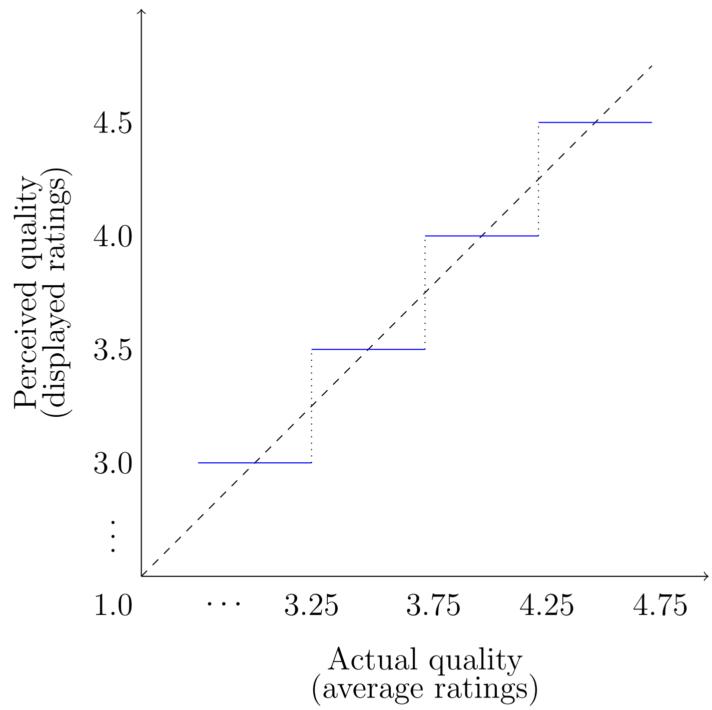


Figure 1: How perceived quality varies with actual quality at TripAdvisor

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 who live in many, widely-dispersed locations, online word-of-mouth is comparatively more important

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

[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.

[5]Small variations in average ratings *away* from the discontinuity thresholds, say from 3.48 to 3.49, can also have a demand-side effect on advertising spending, but by a more circuitous route. While displayed rating is constant for this variation, consumers may still respond because TripAdvisor ranks the 3.49-rated hotel higher than the 3.48-rated hotel.

than 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. Therefore, even apart from what we learn about the relationship between advertising and quality, the hotel industry is of some interest purely as a case-study.

Our empirical analysis is based on four datasets, one comprising TripAdvisor hotel reviews, 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 between 2002 and 2015, 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, therefore, is simply the collection and matching of these datasets, allowing for the first time-series, 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 period of time, but rather the entire experience of an industry over virtually the entire time online reviews have existed.

Our main results are the following. First, total ad spending by hotels in our measured media 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. Extrapolating from the last observation, since our advertising data do not cover several media that came into existence only after 2002—e.g., social media, mobile advertising, and online video—overall ad spending by hotels has probably increased over this fourteen-year period.

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). However, our most important and novel result is the negative causal relationship we observe between *displayed* TripAdvisor ratings and advertising spending. Hotels with higher displayed ratings spend less than hotels with lower displayed ratings around the TripAdvisor rounding thresholds. This effect is observed in all the traditional media as well as in Internet display and search advertising. Furthermore, it operates both at the extensive margin (decision to advertise) as well as at the intensive margin (advertising level given advertising). Together, these results suggest a substitution relationship between online ratings and advertising spending: hotels with "good" ratings treat their ratings as a substitute for advertising spending, while hotels with relatively poor ratings treat advertising as a substitute for their ratings. Effectively, hotels with higher ratings are targeting the informed consumer who gets her information from TripAdvisor, while hotels with lower ratings are targeting the uninformed consumer via advertising.

Two factors moderate the substitution relationship. First, the 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). Apparently, a strong brand name continues to provide some immunity to reviews—as has been found in other contexts as well, such as movies (Dhar and Moorthy 2017). Second, when we examine how the ratings-advertising relationship operates in different markets, we find that less differentiated markets, i.e., markets with hotels' ratings tightly bunched together, show a stronger 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) to what it became in later years (2012-2015), we find that the relationship has strengthened. During this period, review platforms have become more popular, and the proportion of informed consumers has almost certainly grown. Apparently, when this proportion was small, both high- and low-quality hotels probably found it optimal to ignore this consumer and target the uninformed consumer instead. Now, however, the informed segment is too large to ignore, and this, perhaps in conjunction with capacity constraints, causes the advertising strategies of the high- and low-quality hotels to separate (Horstmann and Moorthy 2003, Lei 2015). Ultimately, the basic message that comes out of all this is that firms will respond to their online ratings if and only if consumers respond to them.

2 Background

As noted earlier, survey evidence shows that a large number of people consult online reviews before making purchase decisions. Consistent with this, a number of empirical studies show that online reviews affect sales. This includes Anderson and Magruder (2012) and Luca (2016) who, using the same RD strategy we follow, find that higher-rated restaurants enjoy higher sales than lower-rated restaurants. The demand-side literature also includes: Chevalier and Mayzlin (2006) and Sun (2012) for books, 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 demand effect of online reviews 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 reviews 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.[6] One setting where firms' responses to ratings has been studied is eBay, but even

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

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 duopoly model with vertical differentiation. 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 the two Chen and Xie papers is what happens to advertising strategy as reviews become available, i.e., their focus is on a time-series prediction. 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 is 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).

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 its demand-side effects. Our regression discontinuity design allows us to identify a causal effect, and because the variation we exploit is a "local" variation in average ratings, we can be confident that what we are finding is a demand-side effect, 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: TripAdvisor, Kantar Media, SpyFu, and STR.

TripAdvisor. We begin with TripAdvisor because our unit of analysis is a hotel property reviewed on TripAdvisor in the 2002-20015 period.[7]

Launched in 2000, TripAdvisor is now one of the most popular review platforms on the Internet.[8] In an average month, it has about 350 million unique visitors worldwide.[9] In addition, TripAdvisor's ratings are widely displayed on other travel search platforms such as Hotels.com, Orbitz, Travelocity, and Expedia.com. As of May 2016, TripAdvisor had over 500 million customer reviews on over 6 million accommodations, restaurants, and attractions.

[7]For independent hotels, "hotel property" and "hotel" mean the same thing; for chain hotels, they do not. However, for ease of writing, we will sometimes abbreviate "hotel property" to "hotel" in what follows. As we will discuss shortly, since TripAdvisor ratings are available at the individual property level, while advertising spending numbers are not necessarily so, part of our challenge is how to associate the right spending with the right property.

[8]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).

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

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

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 the literature (e.g., Caves and Greene 1996, Shum 2004, Kim and McAlister 2011, Honka et al. 2017). Our Kantar dataset covers all the traditional advertising media—TV, radio, magazines, newspapers, and outdoor—plus Internet display advertising.[10] Missing are several newer media: search advertising, social media advertising, mobile advertising, e-mail advertising, and online video advertising. (The first of these we partially remedy with data from SpuFu, discussed below.) Kantar's methodology is essentially a bottom-up approach, combining direct monitoring of ads and information supplied by media outlets.[11]

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."[12] Brand ad expenditure is the sum of all product ad expenditures under the brand. We obtained monthly advertising expenditures for 16,852 brand-product pairs in the 2002-2015 period. From this dataset, we remove products containing the string "Combo" because such prod-

[10]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.

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

[12]As this example indicates, "product" may indicate a specific property, but it may also indicate a "type of ad," as well as an advertising partnership.

ucts include advertising spending on 2+ brand-product pairs. This leaves us with 15,973 distinct brand-product pairs.

SpyFu. SpyFu is a company that tracks online search advertising; it provided us our online search advertising data. The company's methodology is to search millions of keywords on Google, Bing, and Yahoo and record the URLs these searches return, along 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.[13] Their reach is extensive and includes even very specific keywords such as "Dockside Inn Fort Pierce" or "hotel New Brunswick." The chief limitation of this data is that only independent hotels' search advertising spending can be examined; chain hotels' property-specific search ad spending numbers cannot be obtained because SpyFu doesn't provide ad spending estimates at the sub-URL level.[14]

To use the SpyFu data we proceeded as follows. First, from TripAdvisor, we obtained each property's homepage URL.[15] 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 had this information. This procedure yielded monthly search advertising spending data on 9,718 independent hotels.

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,

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

[14]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.

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

price category, class, ownership, and capacity, among others;[16] and (3) a panel of hotel prices (average daily rates) at the hotel-year-month level for 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 step in using these four datasets is to match them up, starting with our unit of analysis, a TripAdvisor-reviewed hotel.

Matching Kantar with 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 less precise hotel designation. For chain hotels, especially, this is a serious problem because we don't want to conflate chain-wide advertising (that is unlikely to react to property-specific reviews) with property-specific advertising.

To perform this match at scale, we adopted the following algorithm:

1. Perform a Google search for each Kantar brand-product and examine the URLs of the top-10 search results.

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

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

4. If the TripAdvisor hotel ID extracted is unique, the algorithm returns the pair (TripAdvisor ID, Kantar brand-product ID) as a possible match; if the TripAdvisor hotel ID extracted is not unique, then we cannot match this Kantar brand-product, so we do not consider the advertising numbers associated with it. By this process, we identified 10,470 Kantar brand-products matched uniquely to specific TripAdvisor properties.

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

5. Compute a similarity score between the Kantar brand-product name and the TripAdvisor hotel name and retain only those matches that show "high similarity."

Finally, we manually checked the output of the algorithm for correctness.

This algorithm ends up matching 6,312 Kantar brand-products with 5,666 TripAdvisor hotel properties, which is 40% of the total Kantar sample. The large attrition is due to the fact that the full Kantar sample contains many brand-products that are not uniquely matchable to specific TripAdvisor properties.[17] Out of the 5,666 TripAdvisor hotel properties 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, 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 complete address. This matching yields 3,996 hotels (about 72% of 5,563). Out of these 3,996 hotels, STR provided us with financial information (prices) for 2,810 hotels.

[17] The reason is, even if our algorithm returns a unique TripAdvisor ID at Step 4, that match may not be a "correct" match. This is why we need Step 5 followed by a further manual check at the end. Google's search results may be "wrong" because: (i) the Kantar brand-product we searched for did not have a TripAdvisor page, and Google returned a TripAdvisor URL for a hotel with a similar name, (ii) the hotel we searched for is closed, and Google returned the TripAdvisor URL of the hotel currently open in a similar location, and (iii) the name of the hotel on the TripAdvisor page is different enough from the Kantar brand-product name that the match is discarded even though it is correct. For example our algorithm, for the Kantar brand-product "28th Street Hotel," returned as a potential match the hotel "Hampton Inn & Suites Grand Rapids Airport / 28th St"; however, we rejected this match in Step 5 because the two names were not similar enough. For another Kantar brand-product, "Calabasas Inn," the algorithm returned as a potential match "Good Nite Inn - Calabasas," which again we discarded in Step 5. (It turns out that the original Calabasas Inn closed and was reincarnated as The Hilton Garden Inn Calabasas.) For these reasons, there is a 40% attrition in going from Step 4 to Step 5 of the algorithm.

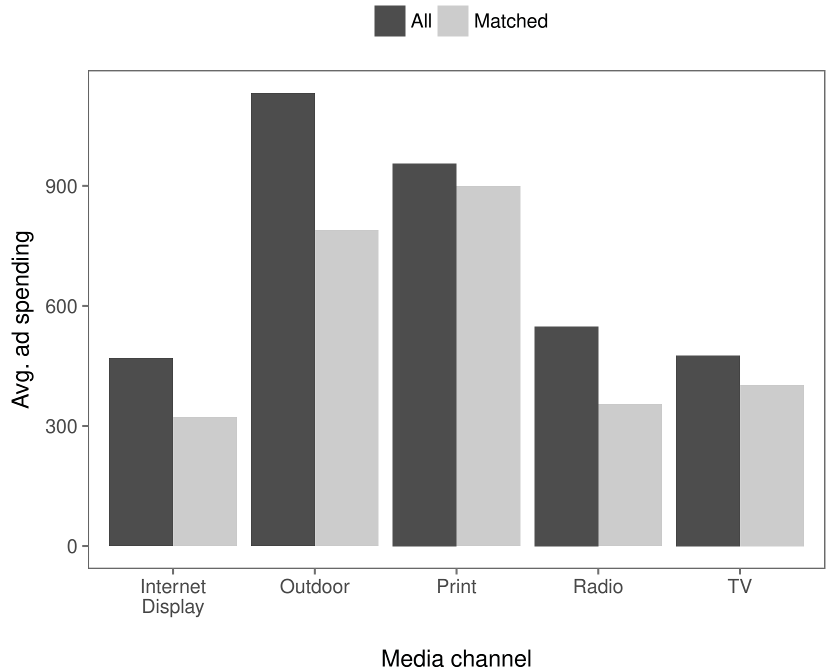


Figure 2: Comparing average advertising spending by media for all hotels in the full Kantar sample and the matched sample.

Adding SpyFu advertising spending. Out of the 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 hotels, then, we have a substantially complete advertising profile: advertising spending in all the traditional media as well as advertising spending in Internet display and Internet search.

Besides this more comprehensive dataset, we also 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 generalize to the larger dataset as well.

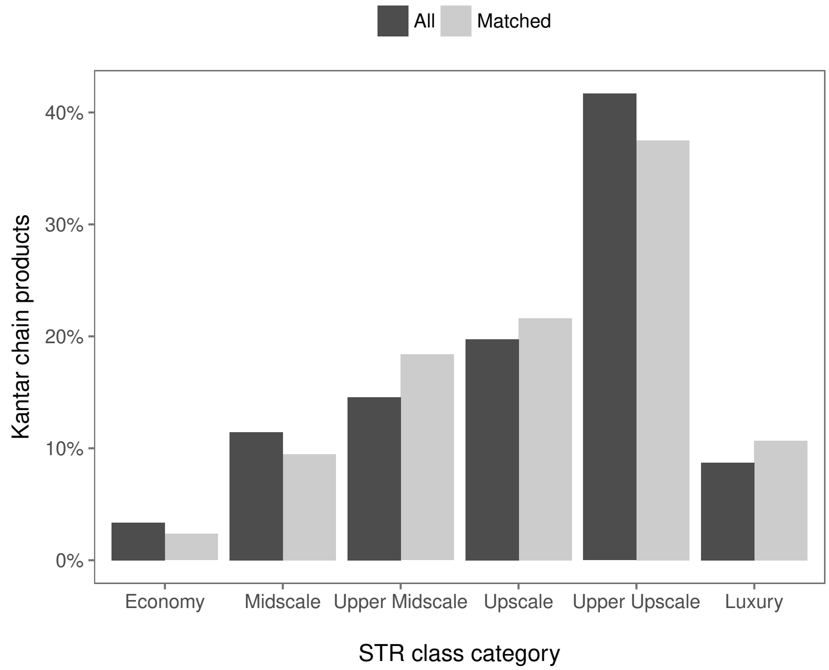


Figure 3: Comparing STR class distribution for hotel chains in the full Kantar sample and the matched sample.

3.2 Checks for matching bias and selection

We compare the matched dataset to the full Kantar dataset to see if our matching process produced any systematic distortions. First, we compare the average advertising spending levels in the different media (Internet display, outdoor, print, radio, and TV) in the matched and full 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 display, radio, and TV are the media where they spend less. Second, we compare the hotel class distributions for hotel chains in the matched and full Kantar samples in Figure 3.[18] These distributions are also 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, and the fraction of chain hotels is 0.27 in the

[18]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.

full sample and 0.26 in the matched sample. These results reassure us that our matched sample is quite similar to the full Kantar sample.

4 Descriptive evidence

In this section we describe the general patterns in the ratings and advertising data, starting in Table 1 with a comparison of summary statistics between 2002 (the start of our data) and 2015 (the end of our data). Comparing these two time periods shows how user ratings and advertising spending have changed over a 14-year period.

Table 1: Summary statistics

	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 per hotel ($)		
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[19]	1,326	1,153

Note: Average hotel rating and average reviews per hotel are based on end-of-year numbers.

First, both average hotel rating 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

[19] As noted earlier, our Internet search advertising data are only for independent hotels, and they do not begin until 2006. Therefore, Column 1, in this case, refers to the year 2006. Comparing the numbers in this row to the more detailed picture presented in Figure 4 we see that the decline in Internet search advertising from 2006 to 2015 is largely spurious.

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.[20] This decrease was particularly 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

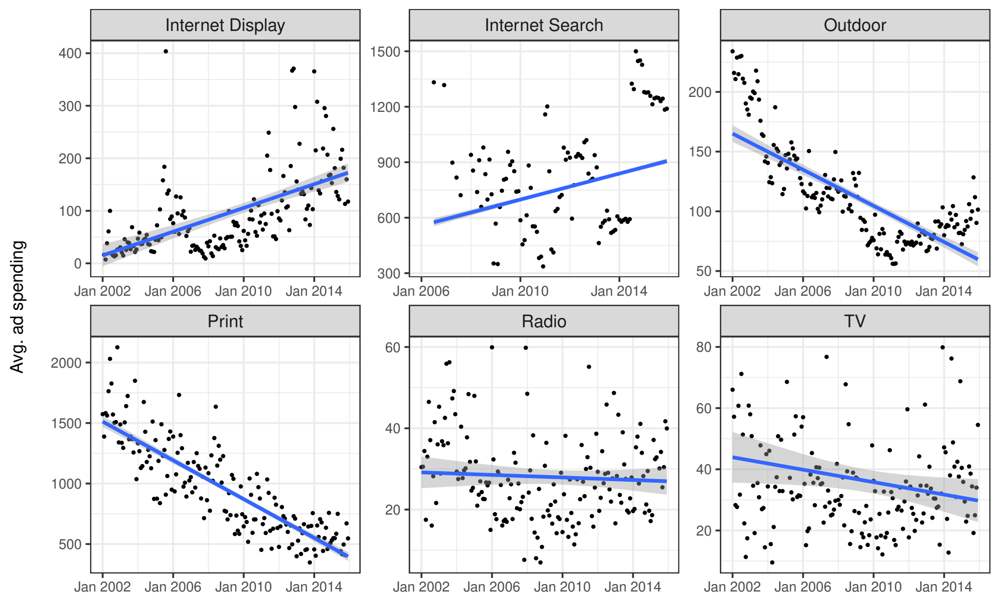


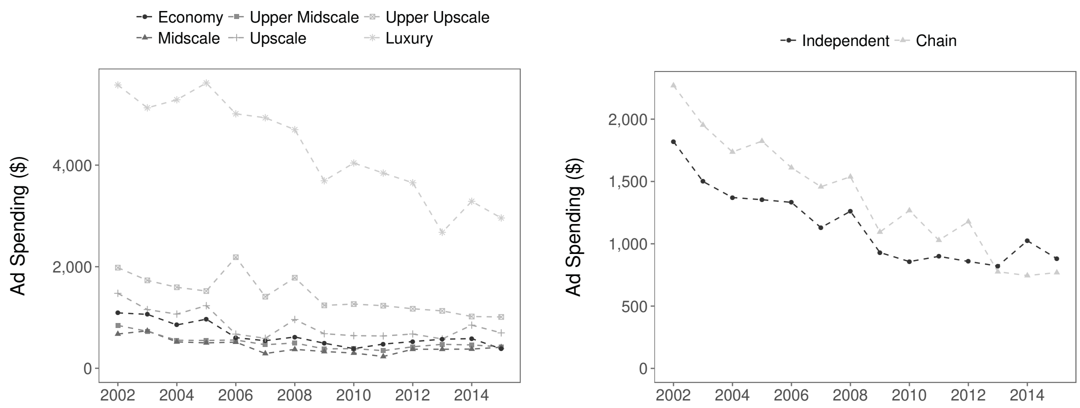
Figure 4: Average year-month ad spending by media channel. The blue line represents a linear fit and the grey shade the 95% confidence interval. Note that Internet search advertising covers independent hotels only.

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—our data don't begin until 2006— to about $1,150 per month by 2015.[21]

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

[21]One can conjecture that ad spending in the newer media that our data do not capture, namely, social media, online video, e-mail, and mobile, also shows an increase, so overall ad spending has probably increased from 2002 to 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 and 11 in Appendix A shows that if search advertising is included, total ad spending does not fall as much.



(a) (b)

Figure 5: Year-over-year average monthly ad spending in Kantar media by hotel class and independent versus chain hotels. Note: Internet search advertising and advertising in several other digital media are 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. 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.[22] And once they start doing so, an

[22]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 from year to year.

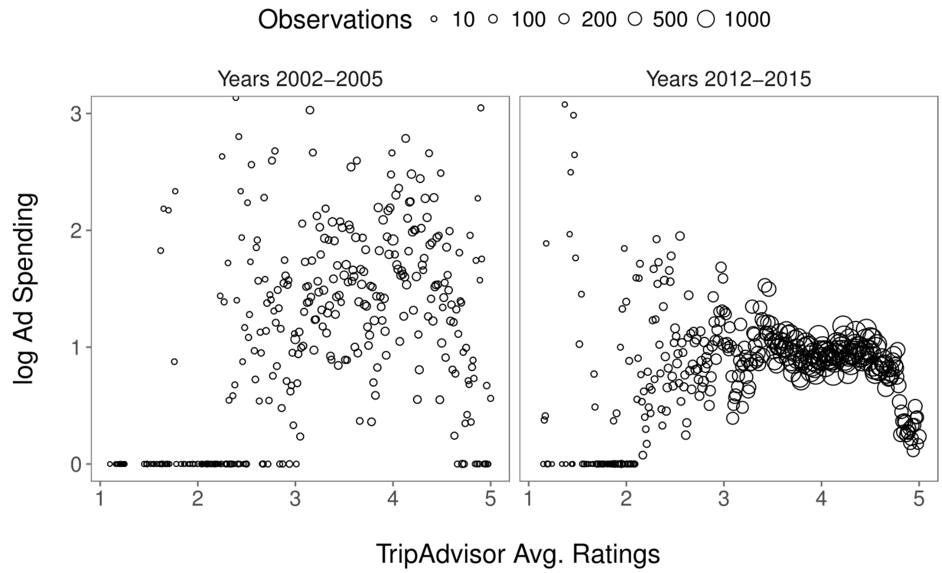


Figure 6: Relationship between advertising spending (excluding search advertising) and hotel ratings: 2002-2005 versus 2012-2015.

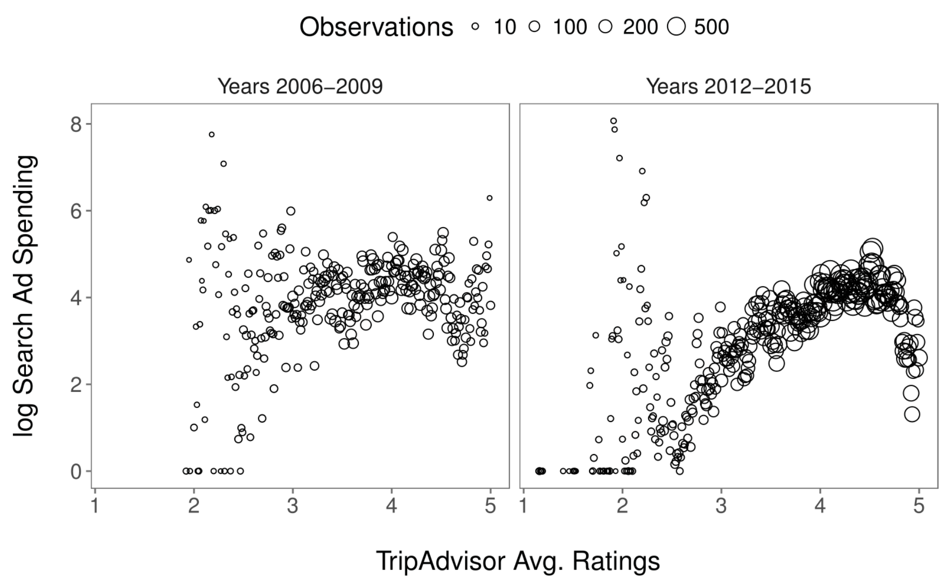


Figure 7: Relationship between search advertising spending and hotel ratings: 2006-2009 versus 2012-2015; independent hotels only, Kantar-matched SpyFu data

inverted-U relationship between ad spending and average ratings emerges. This relationship is reminiscent of previous results in the literature such as Horstmann and Moorthy (2003), and suggests the multiple forces at work. What starts out as a complementarity relationship ends up as a substitutability relationship.

5 Empirical framework

We estimate a causal effect of online ratings on advertising spending by using a regression discontinuity design (RDD). As noted in the Introduction, 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 are an unbiased estimate of the true quality of a hotel, but subject to sampling variation, then this rounding mechanism creates discrete, random variations in perceived quality around the rounding thresholds that is effectively independent of a hotel's true quality (see Figure 1). To the extent hotels' marginal costs depend on quality, they depend on true quality, not the small sampling variations in average ratings around the rounding thresholds. Therefore, any variation in a firm's advertising level that correlates with average ratings around those rounding thresholds represents a causal demand-side effect of perceived quality, not a cost-side effect stemming from variations in true quality.

5.1 RDD specification

We estimate the following specification:

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

Here, t refers to the beginning of a 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 spending by hotel i in the period $[t, t + 6 \text{ months}]$. We do this aggregation because (i) ads are purchased before they are delivered, and the lead-time between purchase and delivery, we conjecture, might be anywhere from 0 to 6 months, depending on the ad medium, and (ii) aggregating over six months reduces noise in the variable.[23]

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 or below it. 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 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.

To estimate equation (1), we made two important design choices with respect to the RDD. First, taking advantage of the large dataset available to us, we limited attention to hotels that are within 0.05 stars of each rounding threshold—a very small bandwidth for a RDD. Second, to reduce sampling variance in the average ratings, we limit all our analyses to hotels with 20 or more reviews. In Section 7, we discuss additional tests to check for the robustness of our estimates to different functional forms, bandwidths, and different aggregation windows of the dependent variable.

5.2 Identification tests

Our RD regression's identification of a causal effect relies on the assumptions that (i) expected advertising spending given average rating, as a function of average rating, is continuous around the rounding thresholds, both under the "treatment"—displayed rating is "high"—and not under the treatment—displayed rating is "low"—and (ii) that hotels do not select into the treatment based on the anticipated advertising effect.[24]

As McCrary (2008) has observed, the latter assumption can fail if hotels could manipulate their average ratings around the rounding thresholds. To check for such a violation McCrary

[23]Ideally, one would observe when ad spending decisions are made and relate those decisions directly to the average and displayed ratings prevailing at the time. However, we do not observe when ad decisions are made, only when advertising money is spent. Aggregating ad spending over the subsequent 6 months may therefore be seen as introducing measurement error in the dependent variable, which only makes it harder to find a ratings-ad spending effect.

[24]For a formal discussion of the identification assumptions underlying RDD see Hahn et al. (2001).

(2008) has proposed a simple test based on continuity of the density of the "running variable" around the rounding thresholds. Intuitively, if there is upward manipulation of ratings, we should see relatively few firms with average ratings just below the thresholds and a clump of firms with average ratings just above the thresholds, and if there is downward manipulation of ratings—say, due to competitors' actions—then we should see the opposite. As Figure 8 shows, however, the density of average ratings in our data is essentially continuous, uniform in fact, with neither bumps nor dips, above or below the rounding thresholds.[25] The visual evidence is confirmed by a formal McCarary test in Table 2; the RD regression with the density of average ratings as the dependent variable shows an insignificant coefficient for "Above Threshold."

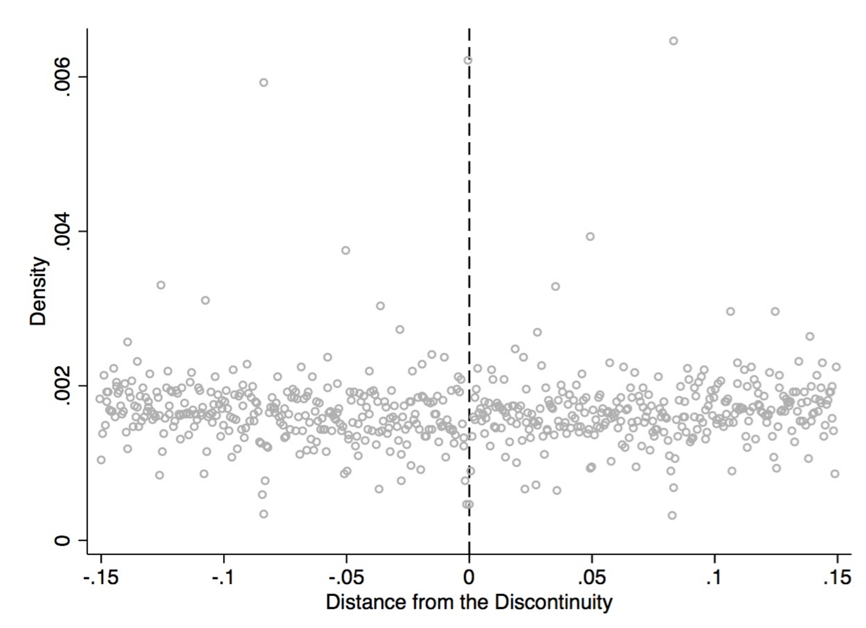


Figure 8: McCrary test: distribution of average ratings near rounding thresholds

Another category of identification tests is based on the idea, formalized in Lee (2008), that if the assignment of hotels to "just above the rounding threshold" and "just below the rounding threshold" is random, then the baseline characteristics of the hotels—characteristics

[25]While round numbers such as the rounding thresholds are more common, they are not more common than other round numbers, such as 4.15 and 4.35. Note that uniform density is not the requirement, continuity is. For an example of a non-uniform density that is still continuous see Figure 16 in Lee and Lemieux (2010).

	(1)
Above Threshold	0.001
	(0.001)
Avg. Ratings	−0.030
	(0.029)
Above Threshold × Avg. Ratings	0.038
	(0.047)
Year-month FE	Yes
Brand FE	Yes
N	251
R^2	0.54

Note: The dependent variable is density of average ratings, computed on a bin size of 0.0004 stars. Pooled RDD with a bandwidth of 0.05 stars. 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

that are determined prior to the realization of the average rating—should have the same distribution just above and below the thresholds. In Table 3 we see just this: hotels just above the threshold do not differ systematically from those just below the threshold on a variety of characteristics—including price, which is arguably not a predetermined characteristic.

Finally, in Appendix B, we provide several additional tests for ratings manipulation based on various correlates of manipulated reviews identified in the past literature (Anderson and Magruder 2012, Luca and Zervas 2016, Mayzlin et al. 2014, Proserpio and Zervas 2017). These include things like number of 5-star reviews, reviewers with "few" versus "many" reviews, independent hotels managed by "large" versus "small" owners, and reviews with and without managerial responses. None of these variables show discontinuities around the rounding thresholds.

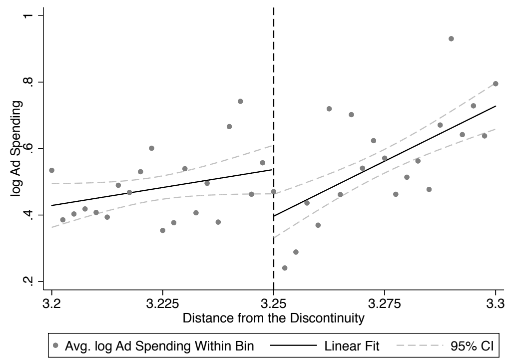
Collectively, then, we don't see any evidence that the identifications assumptions of RDD are being violated here. This doesn't mean, of course, that there are no hotel-manipulated reviews in our data. Rather, what it means is that, notwithstanding any manipulation,

	Below	Above	Difference (SE)
Hotel is Chain	0.38	0.38	0.002 (0.004)
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	−0.004 (0.01)
Hotel Location	4.42	4.41	0.01 (0.01)
Hotel has Meeting Space	0.79	0.79	0.0009 (0.003)
Hotel Brand	18.60	18.50	0.06 (0.23)
Hotel Price	164.60	165.70	−1.16 (0.96)

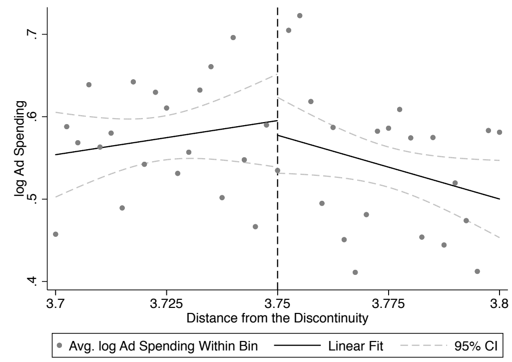
Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

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

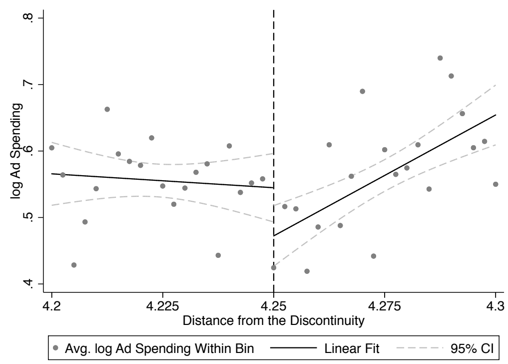
hotels don't seem to have precise control over their average ratings around the rounding thresholds. As Lee and Lemieux (2010) have emphasized, precise control is the issue: "If individuals—even while having some influence—are unable to *precisely* manipulate the assignment variable, a *consequence* of this is that variation in treatment near the threshold is randomized as though from a randomized experiment." With average ratings, the problem of precise control is compounded by the fact that while a hotel may try to raise its ratings by writing fake positive reviews, its competitors are probably trying to do just the opposite by writing fake negative reviews. Ultimately, what moves average ratings in this scenario is the consumer's honest opinion, reflecting her actual experiences with the hotel's quality.



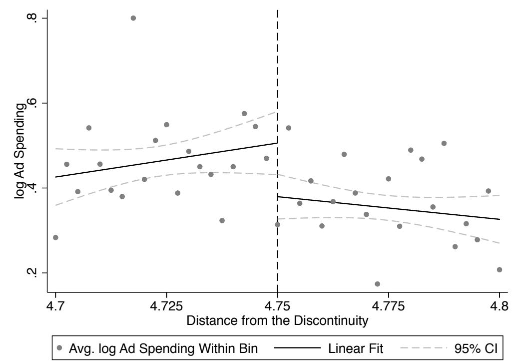
(a)



(b)



(c)



(d)

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

6 Results

6.1 How online ratings affect advertising spending

We start our analysis by presenting visual evidence that hotels' total ad spending is sensitive to their average ratings around the rounding thresholds. 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.[26] There are clear gaps in the intercepts at thresholds 3.25, 4.25, and 4.75, and a less clear one at threshold

[26] For average ratings 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 an average rating below 3 stars in 2015.

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.

	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 total ad spending: RDD estimates

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. These 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 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.

6.2 Effects by media

	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

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 with total ad spending 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.

6.3 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. Are firms responding to their ratings by varying whether they advertise or not, or are they merely adjusting their spending level? To answer this question, 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 with a dummy dependent 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 on both margins. Firms are less likely to advertise if they are above the rounding threshold, and if they do, the amount they spend is significantly lower.

	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

6.4 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. Table 7 presents the results, with columns 2 and 3 excluding spending on search ads, while column 4 shows results for independent hotels with search ad spending included in total ad spending.[27] We observe that the relationship for chains is negative, but not statistically significant, whereas for independent hotels it is both

[27] As we noted in Section 3, Internet search data are only available for independent hotels, and that too, only starting 2006.

negative and significant. 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, these results suggest that the ad response of independent hotels drives much of the overall result. Finally, in column 4, when we repeat the analysis for independent hotels including search ad spending, we find an even stronger result.

	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

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 less sensitive to reviews, hotels have less reason to respond.

We next test how the effect varies within chains, based on size and quality-tier differences. We separately estimate the RDD 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.

	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 hotel type

6.5 Market effects

In this section, we examine how the degree of differentiation in a market affects the relationship between ratings and ad spending. We use the STR definition of market as a Metropolitan Statistical Area (MSA) and operationalize differentiation by computing the standard deviation of average ratings in each market each year. Our hypothesis is that in markets with low standard deviation of average ratings, the boost in displayed ratings that comes from average ratings crossing the rounding threshold from left to right 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.

6.6 Early versus late effects

In Figure 6, 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 influence of TripAdvisor among consumers. 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 8, 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.) Hotels that didn't feel any need to respond to their ratings when consumers were not paying attention, surely had to respond once consumers started paying attention.

We test this hypothesis in Table 10 where two regressions are reported, 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 regression when we aggregated across all periods (Table 7, column 1).

7 Robustness of results

In this section, we examine the robustness of our results to alternative formulations of the RDD: quadratic polynomial functional form, alternative bandwidths, placebo thresholds, inclusion of hotel prices as a control, and alternative aggregation windows for advertising spending.

Quadratic polynomial estimator. Here we test whether our results are sensitive to changing from our previous linear specification to a quadratic polynomial specification. We add to equation (1) a quadratic term for the average ratings variable, and interact it with the Above Threshold dummy to allow for a separate effect of the quadratic term above and below the discontinuity. The specification thus becomes:

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

The results are in Table 11 are qualitatively similar to our earlier results with a linear specification.

Alternative 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 12. The coefficient of interest, Above Threshold, is statistically significant for every bandwidth tested; however, the effects are larger for the 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 RDD on placebo thresholds, i.e., thresholds where there is no jump in the displayed ratings. The results are in Table 13. 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, implying that without a change in displayed ratings there is no effect of local variation in average

	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 11: 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 12: Sensitivity analysis: different bandwidths

ratings. In other words, the theoretical possibility of a demand-side effect arising from such variations, as discussed in footnote 5, does not materialize.

	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 13: Sensitivity analysis: placebo test

Controlling for hotel prices. Theoretically, one may argue that since both prices and advertising affect demand,[28] hotels may respond to variations in displayed ratings both by adjusting advertising spending as well as by adjusting price. While a test of whether prices also adjust to displayed ratings is beyond the scope of this paper, we can include price as a covariate in our regressions to see whether it makes a difference. Along these lines, we reestimate equation 1 on a subset of hotels for which we have average daily rates (ADR) (see Section 3). The results are in Table 14. 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 under the full sample (see Table 7, column 1). Second, when we insert the logarithm of ADR as a control, the coefficient remains negative,

[28]Luca's (2016) results showing ratings affecting restaurant revenue is suggestive of a price effect.

	(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 14: How TripAdvisor average ratings affect ad spending, controlling for hotel prices

	(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 15: Results with different time windows for advertising spending

statistically significant, and similar in magnitude to that reported in column 1, suggesting that our results are not affected by the inclusion of hotel prices.

Alternative 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 15 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.

8 Discussion

What we have shown is that local variations in average ratings that trigger relatively "big" variations in displayed ratings ("perceived quality") have a causal demand-side effect on advertising spending by hotels. The reason it can only be a demand-side effect is that the average ratings variation driving the response is confined to a small neighborhood of the rounding thresholds, more plausible as a sampling variation between hotels similar in quality, than as a variation stemming from an actual difference in quality. Once differences in actual quality are off the table, the cost-side motivation disappears.

The nature of the demand-side effect is, of course, our most interesting result. We observe that hotels with "high" perceived quality—those just to the right of a rounding threshold—spend less on advertising than hotels with "low" perceived quality, those just to the left of the rounding threshold. Why might this be so? We can only speculate, but extrapolating from past theoretical models that have appeared in the literature, such as Horstmann and Moorthy (2003) and Lei (2015), a plausible explanation might be that hotels with a high perceived quality are targeting a different consumer segment than hotels with a low perceived quality. In particular, the former might be targeting the "informed" consumer who gets her information from TripAdvisor's displayed ratings, while the latter is targeting the "uninformed" consumer who gets her information from advertising. While the former's targeting choice might be deemed opportunistic, the latter's targeting choice is arguably born out of necessity: consumers visiting the TripAdvisor site are unlikely to choose a hotel with lower ratings over a hotel with higher ratings, *ceteris paribus*.

For the hotel with high perceived quality, then, a lower ad spending level is justified because (a) advertising will not be productive on informed consumers, and (b) why spend money on advertising when online ratings can do the work for free? On the other hand, precisely because it will be hard to persuade informed consumers, the hotel with low perceived quality must target the uninformed consumer with advertising.

The remaining question to be resolved is why the hotel with high perceived quality doesn't also go after the uninformed segment, for if it were to do so, perhaps we won't see a difference in their advertising strategies. The answer to this must await a fully-worked out theoretical model, but one possibility is "price dilution." Since uninformed consumers hold "average quality" beliefs, the high perceived quality hotel will not be able to charge high prices by pursuing this segment. In other words, pursuing the uninformed segment has the opportunity cost of foregoing the high prices that would be otherwise possible. The fact that the negative relationship between perceived quality and advertising spending has become apparent only in the more recent years of our data, as the proportion of informed consumers has presumably risen, lends credence to this argument. The presence of capacity constraints only makes it stronger.

Is the substitution effect between online ratings and advertising likely to be a universal feature of all industries? We conjecture not. In fact, we know, based on Dhar and Moorthy (2017), that the result does not hold in the movie category. There, in the sub-category of limited-release movies, advertising complements critics' reviews. The reasons for the difference may have to do with a number of things that differentiate the movie industry from the hotel industry: (a) all movies, regardless of reviews, are priced the same, (b) movies can incorporate their critics ratings in advertising copy, and (c) movies, unlike hotels, do not face capacity constraints.

9 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' displayed 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 and 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" displayed ratings seem to be treating their ratings as a substitute for advertising while hotels with "low" displayed ratings seem to be treating advertising as a substitute for their ratings. It is as if the former are targeting the informed consumer while the latter are targeting the uninformed consumer.

Beneath the 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 continues to provide some immunity to reviews. 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 pivotal. Finally, in the time-series, we see hotels becoming more responsive to their ratings with time, just as TripAdvisor's popularity has risen. This tells us that it is not the presence of reviews *per se* that triggers a firm reaction, but rather the recognition that consumers are responding to them.

Our empirical analysis, based on regression-discontinuity designs, wouldn't be possible without the exogenous discontinuities in TripAdvisor's displayed ratings. However, this is also 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 critics' ratings and ad spending for limited-release 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. Paper presented at the Marketing Science Conference at USC.

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

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.

Lee, David S. 2008. Randomized experiments from non-random selection in us house elections. *Journal of Econometrics* **142**(2) 675–697.

Lee, David S, Thomas Lemieux. 2010. Regression discontinuity designs in economics. *Journal of economic literature* **48**(2) 281–355.

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 search advertising data

This appendix presents several additional results using search advertising data. In Figure 10, we replicate panel (b) of Figure 5 for independent hotels with and without search advertising data; in Figure 11 we do the same for all hotels;[29] in Figure 12, we replicate Figure 7 using the full sample of 10,398 independent hotels for which we have search advertising data; and in Table 16 we report the effect of ratings on search advertising for this sample of independent hotels. These results are consistent with those reported in the paper.

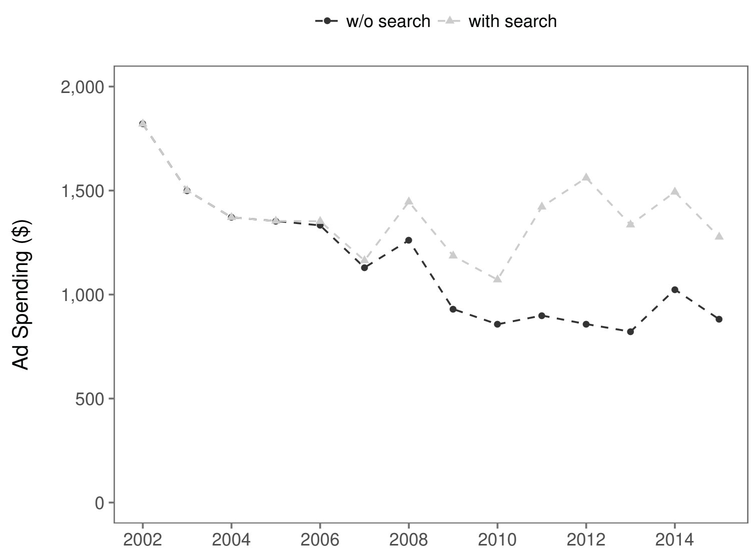


Figure 10: Year-over-year average monthly ad spending for independent hotels with and without search advertising

[29]Note that for chain hotels we do not have individual search advertising spending, but only aggregate spending at the chain level. Therefore, for every chain hotel in our dataset for which we have aggregate data at the chain level, we assign an amount of search advertising spending equal to the total advertising spending at the chain level divided by the number of properties in that chain.

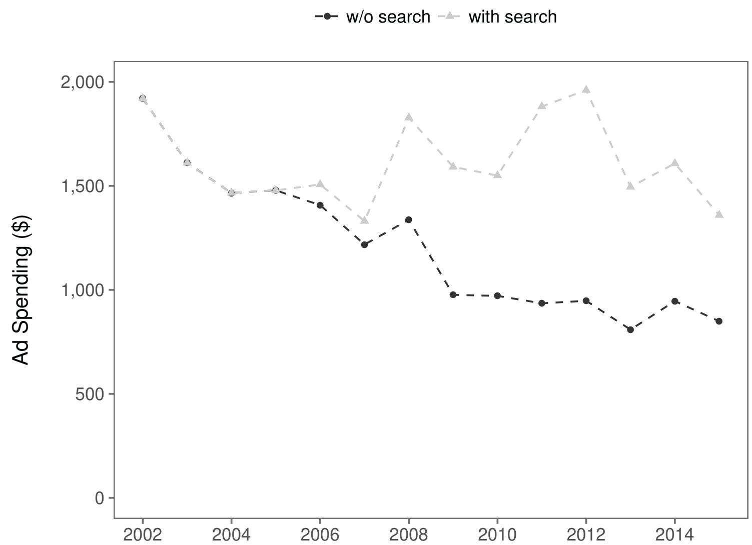


Figure 11: Year-over-year average monthly ad spending for all hotels (independent and chain) with and without search advertising

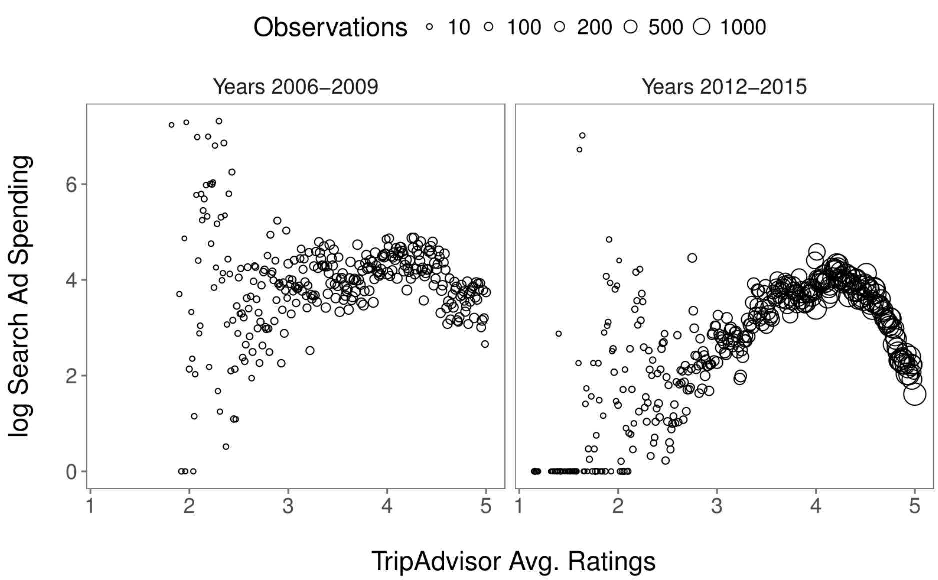


Figure 12: Relationship between search advertising spending and hotel ratings: 2006-2009 versus 2012-2015; all independent hotels in the SpyFu data set

Table 16: 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.

B Additional tests for ratings manipulation

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 17, 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.

	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 × Avg. Ratings	0.1583	−0.7688	0.1321	0.0075	−0.2584**
	(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 17: Test for review manipulation: reviews and reviewers characteristics

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.[30] LIWC's authenticity algorithm is based on a series of studies that tested whether

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

	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 18: 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 19: Test for review manipulation: independent hotels managed by large versus small owners

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). Using LIWC, we check whether review authenticity scores differ above and below the rounding thresholds. The results of this analysis, reported in Table 18, 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 19, 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 17, we check whether the fraction of reviews with responses shows discontinuities around the rounding thresholds. Again, the coefficient of interest is statistically insignificant.