ESSAYS ON ONLINE PRESENCE AND BUSINESS PERFORMANCE

by

ATM Sayfuddin

A Dissertation Submitted in Partial Fulfillment
of the Requirement for the Degree of
Doctor of Philosophy in Economics

Middle Tennessee State University

March 2019

Dissertation Committee

Dr. Adam D. Rennhoff, Chair

Dr. Michael Roach, Member

Dr. Keith Jacks Gamble, Member

This dissertation is dedicated to my late father, Md. Shahabuddin, my mother, Tahera Begum, my caring wife, Sazia Zarin, and all my sisters, without supports from whom I would not be able to be where I am now.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Almighty Allah (SWT), the Most Beneficent and the Most Merciful, for giving me the knowledge, strength, and ability to successfully complete this dissertation.

I am grateful to my dissertation committee chair, Dr. Adam D. Rennhoff, for his guidance and advice. I am thankful to Dr. Michael Roach, Dr. Phillip Seagraves, and Dr. Keith Jacks Gamble for their feedback and support. Parts of this dissertation also benefited by comments from my peers at Middle Tennessee State University (MTSU), including Elvedin Bijelic, Max Deaton, and Mohammad Movahed; from participants at the 2018 American Real Estate Society conference in Bonita Springs, Florida; from the 2018 TEAMFest conference participants at the University of Tennessee, Knoxville; and from the professors at the economics research seminars in MTSU. I render special thanks to the Jones College of Business and particularly to Dr. Keith Jacks Gamble, Chair of the Department of Economics and Finance at MTSU, for providing a generous fund (Dean's Fund for Research Excellence) for data collection for the first chapter of this dissertation. All errors are my own.

ATM Sayfuddin

ABSTRACT

This dissertation focuses on interdisciplinary topics related to Industrial Organization, Environmental Economics, and Real Estate Economics. The first chapter of my dissertation studies the impact of going green on business performance. For this study, I show that a firm can adopt green practices in order to differentiate itself from its competitors. Competition in a market drives down prices, but a firm can be less affected by the competition when its products are differentiated. Hence, going green can have economic implications for businesses. Employing multiple empirical strategies, I find a hotel's location plays a determining role in the effect of going green on its performance (i.e., occupancy rate, price, and revenue). My results suggest while green hotels in small towns and resorts enjoy a price and a revenue premium, with no significant effect on their occupancy rates, green hotels near interstates, airports, and in big cities do not get the economic benefits of adopting green practices. Further investigation reveals that the hotels in less popular cities enjoy the most benefit from becoming green. The results of this study thus point out to the need for asking "when" or "where" going green pays off, instead of "whether" going green pays off.

The second chapter of my dissertation investigates the economic implications of online reviews. I use review data from a leading travel website, TripAdvisor.com, and revenue data for the hotel industry in Texas to examine the causal impact of online customer reviews on hotel revenue. On TripAdvisor.com, the star-rating displayed for each hotel represents a rounded average rating for all the submitted reviews, which results in a hotel having a 0.5-star increase in its displayed rating when its actual average rating crosses a threshold. This allows me using a quasi-experimental approach, regression discontinuity, to study the impact of a 0.5-star increase on the revenue of hotels. My findings show that a 1-star increase in the star-rating of a hotel on TripAdvisor.com leads to approximately a 2.2 - 3 percent increase in monthly

revenue. This is equivalent to a range of additional \$4,593 - \$6263 monthly revenue or \$55,117 - \$75,159 yearly revenue for an average hotel.

The third chapter investigates the financial implications of brand affiliation for businesses. Using a sample of hotels in the state of Texas that had a change of ownership between 2014 and 2017, I explore how a change in brand-affiliation that coincides with the ownership change is associated with hotel revenue. For the sample of hotels included in this study, we find after an independent hotel obtains brand-affiliation, its monthly revenue per available room (RevPAR) increases by 28.8%, on average; but I do not find any statistically significant improvement of monthly RevPAR for hotels that give up their affiliation status and become independent hotels.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I: FIGURES

CHAPTER II: FIGURES

CHAPTER I

WHEN GREEN PRACTICES AFFECT BUSINESS PERFORMANCE: AN INVESTIGATION INTO CALIFORNIA'S HOTEL INDUSTRY

1. Introduction

The relevance of green initiatives in business practices has been a widely discussed issue in recent years due to growing environmental concerns and consumer awareness. The gradual transformation of consumer behavior and their growing interest in the interaction between business organizations and environment have helped many businesses recognize responsible practices as a strategy to gain competitive advantage (Fernie *et al.*, 2010; Jones *et al.*, 2008). In almost every industry, many businesses have undertaken green initiatives in order to act responsibly (Laroche *et al.*, 2001; Trudel and Cotte, 2009). Some of the early researchers in the social sciences investigating the business benefits of green practices are from the tourism literature (Robinson *et al.*, 2016). In the tourism industry, hotels are reported to be the source of 21% of carbon emission (Han *et al.* 2011). Various studies have investigated green features in the tourism and hospitality industry, and a large section of the studies focuses on consumers' willingness-to-pay and attitude for green attributes in hotels. However, we do not have any conclusive evidence regarding the financial implications of green initiatives for businesses.

The ambiguous evidence on hotels' ability to yield revenue premiums and the presence of anecdotal evidence on the increased demand for green hotels warrant further analysis into the impact of green certifications on hotels' performance. This study, therefore, investigates the effect of green practices on business performance in the hotel industry, using the listing of green hotels on TripAdvisor.com. This website categorizes, under its GreenLeaders program, the participating hotels around the world into five levels, such as *Platinum*, *Gold*, *Silver*, *Bronze*, and *GreenPartner,* based on their eco-friendly practices. Using a cross-section of 865 hotels of which 342 are green hotels, this paper investigates whether participation in the GreenLeaders program has any effect on a hotel's performance and whether there is any heterogeneity in the effect of participation. The Key research questions include: do TripAdvisor's GreenLeaders badges have any impact on the participating hotels' occupancy rates (Occ), average daily rates (ADR), and revenue per available room (RevPAR)? Is there any heterogeneity in the effects of participation across badge types? In seeking answers to the research questions, this study differs from the prior research. This study addresses the endogeneity bias arising from hotels' self-selection for obtaining green certifications. Apart from the empirical approach, this paper utilizes a novel dataset from TripAdvisor.com, which makes it possible to identify the exact date when each hotel went green. This unique information, not used in prior studies, enables me to estimate the impact of green labels on the hotel performance more reliably. I discuss this in further detail in section 4.

In the commercial real estate sector, business performance and corporate social responsibility (CSR) are closely associated. Organizations with green agenda are usually willing to pay a premium as tenants of green offices. In the hotel industry, however, there are differences concerning the price premium for green hotels. When it comes to a traveler's hotel choice, price plays a crucial role in both the leisure and business travelers (Lockyer, 2005). Hotel leases are also much shorter (i.e., one or more nights) compared to office leases (i.e., 3 to as long as 20 years). As a result, travelers may not appreciate the benefits associated with CSR for the price premium in green hotels. Nonetheless, travelers that tend to stay longer and travel frequently may have some preference for green hotels (Robinson *et al.*, 2016). Some of the recent surveys indicate a growing awareness for green choices among travelers. In March 2017, Booking.com conducted a global survey of 10,000 travelers and found 42% of the respondents considered themselves sustainable travelers. In another study conducted by TripAdvisor, two-thirds of the travelers said they planned to make more environment-friendly choices over the following years. Despite a growing awareness among travelers, the overall performance of the green hotels depends on the market share of such travelers (Robinson *et al.*, 2016).

It is possible that green certifications signal different quality, such as prestige. Griskevicius *et al.*, (2010) argue that patronizing green products can be construed as altruistic, and consumers may use green purchase as a means to signal "status." If so, hotels may obtain green certifications to differentiate themselves from their competitors. Mazzeo (2002) shows that firms enjoy a significant benefit by offering differentiated products. Competition in a market drives down prices, but a firm can be less affected by the competition when its products are differentiated. Hence, differentiation is the optimum product choice behavior. In the hotel industry, as long as consumers gain different levels of utility from diverse product types, a competing hotel can differentiate itself by offering green choice and charge a price higher than marginal cost in equilibrium without losing the whole market share. A green traveler may be inclined to forego the utility related to the higher price if he/she has a strong preference for a green stay or the associated differential quality. The distribution of travelers' preferences over product types offered by the hotels is important. If travelers' preferences are skewed in favor of a product type, the resulting price elasticity for a hotel offering the popular product type may be smaller, and vice versa. The relative product-space locations of competitors also affect the relevant price elasticity. Overall, hotels' profit-maximizing choices of product space locations will determine the underlying tradeoff between price and market share, in other words, their economic performance.

Although a number of earlier studies suggest travelers show a preference for green hotels, the financial implications of such findings are inconclusive (Han *et al.*, 2009; Han and Kim, 2010; Lee *et al.*, 2010; Manaktola and Jauhari, 2007). One reason for this inconclusiveness is the fact that "saying is one thing; doing is another," as pointed out by Bosson *et al.* (2004) and Pager and Quillian (2005). Walsman *et al.* (2014) report a RevPAR premium for the hotels with LEED certifications compared to the non-LEED hotels, but due to limitations in their data, they pointed out the need for further research in this subject.

An early study conducted by Chan and Lam (2002) points out the inadequacy of measures within the hotel industry in dealing with pollutants produced by electricity consumption. Since then, several international studies laid out the foundation for research on the subject. Rivera (2002) demonstrates that the hotels in Costa Rica experienced a price premium after the adoption of a voluntary environmental program. However, the author points out the study is limited in its ability to infer causation due to the use of cross-sectional data. Surveying 349 hotels in Poland and Sweden, Bohdanowicz (2006) reports an emergence of recognition for environmental protection needs. Tarí *et al.* (2010), through the analysis of variance and cluster analysis of 301 hotels in Spain, report that environmental practices influence hotels performance. Based on a survey of accommodation managers in Spain, Garay and Font (2012) suggest that CSR is mostly altruistically motivated, and environmental responsiveness is a part of it. However, they recognize competitiveness also plays some role in CSR initiatives. Rahman *et al.* (2012) show that chain hotels are more likely to embrace green initiatives compared to independent hotels. In another study, conducted using a sample of Greek hotels, Leonidou *et al.* (2013) show that the sufficiency of physical and financial resources determines green marketing strategies. As competition intensifies in the market, such strategies become stronger.

Many studies in the real estate sector investigate the operational and financial premiums of green buildings (Fuerst & McAllister, 2011; Zhang *et al.*, 2017). These studies examine buildings with green certifications like LEED or Energy Star. Some find green buildings enjoy a price premium, including evidence of heterogeneous price premiums in various value categories (Das and Wiley, 2014; Eichholtz, Kok, and Quigley, 2010; Robinson and McAllister, 2015). A few studies find that green buildings experience higher development and operating costs (Miller *et al.* 2010; Kok and Jennen, 2012; Nikodem and Fuers, 2013). Robinson *et al.* (2016), however, argue the significant high occupancy and rental rates must be the reason of price premiums in green buildings. Likewise, Das *et al.* (2011) show green buildings enjoy a notably higher rental rate (2.4%) during down markets, but during up markets, the rates drop significantly. Robinson and Reichert (2015) report that green certifications marginally affect appraisal values. Kok and Jennen (2012) show that buildings in the Netherlands with no energy-performance certifications experience 6.5% lower rental rates.

One common limitation of the previous studies is the presence of endogeneity bias, stemming from the self-selection of green certifications by businesses. Arguably, businesses may choose to obtain green labels because they expect to enjoy a price premium. It is possible that the unobserved factors, only known to business managers, underpin their expectation. In such cases, the price premium cannot be attributed to the green labels, but to the aptness of the business managers' decision to go green. On the contrary, the literature on the effect of green certifications on businesses other than office buildings is limited. Besides, office buildings and hotels operate in different settings. The closest study to this article conducted by Robinson *et al.* (2016) examined the financial impact of LEED and Energy Star certifications on hotel revenues. As the authors pointed out, their econometric techniques suffer endogeneity bias due to the unavailability of information regarding the exact timing when each hotel went green. Also, the study does not address the bias associated with self-selection of the hotels' green certifications. Due to the limitations of data, econometric techniques, and a limited number of studies in the existing literature, there seems to be a gap in understanding the effect of green labels on business performance, particularly in the hotel industry. This paper seeks to address the gap.

This paper is organized as follows. Section 2 provides some background information regarding the GreenLeaders program. Section 3 illustrates a theoretical model explaining how adopting green practices may be a strategy for product differentiation and its economic implications of green hotels. Section 4 describes the data collection procedure and the process of constructing a sample of hotels for this study. Section 5 then outlines and elaborates on the empirical specifications; and section 6 illustrates the results. Section 7 presents an analysis of the results; and finally, section 8 concludes.

2. GreenLeaders program

In 2013, TripAdvisor commenced the GreenLeaders program in partnership with U.S. Green Building Council's LEED Certification Program, the United Nations Environment Program, the U.S. Environmental Protection Agency's Energy Star program, and other sustainability experts (TripAdvisor, 2018). Under this program, hotels, bed and breakfasts (B&B), and specialty lodgings are awarded for their commitment to the environment and sustainability. The program is available to all hotels, B&Bs,

and specialty lodgings in the U.S., Canada, and some selected countries in the Europe. A hotel interested in obtaining a GreenLeaders badge can participate in the program free of charge but is required to participate in an online survey in order to determine its eligibility.[1] If qualified, the score on the survey determines an appropriate badge level, as shown below. All participating hotels must reapply every year to ensure their continued enrollment in the program and to keep their badges on the TripAdvisor page of their properties. I addition to the initial screening for determining eligibility, all participating hotels are subject to a set of audits conducted every year by independent sustainability organizations. A participating hotel in the GreenLeaders program receives one of the five types of badges (i.e., Platinum, Gold, Silver, Bronze, and GreenPartner) on its listing, a widget for its official website, and a printed certificate. On TripAdvisor.com, travelers can identify GreenLeaders hotels with different levels of badges in their locations of interest. Travelers can also see the full list of practices by clicking a property's GreenLeaders icon on its TripAdvisor page. The different types of GreenLeaders awards a property can receive are as follows:

- *Platinum*: 60 percent or greater score on the Green Practices survey.
- *Gold*: 50 percent score on the Green Practices survey.
- *Silver*: 40 percent score on the Green Practices survey.
- *Bronze*: Meets minimum requirements and achieves a 30% score on the Green Practices survey.
- *GreenPartner*: Meets minimum requirements.

[1]Click ***here*** to view the survey questionnaire.

Figure 1.7 shows a search result on TripAdvisor.com for the hotels, including GreenLeaders hotels, in San Francisco. As illustrated, one can identify a GreenLeaders hotel by its badge on the hotel image, next to its name. The figure includes four GreenLeaders hotels, including Phoenix Hotel, a Joie de Vivre hotel; Best Western Plus Americania, Carriage Inn, and The Good Hotel.

3. Theoretical framework: A product differentiation model for green hotels

In this section, I set up a simple version of product differentiation model (Hotelling, 1929; Dixit, 1979; Vives, 1984; Beath and Katsoulacos, 1991; Anderson *et al.*, 1992; Shy, 1995) to illustrate the economic implications of going green for a hotel. I assume hotels operate in a vertically differentiated market (illustrated in Figure 1.1) where all consumers have their hotels located at any point on the [0, 1] interval. There is a continuum of consumers uniformly distributed on the interval [0, 1]. G and H denote two hotels that are located at points g and h ($0 \leq h \leq g \leq 1$) from the origin, respectively. I also assume G represents a green (or a high quality) hotel that signifies higher quality (i.e., status, altruism, or any other quality) and H denotes a non-green (or a relatively lower quality) hotel. The utility of a consumer located at point n, $n \in [0, 1]$ and staying in hotel $i, i = G, H$ is defined by

$$U_n(i) \equiv \begin{cases} hn - p_H & i = H \\ gn - p_G & i = G \end{cases}$$

where hotel H and G charge the prices p_H and p_G, respectively.

I define a two period game, where hotels choose their locations in the first period, and then determine prices in the second period. Before defining the game, let us solve for the Nash-Bertrand equilibrium in prices, assuming fixed locations.

Let $\hat{n}$ denote a traveler who is indifferent to whether he or she chooses to stay in hotel G or H. Assuming that such a traveler exists, and that the traveler $\hat{n}$ intends to locate anywhere between the two hotels, that is $h \leq \hat{n} \leq g$, the intended location of the indifferent traveler is determined by

$$U_{\hat{n}}(H) = h\hat{n} - p_H = g\hat{n} - p_G = U_{\hat{n}}(G) \tag{1}$$

Thus, the utility of a traveler indexed by $\hat{n}$ from staying in hotel G equals his utility from staying in hotel H. As a result, based on the assumption $h \leq \hat{n} \leq g$, the number of travelers staying in hotel H is $\hat{n}$, whereas the number of travelers staying in in hotel G is (1- $\hat{n}$). Solving for $\hat{n}$ from equation (1) gives

$$\hat{n} = \frac{p_{G-}p_H}{g-h} \quad \text{and} \quad 1 - \hat{n} = 1 - \frac{p_G - p_H}{g-h}$$

Figure 1.2 illustrates how $\hat{n}$ is determined. The left side of Figure 1.2 shows the utility for a traveler intending to locate at any point $0 \leq n \leq 1$ when he stays in hotel G or H, assuming $p_G > p_H$. By definition, a traveler located at $\hat{n}$ derives the same utility from staying in hotel G as the utility from staying in hotel H. In addition, Figure 1.2 illustrates that all consumers located on $[0, \hat{n}]$ gain a higher utility from staying in hotel H than from staying in hotel G. Likewise, travelers located on [$\hat{n}$, 1] gain a higher utility from staying in hotel G (relatively higher quality) than from staying in hotel H.

It should be noted that I assume travelers cannot stay in both hotels, hotel G and H, at the same time. I also assume that travelers with a reservation utility of zero would not choose to stay in any hotel if they derive negative utilities. Hence, on the left side of Figure 1.2, all travelers on [0, m] will not stay in any hotel, reducing the market size for hotel H to the interval $[m, \hat{n}]$. It is also clear from the right-hand side of Figure 1.2 that all travelers choose to stay in hotel G when the price of the nongreen hotel (hotel H) is higher than the price of the green hotel (hotel G).

In the second period, for given locations of hotels, each hotel takes the price set by its competitor as given and determines its price to maximize its profit. Hotel G and H thus solves:

$$\max_{p_H} \pi_H(g, h, p_G, p_H) = p_H \hat{n} = p_H \left[\frac{p_G - p_H}{g - h}\right]$$

$$\max_{p_G} \pi_G(g, h, p_G, p_H) = p_G(1 - \hat{n}) = p_G \left[1 - \frac{p_G - p_H}{g - h}\right] \tag{2}$$

The quadruple $< g^e, h^e, p^e_G(g, h), p^e_H(g, h) >$ is said to be a vertically differentiated market equilibrium if, in the second period, for given locations of hotels (g and h), $p^e_G(g, h)$ and $p^e_H(g, h)$ represent a Nash equilibrium; and in the first period, given the second-period price functions of locations $p^e_G(g, h)$, $p^e_H(g, h)$, and $\hat{n}(p^e_G(g, h), p^e_H(g, h))$, (g^e, h^e) is a Nash equilibrium in location. This is also a subgame perfect equilibrium in which hotels choose their locations in the first stage after accounting for how their location choices will affect the equilibrium prices in the second period and, thereby, profit levels. In the second period, equilibrium actions of the hotels are functions (not scalars) of all the possible given locations of hotels. Solving equation 2, we get:

$$p^e_H(g, h) = \frac{g - h}{3} \quad \text{and} \quad p^e_G(g, h) = \frac{2(g - h)}{3} \tag{3}$$

Note that both the equilibrium prices surpass marginal cost. Equation 3 gives,

Proposition 1: *A green hotel, providing higher quality products (or services), charges a higher price even if the cost for the non-green hotel is the same as the cost of the green hotel.*

Substituting $p^e_H(g,h)$ and $p^e_G(g,h)$ from equation 3 into equation 2 gives,

$$\pi_H(g,h) = \frac{h-g}{9}$$

$$\pi_G(g,h) = \frac{4(h-g)}{9}$$

$\pi_H(g,h)$ and $\pi_G(g,h)$ above show that hotel H and G benefit more as they move further away from each other. This model can be further extended by allowing more than two hotels in the same market to show as more hotels choose to locate near hotel G, its ability to charge a higher price diminishes. Hence,

Proposition 2: *A green hotel's ability to charge a higher price diminishes as more green hotels enter the market and choose to locate nearby.*

4. Data

My data come from two primary sources: TripAdvisor and STR, Inc.[2] By crawling TripAdvisor.com, for 626 different cities in California, I construct a dataset with a cross-section of information on 5,157 hotels. This data set contains all observed hotel characteristics reported by the hotels on their TripAdvisor page. In order to determine when the participating hotels in the GreenLeaders program received their badges, which is not publicly available on TripAdvisor, I use a proprietary dataset that has been collected by personally contacting the TripAdvisor authority. The data contain badge levels and badge award dates for all participating hotels in the GreenLeaders program between June 2013 and March 2017. In California, 824 hotels participated in the GreenLeaders program as of April 2017. Figure 1.3 illustrates the

[2]STR, Inc. is a U.S. based market research company that tracks supply and demand data for multiple market sectors, including the global hotel industry. STR provides market share analysis for major hotel chains and brands in North America, Europe, Asia Pacific, Middle East, and Africa.

numbers of GreenLeaders hotels by their badge types for the 2013-2016 period in the U.S.

From STR, Inc., I obtain a sample of hotels in California that report their performance data, particularly occupancy rate, average daily rate (ADR), and revenue per available room (RevPAR), to STR in daily frequency. STR defines occupancy as the percentage of available rooms sold during a specified period. Daily occupancy rate is calculated by dividing the number of rooms sold by the total number of rooms available on a given day. ADR is a measure of the average price paid for rooms sold, calculated by dividing total room revenue by the number of rooms sold. Lastly, RevPAR is calculated by dividing total room revenue by the total number of available rooms.[3] RevPAR differs from ADR because RevPAR is affected by the number of unoccupied rooms, while ADR shows only the average price of the sold rooms. Of the 5,157 hotels from TripAdvisor's data, STR receives daily performance reports from 3,267 hotels. Because different hotels started reporting to STR from different dates, not all of the 3,267 hotels have performance data for the same length of duration in STR's data set. Besides, in STR's data, a significant number of hotels have missing observations for several months. As a result, after merging TripAdvisor's data with that of STR, I construct a sample of hotels for which there are no missing observations between the period of February 2011 and June 2017, providing a strongly balanced panel data. I merge the two datasets based on the hotels' addresses and names. At this stage, the sample contains 2,446 hotels including 517 GreenLeaders hotels. Next, I construct a number of clusters of hotels by imposing the following condition. I keep a cluster containing hotels in the same zip code if it includes at least one green and one nongreen hotel. I drop the clusters and the hotels within each of them that do not meet the above condition. Grouping hotels by their zip codes has two advan-

[3] https://www.strglobal.com/resources/glossary

tages. First, it allows me to construct a larger sample of hotels, which is not possible if clustered by imposing a distance restriction, say, of one or two miles radius. Second, clustering allows me to control for unobserved hotel trends related to locations. However, one concern of this approach is that some zip codes may be considerably larger than others. Figure 1.8 displays a whisker plot suggesting distance should not be a concern for my sample of hotels because, in each cluster, all nongreen hotels are located within approximately three miles away from the green hotels, and 75% of the nongreen hotels are located within a little above one mile distance from their green counterparts. The average distance between green and nongreen hotels in the final sample is 0.88 mile.

My final data set includes 865 hotels of which 342 are green hotels, including 16 Platinums, 37 Golds, 106 Silvers, 99 Bronzes, and 84 GreenPartners, from 98 cities and 145 zip codes. Figure 1.4 illustrates locations of the hotels included in the final sample. The sample of hotels in the final data represents 16.8% of all hotels and 41.5% of the green hotels in California. The resulting data set contains strongly balanced panel data including daily occupancy rate, daily ADR, and daily RevPAR from February 01, 2011 through June 29, 2017 for each of the 865 hotels. Table 1.1 reports the summary statistics of all dependent and independent variables.

4.1 Estimation sample

A primary principle of any experimental design is that the treated and control units are chosen randomly. This poses a challenge in this study as the hotels' choice to participate in the GreenLeaders program is not randomized. Instead, the hotels endogenously decide to participate in the program. The participating hotels (treated group) might be substantially different from the nonparticipating hotels (control group). I, therefore, limit the analysis within a sample of hotels in which the participating and nonparticipating hotels are similar to each other based on their observable characteristics. I assume if the hotels' observable characteristics are not different from each other, their performance (i.e., Occupancy, ADR, and RevPAR) should be similar. As a result, it does not matter which hotel receives a GreenLeaders badge. Hence, a badge awarded to a hotel would assumably mimic a randomized process. One limitation of the assumption is there might be unobserved hotel characteristics that play a role in the hotels' decision to participate in the program. To address the concern, I use a different empirical specification described in section 5.3. In this section, I use a propensity score matching method, particularly nearest neighbor matching, to select a comparable control hotel for each treated hotel (Becker & Hvide, 2017; Zhang *et al.*, 2017; Ichino *et al.*, 2017). The rationale behind using propensity score matched data is to address the bias arising from self-selection of the participating hotels in the GreenLeaders program. The propensity score refers to the probability of receiving a treatment, which, in this case, is receiving a GreenLeaders badge conditional on pre-treatment characteristics. The idea is to match treated and control units based on their *ex-ante* likelihood of receiving treatment predicted by their pre-treatment characteristics (Rosenbaum and Rubin, 1983). The hotel characteristics shown in Table 1.2 are the pre-treatment characteristics used in the matching process.

I estimate a probit model of participation in the GreenLeaders program on hotel characteristics to estimate propensity scores for all 865 hotels. Next, I use a nearest-neighbor matching method (without replacement) based on the estimated propensity scores to obtain a matched pair of one treated unit and one control unit. To ensure good matches, I impose a caliper of 0.05 so that any treated unit that does not have a control unit within 0.05 of the propensity score of the treated unit is eliminated. I also impose exact matching on the zip-codes of the hotels to control for unobserved time-variant factors related to locations that may affect both the green and non-green hotels similarly. The matching process discards 89 green hotels and 270 nongreen hotels, leaving in total 506 (253 matched pairs) green and non-green hotels.

To check the effectiveness of the probit model in reducing differences between the treated and control units, I estimate the median absolute standardized bias (MASB), as shown in equation 4 below, from Rosenbaum and Rubin (1985):

$$\text{MASB} = \frac{100(\bar{x}_{i1} - \bar{x}_{i0})}{\sqrt{\frac{1}{2}(s_{i1}^2 + s_{i0}^2)}} \tag{4}$$

where $\bar{x}_{i1}$ and $\bar{x}_{i0}$ denote means of covariate x_i in treated and control units, respectively. s_{i1}^2 and s_{i0}^2 denote sample variances of covariate x_i in treated and control units, respectively. Before matching, the MASB estimate was 28.54, which was reduced to 3.27 after matching. According to Rosenbaum and Rubin (1985), an MASB estimate of 20 is "large." It is, therefore, safe to note that the matching procedure has significantly reduced differences between the treated and control groups.

Table 1.2 and Figure 1.5 illustrate how well the characteristics and the propensity scores of control units match that of treated units, respectively, after matching. In Table 1.2, a comparison of the hotel characteristics between "Before Matching" and "After Matching" shows that, on average, differences between the treated and control units are reduced after constructing matched pairs with propensity scores. For instance, before matching, 92.98% of the participating hotels and 75.1% of the nonparticipating hotels had multilingual staffs; but after matching, the difference was reduced significantly. Figure 1.5 illustrates distributions of the propensity scores and suggests that the matching produces a better control group by reducing differences between treated and control units in terms of their estimated propensity scores. Hence, for each treated unit, the matching produces a control unit with similar pretreatment characteristics. Figure 1.6 shows the differences in occupancy rates, ADR, and RevPAR between treated and control units have reduced after the matching procedure. The process also ensures both the treated and control units in a matching pair are located in the same zip-code.

5. Empirical specification

In order to estimate the effects of participation in the GreenLeaders program, I estimate three different models: fixed effects regression, difference-in-differences (DID) regression, and generalized synthetic control (GSC) model. The fixed effects regression outlined in section 5.1 is similar to what Zhang et al. (2017) used as an empirical specification for studying the impact of green certifications on hotel performance in Beijing, China. The estimated results from the model provides a basis for comparison with prior studies. Next, I estimate difference-in-differences regressions that include the additional information regarding timings of hotels' green certifications, an important variable not incorporated in previous studies. The third empirical specification, the GSC model, complements the results produced by the DID model and provides robustness checks.

5.1 Fixed effects regression

Using the propensity-score-matched data, I estimate fixed effects regressions as a baseline model to examine the effects of green certifications. In the regression, I control for observed hotel characteristics, such as whether a hotel offers free breakfast, airport transportation, free parking and various other types of services. In order to control for unobserved characteristics related to locations and market condition, I follow Eichholtz et al. (2010) and Zhang et al. (2017) by including group-fixed effects, which take advantage of the homogeneity within each matched pair (within a 0.88-mile radius on average). Equation 5 specifies the model:

$$Y_{itg} = \beta_0 + \beta_1 Green_{ig} + +\gamma X_{ig} + \sum_{t=2}^{T} \beta_t month_t + \sum_{k=2}^{K} \beta_k year_k + \sum_{d=2}^{D} \beta_d day_d + \epsilon_{itg} \quad (5)$$

where Y_{itg} denotes occupancy rate, or log-transformed ADR, or log-transformed RevPAR of hotel i within group g on date t. *Green* is a dummy variable indicating participation in the GreenLeaders program; X_{ig} denotes observed hotel characteristics; *month*, *year,* and *day* denote dummy variables for month, year, and day, respectively; and ϵ_{itg} denotes robust standard errors, clustered at the group level.

5.2 Difference-in-differences

5.2.1. Estimation

In order to estimate the effect of participation on the three dependent variables (i.e., *Occupancy*, *ADR* and *RevPAR*), I estimate the regression specified by equation 6 with group fixed effects. This model differs from equation 5 as it includes the additional variable regarding certification dates for the green hotels. I estimate the regression for each of the three dependent variables.

$$Y_{itg} = \beta_0 + \beta_1 Treated_{ig} + \beta_2 Post_{itg} + \beta_3 Post_{itg} * Treated_{ig} + \gamma X_{ig} + \sum_{t=2}^{T} \beta_t month_t$$
$$+ \sum_{k=2}^{K} \beta_k year_k + \sum_{d=2}^{D} \beta_d day_d + \epsilon_{itg} \quad (6)$$

In equation 6, the outcome variable Y_{itg} denotes *Occupancy rate,* or log-transformed *ADR,* or log-transformed *RevPAR* for hotel i on date t and in group g. The variable *Post* takes a value of 1 on and after hotel i receives a GreenLeaders badge. *Treated* is a dummy variable if the observation participates in the GreenLeaders program, γ is the coefficient of time-invariant hotel characteristic X; *month*, *year,* and *day* denote dummy variables for month, year, and day, respectively; and ϵ denotes residuals. The coefficient of interest β_3 indicates the effect of participation on *Occupancy*, *ADR*, and *RevPAR*. In order to analyze further, I extend the model in several ways. I examine potential heterogeneous effects of participation in the GreenLeaders program across badges (i.e., *Platinum, Gold, Silver, Bronze,* and *GreenPartner*) and types of locations (i.e., interstate, resort, small metro/town, suburban, and urban).

5.3. Generalized synthetic control

There may be concerns as to the estimated participation effects using the DID model with propensity score matched data. The presence of unobserved time-varying confounders can bias the DID estimates. For example, some hotels might have improved their quality over time and eventually opted in for green certifications to signal better quality. Unobserved and time-variant changes of such nature, if not taken into account, can confound the DID estimates. Because propensity score matching only reduces observable hotel differences, unobservable and potentially time-varying hotel characteristics are left unaddressed in the DID model. To address the concern and to complement the DID results, I use a generalized synthetic control (GSC) method proposed by Xu (2017). The model allows estimation of the treatment effect on the treated for multiple treated groups with multiple treated periods. In principle, this model is analogous to the synthetic control method proposed by Abadie *et al.* (2010) as it essentially reweights the pretreatment treated outcomes for benchmarking while choosing weights for control units, and utilizes cross-sectional correlations between treated and control units in order to predict counterfactuals. However, unlike the synthetic control method, this method uses a dimension reduction procedure before reweighting so that the vectors to be reweighted on are smoothed across control units.

5.3.1. Model Framework

To illustrate the model framework, I adopt the same notations as Xu (2017). Let Y_{it} denote the outcome of interest for unit (i.e., hotel) i at time t. $\mathcal{T}$ and $\mathcal{C}$ denote the sets of units in treated and control groups, respectively. The total number of units is represented by $N = N_{tr} + N_{co}$ in which N_{tr} and N_{co} indicate the numbers of treated and control units, respectively. All units are observed for T periods, from time 1 to time T. Let $T_{0,i}$ denote the number of pre-treatment periods for unit i that is first exposed to the treatment (i.e., enters the GreenLeaders program) at time $(T_{0,i} + 1)$ and later observed for $q_i = T - T_{0,i}$ periods. Over the observed time span, control units are never exposed to the treatment. For notational convenience, let us assume that all the treated units are first exposed to the treatment at the same time, i.e., $T_{0,i} = T_0$ and $q_i = q$; variable treatment periods can also be accommodated. Firstly, the model assumes Y_{it} is given by a linear factor model:

$$Y_{it} = \delta_{it} D_{it} + x'_{it}\beta + \lambda'_i f_t + \varepsilon_{it},$$

where D_{it} denotes the treatment indicator that takes a value of 1 if unit i has been exposed to the treatment prior to time t, or else 0 (i.e., $D_{it} = 1$ when$i \in \mathcal{T}$ and $t > \mathcal{T}_0$, or else $D_{it} = 0$). δ_{it} denotes the heterogeneous treatment effect on unit i at time t; x_{it} represents the observed covariate(s), β denotes a vector of unknown parameters, f_t denotes a vector of unobserved common factors, λ_i denotes a vector of unknown factor loadings, and ε_{it} denotes unobserved idiosyncratic error terms for unit i at time t and has a mean value of zero.[4] Let $Y_{it}(1)$ and $Y_{it}(0)$ be the potential

[4] In interactive fixed effects (IFE) model proposed by Bai (2009), the time varying coefficients are referred to as common factors or latent factors and the unit-specific intercepts are known as factor

outcomes for individual i at time t when $D_{it} = 1$ or $D_{it} = 0$, respectively. Hence, we obtain $Y_{it}(0) = x'_{it}\beta + \lambda'_i f_t + \varepsilon_{it}$, and $Y_{it}(1) = \delta_{it} + x'_{it}\beta + \lambda'_i f_t + \varepsilon_{it}$. We can derive the individual treatment effect on the treated unit i at time t as $\delta_{it} = Y_{it}(1) - Y_{it}(0)$ for $i \in \mathcal{T}$, and $t > T_0$. The key estimate of interest, average treatment effect on the treated (ATT) at time t (when $t > T_0$):

$$ATT_{t,t>T_0} = \frac{1}{N_{tr}}\sum_{i\in\mathcal{T}}[Y_{it}(1) - Y_{it}(0)] = \frac{1}{N_{tr}}\sum_{i\in\mathcal{T}}\delta_{it}$$

5.3.2. Estimation strategy

In the first stage, a GSC estimator is estimated for each treated unit's treatment effect. This is, in essence, based on Bai (2009)'s out-of-sample prediction method. For the treatment effect on treated unit i at time period t, the GSC estimator is given by the difference between an actual outcome and its estimated counterfactual as follows: $\hat{\delta}_{it} = Y_{it}(1) - \hat{Y}_{it}(0)$, where $\hat{Y}_{it}(0)$ is estimated in three steps. The first step involves estimation of IFE model using only control units information to obtain $\hat{F}$, $\hat{\Lambda}_{co}$, and $\hat{\beta}$:

$$\text{Step 1: } (\hat{F}, \hat{\Lambda}_{co}, \hat{\beta}) = \underset{\tilde{\beta},\tilde{F},\tilde{\Lambda}_{co}}{\operatorname{argmin}}\sum_{i\in C}(Y_i - X_i\tilde{\beta} - \tilde{F}\tilde{\lambda}_i)'(Y_i - X_i\tilde{\beta} - \tilde{F}\tilde{\lambda}_i)$$

$$s.t.\ \tilde{F}'\tilde{F}/T = I_r \text{ and } \tilde{\Lambda}'_{co}\tilde{\Lambda}_{co} = \text{diagonal}.$$

The second step involves estimation of factor loadings for each treated unit by minimizing the mean squared error of the predicted treated outcome in pre-treatment periods:

loadings.

$$\text{step 2: } \hat{\lambda}_i = \underset{\hat{\lambda}_i}{\text{argmin}}\ (Y_i^0 - X_i^0\hat{\beta} - \hat{F}^0\tilde{\lambda}_i)'(Y_i^0 - X_i^0\hat{\beta} - \hat{F}^0\tilde{\lambda}_i)$$

$$= (\hat{F^0}'\hat{F^0})^{-1}\hat{F^0}'(Y_i^0 - X_i^0\hat{\beta}),\ i \in \mathcal{T},$$

where $\hat{\beta}$ and $\hat{F}^0$ are estimated in step 1, and the superscript "0" indicates pre-treatment period. In the next step, treated counterfactuals are estimated based on $\hat{\beta}, \hat{F}$, and $\hat{\lambda}_i$.

$$\text{Step 3: } \hat{Y}_{it}(0) = x'_{it}\hat{\beta} + \hat{\lambda}'_i\hat{f}_t\ ,\ i \in \mathcal{T}\ ,\ t > \mathcal{T}$$

Hence, the estimator for ATT_t *is:*

$$\widehat{ATT_t} = (1/N_{tr})\sum_{i\in\mathcal{T}}[Y_{it}(1) - \hat{Y}_{it}(0)]\ \textit{for}\ t > T_0$$

Before estimating causal effect, a cross-validation procedure is used - in case of limited knowledge on the number of factors to be included - to select the right model. This procedure relies on both the treated and control group information in the pre-treatment periods.[5] The idea is to hold back a small portion of data (i.e., treated group's one pre-treatment period) and utilize the remaining data in order to predict the held-back data. The next step is to then select the model that makes the most accurate predictions on average.

To obtain uncertainty estimates of the estimator, GSC uses a parametric bootstrap procedure. Conditional on observed covariates, unobserved factors, and factor loadings, the model provides uncertainty estimates using a parametric bootstrap procedure by resampling the residuals. The goal is to estimate the conditional variance

[5]See Xu (2017) for further details on the cross-validation procedure.

of ATT estimator (i.e., $\mathrm{VAR}_{\varepsilon}(\widehat{ATT}|D, X, \Lambda, F)$. The residuals, ε_i, represent the only random variable that is not being conditioned on because they are assumed to be independent of the treatment assignment, factors, factor loadings, and observed covariates. The model treats ε_i as measurement errors that cannot be explained, but are unrelated to the treatment assignments.

The parametric bootstrap procedure simulates treated counterfactuals and control units based on the following resampling procedure:

$$\tilde{Y}_i(0) = X_i\hat{\beta} + \hat{F}\hat{\lambda}_i + \tilde{\varepsilon}_i, \forall i \in \mathcal{C};$$
$$\tilde{Y}_i(0) = X_i\hat{\beta} + \hat{F}\hat{\lambda}_i + \tilde{\varepsilon}_i^p, \forall i \in \mathcal{T},$$

where $\tilde{Y}_i(0)$ denotes a vector of simulated outcomes in the absence of treatment; $X_i\hat{\beta}+\hat{F}\hat{\lambda}_i$ provides the estimated conditional mean; and $\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_i^p$ represent resampled residuals for unit i, which either belongs to treated or control group. As $\hat{F}$ and $\hat{\beta}$ are estimated based on control group data, $X_i\hat{\beta} + \hat{F}\hat{\lambda}_i$ fits $X_i\beta + F\lambda_i$ better for control units than treated units. As a result, $\tilde{\varepsilon}_i^p$ has a greater variance compared to $\tilde{\varepsilon}_i$. Thus $\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_i^p$ are drawn from disparate empirical distributions. $\tilde{\varepsilon}_i$ is drawn from the empirical distribution of the residuals of IFE model, whereas $\tilde{\varepsilon}_i^p$ is the prediction error of IFE model for treated counterfactuals. Incorporating control group information, GSC uses a cross-validation procedure to simulate $\tilde{\varepsilon}_i^p$. The model is based on the following assumptions: (1) the residuals are independent and homoskedastic across space; and (2) the treated and control groups follow the same factor model (Efron 2012; Xu 2017).

6. Results

6.1. Fixed effects regression estimates

Table 1.3 reports the regression results based on Equation 5. The estimates in column 1 suggest green certifications have no impact on *Occupancy*, as the coefficient for *Green* is not statistically significant. Column (2) results, however, suggest a 3% (exp(0.030)-1$\approx$ 1.030) price premium in the green hotels, and this is statistically significant at 5% level. Finally, in column (3), the coefficient of *Green* is significant at 1% level, suggesting a 5.3% revenue premium for green certifications. Column 4, 5, and 6 report heterogeneous impacts of green certifications across badge types. The estimates indicate only the hotels carrying GreenPartner badges see an 8.2% and a 12.2% increase in their ADR and RevPAR, respectively; and the occupancy rates are again unaffected. Overall, the results at this stage indicate green hotels, on average, command a price and a revenue premium, with no statistically significant impact on their occupancy rates. I investigate further below to examine whether the results remain robust when the empirical specification includes the key information regarding each hotel's green certification date.

6.2. DID and GSC estimates

Table 1.4 presents the DID and the GSC estimates in panel A and B, respectively. Each column shows results for different regressions: column 1, 2, and 3 present results for the regressions with dependent variables *Occupancy*, *ADR,* and *RevPAR*, respectively. The coefficient of interest (β_3) for *Post*Treated* is significant at 10% level for the dependent variable *ADR*, indicating the green hotels experience an increase of 1.8% in their *ADR* relative to the nongreen hotels. Participation does not appear to have any statistically significant effect on *Occupancy* and *RevPAR*.

Some of the other interesting results from the DID model include the hotels with *Babysitting* services charge a 46.5% (exp(0.382)≈1.465) higher price, resulting in a 39% (exp(0.329)≈1.390) higher *RevPAR* relative to their counterparts. *Babysitting* services are perhaps highly correlated with unobserved hotel qualities that help them enjoy a substantial price and revenue premium. Likewise, for the hotels with *Business centers,* there is a 12.2% and 16.8% increase in *ADR* and *RevPAR,* respectively*;* and the hotels with *meeting rooms* also see similar increases in their *ADR* and *RevPAR*. Lastly, hotels with *Breakfast included* services experience a 2.8% percent higher occupancy rate and a 7.3% higher *RevPAR,* with no statistically significant change in their ADR. On the contrary, based on the estimates from the GSC model, presented under panel B, I find participation has no effects on the participating hotels' *Occupancy*, *ADR,* and *RevPAR*.

My presented results from the DID and GSC model provide average treatment effects for all hotels. Perhaps the treatment effects differ for different types of GreenLeaders badges, locations, or other attributes. In the subsequent part, I, therefore, estimate the effects of participation in the GreenLeaders program in several ways. At first, I investigate the potential heterogeneous effects of participation in the GreenLeaders program across hotels' badge types. Next, I examine whether location plays any role in the effect of participation.

To investigate potential heterogeneous effects across different badges of the GreenLeaders hotels, I estimate both the DID and GSC models. In Table 1.5, column 1 to 6 report coefficients of the participation effects by badge types. Column 1 to 3 report results from the DID model, and column 4 to 6 report results from the GSC model. Column 7, 8, and 9 report numbers of the treated units, control units, and total observations for the GSC estimates. Using the DID model, I find only the hotels carrying *GreenPartner* badges have statistically significant effects on their *ADR* and

RevPAR. However, *Occupancy* of the participating hotels is unaffected. Based on the GSC model, however, I find none of the badges have statistically significant effects on *Occupancy, ADR,* and *RevPAR* of the participating hotels.

6.3. Participation effects across location types

All the hotels in my data set can be categorized into six different types of locations, such as resort, small/metro town, airport, suburban, urban, and interstate. I examine if participation in the GreenLeaders program affects the participating hotels differently across location types. Table 1.8 illustrates how the location segments are defined. Table 1.6 reports results produced by the DID and GSC models, under panel A and B, respectively, by location types. Panel A illustrates results for the three dependent variables, *Occupancy*, *ADR*, and *RevPAR*. Each of the variables has been regressed separately by the six location types. The results indicate the effect of participation in the GreenLeaders program varies depending on a hotel's location type. Although *Occupancy* rates of the hotels are not affected, participating hotels located in *Resort* and *Small/metro town* do see an average increase of 7.8% and 4.3% in their *ADR,* respectively. Because the *Occupancy* rate is unaffected, an increase in *ADR* should increase a hotel's *RevPAR*, and this is what I find. Looking at the coefficients of RevPAR, we see the hotels within resorts and small towns experience a 7.4% and 4.6% increase, respectively. None of the participating hotels in other location types has any statistically significant change in their *Occupancy, ADR,* and *RevPAR.*

The results based on the GSC model, reported under panel B of Table 1.6, show similar effects of participation as found using the DID model. Based on this model, the participating hotels in *Resort* and *Small/metro town* experience a 6.3% and a 3.7% increase in their ADR, respectively, which results in a 6.5% and a 3.5% increase in their RevPAR, respectively. The *Occupancy* rates of all hotels are unaffected. Like the DID estimates, the GSC estimates show no effect of participation for hotels in other location types. Overall, the results presented in Table 1.6 suggest locations play an essential role in the effects of participation in the GreenLeaders program.

7. Discussion

Why do we see a price premium in some GreenLeaders hotels? Proposition 1 from the theoretical framework in section 3 explains why and how green hotels could potentially differentiate themselves and charge a higher price. In essence, by differentiating, a hotel does not have to compete as directly with its rivals. Because of less competition, a green hotel can then command a price premium without any significant loss of market share, given that there is sufficient demand for its product in the market. It is important to note here that the increase in ADR of the GreenLeaders hotels does not necessarily mean the hotels charge a higher price, although it is possible. Because ADR, by definition, refers to the average price of the rooms sold, an alternative explanation may be the participating hotels within *Resort* and *Small/metro town* are able to sell more premium rooms. The price premium in the participating hotels have two possible implications: first, the participating hotels do not attract any new segment of customers but are able to charge a higher price after going green. However, the basic economic principles of supply and demand suggest an increase in price should reduce the number of quantity demanded. This means the price premium in the participating hotels should cause a drop in their occupancy rate, which I do not find to be the case. Second, the participating hotels are able to differentiate themselves in a way that they can draw a new segment of customers who are less price sensitive and, hence, are likely more interested in premium rooms. In this case, any decrease in occupancy rate due to the price premium may be compensated by the new segment of less price sensitive customers. As a result, in the participating hotels, overall occupancy rates do not change significantly. Table 1.6 results lend credence to the second implication.

A survey conducted by TripAdvisor reports almost 25% of Americans are consciously trying to make eco-friendly choices when it comes to their hotel stays (Harrison, 2014). Despite prior anecdotal evidence indicating an increasing demand for green choices among travelers, Table 1.6 results indicate the green hotels only in resorts and small towns get the economic benefits of going green. Naturally, the question is then why hotels in the other location types are not impacted. Proposition 2 in the theoretical framework answers the question. According to Proposition 2, a green hotel's ability to charge a higher price diminishes as more green hotels enter the market and choose to locate nearby. Looking at the GreenLeaders hotels by their location types, I find the proportion of GreenLeaders hotels is significantly high in big or popular cities compared to small or less popular cities. Further investigation into TripAdvisor's city-popularity-rank reveals most of the hotels from *Resorts* and *Small/metro Towns* in my data are from less popular cities. Besides, unlike hotels in small cities, hotels in big or more popular cities engage in various marketing and promotional activities in order to differentiate themselves and stay ahead of their competitors (Sharkey 2013). It is possible that these marketing and promotional activities - which are less intense in small cities - distort hotel choice decisions of big-city travelers who would otherwise patronize a green hotel without hesitation. Hotel prices also vary considerably depending on the popularity of a city. For instance, a Courtyard by Marriott standard room in Tampa costs approximately \$109 on a regular weekend, whereas in a relatively more popular city, such as the New York City, a similar room on the same weekend may cost as much as \$409.[6] Consequently, a customer who is relatively less price sensitive in a small (or less expensive) city

[6] Based on search results on TripAdvisor.com on December 10, 2017.

may become very price sensitive in a big (or more expensive) city for having to pay more for a green hotel. In other words, travelers' demand for green hotels is likely more elastic at high prices, because of which green hotels in big cities do not see any statistically significant impact on their *Occupancy*, *ADR,* and *RevPAR*.

Furthermore, a traveler's average length of stay can also be an important factor in explaining my results. Because small-town and resort hotels are mostly driven by leisure travelers, who tend to stay longer, their hotel choice decisions are likely more conscious and careful. As a result, they could be more interested in green hotels. On the other hand, big-city, airport, interstate, and suburban hotels attract business and other kinds of travelers whose average length of stay is relatively shorter than leisure travelers. It is also possible that the majority of these travelers are under strict time constraints, making it costly for them to search and stay in green hotels. Besides, a lot of business travelers have to follow their employers' travel policy for reimbursement of the travel expenses. Consequently, a business traveler may not be able to stay in a green hotel of his or her choice.

The results from Table 1.7 support my findings from Table 1.6. Table 1.7 reports the coefficients of $Post_{itz} * Treated_{iz}$ from the DID model specified by equation 6. Each column between column 1 and column 7 report DID estimates based on regressions using different samples. I estimate the DID model using the matched data by incrementally excluding popular cities (thereby, the hotels located within) from the top of the city-popularity-rank by TripAdvisor. Column 1 excludes no cities, and Column 7 excludes top 60 of the most popular cities. One interesting finding in Table 1.7 is as we move from column 1 to 7 along the ADR and RevPAR rows, we see the coefficients become gradually larger, although not all of them are statistically significant. This phenomenon is absent when *Occupancy* is the dependent variable. For *ADR*, I find statistically significant results when no cities and top 50 or greater number of most popular cities are excluded. However, for RevPAR, I find statistically significant participation effects when 50 or greater number of most popular cities are excluded. These results support my earlier findings in Table 1.6 and imply location plays an important role in the impact of participation in the GreenLeaders program. My findings suggest hotels in less popular cities enjoy the most benefit from adopting green practices.

8. Conclusion

Overall, this paper investigates the effect of participation in TripAdvisor's GreenLeaders program by studying the hotels in California. In particular, using multiple empirical strategies, including difference-in-differences and generalized synthetic control, I have examined whether participation in the GreenLeaders program has any effect on a hotel's occupancy rates, average daily rates (ADR) and revenue per available room (RevPAR). My findings show that the effects of participation depend on the location of a hotel. Based on the full sample of hotels, on average, participation in the GreenLeaders program does not affect the hotels' performance. However, analyses on the hotels by their location types reveal that the hotels located in resorts and small/metro cities see increases in their ADR and RevPAR. Further analysis based on TripAdvisor's city-popularity-rank reveals that hotels in less popular cities get the most benefit from participating in the GreenLeaders program. This supports the participation effects found in hotels within *Resort* and *Small/metro town* locations because all the hotels in my data from within these two types of location are predominantly located in less popular cities.

I argue the degree of competition across location types may explain the results. Green hotels can signal better quality, higher prestige, altruism, and so forth. Hotels may go green in order to differentiate from competitors. But the effect of differentiation depends on how closely other hotels are located in the product space. In big cities, competition is intense. As more hotels try to differentiate themselves to stay ahead of their competitors, they end up locating close to each other in the product space, undercutting each other's market share. Conversely, in less popular cities, due to relatively less competition, GreenLeaders hotels can differentiate themselves sufficiently to have a statistically significant effect on their performance. Besides, price sensitivity of a customer could play an important role here. A traveler who is relatively less price sensitive in a small town may be highly price sensitive in a popular city because of the substantial price differences between the two locations. For a traveler, when the prices are too high, the utility gain from staying in a green hotel may be much less than the disutility from paying the associated high price premium in a popular city. As a result, a green hotel in a popular city may be a less desirable option for the traveler. The results of this study thus point out to the need for asking when going green pays off instead of whether going green pays off.

One limitation of this paper is the absence of analysis into how prices and number of online bookings in TripAdvisor.com changed for the GreenLeaders hotels after receiving their badges. As travelers can conveniently search for green hotels on TripAdvisor.com, an analysis of the price and online booking data directly from the website could provide more accurate results on the effect of participation. Due to unavailability of online booking and price data, such analysis was not possible. Also, a caveat for explaining the results of this study is that the increase in ADR and RevPAR within green hotels does not imply an increase in profitability. This study could not estimate the effect of participation on hotels' profitability due to unavailability of cost data. Although some hotels might have undergone operational changes requiring additional initial investment for becoming green, arguably, their lower operating costs from, say, energy savings could compensate for the initial investment and eventually increase profitability. In the GreenLeaders program, to meet the minimum requirements for obtaining a GreenPartner badge (the lowest level badge), a hotel must demonstrate initiatives like linen and bath towel reuse program. Such practices can save a significant amount of cost associated with energy, water, detergent, labor, and linen or towel replacement (Werntz, 2015). Yet, without cost data, it is impossible to objectively determine the effect of participation on the hotels' profitability. On the contrary, in addition to TripAdvisor's GreenLeaders program, there are a number of other green certification programs that evolved over the past few years. This study does not take into account if a participating hotel has other green certifications. It is also possible that some hotels with no GreenLeaders badge have certifications from other programs. However, with more than 11,000 participants from around the world and almost 6,000 hotels from within the U.S., TripAdvisor's GreenLeaders program is claimed to be the largest green certification program in the hotel industry (Hasek, 2016). Also, the program's collaboration with globally reputed organizations, such as LEED, Energy Star, and UNEP, lends credibility to the authenticity of the program. Nonetheless, the limitations of this study offer opportunities for future research.

This study makes several contributions to the literature of green certifications and business performance. This study corrects for the endogeneity bias arising from businesses' self-selection for green certifications, and estimates causal effects of green certifications on the performance of green hotels. Alongside offering statistical evidence on the role of locations in gaining economic benefits from green certifications, this study presents different perspectives on how and when going green could pay off. For businesses, this paper shows economic benefits of going green and provides managerial insights into when going green could work as a strategy to differentiate in a competitive market.

REFERENCES

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program . *Journal of the American Statistical Association*, 105(490), 493–505.

Anderson, S. P., Palma, A. d., & Thisse, J.-F. (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge, Mass.: MIT Press.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279.

Beath, J., & Katsoulacos, Y. (1991). *The Economic Theory of Product Differentiation*. Cambridge: Cambridge University Press.

Becker, S. O., & Hvide, H. K. (2017). *Do Entrepreneurs Matter?* Retrieved December 20, 2017, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2199645

Bohdanowicz, P. (2006). Environmentalawareness and initiatives in the Swedish and Polish hotel industries-survey results. *International Journal of Hospitality Management*, 25 (4), 662 – 682.

Bosson, J. K., Haymovitz, E. L., & Pine, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, 40(2), 247–255.

Chan, W. W., & Lam, J. C. (2002). Prediction of pollutant emission through electricity consumption by the hotel industry in Hong Kong . *International Journal of Hospitality Management*, 21(4), 381–391.

Das, P., & Wiley, J. (2014). Determinants of premia for energy efficient design in the office market. *Journal of Property Research*, 31(1), 64–86.

Das, P., Tidwell, A., & Ziobrowski, A. (2011). Dynamics of green rentals over market cycles: Evidence from commercial office properties in San Francisco and Washington, DC. *Journal of Sustainable Real Estate*, 3(1), 1-22.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, 67(3), 297-308.

Efron, B. (2012). The Estimation of Prediction Error: Covariance Penalties and Cross-Validation. *Journal of the American Statistical Association*, 99(467), 619–632.

Eichholtz, P., Kok, N., & Quigley, J. M. (2010). Doing well by doing good? Green office buildings. *American Economic Review*, 100(5), 2492–2509.

Fernie, J., Sparks, L., & McKinnon, A. C. (2010). Retail logistics in the UK: past, present and future. *International Journal of Retail & Distribution Management*, 38 (11/12), 894-914.

Fuerst, F., & McAllister, P. (2011). Green Noise or Green Value? Measuring the Effects of Environmental Certification on Office Values. *Real Estate Economics*, 39(1), 45–69.

Garay, L., & Font, X. (2012). Doing good to do well? Corporate social responsibility reasons, practices and impacts in small and medium accommodation enterprises. *International Journal of Hospitality Management*, 31(2), 329 – 337.

Griskevicius, V., Tybur, J. M., & Bergh, B. V. (2010). Going Green to Be Seen: Status, Reputation, and Conspicuous Conservation. *Journal of Personality and Social Psychology*, 98(3), 392-404. DOI: 10.1037/a0017346.

Han, H., & Kim, Y. (2010). An investigation of green hotel customers' decision formation: Developing an extended model of the theory of planned behavior. *International Journal of Hospitality Management*, 29(4), 659 – 668.

Han, H., Hsu, L. J., & Lee, J. (2009). Empirical investigation of the roles of attitudes toward green behaviors, overall image, gender, and age in hotel customers' eco-friendly decision-making process. *International Journal of Hospitality Management*, 28 (4), 519–528.

Han, H., Hsu, L. J., Lee, J., & Sheu, C. (2011). Are lodging customers ready to go green? An examination of attitudes, demographics, and eco-friendly intentions. *International Journal of Hospitality Management*, 30 (2), 345–355.

Harrison, K. (2014). *Green Hotels Are Leading The Pack On TripAdvisor*. Retrieved June 05, 2018, from www.forbes.com: https://www.forbes.com/sites/kateharrison /2014/06/20/green-hotels-are-leading-the-pack-on-tripadvisor/#1cfcbc702917

Hasek, G. (2016). *Three Years After Launch, TripAdvisor GreenLeaders Tops 11,000 Participants*. Retrieved June 05, 2018, from http://www.greenlodgingnews.com/ three-years-after-launch-tripadvisor-greenleaders/

Hotelling, H. (1929). Stability in Competition . *The Economic Journal*, 39(153), 41-57.

Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20(1), 1-24.

Ichino, A., Schwerdt, G., Winter-Ebmer, R., & Zweimüller, J. (2017). Too Old to Work, Too Young to Retire? The *Journal of the Economics of Ageing*, 9(June 2017), 14-29.

Jones, P., Comfort, D., & Hillier, D. (2008). UK retailing through the looking glass. *International Journal of Retail & Distribution Management*, 36(7), 564-570.

Kok, N., & Jennen, M. (2012). The impact of energy labels and accessibility on office rents. *Energy Policy*, 46, 489–497.

Laroche, M., Bergeron, J., & Barbaro-Forleo, G. (2001). Targeting consumers who are willing to pay more for environmentally friendly products. *Journal of Consumer Marketing*, 18(6), 503-520.

Lee, J.-S., Hsu, L.-T. (., Han, H., & Kim, Y. (2010). Understanding how consumers view green hotels: how a hotel's green image can influence behavioural intentions. *Journal of Sustainable Tourism*, 18(7), 901-914.

Leonidou, L. C., Leonidou, C. N., Fotiadis, T. A., & Zeriti, A. (2013). Resources and capabili- ties as drivers of hotel environmental market- ing strategy: Implications for competitive advantage and performance. *Tourism Man- agement*, 35, 94–110.

Lockyer, T. (2005). The perceived importance of price as one hotel selection dimension. *Tourism Management*, 26(4), 529–537.

Manaktola, K., & Jauhari, V. (2007). Exploring consumer attitude and behaviour towards green practices in the lodging industry in India. *International Journal of Contemporary Hospitality Management*, 19 (5), 364–377.

Mazzeo, M. J. (2002). Competitive Outcomes in Product-Differentiated Oligopoly. *Review of Economics and Statistics*, 84(4), 716-728.

Miller, N., Pogue, D., Saville, J., & Tu, C. (2010). The operations and management of green buildings in the United States. *Journal of Sustainable Real Estate*, 2(1), 51–66.

Nikodem, S., & Fuerst, F. (2013). The operating expense puzzle of U.S. green office buildings. *Journal of Sustainable Real Estate*, 5(1), 1-25.

Pager, D., & Quillian, L. (2005). Walking the Talk? What Employers Say Versus What They Do. *American Sociological Review*, 70(3), 355 – 380.

Rahman, I., Reynolds, D., & Svaren, S. (2012). How "green" are North American hotels? An exploration of low-cost adoption. *Inter- national Journal of Hospitality Management*, 31(3), 720–727.

Rivera, J. (2002). Assessing a voluntary environmental initiative in the developing world: The Costa Rican certification for sustainable tourism. *Policy Sciences*, 35(4), 333 – 360.

Robinson, S. J., & McAllister, P. (2015). Heterogeneous price premiums in sustainable real estate: An investigation of the relationship between value and price premiums. *Journal of Sustainable Real Estate*, 7(1).

Robinson, S. J., & Reichert, A. (2015). The application of hedonic grid regression to commercial real estate. *Journal of Real Estate Research*, 37(4), 563-596.

Robinson, S., Singh, A. J., & Das, P. (2016). Financial Impact of LEED and Energy Star Certifications on Hotel Revenues. *The Journal of Hospitality Financial Management*, 24(2): 110-126.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects . *Biometrika*, 70(1), 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, 39 (1), 33-38.

Sharkey, J. (2013). *New Rivals Compete for Convention Business*. Retrieved June 05, 2018, from www.nytimes.com: http://www.nytimes.com/2013/08/27/business /new-rivals-in-the-competition-for-convention-business.html

Shy, O. (1995). *Industrial Organization: Theory and Applications*. Cambridge, MA: The MIT Press.

Tarí, J. J.,´s, E. C.-C., Pereira-Moliner, J., &´ın, J. F.-A. (2010). Levels of quality and environmental management in the hotel industry: Their joint influence on firm performance. *International Journal of Hospitality Management*, 29(3), 500–510.

TripAdvisor. (2018, June 05). *Green Leaders*. Retrieved June 05, 2018, from TripAdvisor.com: https://www.tripadvisor.com/GreenLeaders

Trudel, R., & Cotte, J. (2009). Does It Pay to Be Good? *MIT Sloan Management Review*, 50(2), 61-68.

Vives, X. (1985). On the efficiency of Bertrand and Cournot equilibria with product differentation. *Journal of Economic Theory*, 36(1), 166-175.

Walsman, M., Verma, R., & Muthulingam, S. (2014). The Impact of LEED Certification on Hotel Performance. *Cornell Hospitality Report*, 14(15), 6-13.

Werntz, K. M. (2015). *TripAdvisor scheme helps travellers select a sustainable bed for the night.* Retrieved June 05, 2018, from https://www.theguardian.com/sustainable-business/2015/jan/07/tripadvisor-travel-sustainable-bed-green-hotels-holiday

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, 25(1), 57-76.

Zhang, L., Wu, J., Liu, H., & Zhang, X. (2017). The Value of Going Green in the Hotel Industry: Evidence from Beijing. *Real Estate Economics.*

APPENDICES CHAPTER I

APPENDIX A: TABLES

Table 1.1: Summary statistics

Variables	Definition	Mean	Std. Dev.	Min	Max
Occupancy	Percentage of available rooms sold each day (%)	73.28	23.06	0.50	100
ADR	Average price paid for rooms sold ($)	132.76	77.67	9.65	2314.82
RevPAR	Revenue per Available Room ($)[7]	101.91	74.84	0.41	2245.43
Platinum	Whether the hotel received a Platinum badge	0.02	0.13	0	1
Gold	Whether the hotel received a Gold badge	0.04	0.2	0	1
Silver	Whether the hotel received a Silver badge	0.12	0.33	0	1
Bronze	Whether the hotel received a Bronze badge	0.11	0.32	0	1
Green Partner	Whether the hotel received a Green Partner badge	0.10	0.31	0	1
Banquet room	Whether the hotel has Banquet room(s)	0.44	0.49	0	1
Babysitting facility	Whether the hotel has babysitting facilities	0.05	0.22	0	1
Airport transport	Whether the hotel provides airport transportation service	0.21	0.41	0	1
Breakfast included	Whether the hotel offers free breakfast	0.49	0.5	0	1
Free parking	Whether the hotel offers free parking service	0.57	0.5	0	1
Fitness center	Whether the hotel has fitness center	0.71	0.45	0	1
Business center	Whether the hotel has business center	0.77	0.42	0	1
Multilingual staffs	Whether the hotel has multilingual staffs	0.82	0.38	0	1
Conference facility	Whether the hotel has conference facility	0.46	0.5	0	1
Meeting room	Whether the hotel has meeting room(s)	0.69	0.46	0	1
Franchise	Whether the hotel is a franchise	0.59	0.49	0	1

[7] RevPAR is calculated by dividing total guest room revenue by the total number of available rooms, including sold and unsold rooms.

Table 1.1: Continued

Chain	Whether the hotel is a chain hotel	0.27	0.44	0	1
Independent	Whether the hotel is an independent hotel	0.14	0.35	0	1
Luxury	Whether the hotel is a luxury hotel	0.16	0.37	0	1
Upscale	Whether the hotel is an Upscale hotel	0.2	0.4	0	1
Mid-price	Whether the hotel is mid-priced hotel	0.34	0.48	0	1
Economy	Whether the hotel is an economy hotel	0.18	0.39	0	1
Budget	Whether the hotel is a budget hotel	0.11	0.32	0	1
Resort	Whether the hotel is located in a resort area	0.17	0.37	0	1
Small/metro town	Whether the hotel is located in small/metro town	0.08	0.27	0	1
Airport	Whether the hotel is located near airport	0.12	0.32	0	1
Suburb	Whether the hotel is located in a suburban area	0.38	0.49	0	1
Urban	Whether the hotel is located in an urban area	0.21	0.41	0	1
Interstate	Whether the hotel is located near an interstate	0.04	0.19	0	1
City popularity rank	TripAdvisor's popularity rank of the city where the hotel is located	59.51	84.07	1	527

Notes: This table reports descriptive statistics for the dependent variables (i.e., Occupancy rate, Average Daily Rate, and Revenue Per Available Room) and independent variables for 865 hotels included in my final sample. The dependent variables are daily data and span a period between February 01, 2011 and June 29, 2017. The independent variables are dummy variables. There are five different types of independent variables reported in this table. Variables related to GreenLeaders badges (i.e., *Platinum, Gold, Silver, Bronze,* and *GreenPartner*), hotel amenities (i.e., *Banquet room, Babysitting facility, Airport transport, Breakfast included, Free parking, Fitness center, Business center, Multilingual staffs, Conference facility,* and *Meeting room*), chain affiliation (i.e., *Franchise, Chain,* and *Independent*), hotel class (i.e., *Luxury, Upscale, Mid-price, Economy,* and *Budget*), and hotel location (i.e., *Resort, Small/metro town, Airport, Suburb, Urban, Interstate,* and *City popularity rank*).

Table 1.2: Summary of covariate balance before and after matching

Variables: hotel characteristics	Before Matching		After Matching	
	Participating Hotels	Non-participating Hotels	Participating Hotels	Non-Participating Hotels
Multilingual Staff	0.9298	0.751	0.9224	0.9353
Conference Facility	0.652	0.3314	0.5776	0.5647
Meeting Room	0.8684	0.5728	0.8362	0.8491
Franchise	0.5292	0.6322	0.6293	0.6379
Chain	0.3304	0.228	0.2155	0.2026
Independent	0.1404	0.1398	0.1552	0.1595
Luxury	0.2485	0.0996	0.1552	0.1638
Upscale	0.2865	0.1456	0.25	0.2457
Mid-price	0.3421	0.3467	0.444	0.431
Economy	0.1111	0.2261	0.1336	0.1422
Budget	0.0117	0.182	0.0172	0.0172

Notes: This table illustrates how well the characteristics of the GreenLeaders hotels (treated units) match the characteristics of the non-GreenLeaders (control units) hotels after propensity score matching. In particular, I use the nearest-neighbor matching method (without replacement) to obtain a matched pair of one treated and one control unit. To ensure a good match, I impose a caliper of 0.05 so that the absolute difference between the treated and control unit's propensity score is up to 0.05. A comparison of the hotel characteristics before and after matching shows that, on average, the differences between the treated and control units are reduced after constructing matched pairs using propensity scores.

Table 1.3: Green premiums in hotel Occupancy rate, ADR and RevPAR

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Occupancy	ln(ADR)	ln(RevPAR)	Occupancy	ln(ADR)	ln(RevPAR)
Green	0.987	0.030**	0.052***			
	(0.691)	(0.012)	(0.010)			
Platinum				-0.822	0.016	0.004
				(1.793)	(0.045)	(0.064)
Gold				1.983	0.019	0.068
				(2.044)	(0.046)	(0.074)
Silver				0.662	-0.012	0.005
				(1.407)	(0.031)	(0.041)
Bronze				0.363	0.031	0.043
				(1.295)	(0.027)	(0.037)
GreenPartner				1.907	0.079**	0.115***
				(1.326)	(0.038)	(0.040)
Banquet room	-0.693	0.014	0.002	-0.363	0.017	0.009
	(1.324)	(0.032)	(0.044)	(1.421)	(0.034)	(0.047)
Babysitting	-3.133*	0.381***	0.329***	-3.253*	0.380***	0.324***
	(1.829)	(0.073)	(0.067)	(1.841)	(0.075)	(0.067)
Airport transportation	0.497	-0.061	-0.044	0.603	-0.056	-0.037
	(1.632)	(0.039)	(0.046)	(1.615)	(0.039)	(0.045)
Breakfast included	2.845**	0.021	0.069*	3.043**	0.025	0.078**
	(1.385)	(0.027)	(0.037)	(1.370)	(0.028)	(0.037)
Free parking	0.049	0.016	0.022	0.121	0.019	0.026
	(1.447)	(0.044)	(0.046)	(1.473)	(0.043)	(0.046)
Fitness center	1.058	0.076*	0.101*	0.833	0.071*	0.090*
	(1.949)	(0.039)	(0.051)	(1.939)	(0.039)	(0.051)
Business center	1.792	0.115***	0.155**	1.791	0.109***	0.149**
	(2.165)	(0.042)	(0.061)	(2.150)	(0.042)	(0.062)

Table 1.3: Continued

Multilingual staff	-0.678	0.039	0.020	-0.812	0.038	0.016
	(2.445)	(0.039)	(0.065)	(2.450)	(0.040)	(0.066)
Conference facility	1.094	0.065*	0.087*	1.145	0.072**	0.094*
	(1.516)	(0.037)	(0.049)	(1.539)	(0.036)	(0.049)
Meeting room	2.530	0.107***	0.143***	2.520	0.115***	0.151***
	(2.194)	(0.040)	(0.054)	(2.181)	(0.038)	(0.051)
Occupancy		0.003***	0.020***		0.003***	0.020***
		(0.000)	(0.000)		(0.000)	(0.000)
Time-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,189,228	1,189,228	1,189,228	1,189,228	1,189,228	1,189,228
Adjusted R^2	0.21	0.41	0.70	0.21	0.41	0.70

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. This table reports estimates for the effects of participation in the GreenLeaders program using fixed effects regression specified by equation 5. Using the propensity score matched data, the regression controls for observed hotel characteristics. The regression includes group-fixed effects in order to control for unobserved characteristics related to locations and market condition. The table reports whether participation has any effect on the hotels' *Occupancy, ADR,* and *RevPAR*. Each column reports results for a separate regression. Columns 1, 2, and 3 present results for the overall effects of participation on *Occupancy*, *ADR,* and *RevPAR*, respectively. Column 4, 5, and 6 report the heterogeneous effects of participation by the hotels' badge types. Standard errors reported in parenthesis are robust at the group level.

Table 1.4: Effects of participation in the GreenLeaders program: DID and GSC estimates

Covariates	Occupancy (1)	ln(ADR) (2)	ln(RevPAR) (3)
	Panel A: DID		
Treated	1.145	0.023	0.050**
	(0.724)	(0.018)	(0.022)
Post	0.107	-0.008	-0.003
	(0.480)	(0.009)	(0.012)
Post*Treated	-0.426	0.018*	0.006
	(0.621)	(0.010)	(0.016)
Banquet room	-0.646	0.014	0.000
	(1.309)	(0.032)	(0.044)
Babysitting	-3.025*	0.382***	0.329***
	(1.735)	(0.073)	(0.067)
Airport transportation	0.345	-0.060	-0.044
	(1.584)	(0.039)	(0.046)
Breakfast included	2.776**	0.021	0.070*
	(1.395)	(0.026)	(0.036)
Free parking	0.003	0.017	0.023
	(1.428)	(0.044)	(0.046)
Fitness center	1.167	0.076*	0.100*
	(1.954)	(0.039)	(0.051)
Business center	1.907	0.115***	0.155**
	(2.181)	(0.042)	(0.061)
Multilingual staff	-0.761	0.038	0.020
	(2.451)	(0.039)	(0.065)
Conference facility	1.035	0.065*	0.087*
	(1.523)	(0.037)	(0.049)
Meeting room	2.446	0.107***	0.143***
	(2.195)	(0.040)	(0.054)
Occupancy		0.003***	0.020***
		(0.000)	(0.000)
Zip-code fixed effect	Yes	Yes	Yes
Observation	1,189,228	1,189,228	1,189,228
R^2	0.1760	0.4044	0.8011
	Panel B: GSC		
Participation	0.794	0.027	0.005
	(1.056)	(0.158)	(0.054)
Observation	2,024,965	2,024,965	2,024,965
Treated	342	342	342
Control	523	523	523

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. This table presents estimates for the effects of participation in the GreenLeaders program based on two different models, difference-in-differences (DID), specified by equation 6, and Generalized Synthetic Control (GSC). Panel A reports DID estimates and panel B reports GSC estimates. The table reports the effects of green certifications on hotels' *Occupancy, ADR,* and *RevPAR*. Each column shows results for different regressions: columns 1, 2, and 3 present results for the regressions with dependent variables *Occupancy*, *ADR,* and *RevPAR*, respectively. Under panel A, standard errors reported in parenthesis are robust at the group level.

Table 1.5: Heterogeneous effects across badge types

	Panel A: DID			Panel B: GSC					
Covariates: Badges	Occupancy	ln(ADR)	ln(RevPAR)	Occupancy	ln(ADR)	ln(RevPAR)	Treatment Units	Control Units	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Platinum	-1.302	0.013	0.012	-3.725	0.065	0.081	14	523	1,257,117
	(2.186)	(0.024)	(0.023)	(2.301)	(0.053)	(0.057)			
Gold	-1.770	0.009	-0.002	-2.531	0.024	0.043	36	523	1,308,619
	(1.414)	(0.018)	(0.018)	(1.516)	(0.030)	(0.038)			
Silver	0.0153	0.005	0.007	-0.332	0.022	0.032	104	523	1,467,807
	(0.865)	(0.015)	(0.015)	(0.859)	(0.029)	(0.026)			
Bronze	0.924	0.010	0.006	0.362	0.007	0.050	98	523	1,453,761
	(0.856)	(0.015)	(0.014)	(0.872)	(0.029)	(0.044)			
Green-	-1.975	0.050**	0.042**	-1.102	0.037	0.023	90	523	1,435,033
Partner	(1.909)	(0.021)	(0.020)	(0.950)	(0.026)	(0.026)			
Number of Hotels	506	506	506						
N	1189228	1189228	1189228						
R^2	0.1735	0.4666	0.6070						

Notes: *** p<0.01, ** p<0.05, * p<0.1. Under panel A, standard errors reported in parenthesis are robust at the group level. This table reports estimates for the heterogeneous effects across different badges of the GreenLeaders hotels using both the difference-in-differences (DID) and generalized synthetic control (GSC) models. Column 1 to 6 report coefficients for the participation effects by badge types. Column 1 to 3 report results from the DID model, and column 4 to 6 report results from the GSC model. Column 7, 8, and 9 report the numbers of treated units, control units, and total observations, respectively, for the GSC estimates.

Table 1.6: Participation effects by location types

Dependent variables	Resort	Small/metro town	Airport	Suburban	Urban	Interstate
			Panel A: DID			
Occupancy	-0.335	0.080	-1.018	-1.061	-0.535	2.669
	(0.857)	(0.969)	(0.546)	(0.955)	(1.291)	(3.289)
ln(ADR)	0.075**	0.042**	-0.018	0.009	0.006	0.032
	(0.031)	(0.018)	(0.018)	(0.014)	(0.018)	(0.023)
ln(RevPAR)	0.071**	0.045**	-0.020	0.005	0.008	0.002
	(0.030)	(0.020)	(0.018)	(0.013)	(0.019)	(0.039)
Zip code fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
N	177,916	88,958	168,552	496,292	234,100	23,410
Number of hotels	76	38	72	210	100	10
			Panel B: GSC			
Occupancy	-0.723	2.268	-2.056	-0.851	0.981	4.081
	(1.334)	(2.236)	(2.405)	(1.154)	(1.537)	(3.551)
ln(ADR)	0.061**	0.036**	0.012	0.014	-0.006	0.025
	(0.028)	(0.016)	(0.228)	(0.147)	(0.033)	(0.050)
ln(RevPAR)	0.063**	0.034**	0.032	-0.002	-0.004	0.012
	(0.027)	(0.014)	(0.121)	(0.058)	(0.026)	(0.101)
N	341786	159188	241123	777212	430744	74912
Treated hotels	50	19	38	135	91	7
Control hotels	96	49	65	197	93	25

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Under panel A, standard errors reported in parenthesis are robust at the group level. This table reports the effects of participation in the GreenLeaders program across six locations types (i.e., resort, small/metro town, airport, suburban, urban, and interstate) based on the difference-in-differences (DID) and generalized synthetic control (GSC) models. Panel A reports DID estimates and panel B reports GSC estimates. Each of the dependent variables (i.e., *Occupancy, ADR,* and *RevPAR*) has been regressed separately by the six location types, and every column reports estimates of the participation effects for a particular location type.

Table 1.7: Participation effects by city popularity (with matched data)

Dependent Variable	All	Excluding Top 10	Excluding Top 20	Excluding Top 30	Excluding Top 40	Excluding Top 50	Excluding Top 60
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Occupancy	-0.426	-0.335	-0.422	-0.439	-0.872	-0.722	-1.179
	(0.621)	(0.781)	(0.845)	(0.901)	(0.950)	(1.052)	(1.094)
ln(ADR)	0.018*	0.021	0.031	0.026	0.032	0.041**	0.049**
	(0.010)	(0.014)	(0.024)	(0.018)	(0.024)	(0.018)	(0.022)
ln(RevPAR)	0.006	0.009	0.011	0.007	0.019	0.027*	0.039**
	(0.016)	(0.021)	0.016	(0.016)	(0.018)	(0.014)	(0.018)
N	1,189,228	739,756	664,844	608,660	519,702	397,970	355,832
Number of Hotels	506	316	284	260	222	170	152

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors reported in parenthesis are robust at the group level. This table reports the effects of participation in the GreenLeaders program using the difference-in-differences (DID) model. As we move from column 1 to column 7, I estimate the DID model with the propensity score matched data by incrementally excluding more cities (thereby, the hotels located within) from the top of the city-popularity-rank by TripAdvisor. Column 1 excludes no cities, and Column 7 excludes the top 60 of the most popular cities. The table shows whether the effects of green certifications vary based on the popularity of the city of a GreenLeaders hotel.

Table 1.8: STR's definitions of location segments

Location	Definition
Urban	Densely populated location in a large metropolitan area. (e.g., Atlanta, Boston, San Francisco, London, Tokyo).
Suburban	Suburbs of metropolitan markets. Examples are Sag Harbor and White Plains, NY (near New York City, USA) and Croydon and Wimbledon (near London, UK). Distance from center city varies based on population and market orientation.
Airport:	Hotel in close proximity to an airport that primarily serves demand from airport traffic. Distance may vary.
Interstate/Motorway:	Property in close proximity to major highway, motorway or other major roads with the primary source of business via passerby travel. Hotels located in suburban areas have the suburban classification.
Resort:	Property located in a resort area or market where a significant source of business is derived from leisure/destination travel. Examples: Orlando, Lake Tahoe, Daytona Beach, Hilton Head Island, Virginia Beach.
Small Metro/Town:	Area with either a smaller population or remote locations with limited services. Size varies by market orientation. Suburban locations do not exist in proximity to these areas. In North America, metropolitan small town areas are populated with less than 150,000 people.

APPENDIX B: FIGURES

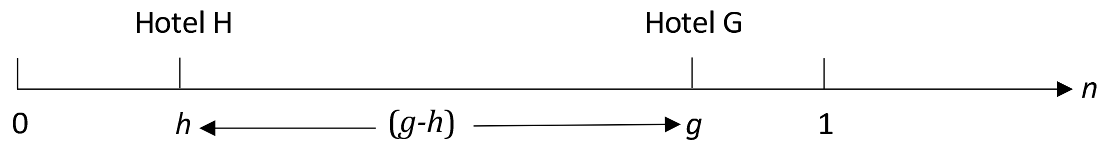

Figure 1.1: Vertical differentiation in a modified Hotelling model (adapted from Shy, 1995)

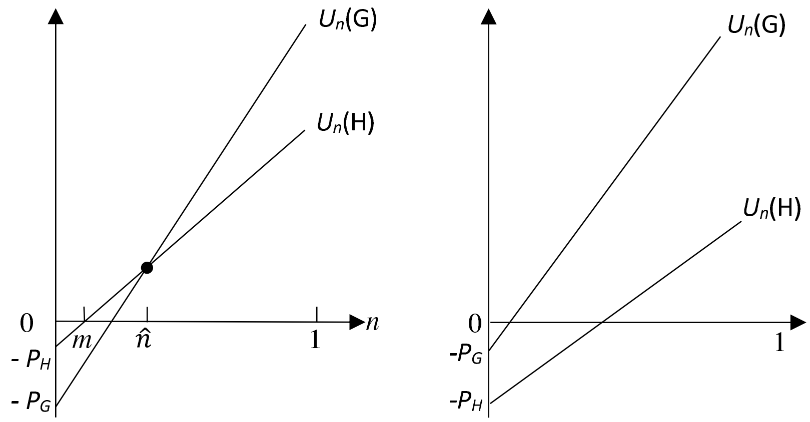

Figure 1.2: Determination of the indifferent consumer among vertically differentiated hotels. Left: $p_H < p_G$, Right: $p_H > p_G$ (adapted from Shy, 1995)

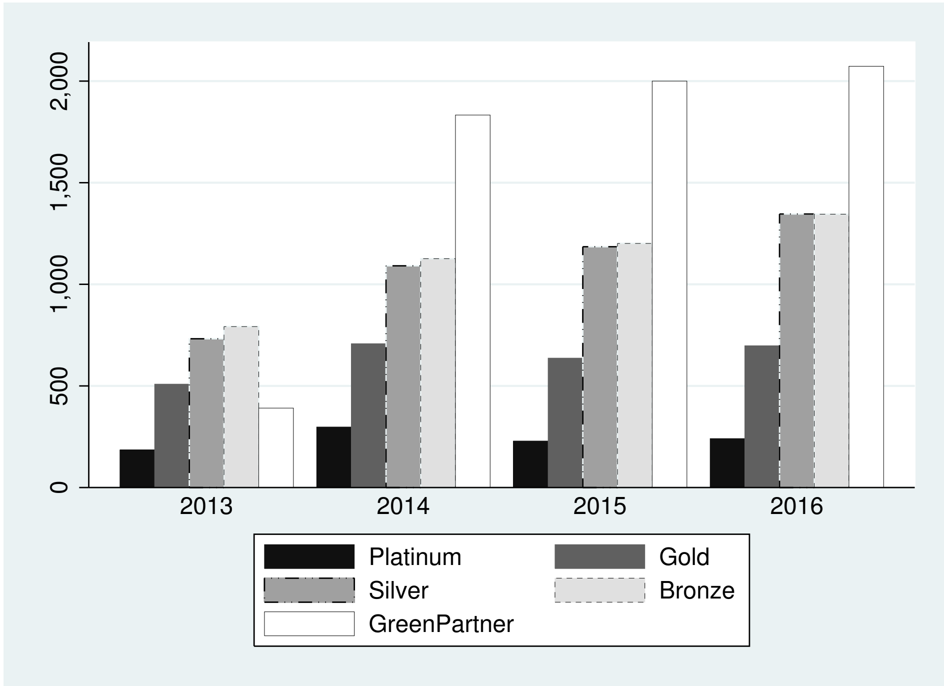

Figure 1.3: Number of participating hotels by their badge types in the GreenLeaders program for the 2013-2016 period.

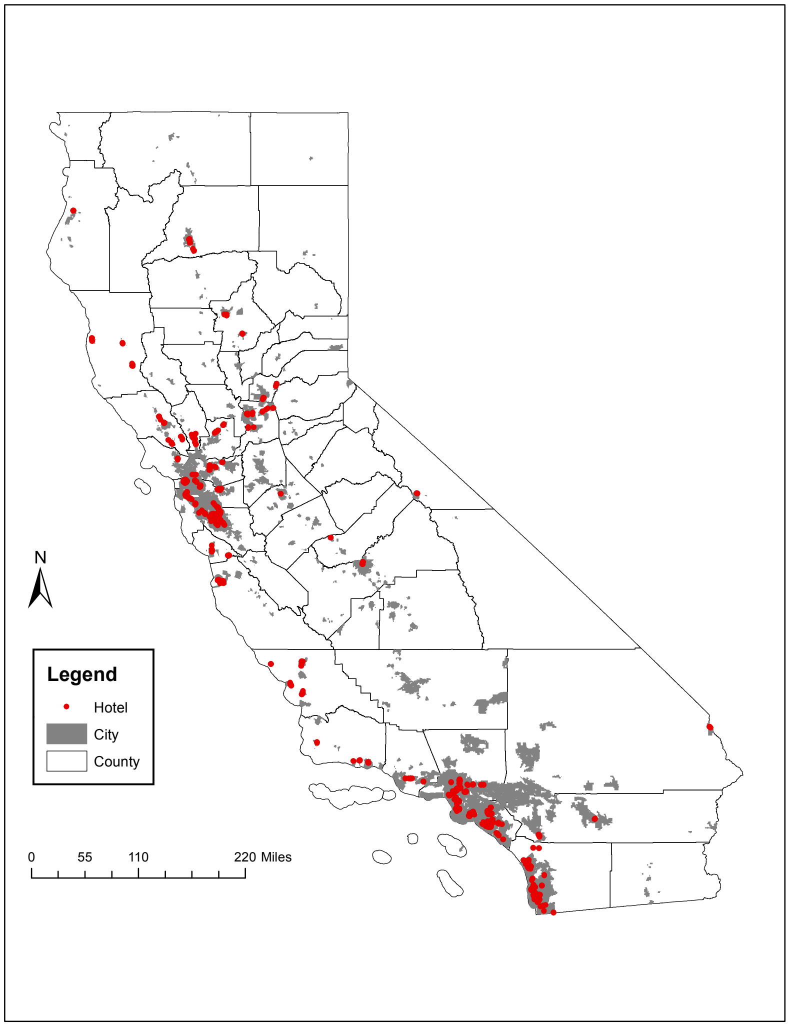

Figure 1.4: Sample of hotels in California

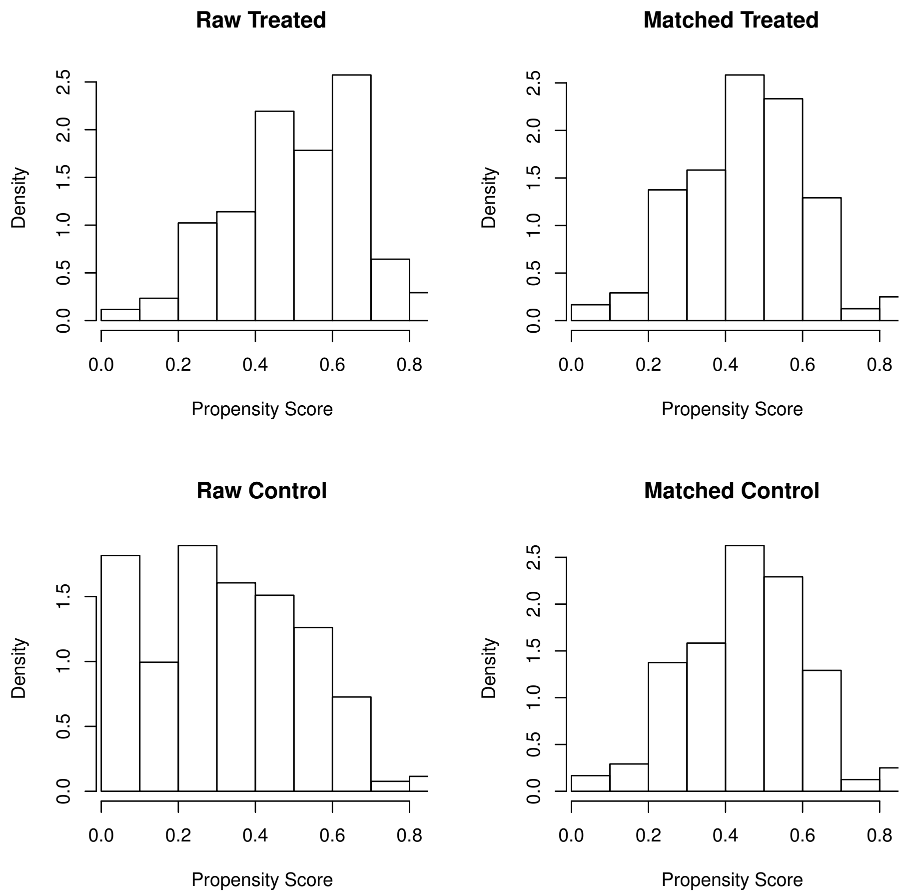

Figure 1.5: Histogram of propensity scores between treatment and control groups: Raw and Matched

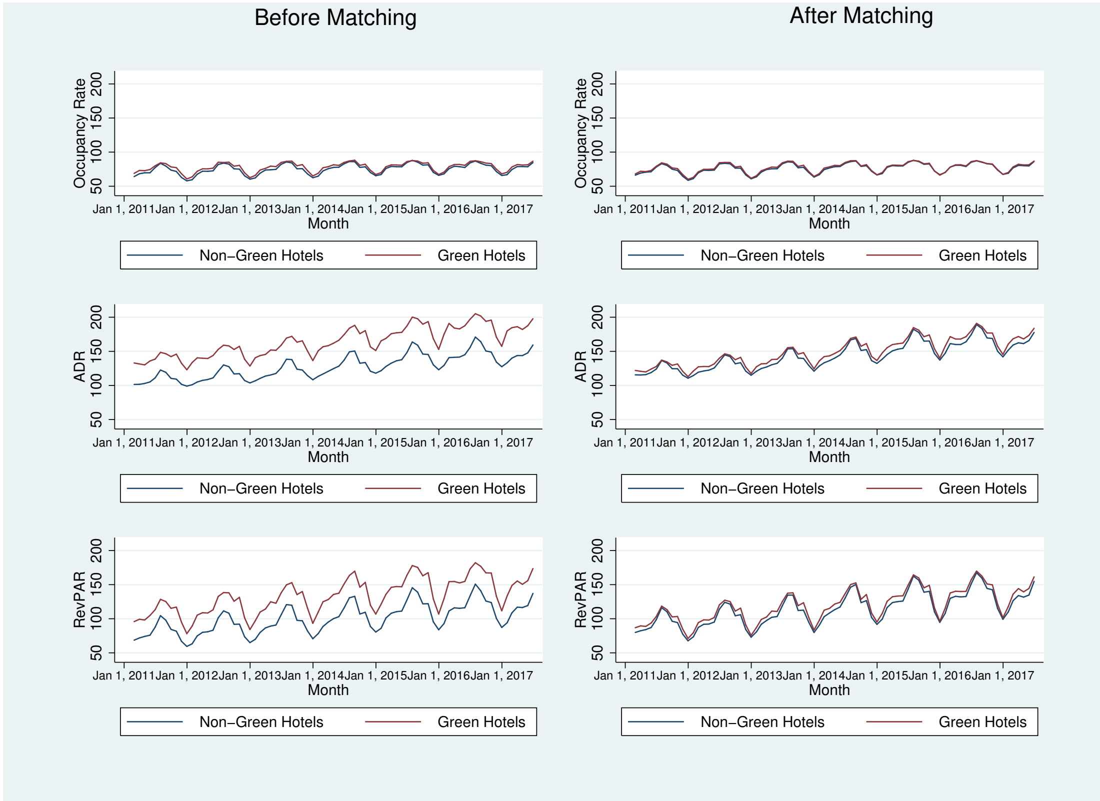

Figure 1.6: Performance of treated and control groups before and after matching

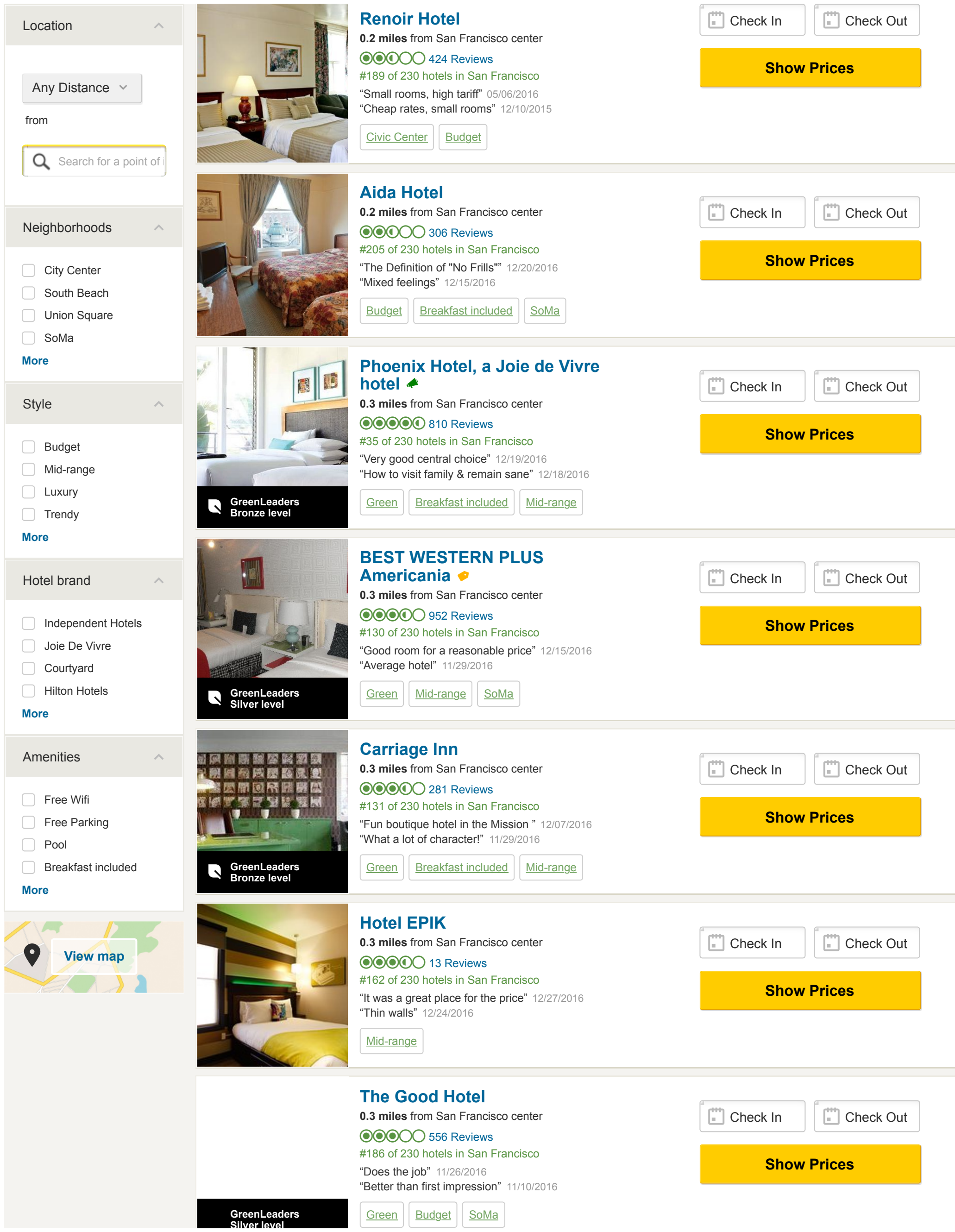

Figure 1.7: GreenLeaders hotels on TripAdvisor.com

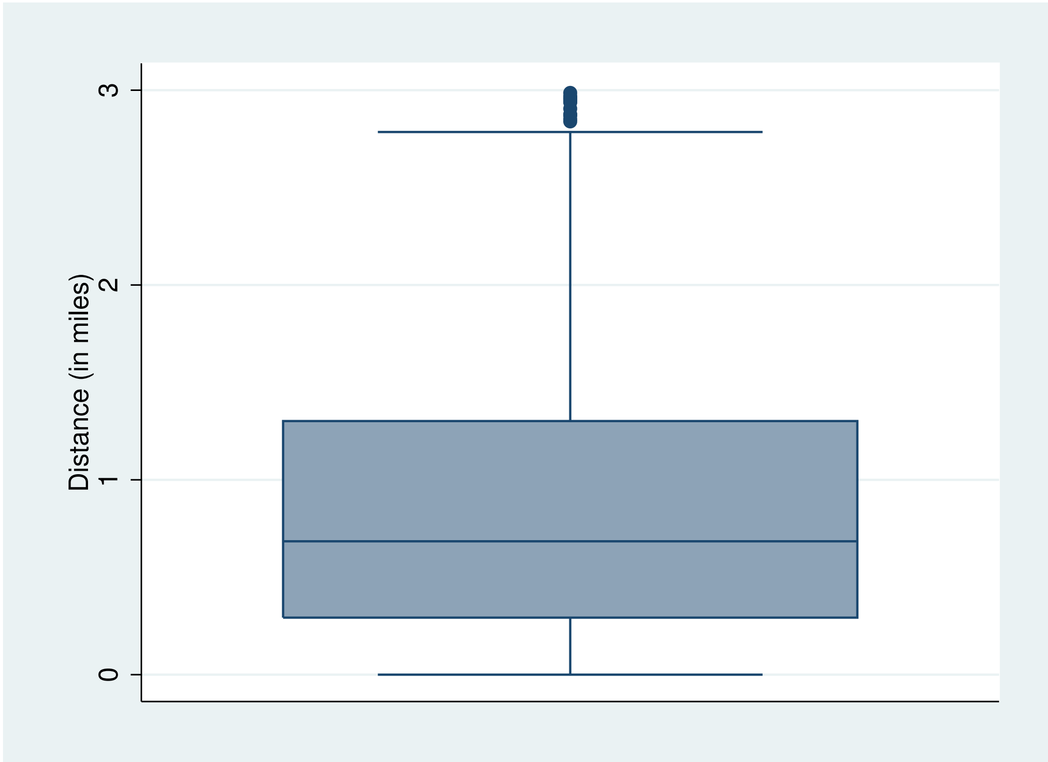

Figure 1.8: Box plot of distances between green and nongreen hotels

Chapter II

THE IMPACT OF TRIPADVISOR'S STAR RATING ON HOTEL REVENUE

1. Introduction

The technological advancement in the twenty-first century has led to the digitization of interactions between firms and consumers on online platforms, such as TripAdvisor, where consumers can share publicly available reviews about products and services. These online reviews, also known as user-generated content (UGC) or electronic word of mouth (eWOM), have become notably important over the past decade as the Internet became increasingly popular. In recent years, online reviews are considered to be more successful compared to the traditional marketing and promotional activities in influencing consumers' product or service choice (Ye et al., 2011; Gretzel and Yoo, 2008; Yang and Mai, 2010; Zhang et al., 2010; Anderson, 2012). The conventional mechanism to inform customers about product quality and features have limitations; for instance, advertising can be expensive and expert reviews often cover small market segments. Online reviews may, therefore, play a crucial role in complementing or substituting traditional source of information (Luca, 2016). However, reviews can be subjective and based on non-representative samples, making it difficult for potential customers to decipher the true quality of products and services.

Many studies have shown the growing influence of customer reviews in business performance. Using data from the video game industry, Zhu and Zhang (2010) show the relationship between online customer reviews and product sales. Their findings indicate online reviews are relatively more influential for less popular games as well as games with more experienced players. Likewise, Forman et al. (2008) report that online sales of products on Amazon are positively associated with reviews containing identity-descriptive information of the reviewers. In the tourism industry also there are studies that examine the association between online reviews and hotel performance.

Although prior studies indicate a strong relationship among business performance, online reviews, and various features of the reviews, there are limited attempts in studying the causal impact of online reviews on businesses performance. In the tourism literature, to the best of my knowledge, there are no papers investigating the causal impact of online customer reviews on the financial performance of hotel businesses. The closest study to this paper focuses on the effect of customer ratings on the restaurant industry. Using Yelp data, Luca (2016), in his working paper, shows that a one-star increase in the restaurants' rating increases their revenue by 5-9 percent. Clearly, restaurants and hotels operate in different competitive environments, despite similarities between the two. Hotel choice is arguably a more careful and conscious decision as travelers spend substantially longer time (i.e., one or more nights) in hotels where they may or may not choose to dine-in. Unlike restaurant customers, travelers consider many more factors, such as room amenities, hotel amenities, free breakfast, and so forth, for choosing hotels. Moreover, majority or almost all the hotel customers are from outside of the town, which is usually not the case for restaurants. As a result, online ratings are likely more important to hotel guests. This motivates me to examine the causal impact of travelers' reviews on hotel revenue. In particular, I have utilized the customer ratings from TripAdvisor.com for a sample of hotels in Texas in order to estimate the causal impact of TripAdvisor ratings on hotel revenue. The key questions I ask are: Do TripAdvisor ratings impact hotel revenue? If yes, what is the impact of a one-star improvement on TripAdvisor.com for a hotel on its revenue?

Figure 2.1 illustrates the boxplots of log revenues across different levels of star-ratings for a sample of 1348 hotels on TripAdvisor.com. The figure shows a positive relationship between TripAdvisor ratings and hotel revenue for the hotels with 3.5 or higher ratings, although between 4.5-star and 5-star, the figure shows a drop in revenue. Further investigation shows almost all the hotels with a 5-star rating have very few customer ratings, which explains the decline in revenue. I discuss this further in the result section. Overall the figure suggests a positive relationship between TripAdvisor's star rating and hotel revenue.

Why TripAdvisor? For the hotel managers, TripAdvisor is usually the first point of call when it comes to the online presence of their businesses (Xie et al., 2014). With many interactive travel forums, TripAdvisor is an early adopter of user-generated contents. The website has been consistently ranked as the most popular website, in terms of the number of unique users, in the USA (Statista, 2018). Unlike other travel websites, TripAdvisor is mainly a hotel review and media site, while other leading travel websites (e.g., Expedia, hotels.com, booking.com, priceline.com, Travelocity.com, kayak.com, orbitz.com, hotwire.com, etc.) are full-service online travel agencies that sell various travel options, allowing travelers to book flights, taxis, and hotels, all at the same place (Frank, 2014). A recent study, conducted by ComScore, reports 70% of the US travelers visit TripAdvisor.com before booking a hotel. The study - analyzing 12 major global markets, including the US, and 325 websites - reports TripAdvisor is the most visited website and app by travelers before booking a hotel (TripAdvisor, 2018). All information presented above and the availability of data for a large number of hotels warrant conducting this study, using TripAdvisor's review data.

This paper is organized as follows. Section 2 conducts a survey of the existing literature; section 3 discusses the data sources and data collection procedure; section

4 then outlines and elaborates on the empirical specifications; section 6 discusses the results; section 6 and 7 present some robustness checks and analyze the data using alternative specifications, respectively; and section 8 concludes.

2. Literature review

Many studies in the tourism literature have illustrated the importance of online reviews in hotel performance. Phillips et al. (2015) use an artificial neural network model, and present evidence suggesting online reviews together with hotel characteristics are the determining factors for hotel performance in Switzerland. Besides online reviews, a hotel manager's responses to online reviews can also influence hotel performance. Xie et al. (2014), studying a sample of 843 hotels on TripAdvisor.com, report that overall ratings, location and cleanliness, and the number of management responses are significantly associated with hotel revenue. In a later work, Xie et al. (2017) investigate the joint effects of online reviews and management responses on hotel revenue, average room prices, and occupancy rates. The study finds timely and detailed management responses to online reviews enhance a hotel's future financial performance.

According to Duan et al. (2008b), for a product or service, the number of reviews from customers is one of the most "critical" attributes of reviews. A number of studies have shown businesses tend to perform better as number of online reviews increase (Viglia et al., 2016; Kim et al., 2016; Zhu and Zhang, 2010; Duan et al., 2008b; Amblee and Bui, 2007; Chevalie and Mayzlin, 2006; Liu, 2006; Blal and Sturman, 2014). Ye et al. (2009) and Torres et al. (2015) report that the number of reviews has a positive effect on online hotel bookings. Kim et al. (2015) report that the number of reviews has a significant effect on hotel revenues. Tuominen (2011) also finds a positive relationship between the number of reviews and a hotel's revenue per available room (RevPAR) and room occupancy. However, because prior studies mainly focus on the correlational relationship between consumer reviews and hotel performance, our understanding as to the causal impact of online reviews on hotel revenue is still limited. This study, therefore, focuses on the causal impact of online hotel ratings on hotel revenue. Why is a quantitative measure for the impact of online rating important? The rising importance of online reputation poses potential opportunities and threats to hotel managers. Having full control over this asset is often a challenging task for any business organization, including the hotel industry, as it is not easy to convert reputation, an intangible form of asset, into monetary value (Roos et al., 2005). According to Castro et al. (2004), information management is the cornerstone for building a good corporate reputation, which requires dynamic information management and making a smooth communication channel between businesses and customers. In this regard, online reviews are a crucial source of information that can shape the reputation of tourism businesses. This paper thus endeavors to do exactly that, which is to measure online reputation in terms of revenue.

3. Data

I have combined monthly revenue data with TripAdvisor's review data for a sample of 1348 hotels in Texas to measure the impact of online reputation on hotel revenue. I have collected data in two parts: hotel data from TripAdvisor.com and revenue data from the Texas Comptroller Office.

3.1. TripAdvisor.com

TripAdvisor.com is a US-based travel and tourism website, well known for travel-related contents such as hotel and restaurant reviews, booking accommodations, and so forth. With many interactive travel forums, the website is an early adopter of user-generated contents. On TripAdvisor.com, a verified traveler can leave a review about his or her experience in an accommodation (i.e., hotel, motel, bed and breakfast, and other lodging facilities). Travelers can also leave reviews about restaurants, locations, and a wide range of other recreational activities available in a location.

For writing a review, a user has to register, which is free of charge, on TripAdvisor.com with a valid email address. A user can then rate (on a scale of 1 to 5) and write a review about his or her experience in a hotel or travel accommodation as long as the accommodation is reserved through TripAdvisor.com. Anyone, with or without an account with TripAdvisor, can read the reviews and any other publicly available information about the hotels and restaurants that are listed on the website. Besides reviews, TripAdvisor contains a wide range of information for the registered hotels, including hotel amenities, address, hotel photos, awards and recognition, price range, things-to-do, nearby attractions, etc.

From TripAdvisor, I have collected data in two stages. In the first stage, for all Texas hotels with active listings on TripAdvisor.com, I collect hotel data containing hotel name, address, amenities (e.g., free parking, shuttle service, pool, free breakfast, and so forth), and any other hotel specific information available on the website. In the next stage, I collect review data, including review dates and review texts, for each hotel. Finally, I combine the review data with the hotel data.

3.2. Revenue data

The dependent variable used in this study is monthly room-revenue for the hotels in Texas. I collected the revenue data from the Texas Comptroller of Public Accounts. The data also contains other basic information about the hotels, such as hotel name, address, and the number of rooms.

For tax purposes, Texas law defines a hotel "...to be any building in which members of the public rent sleeping accommodations for $15 or more per day." As result, the Airbnb properties and any other vacation rentals that comply with the Texas tax code are also reported in the revenue data set. For the purpose of this study, I only include hotel accommodations in my analysis.

3.3. Data aggregation

Using addresses, I have combined the hotel data from TripAdvisor.com with the revenue data from the Texas Comptroller of Public Accounts. One challenge in merging the two datasets is that for many hotels, the formats of addresses differ between the datasets. Also, in some places, I have found multiple hotels with the same address, although this is a rare occurrence. As a result, at this stage, in order to ensure the two datasets are correctly merged, I have utilized Python's FuzzyWuzzy package to match the names of the hotels alongside their addresses. I use an algorithm that produces a score between 0-100 to measure how well two groups of words (or hotel names) match. A score of 100 means the names match perfectly, and a score of 0 means the names do not match at all. Hence, after merging the datasets using addresses, I take a subsample of hotels that get a score of 100.

In the final data set, I have 376,060 reviews for 1348 hotel, with their monthly taxable receipts between the period of January 2014 and December 2017. Table 2.1 reports the summary statistics of the review and revenue data. The average monthly revenue for a hotel in Texas is USD 208,777, and the average customer rating of a hotel on TripAdvisor.com is 3.63. Also, on average, a hotel receives approximately 4 reviews each month on the website.

One challenge in the estimation of empirical specifications is that the revenue data is in monthly frequency. For estimating the OLS regression, I use the rounded-average monthly rating. For the regression discontinuity estimation, however, I assign a treatment variable based on the following condition. If the rating of a hotel crosses a threshold in a given month and stays above the threshold for more than half a month, I consider the hotel's rating as above discontinuity; otherwise, below discontinuity.

4. Empirical Specification

I use two empirical approaches to estimate the causal impact of TripAdvisor ratings on hotel revenue. Using a fixed-effect regression, I at first estimate the relationship between TripAdvisor rating and hotel revenue. Next, I use a regression discontinuity specification in order to estimate the causal impact of TripAdvisor rating.

4.1. Fixed effect regression

In order to study the relationship between TripAdvisor rating and hotel revenue, I estimate a fixed-effect regression with hotel fixed effects. The empirical specification is as follows:

$$\ln(Revenue_{it}) = \beta \ rating_{it} + \gamma_i + \delta_t + \epsilon_{it} \tag{1}$$

where $\ln(Revenue_{it})$ denotes log of revenue for hotel i in month t; $rating_{it}$ denotes the rating for hotel i in month t; γ_i denotes time-invariant unobservables of hotel i; and δ_t denotes time variant unobservables (i.e., year and month fixed effects). The coefficient of interest in equation (1) is β, which, if positive and statistically significant, would indicate TripAdvisor ratings have a positive impact on hotel revenue. However, it is also possible that TripAdvisor ratings are correlated with other unobservable factors that are associated with hotel revenues. In order to address this concern, I use a regression discontinuity design in the following part.

4.2. Regression discontinuity

The TripAdvisor rating for a hotel is an average rating rounded to the nearest half-a-star. This means, for a hotel's actual average rating of 3.74, TripAdvisor rounds down the rating to 3.5, whereas for an actual average rating of 3.75, TripAdvisor rounds up the rating to 4. Hence, the final rating for a hotel, displayed on TripAdvisor.com, is either a rounded-up or rounded-down rating. Because the quality of a hotel with ratings right below the threshold (3.75 in the above example) is not, assumably, different from when its rating is right above the threshold, the rounding of the rating is exogenous to the hotel's quality, which creates the source of variation in this study.

In order to estimate the treatment effect, I use a sample of observations with underlying ratings within 0.1-star of the thresholds to compare the treated hotels (rounded up) as opposed to the control hotels (rounded down). This allows me to estimate the average impact of an exogenous half-a-star increase on the revenue of a hotel. I also use alternative bandwidths to estimate the treatment effects.

4.2.1. Estimation strategy

I use the following regression discontinuity approach to estimate the impact of a half-a-start increase in TripAdvisor ratings on hotel revenue:

$$\ln(Revenue_{it}) = \beta\, T_{it} + \lambda\, r_{it} + \alpha_i + \gamma_t + \epsilon_{it} \tag{2}$$

where $\ln(Revenue_{it})$ denotes log of monthly hotel revenue for hotel i in month t; T is a binary indicator for treatment, which takes a value of one (1) if the actual (unrounded) average rating for hotel i crosses a threshold and, therefore, is rounded up; otherwise, T takes a value of zero (0). The coefficient of interest, β, indicates the discontinuous impact of moving from right below a threshold to above the threshold

- in other words, a 0.5-star increase in the displayed rating- on the outcome variable, $\ln(Revenue_{it})$. Also, r_{it} denotes the underlying average rating of hotel i in month t. The model also controls for hotel specific time-invariant as well as time-variant unobservables.

The main empirical specification incorporates a bandwidth of 0.1 to include only the observations that are up to 0.1-star away in terms of their underlying average ratings. To show that the results are not driven by the choice of bandwidth, I use alternative bandwidths. I also allow for potential non-linear responses to rating by including quadratic and higher order ratings.

4.3. Heterogeneous impacts

After studying the causal impact of TripAdvisor ratings on hotel revenue, I examine whether the magnitude of the impact differs as the number of customer review increases. The idea is if each review contains noisy information related to the quality of a hotel, then the information should be less noisy as more reviews are left for the hotel. Hence, if more reviews translate to more precise information, with an increase in the number of reviews, TripAdvisor ratings should have relatively higher impacts on hotel revenue. To test the hypothesis, I use the following empirical specification:

$$\ln(Revenue_{it}) = \beta\ rating_{it} + \lambda\, rating_{it} * reviews_{it} + \gamma_i + \delta_t + \epsilon_{it} \tag{3}$$

Equation 3 is a modification of equation 1. The new term introduced in the equation is $rating_{it} * reviews_{it}$, an interaction term between *rating* and *reviews,* where *rating* denotes the TripAdvisor rating for hotel i in month t, and *reviews* denotes the number of reviews.

5. Result

5.1. Fixed effect estimates

Table 2.2 reports results for the fixed effect regression represented by equation 1. The result shows that TripAdvisor ratings are associated with a 4.5% increase in hotel revenue, on average. The estimate is also statistically significant at 1% level. One limitation of this empirical specification is that TripAdvisor ratings may be correlated with other unobserved changes in hotel quality. As a result, the estimated coefficient for *rating* may be biased due to unobserved factors that are unrelated to TripAdvisor ratings.

5.2. Regression discontinuity estimates

Because the fixed-effect regression does not account for potential unobserved variables that are correlated with TripAdvisor's star rating, the results based on this model may produce biased estimates for the impact of TripAdvisor's star rating on hotel revenue. In order to address the concern, I use regression discontinuity as an alternative empirical specification. Table 2.3 presents the regression discontinuity estimates of the impact of a half-a-star increase on hotel revenue. The results show a discontinuous jump of 0.5-star increase in the rating leads to a 1.1% increase in the revenue. The estimated impact does not change when I include quadratic ratings under column 2 and higher order ratings under column 3.

Figure 2.2 provides a graphical illustration of de-meaned revenues for restaurants just above and just below the rounding thresholds. Table 2.4 reports regression discontinuity estimates for different bandwidths. Column 1 and 2 report results for the bandwidths of 0.2 and 0.1, respective, around the rounding thresholds. When the bandwidth changes to 0.1 from the original bandwidth of 0.2, I find a 0.5-star improvement in rating causes a 1.5% increase in monthly revenue. A hotel's reputation outside of TripAdvisor should be uncorrelated with whether the average hotel rating has been rounded up or down. The regression discontinuity results thus provide support for the claim that TripAdvisor ratings have a causal impact on hotel revenue.

5.3. Heterogeneous impacts

Based on Equation 3, Table 2.5 presents results for the heterogeneous impacts of TripAdvisor ratings on revenue as the number of reviews increases. If each customer review provides noisy information about the true quality of a hotel, the information should be more precise when the number of reviews is sufficiently large. More reviews should then have greater impacts on hotel revenue. Table 2.5 results support the claim. I find the hotels with more than 50 reviews compared to the hotels with 10 or fewer reviews see a 3.5% higher increase in their revenue for a 1-star increase in their TripAdvisor rating.

6. Robustness Check

In this section, I conduct a few robustness checks.

6.1. Review manipulation

One concern in the regression discontinuity approach is that the underlying average ratings could also be known to hotel managers or owners. This leaves potentials for the manipulation of customer ratings, as pointed out by McCrary (2008). Hotel managers could leave fake reviews in order to improve their ratings, which would ultimately impact their revenues positively. It is also possible that certain types or hotels, such as hotels that earn particularly high or low revenue, are susceptible to manipulating their ratings. In such cases, the regression discontinuity results could be spurious. In this section, I provide statistical evidence to show that there is no evidence of review manipulation in this study.

If a hotel submits fake or inflated reviews to improve its rating, it should stop submitting reviews once the rating crosses a rounding threshold. However, if the hotel stops submitting reviews immediately after jumping above the discontinuity, a subsequent bad review could then bring back the rounded average rating to below the discontinuity. A hotel would, therefore, have to continue submitting good reviews to increase its underlying rating to sufficiently high above the discontinuity (for instance, from 3.2 to 3.4) so as to ensure the rating does not drop below the discontinuity. Although the degree of manipulation is hard to predict, it is a fairly restrictive type of manipulation for the regression results to be spurious.

I conduct a statistical test offered by McCrary (2008) and show that there is no evidence of rating manipulation. The test is based on the idea that if TripAdvisor hotels leave fake, inflated reviews for improving their ratings, there should be a disproportionate number of reviews just above the rounding threshold. In order to test this, the number of reviews and TripAdvisor's underlying rating are the two primary variables of interest. First, I sum the number of reviews for each 0.05-star interval of underlying rating and calculate the probability mass for every interval. Second, I construct a dummy variable indicating intervals that are just above the discontinuity (i.e., 2.25-star - 2.30-star; 2.75-star - 2.80-star). The test utilizes the probability mass, from the first step, as a dependent variable and the dummy variable, from the second step, as an independent variable. Table 2.6 reports the results of McCrary test and shows the numbers of reviews just above the discontinuity is not statistically significantly different or high. Thus manipulation of reviews is not a concern for the regression discontinuity approach.

6.2. Other robustness checks

One limitation of this paper is TripAdvisor ratings are supposed to have a more direct impact on the number of online booking (or revenue from online booking through TripAdvisor.com) instead of monthly hotel revenue. After checking various hotels, hotel amenities, and online reviews at the destination of interest, a traveler can book a hotel room quite a long time ahead of his actual date of arrival. A guest usually provides his/her credit card information for the booking purpose and makes the payment at a later date, usually during or after the hotel stay. As a result, there is a time-lapse between online reservation and the actual payment. Therefore, the number of online hotel bookings, instead of monthly hotel revenue, is supposed to be more directly impacted by discontinuous changes in hotel ratings. Monthly hotel revenue, on the contrary, as a proxy for online hotel booking, may be somewhat noisy because it includes revenues earned from online bookings that are made in the same month as well as any other prior months. Because of the unavailability of information regarding when each guest makes an online reservation, the impact on hotel revenue may be underestimated in this study. I try to address the concern in this section.

TripAdvisor (2018) reports 84% of their visitors make online hotel reservations through their site within the same month of arrival at the hotel. This leaves only 16% of the hotel guests making hotel reservations more than a month before their hotel stay. Therefore, the identification strategy of this study using monthly hotel revenue should not be too much biased. I have conducted further robustness check using quarterly revenue data and estimated the same regression discontinuity model as represented by Equation 2. Table 2.7 reports the regression results and shows the impact of a 0.5-star increase in ratings on quarterly hotel revenue is the same as what I found with monthly revenue data. The regression discontinuity estimate using both the monthly and quarterly hotel revenue data suggests a 1.1% increase in revenue for a 0.5-star improvement in the hotel rating.

7. Conclusion

This paper provides evidence for financial implications of online customer-ratings at a leading travel website, TripAdvisor.com. The study uses a regression discontinuity design to show TripAdvisor ratings have a causal impact on hotel revenue. This study finds an exogenous 1-star improvement in TripAdvisor rating increases a hotel's monthly revenue by 2.2 - 3%. For an average hotel, this is equivalent to range of additional $4,593 - $6263 monthly revenue or $55,117 - $75,159 yearly revenue. The causal impact is robust across different bandwidths. The study also presents results indicating a higher impact of TripAdvisor ratings as the number of reviews increases. The hotels with more than 50 reviews earn a 3.5% higher monthly revenue compared to the hotels with 10 or fewer reviews. I have also attempted to address the concern associated with travelers booking their hotel rooms more than a month before their actual date of arrival, which could bias the regression results due to the use of monthly revenue data. However, by estimating the regression discontinuity model using quarterly revenue data, I find no changes in the regression discontinuity result. Overall, the study contributes to the literature of online reputation and business performance, particularly in the hotel industry, and provides a quantitative measure for the impact of TripAdvisor ratings on hotel revenue.

REFERENCES

Amblee, N. and Bui, T. X. (2007). The Impact of Electronic-Word-of-Mouth on Digital Microproducts: An Empirical Investigation of Amazon Shorts. *AIS Electronic Library, St. Gallen*, pages 36–47.

Anderson, C. K. (2012). The Impact of Social Media on Lodging Performance. Available at: https://vtechworks.lib.vt.edu/bitstream/handle/10919/85353/The%20Impact%20of%20social%20media%20on%20lodging%20performance.pdf?sequence=1&isAllowed=y. Accessed: December 22, 2018.

Blal, I. and Sturman, M. C. (2014). The Differential Effects of the Quality and Quantity of Online Reviews on Hotel Room Sales. *Cornell Hospitality Quarterly*, 55(4):365–375.

Castro, G., Saez, P., and Lopez, J. (2004). The role of corporate reputation in developing relational capital. *Journal of Intellectual Capital*, 5(4):575–585.

Chevalie, J. A. and Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3):345–354.

Duan, W., Gu, B., and Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales - An empirical investigation of the movie indusry. *Journal of Retailing*, 84(2):233–242.

Forman, C., Ghose, A., and Wiesenfeld, B. (2008). Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*, 19(3):291–313.

Frank, B. H. (2014). Word's largest travel site? Expedia and TripAdvisor both lay claim to title. Available at: https://www.geekwire.com/2014/heres-expedia-can-claim-worlds-largest-online-travel-agency/. Accessed: December 15, 2018.

Gretzel, U. and Yoo, K. H. (2008). *Use and Impact of Online Travel Reviews*, chapter :Information and Communication Technologies in Tourism 2008. Springer, Vienna.

Kim, W. G., Li, J., and Brymer, R. A. (2016). The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management*, 41-51:41–51.

Kim, W. G., Lim, H., and Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, 44:165–171.

Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70(3):74–89.

Luca, M. (2016). Reviews, Reputation, and Revenue: The Case of Yelp.Com. Available at: https://ssrn.com/abstract=1928601.

Phillips, P., Zigan, K., Silva, M. M. S., and Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management*, 50:130–141.

Roos, G., Roos, G., Pike, S., and Fernstrom, L. (2005). *Managing Intellectual Capital in Practice*. Oxford: Butterworth-Heinemann.

Statista (2018). TripAdvisor. Available at: https://www.statista.com/study/42474/tripadvisor/. Accessed: June 05, 2018.

Torres, E. N., Singh, D., and Robertson-Ring, A. (2015). Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry. *International Journal of Hospitality Management*, 50:77–83.

TripAdvisor (2018). TripAdvisor Path to Purchase. Available at: https://www.tripadvisor.com/TripAdvisorInsights/w2324. Accessed: December 29, 2018.

Tuominen, P. (2011). The influence of TripAdvisor consumer-generated travel reviews on hotel performance. *University of Hertfordshire Business School Working Paper, Presented at the 19th Annual Frontiers in Service Conference.*

Viglia, G., Minazzi, R., and Buhalis, D. (2016). The influence of e-word-of-mouth on hotel occupancy rate. *International Journal of Contemporary Hospitality Management*, 28(9):2035–2051.

Xie, K. L., So, K. K. F., and Wang, W. (2017). Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management*, 62:101–110.

Xie, K. L., Zhang, Z., and Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43(2014):1–12.

Yang, J. and Mai, E. (2010). Experiential goods with network externalities effects: An empirical study of online rating system. *Journal of Business Research*, 63(9-10):1050–1057.

Ye, Q., Law, R., and Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1):180–182.

Ye, Q., Law, R., Gu, B., and Chen, W. (2011). The Influence of User-Generated Content on Traveler Behavior: An Empirical Investigation on the Effects of E-word-of-mouth to Hotel Online Bookings. *Computers in Human Behavior*, 27(2):634–639.

Zhang, J. Q., Craciun, G., and Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, 63(12):1336–1341.

Zhu, F. and Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(2):133–148.

APPENDICES CHAPTER II

APPENDIX A: TABLES

Table 2.1: Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Max
Revenue (USD)	62,782	208,777	445,583	70	13,315,502
Rating	376,060	3.63	0.97	1	5
Review Count	376,060	4	7.88	0	196

Notes: All statistics are per hotel per month.

Table 2.2: Impact of TripAdvisor rating on hotel revenue

Dependent variable = ln(*Revenue*)	
Rating	0.045***
	(0.002)
Monthly Fixed Effects	Yes
Hotel Fixed Effects	Yes
Observations	62782
Hotels	1348

Note: Robust standard errors are reported in parenthesis. *, **, ** indicate significance at 10%, 5%, and 1% level.

Table 2.3: Regression discontinuity estimate

Dependent variable = ln(*Revenue*)			
	(1)	(2)	(3)
Discontinuity	0.011**	0.011**	0.011**
	(0.005)	(0.005)	(0.005)
Rating	Yes	Yes	Yes
Rating Quadratic		Yes	
Rating Higher Order			Yes
Monthly Fixed Effects	Yes	Yes	Yes
Hotel Fixed Effects	Yes	Yes	Yes
Observations	25085	25085	25085
Hotels	1140	1140	1140

Note: Regressions include all observations within 0.1 stars of a discontinuity. Robust standard errors are reported in parenthesis. *, **, ** indicate significance at 10%, 5%, and 1% level.

Table 2.4: Regression discontinuity for different bandwidths

Dependent variable = ln(*Revenue*)		
	(1)	(2)
Discontinuity	0.011**	.015***
	(0.005)	(0.004)
Rating	Yes	Yes
Monthly Fixed Effects	Yes	Yes
Hotel Fixed Effects	Yes	Yes
Observations	25085	12636
Hotels	1140	939
Bandwidths (stars)	0.2	0.1

Note: Column 1 and 2 report the estimates of regression discontinuity models with 0.2 and 0.1 bandwidths, respectively. Robust standard errors are reported within parenthesis. *, **, *** denote significance at the 10%, 5%, and 1% level.

Table 2.5: Heterogeneous impacts

Dependent variable = ln($Revenue$)	(1)
Rating	0.042***
	(0.002)
Rating×(11-20 reviews)	0.016**
	(0.007)
Rating×(21-30 reviews)	0.017**
	(0.008)
Rating×(31-40 reviews)	0.027*
	(0.014)
Rating×(41-50 reviews)	0.029**
	0.008
Rating×(50+ reviews)	0.035**
	(0.015)
Month Fixed Effects	Yes
Hotel Fixed Effects	Yes
Observations	62782
Hotels	1348

Notes: Robust standard errors are reported within parenthesis. *, **, *** denote significance at the 10%, 5%, and 1% level.

Table 2.6: McCrary Test for Random Reviews

Dependent Variable = Probability Mass of 0.05 Star Bin	
Treatment (0.05 star interval above rounding thresholds)	-0.0009
	(0.0007)
Number of Observations (N)	79

Note: Dependent variable is the probability mass of reviews in each 0.05 star interval. The treatment variable represents intervals just above a rounding threshold.

Table 2.7: Regression discontinuity estimate with quarterly revenue data

Dependent variable = ln(*Revenue*)	
Discontinuity	0.011***
	(0.004)
Rating	x
Monthly Fixed Effects	Yes
Hotel Fixed Effects	Yes
Observations	9502
Hotels	1141

Notes: Table 7 reports the results of the regression discontinuity model estimated using quarterly revenue data. Robust standard errors are reported within parenthesis. *, **, *** denote significance at the 10%, 5%, and 1% level.

APPENDIX B: FIGURES

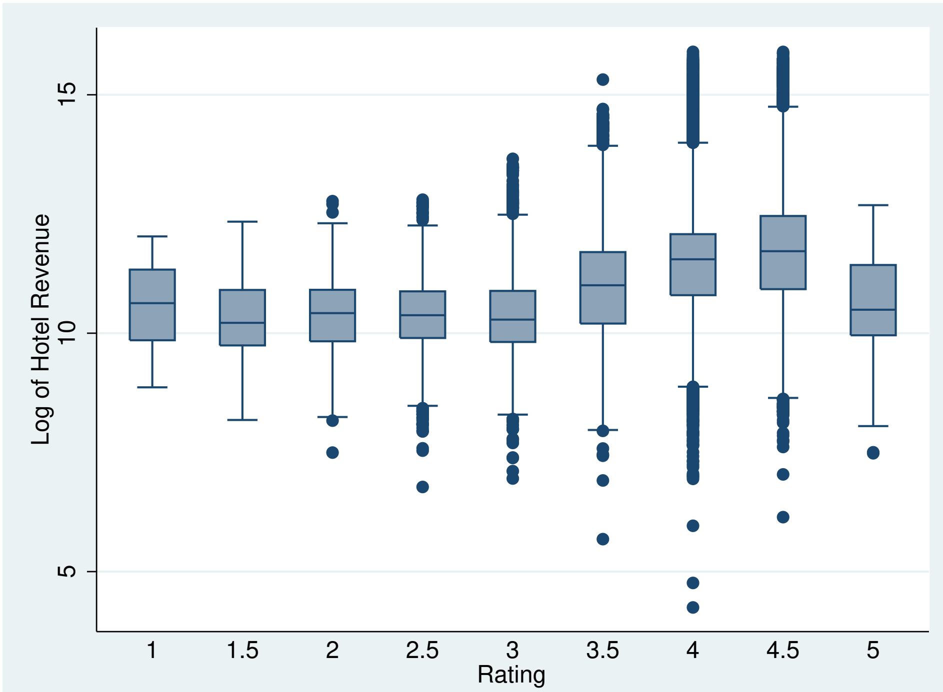

Figure 2.1: Box plot of hotel revenue at different star ratings.

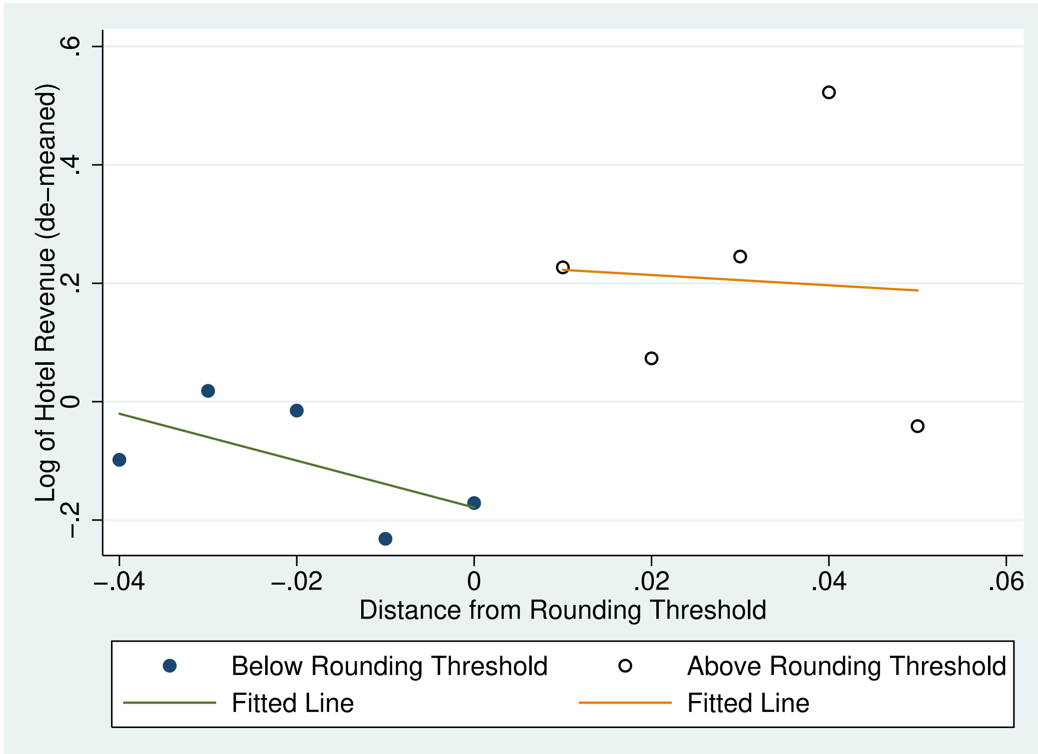

Figure 2.2: Average revenue around discontinuity.

CHAPTER III

THE ROLE OF BRAND AFFILIATION IN BUSINESS PERFORMANCE: AN INVESTIGATION INTO THE HOTEL INDUSTRY

1. Introduction

The existing literature regarding the lodging industry suggests that the brand affiliation of a hotel property is one of the important factors in its financial performance. With the help of branding strategies, both the brand-owning hotel companies and individual hotel operators are able to run viable businesses and foster growth. Because brand affiliation is a form of strategic alliance, value creation is a vital element when it comes to being affiliated with a brand (O'Neill and Xiao, 2006; Carvell et al., 2016).

Previous studies have indicated popular brands render consumers with a range of emotional and functional benefits, which positively impact consumer behavior and perceptions related to the brand. Research has also demonstrated that a brand can be an intangible asset, providing measurable financial values (Keller, 2002). Aaker (1991), using the notion of brand equity, views that both the brand-affiliated companies and consumers attach considerable value to brands. According to the view, brand equity facilitates product differentiation and offsets competition, which allows a brand-affiliated firm to maintain customer loyalty while charging a premium. Various studies have suggested the growth of the brand value is imperative in the successful operation of a business (Kapferer, 1997; Keller, 1998; Aaker, 1991; Aaker, 1996). For instance, Prasad and Dev (2000) assert brand equity is a major determinant for success in the lodging industry. There is empirical evidence that supports their claim. Kim and Kim (2005), for example, have investigated the luxury hotels and reported a significant positive association between sales and brand equity.

If brand affiliation provides value to hotels owners by means of reduced competition, increased prices, and loyal customers, all else equal, the brand affiliated hotel owners should observe a better financial return relative to their unaffiliated counterparts. Empirical findings in this regard, however, show mixed results. Ingram and Baum (1997) report that brand affiliated hotels tend to have a higher survival rate compared to unaffiliated hotels. According to Love et al. (2012), when unaffiliated hotels obtain affiliation, their revenue per available room (RevPAR) index improves. Hanson et al. (2009) suggest hotels rebranding to an upper market segment improve their performance. O'Neill and Carlback (2011) find that the occupancy rates of brand affiliated hotels are significantly higher, on average, compared to their unaffiliated counterparts. Conversely, research also shows unaffiliated hotels enjoy a significantly higher RevPAR and average daily room rates (ADR) (O'Neill and Carlback, 2011).

In the existing literature of brand affiliation and hotel performance, the wide-ranging opinions and contradictory findings warrant further investigation into the role of brand affiliation in hotel performance, particularly by comparing between affiliated and unaffiliated hotels. In addition, we find the current literature mostly utilizes cross-sectional hotel data; and their analysis mainly includes hypothesis testing and analysis of variance (ANOVA) to study the relationship between hotel brands and performance. Our endeavor, therefore, is to contribute to the literature by using richer data (i.e., longitudinal data) and more sophisticated empirical approach in order to present conclusive results.

In this paper, we present a comparative analysis between brand affiliated and unaffiliated hotels by studying 450 hotels in Texas that had a change of ownership between 2014 and 2017. We examine whether a hotel had a statistically significant difference in its revenue following a change of ownership, which may or may not have coincided with a change of brand affiliation. This means ownership change can happen in one of the following four different ways: independent to independent (hence, remains unaffiliated), independent to affiliation, affiliation to independent, and affiliation to affiliation (which means the hotel either keeps or changes its original brand). In particular, we compare the financial implications of brand affiliation by inspecting all the four scenarios above. We ask: do new hotel owners generate higher revenues when they obtain brand affiliation for their previously unaffiliated hotels, and vice versa?

This paper is organized as follows. Section 2 discusses the data sources and data collection procedure, section 3 outlines and explains the empirical approach, section 4 discusses the results, section 5 elaborates on the limitations of this study, and section 6 concludes.

2. Data

The data for this study are collected in two parts. We have collected revenue data from the Texas Comptroller of Public Accounts. This data set contains information on when, if any, each hotel had an ownership change, alongside other basic information, such as hotel name, address, and the number of rooms. For tax purposes, Texas law defines a hotel ". . . to be any building in which members of the public rent sleeping accommodations for $15 or more per day." As a result, the Airbnb properties and any other vacation rentals that comply with the Texas tax code are also reported in the revenue data set. For this study, I only include hotel accommodations that had a change of ownership in the 2014-2017 period. The number of such hotels at this stage is 499.

We have also collected another data set from STR Inc. that includes information related to hotels' address, phone number, open date, brand affiliation, market segment, price segment, and other hotel characteristics. Note that all variables in this data are time invariant.

After collecting both the data set, we combine them based on their property address. The final data set contains 450 hotels, as 49 of the hotels from revenue data were missing in the STR data set.

Table 3.1 reports summary statistics for the sample of hotels in our data. The average monthly RevPAR for a hotel is $1346.08. Table 3.1 also reports summary statistics for hotel categories based on their room price. STR categorizes hotels into five classes based on their average room prices - also known as average daily rates (ADR) - compared to other hotels in the same market. These categories are:

- Luxury: Top 15% average room rates
- Upscale: Next 15% average room rates
- Mid-Price: Middle 30% average room rates
- Economy: Next 20% average room rates
- Budget: Lowest 20% average room rates

The average monthly RevPAR for luxury hotels is $3550.27, whereas the budget hotels' average monthly RevPAR is only $528.26.

3. Empirical Specification

We begin with the following regression with hotel fixed-effects to initially examine the impact of change of ownership, regardless of brand affiliation status, on hotel revenue:

$$\ln(RevPAR_{it}) = \beta \ ownership_{it} + \gamma_i + \delta_t + \epsilon_{it} \tag{1}$$

where $\ln(RevPAR_{it})$ denotes log of RevPAR for hotel i in month t; $ownership_{it}$ is a dummy variable denoting a change of ownership for hotel i in month t; γ_i denotes time-invariant hotel characteristics; and δ_t denotes time-variant factors (i.e., year and month). The coefficient of interest in equation (1) is *ownership*, which, if positive and statistically significant, would indicate a change of ownership has a positive impact on hotel revenue. However, it is also possible that ownership change is correlated with other unobservable factors that are associated with hotel revenues. This empirical strategy, therefore, suffers endogeneity bias, which we intend to address as we continue to collect more data for further investigation in the near future.

Next, we estimate the impact of brand affiliation on hotel revenue. In doing so, we use the same empirical specification as above but with different subsets of hotels. For brand affiliated hotels, we construct the following two samples: (a) hotels that do not change brand affiliation after an ownership change and (b) hotels that switch to independent or unaffiliated status following its ownership change. Likewise, for initially independent or unaffiliated hotels, we construct the following samples of hotel: (c) hotels that maintain unaffiliated status followed by an ownership change and (d) hotels that obtain brand affiliation immediately after an ownership change. It is important to note here that in each of the four constructed samples above, all hotels undergo a change in ownership, which may or may not coincide with a change of their initial affiliation status.

4. Results

Table 3.2 reports regression results based on equation (1). Each column reports the same empirical specification estimated using different samples of hotels. Column 1 reports regression results based on the total sample of hotels, but the results between column 2 and 4 are estimated using samples of hotels depending on how their affiliation changed followed by an ownership change. Column 2, 3, 4, and 5 report ownership change from, respectively, affiliated hotels to affiliated hotels, affiliated hotels to independent hotels, independent hotels to independent hotels, and independent hotels to affiliated hotels. We find, on average, a hotel's RevPAR increases by 11.8% after its ownership change. Looking at the sample of hotels that remain affiliated, we find an ownership change increases their RevPAR by 15.5% on average. The RevPAR increase is highest (28.8%) when an independent hotel becomes an affiliated hotel after its ownership change. In all of the above cases, the coefficients are statistically significant at 1% level. However, for hotels that remain independent (or unaffiliated) or convert from affiliated to independent status, a change of ownership does not have any statistically significant effect on their RevPAR. Overall, the results indicate brand affiliation does have a positive effect on the revenue of a hotel.

5. Limitation

Our results suffer endogeneity bias due to unobserved factors that lead to some hotels undergoing an ownership change as well as a change of affiliation status. We are continuously looking to improve this study by incorporating more sophisticated analysis and completing the existing data set with more data, such as review data. We believe review data may provide necessary insights related to how a hotels quality changes over time, including before and after an ownership change, and whether the hotel underwent a renovation during the ownership change. At this stage, our results only indicate a positive association, not causation, between brand affiliation and hotel revenue.

6. Conclusion

Overall, in this paper, we have investigated 450 hotels in Texas that had a change of ownership between 2014 and 2017. Alongside the ownership change, some hotels changed their affiliation status, becoming an independent hotel, and vice versa. Other hotels maintained their original status- independent or affiliated - after their ownership change. This study investigated whether changes in affiliation during or after an ownership change has any impact on the hotels' revenue. By estimating fixed-effects regressions, our results suggest brand affiliation enhances hotel revenue. For instance, within our sample of hotels, we find when an independent hotel becomes an affiliated hotel after its ownership change, its monthly RevPAR increases by 28.8% on average.

On the other hand, we do not find any statistically significant improvement of monthly RevPAR for hotels that give up their affiliation status and become independent hotels. Our results support previous findings that brand affiliation boosts the financial performance of a business. Although the empirical strategy used in this paper suffers endogeneity bias due to unobserved factors not being accounted for, the results indicate necessary insights to further the study and contribute to the existing literature. We plan to collect more data and incorporate more sophisticated empirical approach in order to address the limitations of this paper.

REFERENCES

Aaker, D. A. (1991). *Managing Brand Equity.* New York, NY: Free Press.

Aaker, D. A. (1996). *Building strong brands.* New York, NY: Free Press.

Carvell, S. A., Canina, L., and Sturman, M. C. (2016). A Comparison of the Performance of Brand-Affiliated and Unaffiliated Hotel Properties. *Cornell Hospitality Quarterly*, 57(2):193–201.

Hanson, B., Mattila, A. S., O'Neill, J. W., and Kim, Y. (2009). Hotel Rebranding and Rescaling. *Cornell Hospitality Quarterly*, 50(3):360–370.

Ingram, P. and Baum, J. A. C. (1997). Chain Affiliation and the Failure of Manhattan Hotels, 1898-1980. *Administrative Science Quarterly*, 42(1):68.

Kapferer, J.-N. (1997). *Strategic Brand Management: Creating and Sustaining Brand Equity Long Term.* London, England : Kogan Page.

Keller, K. L. (1998). *Strategic brand management: Building, mea- suring and managing brand equity.* Upper Saddle River, NJ : Prentice Hall.

Keller, K. L. (2002). *Branding and Brand Equity.* SAGE Publications Ltd.

Kim, H. B. and Kim, W. G. (2005). The relationship between brand equity and firms' performance in luxury hotels and chain restaurants. *Tourism Management*, 26(4):549–560.

Love, A. S., Walker, B. H., and Sutton, D. W. (2012). New Option in Hotel Appraisals: Quantifying the Revenue Enhancement Value of Hotel Brands. *The Appraisal Journal*, 80(3):223–234.

O'Neill, J. W. and Carlback, M. (2011). Do brands matter? A comparison of branded and independent hotels' performance during a full economic cycle. *International Journal of Hospitality Management*, 30(3):515–521.

O'Neill, J. W. and Xiao, Q. (2006). The Role of Brand Affiliation in Hotel Market Value. *Cornell Hotel and Restaurant Administration Quarterly*, 47(3):210–223.

Prasad, K. and Dev, C. S. (2000). Managing Hotel Brand Equity. *Cornell Hotel and Restaurant Administration Quarterly*, 41(3):22–31.

APPENDICES CHAPTER III

APPENDIX A: TABLES

Table 3.1: Summary Statistics

Hotel	Mean	Std.Dev	Min	Max	No.of Hotels	N
All	1346.08	1195.59	0.27	19846.99	450	16884
Budget	528.26	416.54	0.27	5383.34	152	5808
Economy	1073.83	603.02	29.25	3573.20	106	3828
Midprice	1781.71	1192.55	129.37	19846.99	111	4080
Upscale	2234.60	713.45	169.25	5497.90	58	2256
Luxury	3550.27	1996.33	179.05	14722.94	23	912

Table 3.2: Effect of Changing Ownership on the Revenue

Dependent Variable = ln($RevPAR$)					
Independent Variables	All Data	Affiliation to Affiliation	Affiliation to Independent	Independent to Independent	Independent to Affiliation
	(1)	(2)	(3)	(4)	(5)
Ownership	0.112 ***	0.144 ***	0.001	0.399	0.253 ***
	(0.015)	(0.018)	(0.029)	(0.264)	(0.044)
Year Fixed Effects	x	x	x	x	x
Month Fixed Effects	x	x	x	x	x
No. of hotels	450	269	105	7	75
N	16884	9780	4068	240	2568

Notes: *, **, ** indicate significance at 10%, 5%, and 1% level. Robust standard errors are reported in parenthesis.