Online Reputation Management through Investments in Quality: A study of the Lodging Industry .

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE.

Sergio Gárate*
Peter Newberry †

Abstract

Online reputation systems provide information to consumers about the quality of firms on the platform, thereby generating incentives for firms to invest in their quality in order to maintain their reputation. We study this relationship using data on revenue, occupancy, and investment decisions for 1,815 hotels that are linked to a dataset of 895,768 online consumer reviews. We show that a drop of 0.5 star rating increases the probability of investing in Capital Expenditures by 2.7%. However, the impact of changes in reviews is not homogeneous across locations or building characteristics, as hotels in more competitive areas respond more aggressively to reviews. Subsequently, we show that hotels improve their reputation after investing in capital expenditures. Taken together, we conclude that information flow among consumers and managers through these reputation systems makes buildings' quality salient, thereby providing incentives to invest.

1 Introduction

Most online platforms feature reputation systems that provide information about the quality of the products on the platform. Examples include AirBnB (Bed and Breakfast), TripAdvisor (Hotels/Restaurants), Lending Tree (Mortgages), and other websites that allow consumers to post a review of a product or service (i.e Amazon, and ebay). Many recent studies find that positive consumer reviews increase customers' patronage of stores. For example, Luca (2016) and Anderson

*University of Mississippi, sgarate@bus.olemiss.edu
†Pennsylvania State University, pwnewberry@psu.edu

and Magruder (2012) show that a higher score on Yelp.com increases restaurants revenues and bookings, respectively. While these studies link online reputation to demand, little is known about the supply side effects of it. In this paper, we study the relationship between a firm's online reputation and its decisions to invest in quality, using the lodging industry as a case study.

Specifically, we use hotel reviews from TripAdvisor.com to relate changes in the reputation of a hotel to the decision to invest in capital expenditures. The underlying assumption is that consumer reviews are a noisy signal of buildings' quality and that they reveal information to consumers, thereby affecting their consumption pattern. As a result, landlords recognize potential future losses due to reputation deterioration and respond by increasing investment in capital improvements.

A firm may respond to reviews for two reasons: first, managers do not have access to consumers' beliefs of building's quality and online reviews act as a proxy to those beliefs; second, managers know the quality of the building but given the randomness and discreteness of reviews[1], managers require to maintain a certain level of quality that differs from the level of quality if the signal consumer reviews provide was perfect. The first story has beneficial effect on efficiency due to information provision of reviews to managers, while the latter story is a story we describe as over investment due reviews random component.

We use a dataset with more than 850,000 online reviews from TripAdvisor.com and a panel dataset of capital expenditures from more than 1,800 buildings to test the hypotheses. We analyze the type of information embedded in these online reviews and provide summary statistics, then show that messages posted by consumers on these websites affect the vacancy of corresponding hotels, suggesting that reviews are informative to consumers and that they alter their consumption pattern. Subsequently, we show that after controlling for market conditions, by using a combination of fixed effects and a time variant fixed effect, the probability of landlord investment in capital expenditures increases as facility reputation falls. Finally, by using an event study around the investment decision, we show that capital expenditures help restore facility reputation. Taken together, we conclude that information flow among consumers through message boards helps mitigate asymmetries of information between a consumer and the hotel suggesting an improvement in market efficiency. We also show that the impact of these mechanisms

[1]Reviews in these systems vary by 0.5 stars, therefore if a building receives a couple of bad reviews the star rating could result in full 0.5 star difference rather than a couple decimal points difference.

is not ubiquitous and depends on the characteristics of the competition and of the hotel itself.

The remainder of the paper proceeds as follows. In the next section, we provide an overview of the relevant literature and discuss the main findings of the optionality of capital expenditures, the impact of online reviews on hotel bookings and the link between information disclosure and firms' quality. After describing the data used, we present the main empirical model to test the main hypotheses of the paper. Finally, we provide a detailed discussion of the results.

2 Prior Literature

The theoretical foundation of this paper is built on the literature studying information asymmetry, pioneered by Akerlof (1970), and the reputation and quality decision models, Shapiro (1983) and by Rogerson (1983). This paper is also informed by the empirical literature on verifiable disclosure information (Jin and Leslie, 2003; Lewis, 2011) and the studies on the influence reviews have on consumption patterns (Anderson and Magruder, 2012; Sparks and Browning, 2011). Finally, the paper fits within the literature on information asymmetries in real estate (Garmaise and Moskowitz, 2003; Levitt and Syverson, 2008) and the strand of literature that studies capital improvements in this industry (Bond et al., 2014; **?**).

In his seminal work, Akerlof (1970) emphasizes the presence of adverse selection in markets with information asymmetries. Shapiro (1983) relaxes Akerlof's assumption of one-time transaction related to consumer sole dependence on a market statistic to assess the quality of a product and argue that reputation helps mitigate the adverse selection and provide incentive for firms to market high quality products in order to avoid potential future loses. Our contribution to this literature is to provide empirical evidence on whether consumer asses firms quality with on-line reviews and if firms respond to reputation changes.

In the real estate industry Garmaise and Moskowitz (2003) and Levitt and Syverson (2008) study market distortions related to information asymmetries. Garmaise and Moskowitz (2003) argue that agents mitigate information asymmetries by buying properties at closer distances, prefer buildings with a longer history of net operating income, and most importantly decrease transactions with counterparts with information advantages; while Levitt and Syverson (2008) shows that real estate agents sell their own houses at a higher price and wait longer to sell these properties. Our contribution to this strand of literature is to study whether real estate managers respond to reputation changes.

The model proposed by Rogerson (1983) uses a more flexible approach to the ability of consumers to assess quality. In Rogerson's model, consumers have information asymmetries and can only imperfectly assess the quality of the product ex-post. More specifically, the quality is only imperfectly observed and therefore reputation effects result from a probability of consumers correctly assessing the true quality of the product. This feature of the model relates to consumer reviews as the information provided in the reviews is an imperfect observation of customers that allows the next consumer to make a purchase decision based on the information in reviews. Customers recognize that the review is a noisy signal of what the underlying quality is, and make decisions under this condition. In our paper we study what are the consequences in investment due to reviews noisy signal.

There are several online services that allow customers to book hotels and read consumers' comments on a message board. The evidence found in the literature to date indicates that consumers base their decision to book a hotel on word-of-mouth WOM, specifically, online reviews (Sparks and Browning (2011)). Likewise, in the empirical literature, Anderson and Magruder (2012) use a regression discontinuity design to estimate the impact of half a star rating on restaurant booking. Their results suggest that changes of half a star are associated with a 49% increase in the times a restaurant sells out. The effect of a change in stars exacerbated when other sources of information are scarce. Along the same line and using the same methodology, Luca (2016) finds that 1 star translates into an increase of 5% to 9% in revenue. This paper provides empirical evidence on the impact that on-line reviews have on hotel's occupancy and revenue.

One particular characteristic of reviews is that they make the quality of hotels salient, much like the grade cards from an inspector in the restaurant industry. Jin and Leslie (2003) studied the introduction of hygiene grade cards in Los Angeles California and found that displaying grade cards informs the market and provides incentive for restaurant managers to invest in hygiene, improving the quality of the service and decreasing the number of cases of food poisoning in the area. Another case of information disclosure is the one studied by Lewis (2011). The author argues that given the enforceability of claims made in the description provided in online car auctions, photos and text posted by the seller of the car has a strong influence on price. It is clear then that reviews in the hotel industry may not only affect future bookings, but may also affect hotels' financial fundamentals, like future revenues and occupancy. Ultimately, they may affect the investment decisions of managers to improve the quality of the building.

Similar to the options available to consumers, managers have real options when it comes to capital investment decisions. Flexibility in capital improvements could have significant implications on various performance measures of the firm (**?**) because the flexibility to invest or not in capital expenditure can skew cash flow as managers may invest differently during stronger economic conditions rather than weaker ones. For instance, Titman et al. (2004) suggest that capital investment flexibility alters volatility as well as skewness of the project's value. Bond et al. follow the work of **?**, Titman et al. (2004) and Childs et al. (2004) to set up a model that depends on a long term and a noisy short-term lease rate, and their findings suggest that capital expenditures lead to higher incomes. The authors also find that investments conducted during periods of low lease rate perform better than capital investment during high lease rate periods. On this same line **?** study the disposition effect of properties after a capital expenditure takes place. Their findings suggest owners are encouraged to increase capital investment during periods of high rent in order to capture larger profits. Alternatively, during periods of lower rent or if owners foresee lower rent in the future, the incentive is to delay capital expenditures.

Overall, the literature on capital expenditure provides evidence that managers exercise their option to invest in capital expenditures depending on market conditions as well as their expectation about the future. At the same time, the evidence from consumer reviews suggests that these have the power to influence the purchasing patterns of future customers through a reputation effect. This paper explores information considerations at the moment of investment in capital expenditures after controlling for market conditions using a mix of property and location fixed effects as well as location time varying fixed effect. In the next section I develop a reputation model using the insights from Shapiro (1983) and Rogerson (1983) to derive the paper's hypotheses. I then develop the empirical model to estimate the impact that consumer reviews have on the decision to invest in capital expenditures after controlling for market conditions.

3 Data

Consumer generated Internet content in the travel industry is particularly important at the moment of booking a hotel. Prior to booking a room, independent of the site used to do so, the consumer has information on the star rating given by prior guests and a brief description of the experience during the stay. Although this study uses only hotels in the U.S.A, TripAdvisor.com has consumer reviews for more than 7 million accommodations, attractions and restaurants in 49 markets[2]. With 535 million reviews and more than 415 million visits per month, TripAdvisor.com is the leading provider of consumer generated content in the travel industry. The main data source for this project is a collection of 8,663,790 reviews for hotels within the U.S.A. from this website; Table **??** provides summary statistics for the full sample of reviews.

The dataset for this project comes from various sources. First, the focus of the paper is on investment decision of real estate managers and the relation to information provided in the reviews. Investment as well as performance information of buildings comes from the SNL property report. Second, in order to build a list of competitors with number of employees, number of rooms, square footage, franchise status, year established, and brand information, we use the AcNielsen business directory. Finally, to geocode all the hotels, we use Google Maps Geocoding API. To match information from TripAdvisor.com to investment data from SNL, we use a two steps procedure. In the first step, we match by hotel name and address; and in the second step, we use the geo location of the remaining hotels to match by distance. We consider that the minimum distance between the hotels of two samples are a match. We drop hotels for which the minimum distance was larger than 150 feet[3].

Investment and performance data comes from SNL, a company that collects information on publicly traded REITs. SNL is a subsidiary of S&P Global Market Intelligence. The information available is an unbalanced panel of hotels and other property types in the USA and around the world. We focus on hotels located in the U.S.A. that have data available for subsequent building improvements. The capital expenditure information comes from the 10-k reports in section III, in which REITs provide a detail by property of the subsequent improvement after acquisition. In Table **??** in the appendix is a sample of the type of information available for each hotel. Figure **??** shows the distribution of capital expenditure as a percentage of the book value of the hotels in our sample. We create a binary variable for the capital expenditure, whenever investment exceed 3.5% of the book value of the building we assigned a 1. Figure **??** shows the distribution of this binary variables and approximately 32% of the time capital expenditures exceeds the 3.5% threshold. We use this binary variable to estimate the linear probability model of the decision of managers to investment larger amount in building improvements. Although SNL provides information for more than 4,900 hotels which represents 10% of the total number of hotels in the USA, we use a subsample of

[2]https://tripadvisor.mediaroom.com/us-about-us (Accessed, August 2017)

[3]For a distance greater than 150 feet the matching started to present some difficulties, for example names of hotels matched differ.

1,815 buildings with information on Capital Expenditures and that have consumer reviews. Our final sample has 6,416 property year observations.

Figure **??** shows the distribution across space of the hotels in the sample. The background of the map is a heat map that has the density of the economic activity for Real Estate and Rental and Leasing industry (NAICS 53). It seems by looking at the graph that hotels in the sample are not biased in terms of location and have similar spatial distribution as the economic activity of interest. To further review this claim, we use the list of competitors and perform a T-test on the sample matched, and the entire list of competitors; Table **??** shows this analysis. Although the buildings in the sample are spatially distributed according to the entire population of hotels, there is an over representation of buildings in the MidAtlantic (+4%), and an under representation Southeast (-3%) and West(-5%). Unsurprisingly given that we use REITs properties, other hotel characteristics differ in attributes that describe size, brand[4], and competitive environment.

Importantly, the goal of this paper is to study ratings and their impact on investment. While attributes that refer to size may differ, the attributes that refer to reviews are not different from the entire population of hotels. Hotels in our sample receive on average 5.29 reviews per room and the entire population receives 5.29 with no economical or statistical significance for the difference. Also the ratings in our sample receive on average 3.75 star rating while the population of hotels receives a 3.73. Similarly, in Mayzlin et al. (2014) the average rating of their sample was 3.52 for TripAdvisor.com and 3.95 for Expedia.com. Figures **??** and **??** show examples of 5 stars(good) or 1 star(bad) reviews.

TripAdvisor.com, as well as most sites that offer consumer reviews and ratings, rounds the star rating to the nearest 0.5 star. For example, if the average ratings of reviews received by the hotel up to the date the customer visits the site is 3.24, TripAdvisor.com will round and show consumers that the hotel is a 3 star. By the same token, if the hotel has an average of 3.25, it will show consumers the hotel is a 3.5 star. Following the discussion by Angrist and Pischke (2010) and by Anderson and Magruder (2012) this difference between the 3.24 hotel and the one that has an average of 3.25 creates exogenous variation that allows recognition of the impact of rating changes on performance data. We use this methodology to relate changes in ratings to performance and investment decisions. Therefore, for each year we create a cumulative average up to the date of interest and then round to the nearest 0.5 star. Figure **??** provides the density of ratings in our

[4]The brand variable is a dummy variable that takes the value of one if the hotel has one of the brands described in Table **??**.

sample, and Figure ?? provides the average hotel ratings overtime. The graph at the top of the figure shows the average rating with a band of one standard deviation, and the bottom graph shows examples of how hotels ratings change overtime.

As pointed out by Anderson and Magruder (2012), hotels may have incentives to manipulate reviews, given that a change of 0.5 stars may have repercussions on their occupancy and revenue. Although it is true that hotels may manipulate reviews to profit from this behavior in the short run, the benefits in the long run are not clear as most systems filter reviews and the risk of getting caught cheating can be quite significant[5]. Nevertheless, it is likely that hotels may manipulate reviews and therefore this needs to be controlled for. In order to do so, we use the AcNeilsen Business listing, which contains information of brand, franchise status, estimated number of employees, address, and year established. This information allows us to estimate subsamples and run the analysis for subset of interest.

Although most of the properties provide latitude and longitude information, we use Google Maps API to geocode buildings with missing data or misleading data (i.e Hotel Latitude and Longitude reflects the centroid of the city rather than the centroid of the lot). The location allows us to build rings around each hotel and estimate the number of competitors and average size in terms of number of employees within the distance of analysis. Table ?? provides information of the average size of competitors for the matched sample. The number of employees of competitors in our sample is larger than the number of employees of competitors for the entire population. This helps to recognize hotels that, given their competitive environment, may suffer from review manipulation and therefore see a decrease in the quality of the information provided on the reviews.

As result, we have a dataset with 6,416 property-year observations for 1,815 buildings, with investment information for a time frame between 2006 and 2015. Reviews collected range from the year 2001 to 2015 and there is a total of 895,768 reviews with ratings information associated to those buildings.

4 Empirical Specification

The primary interest of this project is to examine firm responses to online reviews by undergoing investment on capital improvements. A firm may react for two reason: first, managers do not have access to consumers' true belief about building's quality and reviews reveal that belief; second, they know the true quality of the

[5]http://www.telegraph.co.uk/travel/travelnews/8798854/TripAdvisor-upmarket-hotelier-faces-ruin-after-website-red-flags-hotel.html

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Year	6,416	2,011	2.870	2,005	2,015
Occupancy Rate	1,071	71.24	10.38	24.80	100.00
Daily Rate	1,045	146.02	71.79	34.70	688.25
Revenue per Room	1,031	106.96	62.01	21.04	470.93
Net Book Value	6,399	25,042.01	42,665.74	396	753,002
Age	6,416	17.44	15.72	−1	163
Investment in Capex	6,416	0.32	0.47	0	1
Time since Last Capex	6,416	1.64	2.44	0	19
Star Rating	6,414	38.08	6.10	10.00	50.00
Limited	6,416	0.36	0.48	0	1
Budget	6,416	0.34	0.47	0	1
Full Service	6,416	0.04	0.20	0	1
Extended	6,416	0.06	0.24	0	1
Unspecified	6,416	0.26	0.44	0	1
East North Central	6,416	0.01	0.04	0	1
Mountain	6,416	0.12	0.33	0	1
Mid Atlantic	6,416	0.15	0.36	0	1
North East	6,416	0.07	0.25	0	1
Pacidic	6,416	0.16	0.37	0	1
South East	6,416	0.15	0.36	0	1
South West	6,416	0.16	0.37	0	1
West North Central	6,416	0.13	0.34	0	1

Note: For the models using Investment as a dependent variable, there are 5,757 observations; for performance regressions, the observations available decrease to 952. Year represents the year of the observation. Net Book Value is the book value of the building. Age is the age of the building. Investment in Capex takes the value of 1 if the investment is in the top 20% of capital expenditures. Time since last Capex is the number of years since the last capital expenditure exceeded the top 20%. Star Rating is the average rating the hotel receives, from 10 to 50 (1-5 stars). Limited, Budget, Full Service and Extended are indicator variables that take the value of 1 if the hotel is of that Type (Ex. Budget Hotel or Limited Service Hotel) or 0 if otherwise. Location variables (East North Central, Mountain, Mid Atlantic, North East, Pacific, South East, South West, West North Central) are indicators that take the value of 1 if the building is located in that NCREIF declared region.

building but because there is randomness and discreteness in reviews, managers maintain an investment level in excess to avoid falling into a lower star level due to bad random reviews. The former reason, provides a positive outcome through information revelation. The latter reason results in wasteful investment, managers respond to the noise of signal to avoid potential future loses due to a random bad review (or reviews).

Online reviews reveal information to consumer and the information provided has the power to alter demand (Anderson and Magruder, 2012; Luca, 2016).In order to further study the impact of a 0.5 star rating change on performance, we use a model similar to the one proposed by Anderson and Magruder (2012). The idea is that consumers observe only an approximation of the average star rating to the nearest 0.5 star. Therefore, we create eight thresholds that identify when a hotel increases to the next 0.5 star. For example, if the average rating of the hotel is 3.24 stars, the approximation shows the hotel rating as 3.00 stars, but for a 3.25 rating the approximation shows the hotel rating as 3.5 stars. In this example, 3.25 stars is the threshold. This discontinuity allows the identification of the impact of reviews on performance. The thresholds we use for this specification are 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, and 4.75. We use the following empirical model:

$$Perf_{it} = \alpha_i + \vec{\beta_1} f(\bar{R_{it}} - R_{thre}) + \beta_2 I(\bar{R_{it}} > R_{thre}) + \vec{\beta_1}2 f(\bar{R_{it}} - R_{thre}) I(\bar{R_{it}} > R_{thre}) + \vec{\beta_3} \vec{F.E.} + \varepsilon_t \tag{1}$$

where $Perf_{it}$ represents the log value of a performance measure of interest(occupancy rate, revenue per room and average daily rate), $\bar{R_{it}}$ is the average rating of the hotel i at time t, and R_{thre} is the closest threshold to $\bar{R}$. The vector of coefficients $\vec{\beta_1}$ are the coefficients associated to the polynomial function, $f(T)$, to describe the trend with respect to the distance to the threshold. The coefficient of interest is β_2 which represents the impact of an increase of 0.5 stars when $\bar{R}$ passes the threshold R_{thre}. The coefficient in $\vec{\beta_1}2$ are the ones associated with the interaction term between the polynomial describing the trend $f(\bar{R_{it}} - R_{thre})$ and the dummy variable $I(\bar{R_{it}} > R_{thre})$ that indicates that the average rating surpassed a threshold. Finally, $\vec{F.E.}$ is a vector containing other controls like year, property type, and region where the hotel is located.

The second model of interest is the impact of reviews on investment decisions. A characteristic of capital expenditure is that its amount varies significantly from period to period, ranging from a relatively small percentage of the accounting value of the property, to percentages that exceed 20% of its value. Since there are periods where capital improvements are sufficiently small, we dichotomize the values by assigning a 1 to the top 20% of the capital expenditure measured as a percentage of the net book value of the property and zero otherwise[6]. This allows us to estimate a linear probability model that associates the value of investment to consumer reviews, building fixed effects, time varying fixed effects at various locations, property type fixed effects, and market levels. The idea of the introduction of year fixed effect is to control for unobservables specific to a year. The location fixed effect allows us to control for invariant location or market characteristics, and finally the time varying location interaction captures changes over time at the location level or property type level in an effort to control for lease rates and changes in lease rates at the market level.

$$Pr(Capex_t = 1) = \alpha + \vec{\beta_1} f(\bar{R_{it}} - R_{thre}) + \beta_2 I(\bar{R_{it}} > R_{thre}) + \vec{\beta_{12}} f(\bar{R_{it}} - R_{thre}) I(\bar{R_{it}} > R_{thre}) + \vec{\beta_3} \vec{Controls}_{it} + \varepsilon \tag{2}$$

Here the dependent variable represents the probability that capital expenditure exceeds the 80^{th} percentile of the relation capital expenditure over net book value. The constant α and variable R_{it} represent the same as in Equation 1. $\vec{Controls}_{it}$ includes: C_t, the time since the last investment in capital expenditure; A_{it}, the age of the building; Y_t, year fixed effect; L_i, location fixed effect; PT_i, property type fixed effect; and $PT_i : Y_t$ which is the interaction term between property type and year. ε_{it} is the residual. We use some independent variables including the textual analysis of the consumer review (i.e., word count of messages in the review), building characteristics (i.e., chain affiliation) to subsample the data and run models 1 and 2.

First, we test whether or not hotels are affected in terms of performance metrics (i.e., occupancy rate; average daily rate; and revenue per available room) by reviews.

Hypothesis 1: Hotels that receive negative reviews have lower occupancy, average daily rate and revenue per room in the subsequent period.

[6]We divide the capital expenditure for a given year by the book value, then we rank these values and create an indicator variable that takes the value of 1 if the value was above the 80th percentile. In terms of revenue, these capital expenditures exceed 5% of the revenues of the hotel for a given year and on average are equal to 52% of the revenues of a hotel for a year. This is important since 5% of sales is the standard Capex required by hotel brands in the management contract or franchise agreement.

If reviews have an impact on the performance of the firm, then we should expect that managers will respond to those reviews in various ways, with investing in the properties being one option.

Hypothesis 2: A change in the rating received by a hotel changes the probability of investing in the property.

The most important implication of reviews is that they help consumers distinguish some quality characteristics of hotels, and with this information, customers can make a choice conditional on the availability of alternative hotels. Our next hypothesis has to do with the competition available to consumers. More competitive areas will perceive a greater impact of reviews and therefore on the incentives to invest in the buildings.

Hypothesis 3: Reviews of hotels in areas with larger numbers of competitors have a stronger effect on investment decisions.

The question that arises when hotels invest in capital expenditures is whether the investment actually affects the reviews in subsequent periods. The next subsection introduces a methodology to estimate the impact, if any, that capital expenditure has on consumer ratings.

4.1 Treatment Effect Capex on Reviews

So far the methodology establishes the impact that online reviews have on performance measures, as well as the decision to invest in capital expenditures. In this subsection, we set up a methodology to identify the impact capital expenditure has on reviews after investment takes place. For an investment to make sense to the managers, it needs to influence consumer reviews after the investment. To test whether or not this is the case, we develop a regression discontinuity before and after a significant investment, in other words investment is the treatment received.

Most of the hotels have more than one capital expenditure. In order to determine if the building is treated or not, we select hotels where we observe three or more years of data and select the maximum capital expenditure as the year of treatment. We use the year of treatment to construct the pre and post period analysis for our regression. We use the reviews of competitors as control. To do so, we test 2 specifications using competitors within 14 buffers and the closest competitor as control groups. We also test 3 specifications using only competitors of the same property type. We follow a difference in difference approach of the following type

$$Rating_t - Rating_t^{Comp} = \beta_0 + \vec{\beta_1} f(T) + \beta_2 I(T > T_0) + \vec{\beta_1 2} f(T) I(T > T_0) + \vec{\beta_3} \vec{F.E.} + \varepsilon_t \quad (3)$$

where $Rating_t$ describes the average rating of a hotel during a quarter t and $Rating_t^{Comp}$ is the average rating of competitors use as controls in period t. The vector of coefficients $\vec{\beta_1}$ are the coefficients associated to the polynomial function, $f(T)$, to describe the trend. The coefficient β_2 is the coefficient of interest and represents the change in ratings after investment; this is the treatment effect. $\vec{F.E.}$ represents a vector with fixed effect including year of treatment, property fixed effect, and in order to control for potential seasonality of reviews, we use quarter fixed effect.

Hypothesis 4: Investment in capital expenditures leads to rating improvements.

5 Empirical Results

In this section, we present a set of results for the estimation of the model discussed in equations 1 and 2. We provide some evidence for the impact of reviews on performance and investment decisions of hotels. The results, though, are not ubiquitous and there is indication that information effect may be heterogeneous across location characteristics as well as building characteristics. We use clustered-robust standard errors at the state level throughout the analysis to correct for standard error correlation within state.

5.1 The Effect of Reviews on Hotels' Performance

We start by exploring whether there is a relation between reviews and the performance of the hotels. We test whether reviews have any impact on occupancy rate, average daily rate, and/or revenue per available room. We begin by showing the results for the regression discontinuity analysis for the occupancy of the building.

Table 2 shows the results for equation 1. The coefficient of interest is the one associated to Dummy Threshold; this indicates how occupancy changes with the increase of 0.5 stars. The coefficients are economically and statistically significant. The base case scenario suggests that an increase of 0.5 stars translate into a 5% increase in occupancy. The estimate seems consistent with the model that includes property type fixed effects, but becomes less statistically significant with the inclusion of other fixed effects. The last model of Table 2 includes Property Type, Region, and Building level fixed effects.

Table 3 presents the results for the same model but in this case the dependent variable is revenue per room. The change in magnitude and statistical significance in both tables when including property fixed effect suggests that part of the results from the base case scenario are related to building specific characteristics not controlled for. To further analyze this possibility, we run the fixed effect model that includes Property Type, Region, Building level fixed effects, as well as interaction term with year fixed effect to control for market characteristics that change over time. This model is the last column in each table.

5.2 The Effect of Reviews on Hotels' Investment Decision

The main goal of this paper is to establish the relation between online consumer review and investment decisions in capital expenditure. From the literature on the optionality of capital expenditure, we know that managers would delay or accelerate investments depending on the market conditions, specifically rent and volatility of rent. We run various specifications of Equation 2 which take different definitions of market. Table 4 presents the estimates for linear probability models (LPM), and We show the estimates for the probit models as a robustness test. The dependent variable for all the models in this subsection is an indicator variable that takes the value of 1 if the capital expenditure exceeds the 3.5% of the book value of the building. Each column has a combination of other fixed effects to control for building, year and location characteristics. We also control time since last investment of significant capital expenditure to rule out that the results are driven from just pure obsolescence of the building or by recurring investment required by brands in order to keep building up to standard.

Table 2: Regression Discontinuity Model Occupancy

	Dependent variable:			
	Log Occupancy			
	Base	Property Type F.E	NCREIF F.E.	Property ID F.E.
Dummy Thrshold	0.050***	0.048***	0.036**	0.018*
	(0.017)	(0.017)	(0.017)	(0.011)
Distance to Threshold	−0.013	−0.013	−0.005	−0.001
	(0.008)	(0.009)	(0.010)	(0.006)
Interaction	0.002	0.004	−0.007	−0.003
	(0.015)	(0.016)	(0.017)	(0.010)
Constant	4.089***	4.003***	4.174***	4.335***
	(0.067)	(0.074)	(0.069)	(0.106)
Year F.E.	Yes	Yes	Yes	Yes
Property type F.E.	No	Yes	Yes	Yes
NCREIF F.E.	No	No	Yes	Yes
Standard Error	Clustered State Level	Clustered State Level	Clustered State Level	Clustered State Level
N of Buildings	339	339	339	339
Observations	1,016	1,016	1,016	1,016
Adjusted R^2	0.200	0.211	0.313	0.814

Note: We run a regression discontinuity model where the threshold is the cutoff point at which the rating increase by 0.5 star. We use 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, and 4.75 as the cutoffs. If the average rating is above the threshold the **Dummy Threshold** takes the value of one and zero if below. **Distance to Threshold** is the distance to the closest threshold. **Interaction** is the interaction term between **Dummy Threshold** and **Distance to Threshold**. All standard errors are clustered robust at the state level. Base column has the base model of regression discontinuity model. The column Property Type includes property type fixed effects in the regression discontinuity model. The column NCREIF includes Property Type, as well as, NCREIF region fixed effects. Finally, column Property ID includes Property Type, NCREIF region fixed effects and Property ID fixed effect. *p<0.1; **p<0.05; ***p<0.01

Table 3: Regression Discontinuity Model Revenue

	Dependent variable:			
	Log Revenue			
	Base	Property Type F.E	NCREIF F.E.	Property ID F.E.
Dummy Thrshold	0.069***	0.078***	0.057***	0.024**
	(0.022)	(0.024)	(0.022)	(0.012)
Distance to Threshold	−0.025*	−0.033**	−0.015	−0.007
	(0.013)	(0.015)	(0.014)	(0.008)
Interaction	0.029	0.025	0.013	0.002
	(0.028)	(0.029)	(0.026)	(0.013)
Constant	3.567***	2.550***	2.888***	3.331***
	(0.112)	(0.050)	(0.080)	(0.141)
Year F.E.	Yes	Yes	Yes	Yes
Property type F.E.	No	Yes	Yes	Yes
NCREIF F.E.	No	No	Yes	Yes
Standard Error	Clustered State Level	Clustered State Level	Clustered State Level	Clustered State Level
N of Buildings	339	339	339	339
Observations	1,016	1,016	1,016	1,016
Adjusted R^2	0.823	0.821	0.859	0.973

Note: We run a regression discontinuity model where the threshold is the cutoff point at which the rating increase by 0.5 star. We use 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, and 4.75 as the cutoffs. If the average rating is above the threshold the **Dummy Threshold** takes the value of one and zero if below. **Distance to Threshold** is the distance to the closest threshold. **Interaction** is the interaction term between **Dummy Threshold** and **Distance to Threshold**. All standard errors are clustered robust at the state level. Base column has the base model of regression discontinuity model. The column Property Type includes property type fixed effects in the regression discontinuity model. The column NCREIF includes Property Type, as well as, NCREIF region fixed effects. Finally, column Property ID includes Property Type, NCREIF region fixed effects and Property ID fixed effect. *p<0.1; **p<0.05; ***p<0.01

Table 4: Regression Effect Reviews on Investment

	Dependent variable:					
	Invest in Capex Yes(1)/No(0)					
	OLS			*Probit*		
	Hotel Type	State	NCREIF Region	State	County	Zipcode
	(1)	(2)	(3)	(4)	(5)	(6)
Star Rating Lag	-0.200^{***}	-0.243^{***}	-0.216^{***}	-0.448^{***}	-0.624^{***}	-0.622^{***}
	(0.069)	(0.079)	(0.081)	(0.126)	(0.131)	(0.209)
Building Age	0.023	-0.058^{***}	-0.011	-0.006^{***}	-0.004^{**}	-0.002
	(0.017)	(0.008)	(0.032)	(0.002)	(0.002)	(0.004)
Time Since Capex	-0.158^{***}	-0.150^{***}	-0.143^{***}			
	(0.029)	(0.027)	(0.021)			
Constant	0.087	1.913^{***}	1.092	-8.499^{***}	-4.414^{***}	-4.661
	(0.549)	(0.293)	(1.022)	(0.547)	(0.597)	(199.665)
Property F.E.	Yes	Yes	Yes	No	No	No
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
NCREIF F.E.	No	No	Yes	No	No	No
County F.E.	No	No	No	No	Yes	No
State F.E.	No	Yes	No	Yes	No	No
Type Hotel F.E.	Yes	No	No	Yes	Yes	No
Time Variant F.E.	Type Hotel	State	NCREIF	Type Hotel	Type Hotel	Type Hotel
Cluster Std. Errors	State Level	State Level	State Level	State Level	State Level	State Level
N of Buildings	1815	1815	1815	1815	1815	1815
Observations	5,884	5,884	5,884	5,884	5,884	5,884
Adjusted R^2	0.389	0.375	0.364			

Note: We run two models: a Linear Probability Model (OLS) and a Probit model with the dependent variable Invest in Capex, which is an indicator that takes the value of 1 if the capital expenditure is within the top 20% of capital expenditures. All standard errors are clustered robust at the state level. *Star Rating Lag* is the log of average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. *Building Age* represent the age of the building. *Time Since Capex* is the number of years since the last capital expenditure exceed the top 20% of all observation; We also include a quadratic term and cubic term of this variable in the regression, both statistically significant, and not shown here. Column **Hotel Type** is the regression that includes property type fixed effect, as well as property type time variant fixed effects. **State** column in the OLS columns includes time variant fixed effects at the state level. **NCREIF Region** column includes time variant fixed effects for NCREIF regions. The Probit models include fixed effect at the **State**, **County** and **Zipcode** levels and time variant fixed effects at the hotel type level. *p<0.1; **p<0.05; ***p<0.01

In the first column in Table 4, we use the hotel type as our market definition. We use fixed effect for Full Service, Limited Service, Extended Stay and Budget Hotel, and time varying fixed effects at the hotel type level. The results indicate that a 1% drop in the star rating increases the probability to invest in capital expenditure above 3.5% of the book value by approximately 0.200% of investment and is statistically significant at the 1% level. This gives strong evidence for Hypothesis 2; hotels change the probability of investing in capital expenditures depending on star ratings. Other LPM specifications are consistent with this finding and all of them imply that changes in star ratings affect the decision to invest, even after controlling for building characteristics, location fixed effect and time varying fixed effect. The results are also robust to specifications using the Probit model. However, for this last robustness check the inclusion of a property fixed effect does not allow the convergence of the Probit model; therefore, we drop the fixed effect at the property level. We also drop time since capex variable as it gives us the same problem. We do include location fixed effect at various levels, state, county and zip code, property type fixed effects and a time variant fixed effect to control for market conditions. The results still hold, estimates are statistically significant, and the take away from this table is that changes in star rating are correlated with the decision to invest.

To further study the impact of reviews on managers' decision, we estimate the same model of Equation 2 but in this case we use the distance of the cumulative average of star ratings of a hotel to a threshold. We define the thresholds which a manager may be interested in as 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25 and 4.75 stars. Managers may want to invest in a building that has the risk of randomly falling in a lower star category if the cumulative average star rating falls below one of these threshold. The results of Table 5 suggest that there is statistical evidence that as the cumulative average comes close to the threshold managers are more likely to invest in capital expenditure.

Column Log Stars (1) in Table 5 shows that an increase in the logarithm of the average star rating translate into a decrease on the probability of investment in Capex. Column Distance uses the distance of the cumulative average to the threshold as a control, and the result indicates that a distance of 0.1 star from the threshold decreases the probability of investment by 1.3 percentage points. In column 3 (Dist.+Top), we include a dummy that indicates that the cumulative average is now closer to the top threshold and we also include an interaction term. In other words, column 3 accounts for the possibility of a change of slope. The results indicate that as you move further away from the bottom threshold the probability of investment decreases, but as you start moving closer to the top threshold the

Table 5: Regression Model Investment Decision

	Dependent variable:			
	Invest in Capex Yes(1)/No(0)			
	Log Stars	Distance	Dist.+Top	Dummy
	(1)	(2)	(3)	(4)
Star Rating Lag	-0.223^{***}			
	(0.069)			
Distance		-0.013^{*}	-0.032^{**}	
		(0.008)	(0.015)	
Dummy High			-0.080	
			(0.055)	
Distance*Dummy High			0.034^{*}	
			(0.020)	
Close to Low Edge				0.062^{***}
				(0.023)
Close to High Edge				-0.024
				(0.027)
Constant	-0.817^{***}	-1.276^{***}	-1.395^{***}	-1.431^{***}
	(0.259)	(0.209)	(0.254)	(0.189)
Year F.E.	Yes	Yes	Yes	Yes
Property type F.E.	Yes	Yes	Yes	Yes
NCREIF F.E.	Yes	Yes	Yes	Yes
Property ID F.E.	Yes	Yes	Yes	Yes
Threshold F.E.	Yes	Yes	Yes	Yes
Standard Error	Clustered State Level	Clustered State Level	Clustered State Level	Clustered State Level
N of Buildings	1,956	1,956	1,956	1,956
Observations	5,884	5,884	5,884	5,884
Adjusted R^2	0.400	0.400	0.400	0.401

Note: We run a Linear Probability Model (OLS) with the dependent variable Invest in Capex, which is an indicator that takes the value of 1 if the capital expenditure is within the top 20% of capital expenditures. All standard errors are clustered robust at the state level. *Star Rating Lag* is the log of average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. All regressions include *Time Since Capex* which is the number of years since the last capital expenditure exceeded the top 20% of all observation, Year, Property Type, NCREIF region, and Property ID Fixed Effects. The regressions also include interactions terms between Year, Property Type and Region. Column **Log Stars** is the regression that includes the logarithm of the cumulative average star ratings. **Distance** column in the OLS columns includes distance in terms of stars to the bottom threshold before it drops to a 0.5 star lower. **Dist.+Top** column includes a distance to the lower threshold and a dummy to indicate that the distance is closer to the higher threshold. **Dummy** column includes dummy variables to indicate that the average rating is within 0.075 stars to the bottom threshold (Close to Low Edge) and a dummy to indicate the average is close to the top threshold (Close to High Edge). $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

probability of investment increases again. Finally, we also test an specification in which we create dummy variables that takes the value of 1 if the cumulative star rating average is within 0.075 stars to the bottom (*Close to Low Edge*) or 0.075 stars from the top threshold(Close to High Edge). The results further indicate that manager invest if average rating is close to the bottom edge. If the Close to Low Edge takes the value of 1 managers probability of investment increase by 6.2%.

To further study these results, we run sub-sample analysis to see if there is any specific type of hotel driving them. Tables 6 and 7 estimate specification 4 in Table 5 for to sub-samples. Table 6 divides the observations by average star rating and Table **??** by the number of competitors. The results suggest that higher rating hotels are driving these results and areas with more competition make these result more statistically significant. Finally to further analyze we run a falsification on the investment decision. Table 8 present the results that randomize the year the investment takes place for each hotel. The results indicate that there is no significant correlation between the star ratings and the random decision to invest.

Table 6: Regression Model Investment Decision

	Dependent variable:	
	Invest in Capex Yes(1)/No(0)	
	Low Average Star Rating	High Average Star Rating
	(1)	(2)
Close to Low Edge	0.053	0.058*
	(0.051)	(0.032)
Close to High Edge	−0.053	−0.004
	(0.035)	(0.045)
Constant	0.546***	0.463
	(0.067)	(0.310)
Year F.E.	Yes	Yes
Property type F.E.	Yes	Yes
NCREIF F.E.	Yes	Yes
Property ID F.E.	Yes	Yes
Standard Error	Clustered State Level	Clustered State Level
N of Buildings	852	1315
Observations	2,179	3,705
Adjusted R^2	0.448	0.383

Note: We run a Linear Probability Model (OLS) with the dependent variable Invest in Capex, which is an indicator that takes the value of 1 if the capital expenditure is within the top 20% of capital expenditures. All standard errors are clustered robust at the state level. All regressions include *Time Since Capex* which is the number of years since the last capital expenditure exceeded the top 20% of all observation, Year, Property Type, NCREIF region, and Property ID Fixed Effects. The regressions also include interactions terms between Year, Property Type and Region. Column **Low Average Star Rating** is the regression that includes only Hotels with low average star ratings. **High Average Star Rating** column is the regression that includes only Hotels with high average star ratings. Dummy variables, *Close to High Edge and Close to Low Edge*, indicate that the average rating is within 0.075 stars to the bottom threshold (Close to Low Edge) or close to the top threshold (Close to High Edge). $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 7: Regression Model Investment Decision

	Dependent variable:	
	Invest in Capex Yes(1)/No(0)	
	Low Competition(1 or 0 Comp.)	High Competition(+2 Comp.)
	(1)	(2)
Close to Low Edge	0.120*	0.063**
	(0.064)	(0.028)
Close to High Edge	0.006	−0.030
	(0.072)	(0.025)
Constant	1.365**	−1.569***
	(0.552)	(0.376)
Year F.E.	Yes	Yes
Property type F.E.	Yes	Yes
NCREIF F.E.	Yes	Yes
Property ID F.E.	Yes	Yes
Standard Error	Clustered State Level	Clustered State Level
N of Buildings	115	1841
Observations	868	5,016
Adjusted R^2	0.432	0.404

Note: We run a Linear Probability Model (OLS) with the dependent variable Invest in Capex, which is an indicator that takes the value of 1 if the capital expenditure is within the top 20% of capital expenditures. All standard errors are clustered robust at the state level. All regressions include *Time Since Capex* which is the number of years since the last capital expenditure exceeded the top 20% of all observation, Year, Property Type, NCREIF region, and Property ID Fixed Effects. The regressions also include interactions terms between Year, Property Type and Region. Column **Low Competition** is the regression that includes only Hotels with one or less number of competitors within 1 mile radius. **High Competition** column is the that includes only Hotels with two or more number of competitors within 1 mile radius. Dummy variables, *Close to High Edge and Close to Low Edge*, indicate that the average rating is within 0.075 stars to the bottom threshold (Close to Low Edge) or close to the top threshold (Close to High Edge). *p<0.1; **p<0.05; ***p<0.01

Table 8: Regression Falsification Test

	Dependent variable:
	Invest in Capex Yes(1)/No(0) Random Investment Capex
Star Rating Lag	0.062
	(0.075)
Treatment:Star Rating Lag	0.005
	(0.099)
Building Age	−0.015
	(0.016)
Constant	0.670
	(0.594)
Property F.E.	Yes
Year F.E.	Yes
Property Type F.E.	Yes
Time Variant F.E.	Pro. Type
Standard Error	Clustered
	State Level
N of Buildings	1812
Observations	5,875
Adjusted R^2	0.009

Note: I run the model from Table 4in column 1, but in this case, I falsify the investment variable. I recreate a random investment decision (0 or 1). The *Treatment* takes the value of 1 in hotels that are in both the top 80% of hotels in terms of number of competitors within 10 miles and in the top 80% of hotels in terms of the number of reviews received. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after visit, rounded to the nearest 0.5. *Building Age* represent the age of the building. *Time Since Capex* is the number of years since the last capital expenditure exceed the top 20% of all observation, I also include a quadratic term and cubic term of this variable in the regression. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

5.3 The Impact of Investment on Reviews

Our last analysis is an event study around the period during which the investment takes place. Whenever we observe an investment greater than 3.5%, we construct the average rating per quarter for the 8 quarters pre and post the year of the event. We estimate Equation 3 and test different trends in the pre and post period for various subsamples. We drop reviews from the year the investment takes place and focus only on the pre and post period, as observations of ratings within that year are affected by construction in the building due to capital expenditures.

Table 9 shows the estimated coefficients for Equation 3. Columns 1 and 2 show estimates of the regression using the full sample. The trend variable is positive and significant; the ratings difference between the hotel and the closest competitor increase 0.038 stars per year. The trend coefficients for the specifications using all competitors within 14 kilometers buffers are statically significant and of the same sign. Treatment represents the period after the investment in capital expenditure takes place. The coefficients in the full sample regressions suggest that after investment takes place, star ratings increase by approximately 0.18 star relative to the closest competitor. Although not statistically significant, the interaction term between trend and treatment suggest that after investment takes place the effects of capital expenditure on ratings start decreasing. The relation between treatment and interaction term suggest that the effects of capital expenditure disappear after approximately 18 quarters relative to the closest competitors. Figure 1 shows the fitted values of the regression in column 1 of Table 9, the bottom graph shows the detrended values and leaves only the treatment effect of investment and the change in trend after treatment. We run several robustness tests that use a subsample for different definitions of competitors to study the effect of potential review manipulation; columns 3 to 4 show this analysis. The impact of treatment for Hotels that are less likely to have manipulated reviews.

For the columns under the Filter group, we filter out hotels not associated with a brand, we leave out hotels with small competitors, and, finally we drop hotels with more bed and breakfast accommodations within a 10 miles radius[7]. The results from the subsample analysis suggest that investment affects the ratings in subsequent periods. Moreover, the results for treatment hold even after using various buffer analyses to retrieve the close competitors. In the case when we only use the closest hotel as control group, the treatment implies an increase of 0.3098 star.

[7]By "more" we mean hotels that are above the 20th percentile ranked by the number of bread and breakfast accommodations. In other words, we drop observations in the lowest quartile

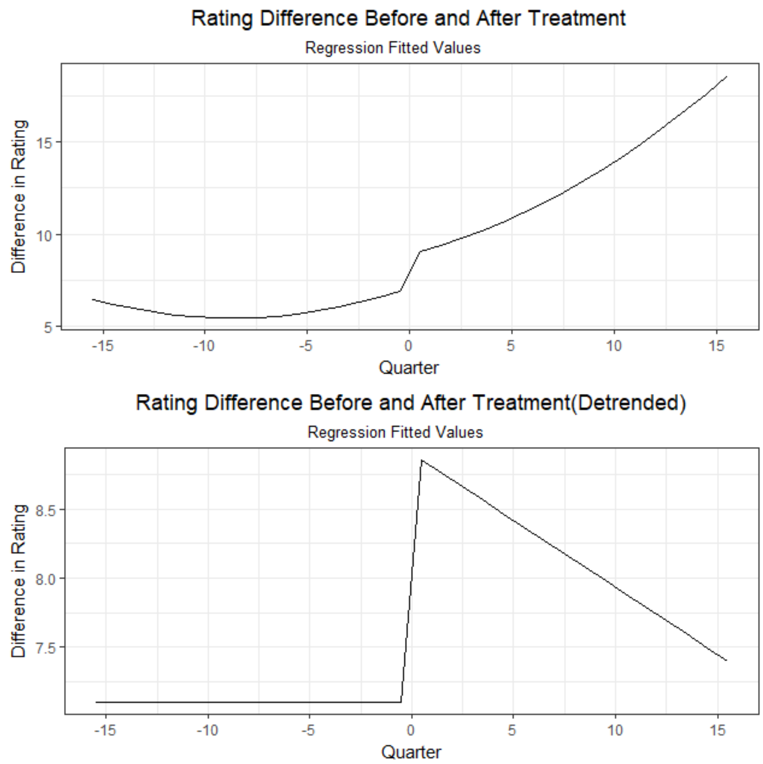

Figure 1: Regression Discontinuity Design.

The graph shows the regression model to estimate treatment effect of capital expenditure. The line represents the fitted values from regression in column 4 of Table 9. Difference in rating describes the difference in rating between the hotel that invest and its closest competitor. The variable has a range of -40 minimum to 40 maximum with a mean of -1.9. The grey line at Quarter 0 represents the treatment period, and indicates the year of the maximum capital expenditure observed for the building. The analysis looks at the 8 previous quarters and 8 quarters subsequent to the year of treatment.

Table 9: Regression Reviews

	Dependent variable:			
	Neighbor	14k Competitors	Neighbor Filter	14k Comp. Filter
	(1)	(2)	(3)	(4)
Trend	0.382***	0.386***	0.613***	0.694***
	(0.084)	(0.079)	(0.130)	(0.130)
Treatment	1.805***	1.674***	3.098***	3.583***
	(0.652)	(0.564)	(0.575)	(0.590)
Interaction	−0.097	−0.036	0.101	0.180
	(0.188)	(0.179)	(0.243)	(0.233)
Constant	8.165**	7.495***	6.462***	6.699***
	(3.195)	(2.248)	(1.323)	(1.012)
Property F.E.	Yes	Yes	Yes	Yes
Quarter F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Standard Error	Clustered	Clustered	Clustered	Clustered
N. Building	353	377	152	161
Observations	5,910	7,118	2,544	3,003
Adjusted R^2	0.484	0.401	0.447	0.400

Note: This table examines the OLS changes in Ratings after investment takes place. The *dependent variable* for these models is the average rating within a quarter in a scale from 10 to 50 (10 representing 1 star and 50 5 stars) relative to competitors in the area. Column Neighbor uses the ratings of the closest competitor to the hotel as a counter factual. Column 14k uses the average rating of competitors within 14 kilometers. Variable Treatment takes the value of 1 if the quarter of the ratings is after the year the hotels invested in capital expenditure, and 0 if the quarter is prior to the investment. The columns under Filter use only observations of hotels that may have less review manipulation. These hotels are the ones that are associated with a brand. We use the top 30 brands in terms of numbers of hotels. We also filter out hotels that have small competitors. We define the small competitors as the ones within the 20th percentile ranked by average number of employees. We filter out hotels that have more bed and breakfast in the area. We use only hotels that are below the 80th percentile ranked by the number of bed and breakfast in the area. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

We also test different definitions of competition. Table 10 shows the analysis using only competitors of the same property type. For example: if a building is a Full-Service hotel, We use only Full-Service hotels within 14 kilometers as competitors. The results suggest a rating improvement after capital expenditure takes place. The only test that yields that the improvement in ratings is not statistically significant is for the Limited-Service hotels.

Table 10: Regression Reviews Property Type

	Dependent variable:		
	Extended Stay	Full-Service	Limited-Service
	(1)	(2)	(3)
Trend	2.326^{***}	0.827^{***}	-0.271^{***}
	(0.136)	(0.136)	(0.101)
Treatment	9.470^{***}	3.605^{***}	0.280
	(0.893)	(1.031)	(0.429)
Interaction	-0.037	-0.607	-0.617^{***}
	(0.291)	(0.403)	(0.181)
Constant	80.014^{***}	18.808^{***}	-11.814^{***}
	(1.839)	(4.316)	(0.947)
Property F.E.	Yes	Yes	Yes
Quarter F.E.	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes
Standard Error	Clustered	Clustered	Clustered
N. of Buildings	185	42	111
Observations	3,389	811	2,145
Adjusted R^2	0.414	0.244	0.271

Note: This table examines the OLS changes in Ratings after investment takes place. The *dependent variable* for these models is the average rating within a quarter in a scale from 10 to 50 (10 representing 1 star and 50 5 stars) relative to competitors in the area. Column Extended Stay uses the ratings of the competitors within the same property type to the hotel as a counter factual. Column Full-Service uses the average rating of competitors that are Full-Service and Column Limited-Service uses the competitors that are limited service. Variable Treatment takes the value of 1 if the quarter of the ratings is after the year the hotels invested in capital expenditure, and 0 if the quarter is prior to the investment. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

As we mentioned in Section 5.1, an improvement in the reputation of the hotel leads to increases in revenue and occupancy. These results relate to the findings in Bond et al. (2014), where the authors conclude that capital improvement leads to higher income. In this paper, we relate revenue increases to and improvement in the hotel reputation online.

5.4 Alternative Hypothesis

An alternative explanation for the negative relation between ratings and capital expenditure decisions is that during the time frame of this study REITs acquired operating hotels and therefore there was a change of ownership. Most management contracts and franchise agreements require new owners to bring the hotel to the current brand standard and therefore invest in property improvements. If the former owner sold the hotel due to declining performance and ratings, the new owner independent of previous performance is required by contract to invest in bringing the property to the new brand standard. Therefore, we should expect a negative relation between ratings and investment in capital improvements.

To test this alternative hypothesis, we use Equation 2 and add a dummy variable to control for the period right after the acquisition of the building. Table 11 shows this analysis where variable Brand Standard Dummy takes the value of 1 if the observation is within 2 years of the acquisition date. The results suggest that in fact new owners are 11.4% more likely to invest during this time period. Nevertheless, the negative relation between ratings and capital expenditures still holds and are statistically significant. In addition, the impact of ratings on investment increased by 7% with respect the base model in Table 4 column 1. In other words, this finding shows that the results on the negative relation is not due to capital expenditure required by contract to bring hotel to new brand standards.

Overall, the results from Subsections 5.1 , 5.2 and 5.3 indicate that the reputation mechanism of online consumer reviews leads to changes in the industry. Consumers alter their consumption patterns as reviews reveal the quality of hotels. I show that these changes in consumption patterns create the incentive for hotels to invest, and finally, that the investments lead to more positive consumer reviews in subsequent periods.

Table 11: Alternative Hypothesis

	Dependent variable:
	Invest in Capex Yes(1)/No(0)
Star Rating Lag	-0.214^{***}
	(0.068)
Building Age	0.034^{*}
	(0.018)
Brand Standard Dummy	0.114^{***}
	(0.029)
Time Since Capex	-0.154^{***}
	(0.029)
Constant	-0.184
Property F.E.	Yes
Year F.E.	Yes
Property Type F.E.	Yes
Time Variant F.E.	Pro. Type
Standard Error	Clustered
	State Level
N of Buildings	1760
Observations	5,715
Adjusted R^2	0.397

Note: We run the model from Table 4 in column 1, but in this case, We include a dummy to control for the first two year since acquisition. The *Brand Standard Dummy* takes the value of 1 in hotels that are in within two years since acquisition date. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after visit, rounded to the nearest 0.5. *Building Age* represent the age of the building. *Time Since Capex* is the number of years since the last capital expenditure exceed the top 20% of all observation, we also include a quadratic term and cubic term of this variable in the regression. All standard errors are clustered robust at the State level. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

6 Conclusion

In this paper we propose a novel question regarding online reviews in the real estate industry. In particular, we examine the impact of online reviews on firm investment decisions. We link online reviews of hotels from TripAdvisor.com to financial information on those hotels. We control for buildings and location characteristics using various fixed effect controls as well as time varying fixed effects. Overall, the results suggest there is a link between consumer generated content and a firm's decision to invest in capital improvements. We find that consumers alter their consumption based on information available in the form of online reviews. The findings suggest that an extra 0.5 star increases the occupancy rate by 1.9% and revenue by 1.6%. The information disclosed by previous consumers has a significant impact on the market by altering consumption as well as quality.

Our research contributes to the literature of information asymmetry in real estate by studying the impact that reputation mechanism has on the quality of buildings. Prior literature on information disclosure looks at mandatory and voluntary disclosure and its impact on price and quality. In this paper, we look at consumer generated information and the impact it has on the quality of buildings. We find that managers alter their decision to invest in a building depending on consumer reviews. If the average star rating of a building is close to a threshold that may affect future revenues managers are 6% more likely to invest in Capital improvements. Following the literature on capital expenditure in real estate, this paper's contribution resides in the impact of information considerations at the moment of making investment decisions in quality. This assumes that capital expenditure is a good proxy for quality investment in a building.

References

Akerlof, G. A. (1970), 'The market for "lemons": Quality uncertainty and the market mechanism', *The Quarterly Journal of Economics* **84**(3), 488–500.
URL: *http://www.jstor.org/stable/1879431*

Anderson, M. and Magruder, J. (2012), 'Learning from the crowd: Regression discontinuity estimates of the effects of an online review database*', *The Economic Journal* **122**(563), 957–989.
URL: *http://dx.doi.org/10.1111/j.1468-0297.2012.02512.x*

Angrist, J. D. and Pischke, J.-S. (2010), 'The credibility revolution in empirical economics: How better research design is taking the con out of econometrics', *Journal of Economic Perspectives* **24**(2), 3–30.
URL: *http://www.aeaweb.org/articles?id=10.1257/jep.24.2.3*

Bond, S., Shilling, J. and Wurtzebach, C. (2014), Commercial real estate market property level capital expenditure: An options analysis, Technical report, Working Paper.

Childs, P. D., Ott, S. H. and Riddiough, T. J. (2004), 'Effects of noise on optimal exercise decisions: The case of risky debt secured by renewable lease income', *The Journal of Real Estate Finance and Economics* **28**(2), 109–121.

Garmaise, M. J. and Moskowitz, T. J. (2003), 'Confronting information asymmetries: Evidence from real estate markets', *The Review of Financial Studies* **17**(2), 405–437.

Jin, G. Z. and Leslie, P. (2003), 'The effect of information on product quality: Evidence from restaurant hygiene grade cards*', *The Quarterly Journal of Economics* **118**(2), 409–451.
URL: *+ http://dx.doi.org/10.1162/003355303321675428*

Levitt, S. D. and Syverson, C. (2008), 'Market distortions when agents are better informed: The value of information in real estate transactions', *The Review of Economics and Statistics* **90**(4), 599–611.

Lewis, G. (2011), 'Asymmetric information, adverse selection and online disclosure: The case of ebay motors', *The American Economic Review* **101**(4), 1535–1546.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.101.4.1535*

Luca, M. (2016), 'Reviews, reputation, and revenue: The case of yelp. com'.

Mayzlin, D., Dover, Y. and Chevalier, J. (2014), 'Promotional reviews: An empirical investigation of online review manipulation', *The American Economic Review* **104**(8), 2421–2455.

Rogerson, W. P. (1983), 'Reputation and product quality', *The Bell Journal of Economics* **14**(2), 508–516.
URL: *http://www.jstor.org/stable/3003651*

Shapiro, C. (1983), 'Premiums for high quality products as returns to reputations*', *The Quarterly Journal of Economics* **98**(4), 659–679.
URL: *http://dx.doi.org/10.2307/1881782*

Sparks, B. A. and Browning, V. (2011), 'The impact of online reviews on hotel booking intentions and perception of trust', *Tourism Management* **32**(6), 1310–1323.

Titman, S., Tompaidis, S. and Tsyplakov, S. (2004), 'Market imperfections, investment flexibility, and default spreads', *The Journal of Finance* **59**(1), 165–205.