The Pennsylvania State University
The Graduate School
Smeal College of Business

ESSAYS ON REPUTATION AND CAPITAL EXPENDITURE IN REAL ESTATE

A Dissertation in
Business Administration
by
Sergio Rodolfo Gárate Alvarez

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2018

The dissertation of Sergio Rodolfo Gárate Alvarez was reviewed and approved* by the following:

Brent Ambrose
Smeal Professor of Risk Management
Head of the Graduate Program
Dissertation Adviser

Jiro Yoshida
Associate Professor of Risk Management

Liang Peng
Associate Professor of Risk Management

Peter Newberry
Assistant Professor of Economics

Austin Jaffe
Sieg Professor of Business Administration (Emeritus)

*Signatures are on file in the Graduate School.

Abstract

Real estate has multiple real options embedded, which include managers' decisions to prepay a loan, delay investment, or expand or close sections of a building. The option to invest in capital improvement in a building is also a real option and understanding what triggers such improvements is critical to the industry.

The following essays relate to three areas in real estate. In the first essay, I develop a real option model to address the impact an alternative use of a property and its market conditions on capital expenditures. I show using actual building capital expenditure information how alternative use market conditions affect improvement decision of an operating building. In the second essay, I analyze how online reputation systems affect the decision to invest in building improvements. I use online consumer reviews to show that reputation affects capital improvement decisions and I also provide evidence that capital expenditures improve the building's reputation. In the final essay, I create an index using consumer-generated content and show that the index is a substantial source of information for managers to include in prepayment and default models.

Table of Contents

List of Figures

List of Tables

Dedication

This dissertation is dedicated to my brilliant and supportive wife, Alissa Blair, and to our sweet and smart children, Sofía and Benjamín.

I also dedicate this work to my parents, Miriam and Sergio, my sister, Marcela, and my extended family for their support throughout this journey.

Finally, I dedicate this work to my friends with whom I had wonderful and inspiring conversations.

Chapter 1 Introduction

1.1 Motivation

Real estate property managers are constantly making decisions to maximize the value of a project. These are real options that include expansions, delay improvements, prepayment of a loan, among others. Empirical evidence on this topic relates capital expenditure to market conditions of the highest best use, but leaves out the conditions of alternative use. It also leaves out reputation considerations of the buildings among consumers. Today, consumer information is easily available for decision making. For example, when buying products online consumers can read about other consumers' experiences with the product and observe a rating. In real estate you can find the same phenomenon: consumers evaluate hotel buildings or apartment complexes and new potential tenants read those comments before leasing an apartment or staying at a hotel. In this work, I address these topics by providing empirical evidence that market conditions of alternative building use and online consumer reviews have a statistically significant impact on the investment decision of capital expenditure.

Real options models on capital improvements generally focus on market conditions of the highest best use (Bond et al., 2014) or the cost of operating an asset (Mauer and Ott, 1995). This leaves out the uncertainty of the alternative use of the property or the alternative use of the asset. In this work, I propose an option price model that includes the uncertainty of the alternative use. I then provide empirical evidence for the hypotheses derived from the model. Although the model considers price of the highest best use and second highest best use, it leaves out consumer considerations.

Nowadays a number of websites provide information about products, services, buildings, or buildings' location. In the real estate industry these online services focus primarily on restaurants (i.e. Yelp), hotels (i.e. TripAdvisor), apartments complexes (i.e. apartment.com or google.com for examples of reviews). This information may help consumers to make informed decision and there is empirical evidence that suggest that this is the case(Anderson and Magruder, 2012), but there is no empirical evidence that managers would respond by altering the capital expenditure. In this work, I address this and also test whether capital expenditure affects future buildings' reputations. This information transfered among consumers may have other potential uses which I also explore.

In online reviews, consumers evaluate locations of buildings which contain information about areas that go beyond the building itself. For example, there is evidence that google searches about mortgage default leads economic data and can be used to asses mortgage riskChauvet et al. (2016). I use location reviews to build a location index that allows me to measure mortgage risk at more refined geographies, including at the 5-digit zip code level and at various frequencies (i.e. weekly or monthly). This is an advantage to other public sources of economic information that frequently become available with lags and at larger geographic levels.

1.2 Contributions

This work makes the following contributions to the literature:

1. It extends the option pricing literature by providing an analytical solution to the option price model of capital expenditure under the presence of an stochastic alternative use. The model uses market parameters such as capitalization rate, price volatility and outside price index.

2. It provides empirical evidence on the impact an alternative use has on the decision to invest in capital expenditure in an operating building by using a panel data with capital expenditure information for more than 20,000 buildings.

3. It provides empirical evidence on the impact that consumer reviews has on the

decision to invest in capital expenditure by using a novel dataset of consumers' online reviews and capital investment decision in hotels.

4. It shows how capital expenditure alters the reputation of a hotel among consumers.

5. Introduces a new index of consumer location evaluations that captures local economic conditions.

1.3 Organization

This work is organized into three Chapters. Chapter 1 provides evidence on the impact that alternative use has on the decision to invest in capital expenditure. I use option price theory to estimate value of the option to invest in capital expenditure. I then test the hypotheses derived from the model using actual data on capital expenditure for more than 20,000 buildings held by Real Estate Investment Trusts. In Chapter 2, I study the impact consumer reviews have on the decision to invest in capital expenditures. I use the hospitality industry as my experimental setting and merge capital expenditure of hotels and consumer reviews to test the hypotheses. In Chapter 3, I provide further evidence of how consumer-generated content helps explain other real options available to managers like loan prepayment or default.

This work also includes two appendices. Appendix A describes the derivations for equations I use in Chapter 1. Appendix B describes the derivation for a reputation model used to derive hypotheses in Chapter 2. I provide a discussion of the findings at the end of each chapter.

Chapter 2 Capital Improvements and the Role of an Alternative Use for Commercial Properties: An Option Analysis

Option pricing literature links capital improvement to market conditions; higher or growing lease rates trigger capital expenditure. Recent studies provide evidence of this relationship using option pricing models to explain this phenomenon. These models, however, leave out the alternative use allowed by land use regulation. In this paper, I develop an option price model that takes into account the impact that alternative land use has on the decision to invest in capital expenditure. Appealing to the idea introduced by Geltner et al. (1996) on the impact that land use regulation has on land value, I focus on the alternative use of a building that is operating. I introduce to the model a depreciating building that requires capital expenditure in order to stay functional. I use a free boundary representation of the problem to estimate the value of the option. I then estimate an ordinary least square model using capital expenditure data from 20,929 buildings owned by REITs. I show that parameters from the option price model related to the alternative use such as volatility, capitalization rate, and price have a significant impact on the decision to invest in capital expenditure.

2.1 Introduction

Each real estate firm has the option to invest in a depreciating building in order to keep the property functional and up to standard. Buildings can suffer physical and functional, as well as economic, depreciation. A good example of the physical decay of a property due to use and time would be a leaking roof. Functional obsolesce is attributed to changes in standards for highest best use, in other words, changes in taste of potential tenants or technological advances in the industry. An example of functional depreciation is the introduction of WiFi in hotels. 20 years ago it was normal and acceptable to not offer a wireless network for guests to access; today, this is not acceptable to most clients. Finally, economic depreciation refers to changes in the conditions that made the property suitable for the actual highest best use. For example, the highest best use of a property may evolve from an office building to retail or hotel.

Real estate managers are constantly making decisions about properties in their portfolios such as which property to sell, remodel, expand, or redevelop. This paper takes a close look at the option to use capital expenditures (Option CAPEX) to make a depreciating property functional again. I provide insight as to the optimal decision to exercise the option on a depreciating building under the presence of an outside option to sell the property. I assume the firm that owns a building has the expertise in a property type that represents the highest best use (HBU) for the building. If the HBU changes, the REIT can sell the property to a developer who then transforms the building into a new HBU. This opens the question of what impact this outside option to sell the building would have on capital expenditure decisions. More specifically, I ask: what is the impact of having a brother alternative land use on capital expenditures of an existing building?

This paper expands the literature on real options by introducing the Capex optimal decision under the presence of construction cost, rent and alternative land use price risk. Previous literature on real options has focused on land value under uncertainty. This is the case in Titman (1985) where he focuses on a binomial option value to price land and the optimal development policy. Capozza and Helsley (1990) introduce stochastic demand along with rents and price as a source of risk to the determined land values. Given the irresistibility of the investment in the model developed by Capozza and Helsley (1990), uncertainty generates a premium that translates into a delay in conversion of land, an increase in the value of agricultural land at the edge of the city, and a reduction of the city size. One drawback of these models is that they rest on the assumption of a finite expiration of the development option.

A solution to the finite expiration problem is the model proposed by Williams (1991). The author uses a close form solution developed by Samuelson and McKean (1965) to price a perpetual real option of a developer that chooses the scale of the project in which both rents and construction costs are uncertain. Williams (1991)' findings suggest that uncertainty affects both the timing and scale of the development. Geltner et al. (1996) proposed a model in which developers have the option not only to select the scale but also the land use. The departure from previous models is that a developer now has the option to select from two stochastic rent processes. In other words, the developer has the option to select the land use of the project along with the scale. Their analysis suggests that the option to select the land use has significant contributions to the value of the property. Moreover, the correlation between the underlying stochastic processes of rents influences the decision to invest where fewer correlated processes result in a delay of the development. The models presented so far focus on the development option a real estate manager has, but leave out the options present after the project is installed.

Property managers have multiple options after a project has been developed. For example they could close or expand a project [1] or reinvest in it to replace the depreciated part of the project. Mauer and Ott (1995) introduce a model in which the cost of operating an asset is stochastic and increases as the asset depreciates. Mauer and Ott find that cost uncertainty discourages replacement investment and therefore increases the cycle in between replacements. Other important findings are that the replacement cycle is increasing in the price of the asset and decreasing in the salvage price. Depreciation impact is inconclusive as it could increase or decrease the replacement cycle. The setup proposed by Mauer and Ott relates primarily with cost of machinery and assets used in the production of services. Real estate assets may behave relatively

[1] Dixit and Pindyck (1994) present a survey on capital budgeting using contingent claims in a real option framework.

differently as the focus is not usually on the cost but rather capacity of generating rent.

Chetty (2007) establishes a framework that relates the impact of interest rate to the capitalization rate. This framework suggests that interest rates have two effects on investment. The first effect is that lower interest rates decrease the cost of capital and therefore stimulate investment. The second effect is that lower interest rates reduce interest expenses thus making more valuable the option to delay investment. Peng and Thibodeau (2018) provide empirical evidence using 12,000 commercial properties across time and metro from the National Council of Real Estate Investment Fiduciaries. Their findings suggest that decreasing the interest rate has a weaker impact on stimulating investment when interest rates are low and capitalization rates are low.

Real options models on capital improvement for real estate properties have received relatively little attention, with the exceptions of Bond et al. (2014) and Ambrose and Steiner (2017). Bond et al. propose a model in which there are two sources of risk: long term and short term lease rates, where the short term follows an Ornstein-Ihlenbeck (mean reverting) process that reverts back to a long term lease rate. The property in this model produces a revenue that depends on the interaction of a short term lease rate and the quality of the property which depends on the stock of maintenance minus a constant depreciation. The value of the project then is the expected present value of the profit function of the interaction of quality and lease rate minus maintenance for the period. The findings suggest that high or rising lease rates incentivize capital expenditure in order to capture profits and defer investments during periods of low or decreasing lease rates. One caveat of their findings is that during low or decreasing lease rates, the capital investments are as profitable as in the cases of low lease rates. The model presented, although informative, leaves an interesting question to be answered as to role of depreciation and eventually the impact of having an outside option to sell the property for redevelopment.

The next section introduces the real option model for a depreciating property that is exposed to the risk of lease rates from the current use of the property as well as the risk of an outside option, depreciation and construction cost. Then I present a brief description of the computational methodology to find a numerical solution for the price of the option.

2.2 Model Development

There are three potential risks associated with the option. First, the rental rate of the HBU is represented by R. Second, CAPEX is considered uncertain and is described by C. Finally, the value of a building for an alternative use for the property is also uncertain and is represented by the value of S.

$$\begin{aligned} \frac{dC}{C} &= \quad \theta dt + \sigma_C dz_C \\ \frac{dR}{R} &= \quad \alpha dt + \sigma_R dz_R \\ \frac{dS}{S} &= \quad \gamma dt + \sigma_S dz_S \end{aligned} \tag{2.1}$$

where dC describes the process followed by CAPEX, dR and dS describe the processes followed by the highest best use rent and the second highest best use value respectively[2]. The rent of the HBU, the value of the alternative use building and CAPEX investment follow geometric Brownian motion processes with a drift. The value of $[\theta, \alpha, \gamma]$ are the respective drift parameters and the values of $[\sigma_C, \sigma_R, \sigma_S]$ are the variance parameters. The values of $[dz_C, dz_S, dz_R]$ represent the increment of a Weiner process, where $dz_i = \epsilon_{it}\sqrt{dt}$ for all $i \in [R, S, C]$. The distributions of ϵ_{it} are the standard normal distribution, therefore $\varepsilon(dz_i) = 0$ and $\varepsilon((dz_i)^2) = dt$. The correlation between dz_R and dz_C is ρ_{RC}[3].

The options available to the portfolio manager given the risks faced can be simplified to three: the sell option or outside option (O); the option to keep the depreciating property and not invest in CAPEX (woCapex); and the option to keep and invest CAPEX on the property (Capex). For the lower bound of the option value the assumption is that the portfolio manager sells a call option on the building which depends on the price of the SHBU price that will be exercised whenever the value

[2] Here for simplicity the highest best use and the second highest best use (SHBU) represent one type of property, example office and retail respectively.

[3] I will define the correlation between the HBU and SHBU in the next section.

of the option to keep the property is lower than the outside option. Let's represent the rent threshold when the sell option is executed as R_L. Whenever the rent of the highest best use drops below R_L, the property manager sells the property. Another important threshold is when the rent R increases up to a rent R_H which represents the threshold when CAPEX investment is the optimal decision. Whenever the rent is between R_H and R_L it is best to keep the property and not invest in CAPEX, and to keep the option to invest alive.

In order to solve the problem, I will divide the options into two different problems. The first one is the outside option ($O(S)$) and second one is when the building is above the R_L and is the depreciated building with the option to invest in Capex ($V(R,C)$). The real estate manager problem can then be described by:

$$F(V,O) = \max[V(R,C), O(S)] \tag{2.2}$$

where $V(R,C)$ represents the value of the project with the option to invest in Capex in order to recover a property and $O(S)$ represents the value of the option to invest in an alternative building use (S).

Following the options available from low rent to high rent R, I start valuing the $O(S)$ option, that is, when the rent is below the R_L threshold. The call option on the building associated to the rent of SHBU represents a lower bound of the options available to the portfolio manager. The $O(S)$ investment opportunity is then evaluated using the price process described by dS. The value of such an option is given by the following formula[4]:

$$O(S) = \frac{(\beta_s - 1)^{(\beta_s - 1)}}{\beta_s^{\beta_s}} \left[\frac{S}{D}\right]^{\beta_s} \tag{2.3}$$

with

$$\beta_s = \frac{1}{2} - \frac{r_f - \delta_s}{\sigma_S^2} + \sqrt[2]{\left[\left(\frac{r_f - \delta_s}{\sigma_S^2}\right) - \frac{1}{2}\right]^2 + 2\frac{r_f}{\sigma_S^2}} \tag{2.4}$$

where r_f is the risk free rate and δ_s is the discount rate for the project (ϱ_S) minus the growth of S, γ. Interestingly enough, the option to invest in the alternative use,

[4]Appendix A.0.0.1 provides the derivation for equations 2.3, 2.4, 2.5.

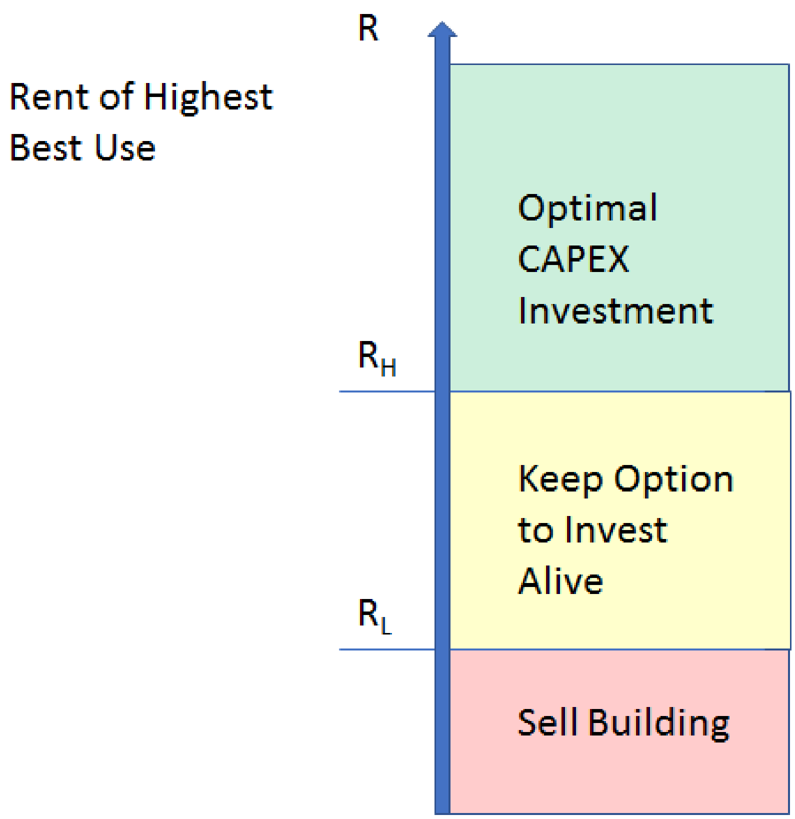

Figure 2.1. Problem Structure

This is a graphical representation of the options the manager has when operating a property. R_L is the trigger rent at which the manager decide to maintain the depreciated property rather than sell it for a transformation to the second highest best use. The middle area between R_L and R_H represents the area in which the manager keep the option to invest in capital expenditure alive instead of selling the property or investing in capital expenditure to recover the property. The area above the trigger rent R_H is where the manager invest in capital expenditure to recover the property.

$O(S)$, also evolves as geometric Brownian motion with the following characteristics:

$$\frac{dO}{O} = \gamma_O dt + \sigma_O dZ_S \tag{2.5}$$

where $\gamma_O = [\beta_S \gamma + \frac{1}{2}\beta_S(\beta_S - 1)\sigma_S^2]$ and $\sigma_O = \beta_S \sigma_S$.

In addition, the option to invest capital expenditure on a depreciated building is $V(R, C)$, where the value of such an option is[5]:

$$V(R, C) = C\frac{(\beta_V - 1)^{\beta_V - 1}}{\beta_V(\lambda + \delta_R)^{\beta_V}}[\frac{R}{C}]^{\beta_V} \tag{2.6}$$

with

$$\begin{aligned}\beta_V = &\frac{1}{2} - \frac{\delta_C - \delta_R}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}} + \\ &+ \sqrt[2]{\left[\left(\frac{\delta_C - \delta_R}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}}\right) - \frac{1}{2}\right]^2 + 2\frac{\delta_C}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}}}\end{aligned} \tag{2.7}$$

where δ_C and δ_R are equal to the discount rates for Capex (ϱ_C) and the project (ϱ_R) minus the corresponding growth rates of C and R, θ and α. The value of V follows the following geometric Brownian motion:

$$\frac{dV}{V} = \gamma_V dt + \sigma_V dz_V \tag{2.8}$$

where $\gamma_V = [\beta_1\alpha + (1 - \beta_1)\theta + \frac{1}{2}\beta_1(\beta_1 - 1)(\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc})]$ and $\sigma_V = \beta_1\sigma_R dz_R + (1 - \beta_1)\sigma_C dz_C$. Equation 2.8 simplifies the problem faced by the manager in which the manager needs to pick the highest value option to hold.

2.2.1 Manager's Problem

In order to solve the problem and find an analytical solution, I define the option in a way that will reduce the problem to one decision variable, that is, the ratio between the two processes (V/O), as described in Figure 2.2. Whenever the ratio exceeds a threshold, the managers will prefer to hold the option to invest in the property rather

[5] Appendix A.0.0.2 provides the derivation for equations 2.6, 2.7, 2.8.

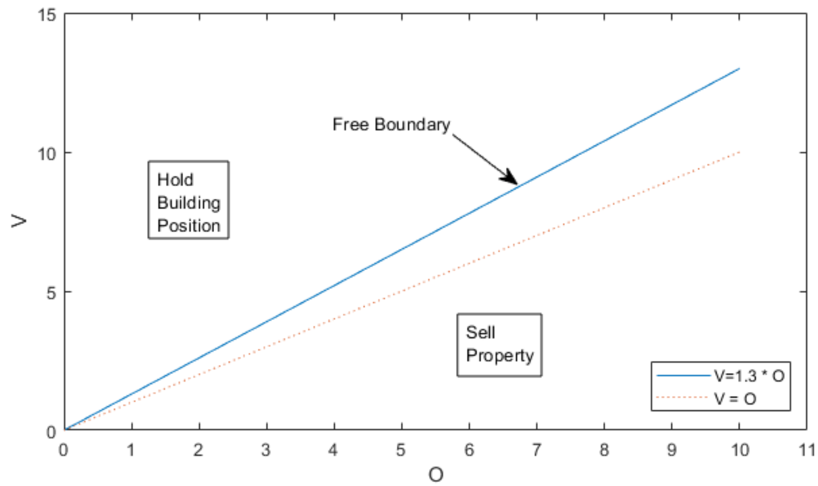

Figure 2.2. Holding Decision with Alternative Use

This is the free boundary problem face by the manager. When ever the value of V exceed O by a certain ratio the manager keeps the property rather than selling the property to the second highest best use. The diagonal represents the free boundary condition that represents the ratio at which manager exercise the option to maintain the building.

than holding the option to invest in the second best use. It is clear that the problem described in Figure 2.2 is homogeneous; whenever the value of V exceeds O by a certain ratio, the manager holds the building option rather than the outside option. Now if both V and O triple in value, the ratio between them would remain the same, and the decision of which option to hold would not change. Now for simplicity lets call the option to invest in either investment opportunity F which depends on the values of V and O, $F(V,O)$.

To find the value of $F(V,O)$, I construct a portfolio(π_{VO}) that holds one unit of F, and holds a short position of $\partial F/\partial V$ units of V and $\partial F/\partial O$ units of O[6]. By Ito's lemma the value $d\pi_{VO}$ can be represented by the following equation:

[6]This methodology assumes that the value of a building can be spanned or synthesized by other securities in the market. In other words, I am assuming complete markets Cochrane (2009). This is a strong assumption and in the presences of incomplete markets this portfolio would not be feasible. For a complete discussion on incomplete markets please refer to Magill and Quinzii (2002)

$$
\begin{aligned}
d(F-\frac{\partial F}{\partial V}V-\frac{\partial F}{\partial O}O) &= \frac{1}{2}[\frac{\partial^2 F}{\partial V^2}V^2\sigma_V^2+\frac{\partial^2 F}{\partial O^2}O^2\sigma_O^2+2\frac{\partial^2 F}{\partial O\partial V}OV\sigma_O\sigma_V\rho_{rc}]dt+ \\
&+\frac{\partial F}{\partial V}V+\frac{\partial F}{\partial O}O-\frac{\partial F}{\partial V}V-\frac{\partial F}{\partial O}O \\
&= \frac{1}{2}[\frac{\partial^2 F}{\partial V^2}V^2\sigma_V^2+\frac{\partial^2 F}{\partial O^2}O^2\sigma_O^2+2\frac{\partial^2 F}{\partial O\partial V}OV\sigma_O\sigma_V\rho_{vo}]dt
\end{aligned}
\tag{2.9}
$$

where ρ_{vo} is the correlation between the V and O. Equation 2.9 shows that return of the portfolio during a short interval of time (t, dt) would result in a capital gain equal to:

$$
\frac{1}{2}[\frac{\partial^2 F}{\partial V^2}V^2\sigma_V^2+\frac{\partial^2 F}{\partial O^2}O^2\sigma_O^2+2\frac{\partial^2 F}{\partial O\partial V}OV\sigma_O\sigma_V\rho_{vo}]dt
\tag{2.10}
$$

If the manager wants to hold this portfolio, she will be required to pay the convenience yield on capital and output to the counterpart holding the long positions. Therefore, the resulting portfolio would have the following return:

$$
\begin{aligned}
r_f\pi_{VO}dt &= \frac{1}{2}[\frac{\partial^2 F}{\partial V^2}V^2\sigma_V^2+\frac{\partial^2 F}{\partial O^2}O^2\sigma_O^2+2\frac{\partial^2 F}{\partial O\partial V}OV\sigma_O\sigma_V\rho_{vo}]dt-\delta_V\frac{\partial F}{\partial V}Vdt- \\
&-\delta_O\frac{\partial F}{\partial O}Odt \\
0 &= \frac{1}{2}[\frac{\partial^2 F}{\partial V^2}V^2\sigma_V^2+\frac{\partial^2 F}{\partial O^2}O^2\sigma_O^2+2\frac{\partial^2 F}{\partial O\partial V}OV\sigma_O\sigma_V\rho_{vo}]+(r_f-\delta_V)\frac{\partial F}{\partial V}V+ \\
&+(r_f-\delta_O)\frac{\partial F}{\partial O}O-r_fF
\end{aligned}
\tag{2.11}
$$

where r_f is the risk free rate[7]. The solution to this last second order partial differential equation is the value of the investment opportunity with two potential uses.

The solution requires some assumptions on what the boundary conditions ought to be. The following conditions must be satisfied for F to have a valid solution:

[7]Since this is a perfectly hedged portfolio, the maximum expected return for the portfolio during a short interval of time (t, dt) would be the risk free rate. Thus, $r_f\pi_{VO}dt$.

Value matching condition:

$$F(V,O) = V - O \tag{2.12}$$

Smooth Pasting Conditions:

$$\frac{\partial F(V,O)}{\partial V} = 1 \tag{2.13}$$

$$\frac{\partial F(V,O)}{\partial O} = -1 \tag{2.14}$$

The value matching conditions indicate that over the region where the manager decides to hold the option to invest in the building rather than the outside option, the value of the option is equal to the investment opportunity in HBU building minus the value of the outside option that the manager forgoes. Equations 2.13 and 2.14 indicate the smooth pasting condition with respect to V and O.

As I mentioned before, the decision to hold one investment opportunity over the other depends on the ratio between the two, $v = V/O$. The value of the option should be homogeneous of degree one in (V,O), which allows the representation of F in the following way:

$$F(V,O) = Of(\frac{V}{O}) = Of(v) \tag{2.15}$$

where f is the new function that I need to find. With this transformation we can rewrite Equation 2.11 using the following derivatives:

$$
\begin{aligned}
\frac{\partial F(V,O)}{\partial V} &= f'(v) \\
\frac{\partial F(V,O)}{\partial O} &= f(v) - rf'(v) \\
\frac{\partial^2 F(V,O)}{\partial V^2} &= \frac{f''(v)}{O} \\
\frac{\partial^2 F(V,O)}{\partial O^2} &= r^2\frac{f''(v)}{O} \\
\frac{\partial^2 F(V,O)}{\partial V \partial O} &= -r\frac{f''(v)}{O}
\end{aligned}
\tag{2.16}
$$

The resulting equation then is:

$$
\begin{aligned}
0 = &\frac{1}{2}[\frac{\partial^2 F}{\partial V^2}V^2\sigma_V^2 + \frac{\partial^2 F}{\partial O^2}O^2\sigma_O^2 + 2\frac{\partial^2 F}{\partial O \partial V}OV\sigma_O\sigma_V\rho_{vo}] + (r_f - \delta_V)\frac{\partial F}{\partial V}V + \\
&+ (r_f - \delta_O)\frac{\partial F}{\partial O}O - r_f F \\
0 = &\frac{1}{2}[\frac{f''(v)}{O}V^2\sigma_V^2 + v^2\frac{f''(v)}{O}O^2\sigma_O^2 - 2v\frac{f''(v)}{O}OV\sigma_O\sigma_V\rho_{vo}] + \\
&+ (r_f - \delta_V)f'(v)V + (r_f - \delta_O)(f(v) - vf'(v))O - r_f Of(v) \quad divide \quad by \quad O \\
0 = &\frac{1}{2}[f''(v)v^2\sigma_V^2 + v^2 f''(v)\sigma_O^2 - 2v^2 f''(v)\sigma_O\sigma_V\rho_{vo}] + (r_f - \delta_V)f'(v)v + \\
&+ (r_f - \delta_O)(f(v) - vf'(v)) - r_f f(v) \\
0 = &\frac{1}{2}[\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}]f''(v)v^2 + (r_f - \delta_V)f'(v)v + (r_f - \delta_O)(f(v) - vf'(v)) - \\
&- r_f f(v) \\
0 = &\frac{1}{2}[\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}]f''(v)v^2 + (\delta_O - \delta_V)f'(v)v - \delta_O f(v)
\end{aligned}
\tag{2.17}
$$

The new boundary conditions given the transformation recently described are:

Value Matching condition

$$\begin{aligned} F(V,O) &= V - O \\ Of(v) &= V - O \\ f(v) &= v - 1 \end{aligned} \tag{2.18}$$

Smooth pasting with respect to V and O

$$\begin{aligned} \frac{\partial F(V,O)}{\partial V} &= 1 \\ f'(v) &= -1 \\ \frac{\partial F(V,O)}{\partial O} &= -1 \\ f(v) - vf'(v) &= -1 \end{aligned} \tag{2.19}$$

Only two of the boundary conditions are independent; the third one can be derived from the other two. I use the value matching condition and smooth pasting with respect to V to find the solution to the differential equation. I also use the condition that as v tends to 0, the value of the option to select V over O also tends to zero. Therefore, the function f should have the following structure:

$$f(v) = A_f v^{\beta_f} \tag{2.20}$$

where A_f is a constant and β_f is the largest root from the fundamental quadratic equation Q. Using the solution for f in Equation 2.20 then:

$$\begin{aligned} Q &= \frac{1}{2}[\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}]f''(v)v^2 + (\delta_O - \delta_V)f'(v)v - \delta_O f(v) \\ Q &= \frac{1}{2}[\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}]A_f(\beta_f - 1)\beta_f v^{\beta_f - 2}v^2 + (\delta_O - \delta_V)\beta_f A_f v^{\beta_f - 1}v - \delta_O A_f v^{\beta_f} \\ 0 &= \frac{1}{2}[\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}](\beta_f - 1)\beta_f + (\delta_O - \delta_V)\beta_f - \delta_O \\ 0 &= \frac{1}{2}(\beta_f - 1)\beta_f + \frac{\delta_O - \delta_V}{\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}}\beta_f - \frac{\delta_O}{\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}} \\ 0 &= \frac{1}{2}\beta_f^2 + [\frac{\delta_O - \delta_V}{\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}} - \frac{1}{2}]\beta_f - \frac{\delta_O}{\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}} \end{aligned} \tag{2.21}$$

where β_f is the largest root to Equation 2.21.

$$\begin{aligned}\beta_f = \frac{1}{2} - \frac{\delta_O - \delta_V}{\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}} + \\ + \sqrt[2]{\left[\left(\frac{\delta_O - \delta_V}{\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}}\right) - \frac{1}{2}\right]^2 + 2\frac{\delta_O}{\sigma_V^2 + \sigma_O^2 - 2\sigma_O\sigma_V\rho_{vo}}}\end{aligned} \tag{2.22}$$

The last part required to estimate the function f is the constant A_f. To do so, I will use the boundary conditions previously described.

$$\begin{aligned} f'(v) &= 1 \\ A_f\beta_f v^{\beta_f-1} &= 1 \\ A_f v^{\beta_f} &= \frac{v}{\beta_f} \\ f(v) &= \frac{v}{\beta_f} \end{aligned} \tag{2.23}$$

Using the value matching condition and Equation 2.23 we get the following:

$$\begin{aligned} &\frac{v*}{\beta_f} = v * -1 \\ &v* = \beta_f v * -\beta_f 1 \\ &v* = \frac{\beta_f}{\beta_f - 1} \end{aligned} \tag{2.24}$$

This ratio separates the area where managers prefer to hold the outside option from the option to invest in capital expenditures.

$$
\begin{aligned}
f(v*) &= \frac{v*}{\beta_f} \\
A_f v*^{\beta_f} &= \frac{v*}{\beta_f} \\
A_f &= \frac{v*^{1-\beta_f}}{\beta_f} \\
A_f &= \frac{(\beta_f - 1)^{\beta_f - 1}}{\beta_f^{\beta_f}}
\end{aligned}
\tag{2.25}
$$

Thus,

$$
F(V, O) = \frac{(\beta_f - 1)^{\beta_f - 1}}{\beta_f^{\beta_f}} O[\frac{V}{O}]^{\beta_f} \tag{2.26}
$$

This last equation provides the value of having two investment opportunities available. The value depends on the ratio between the opportunity to invest in Capex to recover a depreciated building and the value of the outside option. In addition, the value of β_f is critical to the value of the option, which at the same time depends on the convenience yields of both options available to the manager, the volatility of the O and V, and their correlation.

2.2.2 Numerical Example

It is useful to examine how the value of the option changes with respect to changes in different parameters, specifically, how the value of the option changes with respect to the ratio between the HBU and the outside option, and how it changes with correlation between the two alternative investment opportunities.

Figure 2.3 illustrates the impact of correlation ρ_{vo} on the value of β_f. To estimate the effect of ρ_{vo} I hold constant the other parameters in the β_f function. Figure 2.3 shows β_f as a function of ρ_{vo} with δ_O=0.09, δ_V=0.10, and σ_O=0.50. The different curves show the value of β_f for different values of σ_O. The dotted-dashed line shows the value of β_f when the volatility of the outside option is zero, the dashed line shows the value when $\sigma_O = 0.4$, and the solid line when $\sigma_O = 0.4$. As the value of ρ_{vo}

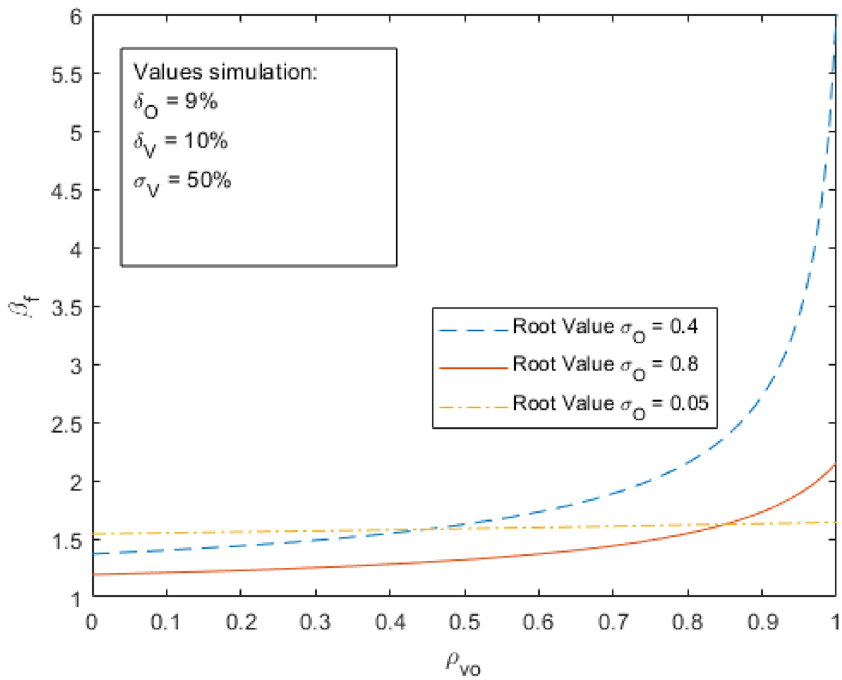

Figure 2.3. Root Value to Fundamental Quadratic Function

This graph shows the impact of the correlation between V and O, ρ_{vo}, on the positive root from Equation 2.21. δ_O and δ_V represents the value the capitalization rate of the alternative use(second highest best use) and the highest best use. σ_O and σ_V are the volatilities of the outside option and the highest best use of the property. The different lines represent different values of the outside option volatility.

increases, the value β_f also increases. The slope of the curve when ρ_{vo} approaches 1 increases significantly, indicating that the relation between ρ_{vo} and β_f is not linear. Also, when the value of σ_O is greater than 0, increases in the volatility of the outside option increase the value of β_f. It is clear that these shifts in the parameters of the outside option have an impact on the value of $F(V, O)$.

In order to further explore the impact that other parameters have on the value of the β_f, instead of varying σ_O , I vary δ_O. Figure 2.4 shows the value of β_f as a function of the correlation parameter ρ_{vo}. This time each curve represents a different value for δ_O. The dotted-dashed line shows the curve when δ_O is equal to 0, the dotted line represents δ_O equal to 0.07, the solid line indicates the case in which δ_O

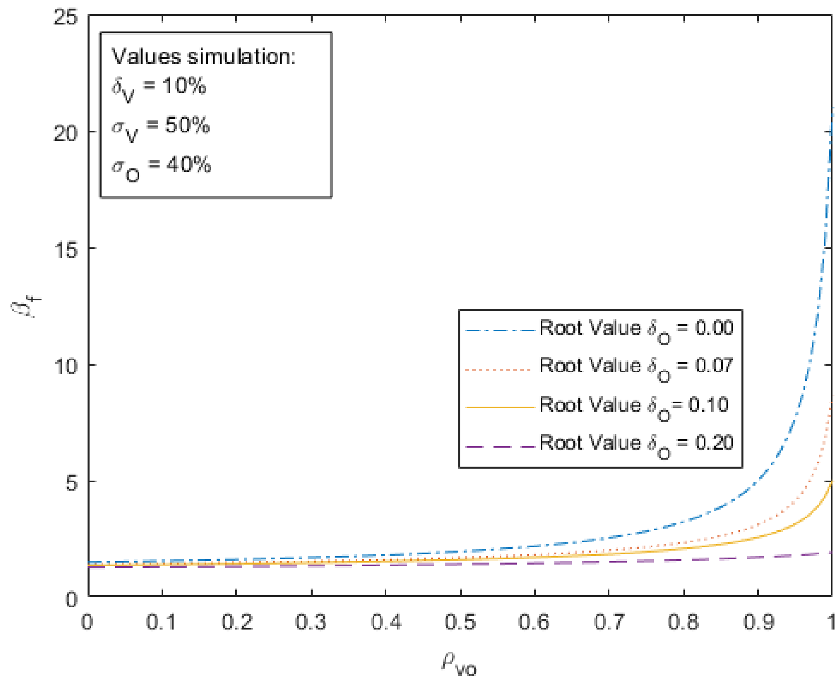

Figure 2.4. Root Value to Fundamental Quadratic Function

This graph shows the impact of the correlation between V and O, ρ_{vo}, on the positive root from Equation 2.21. δ_O and δ_V represents the value the capitalization rate of the alternative use(second highest best use) and the highest best use. σ_O and σ_V are the volatilities of the outside option and the highest best use of the property. The different lines represent different values of the root depending on the value of the capitalization rate of the outside option δ_O.

equals 0.10, and finally the dashed line indicates the case for δ_O equal to 0.20. The pattern indicates as it did before that a higher correlation coefficient ρ_{vo} increases the value of β_f. The figure also shows that increases of δ_O decrease the value of β_f. Although, the value of β_f is of interest in the analysis, the final goal is to analyze how $\rho_v o$ affects the value of the options held by the manager.

Figure 2.5 shows the impact of β_f and the ratio V/O have on the value of the option $F(V,O)$. From the picture we observe that higher values of β_f translate in lower values of $F(V,O)$. This result in conjunction with the impact that ρ_{vo} has on β_f suggest that higher correlation (ρ_{vo}) translates in lower values of the option $F(V,O)$.

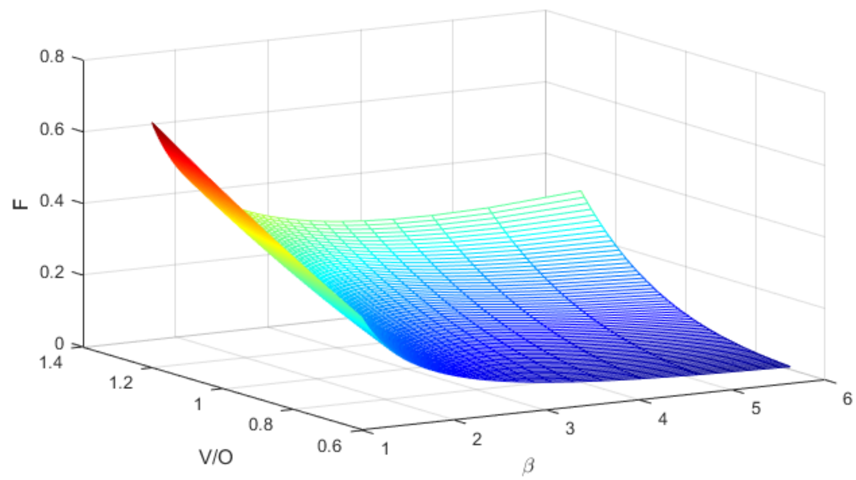

Figure 2.5. Surface of Option Value versus Root and Ratio Value

This 3 dimensional graph shows the value of the option as a function of the value of β_f and the relation V/O as in Equation 2.26. The value of the β_f is a function of the parameters described in Equation 2.22. Each curve represents a different value of the relation V/O, higher values of V/O increase the value of the option to hold the depreciated building.

At the same time higher values of the relation between V and O also translate into increases in the value of F.

Finally, as I described before, σ_O and δ_O affect the value of β_f and therefore indirectly affect the value of the option. Higher values of δ_O would decrease the value of the β_f and consequently would increase the value of F. In the case of σ_O, higher values in the volatility of the outside option translate into lower beta and therefore increase the value of the option F.

Overall, the analysis suggests that the value of the outside option impacts the value of F. The impact on the value of F would affect the optimal time of holding the option of the depreciated building. This relation between F and O affects the decision on the investment in capital expenditure of the depreciated building. In the next section, I develop the hypotheses and the empirical strategy to test them.

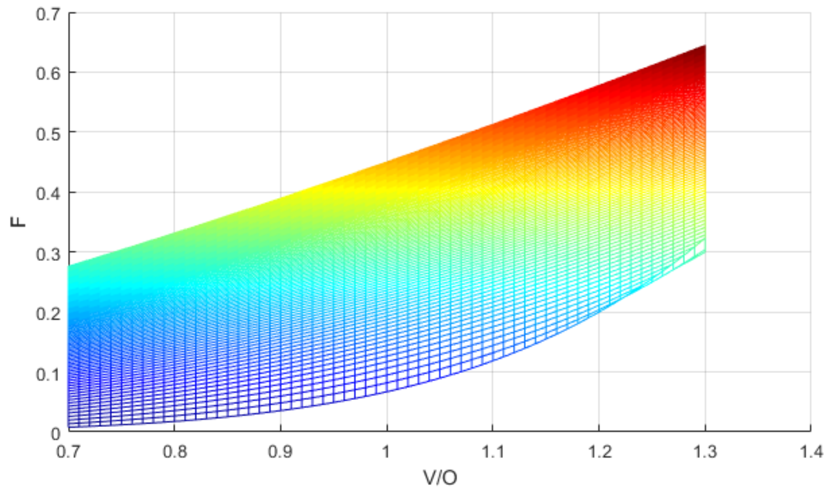

Figure 2.6. Value Option versus Ratio

This graph shows the value of the option as a function of the relation V/O the value of β_f as in Equation 2.26. The value of the β_f is a function of the parameters described in Equation 2.22. Each curve represents a different value of the relation β_f, higher values of β_f reduce the value of the option to hold the depreciated building.

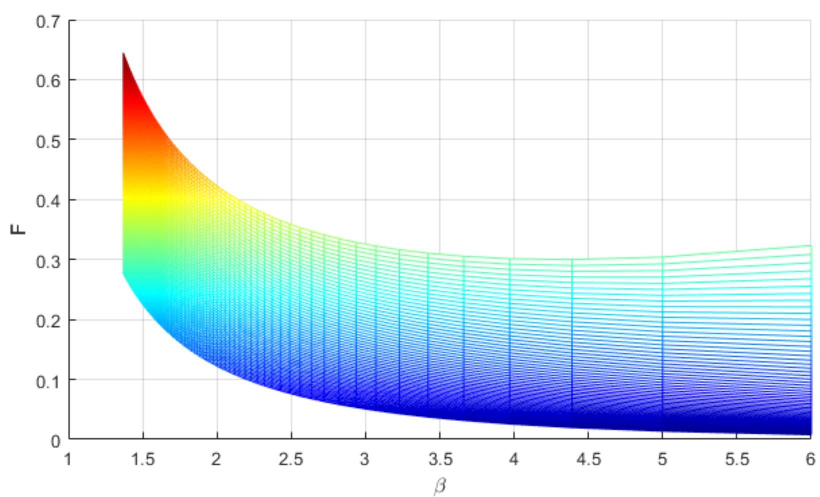

Figure 2.7. Value of Option versus Root Value

This graph shows the value of the option as a function of the value of β_f from Equation 2.26. The value of the β_f is a function of the parameters described in Equation 2.22. Each curve represents a different value of the relation V/O, higher values of V/O increase the value of the option to hold the depreciated building.

2.3 Hypotheses

Following the model previous described it is of interest to test whether the outside option has a statistically significant impact on the decision to invest in capital expenditure. From the option pricing model, it is clear that the value of the option increases as the value of β_f increases(see Figure 2.7), holding the ratio constant. The following hypotheses refer to changes in β_f and consequently to changes in $F(V, O)$.

Hypothesis 1: An increase in the correlation between the highest best use and the second highest best use translates into higher capital expenditure in subsequent period.

Figure 2.3 shows that higher volatility of the outside option reduces the value of the β_f, thus increasing the value of $F(V, O)$. The increase on the value of $F(V, O)$ implies that keeping the option alive is more valuable than investment in capital expenditure; therefore, the threshold in which capital expenditure is optimal increases.

Hypothesis 2: An increase in the volatility σ_O reduces the investment in capital expenditure.

Figure 2.4 shows that higher levels of convenience yield δ_O translate into lower levels of β_f, therefore increasing the value of the option F.

Hypothesis 3: An increase in the convenience yield δ_O reduces the investment in capital expenditure.

To test these hypotheses I will use the value of the option estimated in Equation 2.26 and estimate the log for simplification of the problem.

2.4 Empirical Model

The goal of the paper is to analyze the impact that the outside option has on capital expenditure decisions. The empirical strategy is to use the value of F to explain changes in the amount of capital expenditures; in other words, I use the amount of capital expenditures as my dependent variable and the value of the option as my independent variable. To define the empirical model, I will use the log transformation of Equation 2.26 as follows:

$$Log(F) = (\beta_f - 1)log(\beta_f - 1) - \beta_f log(\beta_f) + log(O) + \beta_f log(V) + \beta_f log(O) \quad (2.27)$$

$$Log(F) = H(\beta_f(\delta_O, \delta_V, \sigma_O, \sigma_V, \rho_{vo})) + log(O) + \beta_f log(V) + \beta_f log(O) \quad (2.28)$$

where β_f is a function of $\delta_O, \delta_V, \sigma_O, \sigma_V$ and ρ_{vo}. I use the equation 2.27 to set the empirical model. For simplicity, I will use a linear function to represent the value of $Log(F)$. I define $\vec{\omega_1}$ as vector with the parameters and ϕ_{imt-1} as vectors that contains the values of δ_{Oimt-1},δ_{Vimt-1}, σ_{Oimt-1}, σ_{Vimt-1}, and $\rho_{voimt-1}$. I use a linear representation of the interaction terms $\beta_f log(V)$ and $\beta_f log(O)$, where $\vec{\omega_2}$ are the coefficients for the interaction $\beta_f log(V)$ and $\vec{\omega_3}$ are the coefficients for the interaction $\beta_f log(O)$. The resulting empirical representation I use is as follows:

$$\begin{aligned} Capex_{it} &= Log(F_{it}) + \vec{X_{it}} + \varepsilon_{it} \\ Capex_{it} &= \beta_0 + \beta_1 O_{imt-1} + \vec{\omega_1}\vec{\phi}_{imt} + \vec{\omega_2}\vec{\phi}_{imt} V_{imt-1} + \vec{\omega_3}\vec{\phi}_{imt} O_{imt-1} + \vec{X_{it}} + \varepsilon_{it} \end{aligned} \quad (2.29)$$

where $\vec{X_{it}}$ is a vector of building characteristics(i.e. age and size), β_0 is a constant and β_1 is the coefficient which measures the impact that the value of the outside option O has on F. V_{imt-1} is the value of a price index for the property type of building i, in market m at time $t-1$[8], and O_{imt-1} is the value of a price index for the alternative use to the property type of building i, in market m at time $t-1$.

I use lag values for the different parameters as I expect that the decision to invest in capital expenditure is not contemporaneous, but depends on the market conditions managers observe at the moment of budgeting for the following year. I do not observe zoning conditions for each building, so the outside option is an average of all the other property type price indexes for a given year.

[8] I do not observe the prices of the property at every point in time, therefore my second best option is to use a repeated sales index for the property type.I define markets as one of the NCREIF regions in the country.

For Hypothesis 1 the coefficients of interest are the coefficients related to ρ_{vo} in $\vec{\omega_1}$, $\vec{\omega_2}$ and $\vec{\omega_3}$. For example, if the coefficient of $\rho_{voimt-1}$ is positive and statistically significant then this would validate Hypothesis 1. The other coefficients of interest are the coefficients associated with the volatility σ_{Oimt-1} in order to test Hypothesis 2 and the δ_{Oimt-1} in order to test Hypothesis 3.

2.5 Data

To test the hypotheses previously described, I use data from 2 sources. The main data on capital expenditures comes from SNL. These are yearly observations at the building level. I also use data from the underlying assets in commercial mortgage backed securities provided by TREPP. I specifically use appraisal data and transaction data for the construction of price indexes (CPPI) and operation data to estimate cap rates from properties backing the commercial mortgages.

To estimate the proxy for property prices, I follow the methodology proposed by Bailey et al. (1963) and Case et al. (1989) to construct the repeated sales index and use more than 25,000 repeated observations including appraisals from TREPP data and transactions of commercial properties from SNL. I estimate the index by property type, NCREIF region and month. The index includes Multi-Family, Retail, Office, Industrial, Lodging and Other property type buildings. The regions cover all of the United States of America, and are divided into East North Central (EN), Mideast (ME), Mountain (MT), Northeast (NE), Pacific (PC), Southeast (SE), Southwest (SW), and West North Central (WN). Table 2.1 provides the distribution of properties by NCREIF region.

The price indexes allow me to estimate a proxy for V and O. I estimate each index by property type and region for each month from 2003 until 2015. For example, I estimate the monthly repeated sale index for the retail buildings in the north-east of the country. The index relates to the property use represented by V; in my example this is retail CPPI for the north-east. O represents the average index of all the alternative uses to that of V, in this case including lodging(LO), industrial(IN), mixed-use(MU), multi-family(MF), office(OF), self-storage(SS) and other(OT). The

Table 2.1. Distribution of Properties by Regions

Region	Frequency	Percentage
EN	3,117	14.9%
ME	3,293	15.7%
MT	1,050	5.0%
NE	3,307	15.8%
PC	2,385	11.4%
SE	3,773	18.0%
SW	2,943	14.1%
WN	1,061	5.1%
Total	20,929	100.0%

Note: This table provides the distribution of commercial properties by region. EN includes the following states: OH, IL, IN, MI and WI. ME includes SC, MD, DE, NC, DC, VA, KY and WV. MT contains AZ, UT, CO, MT, NV, WY and ID. The states NJ, PA, NY, CT, MA, RI, ME, VT and NH are part of the NE region. PC includes states in the Pacific Ocean such as CA, WA, OR, HI and AK. SE are the states located in the south east, including FL, AL, GA, TN and MS. SW includes TX, LA, A. Finally WN includes MO, KS, NE, MN, ND, IA and SD.

estimated indexes then allow me to construct the volatility of prices V and O, as well as the correlation between them.

Another important parameter of the option price model is the convenience yield. I create property type average cap rates at the regional level to proxy the convenience yield. I use net operating income information from the properties in TREPP and SNL along with appraisal data or transaction prices to estimate the market capitalization rate. Figures 2.8 and 2.9 provide examples of the cap rates for the north-east and pacific regions. The graph shows the increase in cap rates during 2008 to 2009 and a subsequent decline. Multi-family in both regions experienced a lower spike and then a smoother decline suggesting a lower volatility σ_V.

As I mentioned before, I use capital expenditures for REIT buildings reported in SNL data. REITs provide a detail of subsequent improvements by building in section III of their 10-k reports. I use only the properties reported within the USA. The resulting sample includes information from 20,929 buildings. Table 2.1 provides the distribution across regions and Table 2.2 provides the distribution across property type.

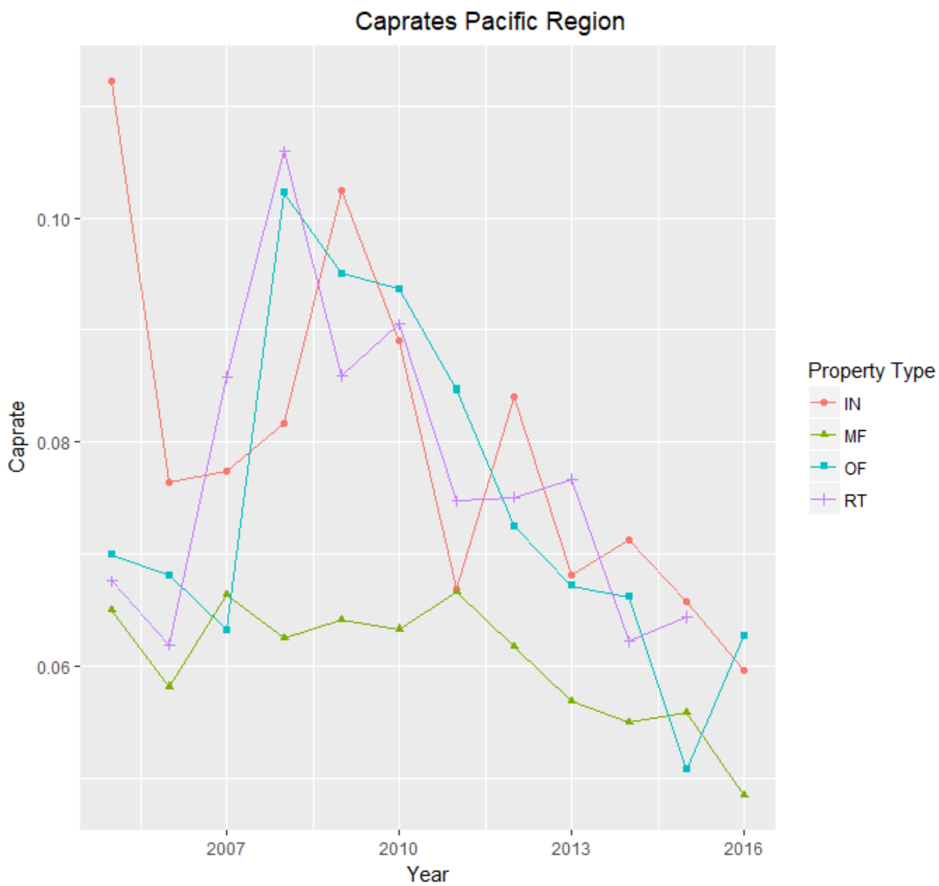

Figure 2.8. Capitalization Rates Pacific Region

This graph shows the average capitalization rate by property type, NCREIF region and by year. In this example, I show the average capitalization rate by property type in the Pacific NCREIF region. The acronyms described lodging(LO), industrial(IN), mixed-use(MU), multi-family(MF), office(OF), self-storage(SS) and other(OT). PC includes states in the Pacific Ocean such as CA, WA, OR, HI and AK.

In the 10-k report REITs disclose the cumulative subsequent improvement for each building by year. In order to estimate the amount of Capex for any given year, I estimate the difference from one year to the next. I then estimate the capital expenditure as a percentage of the reported book value. Figure 2.10 provides a glimpse of the distribution of improvements as a percentage of the book value. This

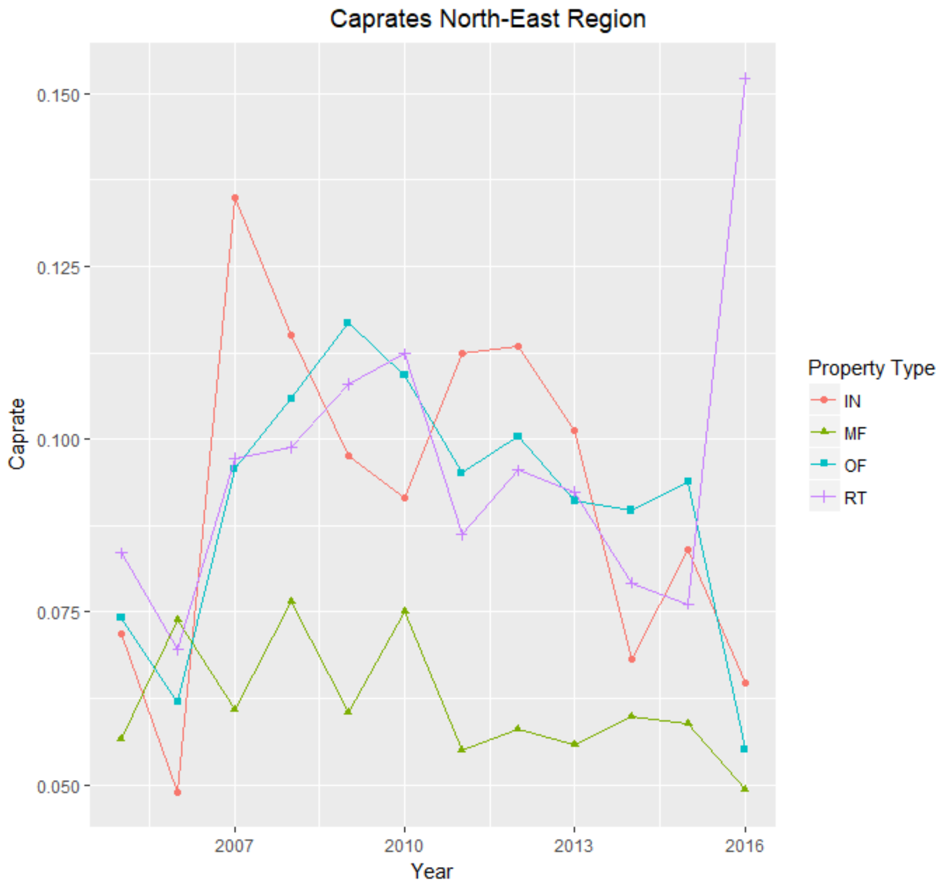

Figure 2.9. Capitalization Rates North-East Region

This graph shows the average capitalization rate by property type, NCREIF region and by year. In this example, I show the average capitalization rate by property type in the North-East NCREIF region. The acronyms described lodging(LO), industrial(IN), mixed-use(MU), multi-family(MF), office(OF), self-storage(SS) and other(OT). The NE region includes states in the north east of the USA including NJ, PA, NY, CT, MA, RI, ME, VT and NH.

figure only shows improvements that are greater than zero and less than 0.3[9]. I also trim the data 2% (1% in each tail) from the extremes of the distribution to decrease the impact of outliers in the analysis.

The other controls I use to estimate the empirical model are REIT id fixed effect, year of observations fixed effect, square footage, and age of the building. SNL provides

[9]There are a 29,978 observation out of 101,356 with capital expenditure equal to 0.

Table 2.2. Distribution of Properties by Type

Property Type	Frequency	Percentage
IN	3,683	17.6%
LO	53	0.3%
MF	2,690	12.9%
MU	82	0.4%
OF	4,441	21.2%
OT	287	1.4%
RT	8,716	41.6%
SS	977	4.7%
Total	20,929	100.0%

Note: This table provides the distribution of commercial properties by Property Type. IN includes all industrial properties. LO are properties in the lodging industry. MF stands for multi-family properties. MU are properties with mixed use. OF are Office buildings. RT are all retail buildings. SS are self-storage buildings. Finally, OT are specialty buildings.

information on these property characteristics as well as the year the property was built. I use this information to estimate the age of the property.

As a result, the final sample has 101,356 observations from 20,929 buildings. Table 2.3 provides the summary statistics for the sample. Observations range from 2003 to 2015 with an average age for the buildings of 20.6 years. The average cap rate for the properties is approximately 9%. The correlation between the price index for highest best use and the price of the outside option is 0.14. Finally, the average size of the buildings is 192,508 square feet. With this data, I then estimate the empirical model described in Equation 2.29.

2.6 Results

Table 2.4 presents the estimates for the various specifications of equation 2.29. Specification 1 uses only proxies for the parameters from equations 2.26 and 2.22. Specification 2 shows results when introducing fixed effects to control for REIT, region, and proper-ty type characteristics invariant over time. Finally, in specification 3, I control for year fixed effects that take into account invariant effects within each

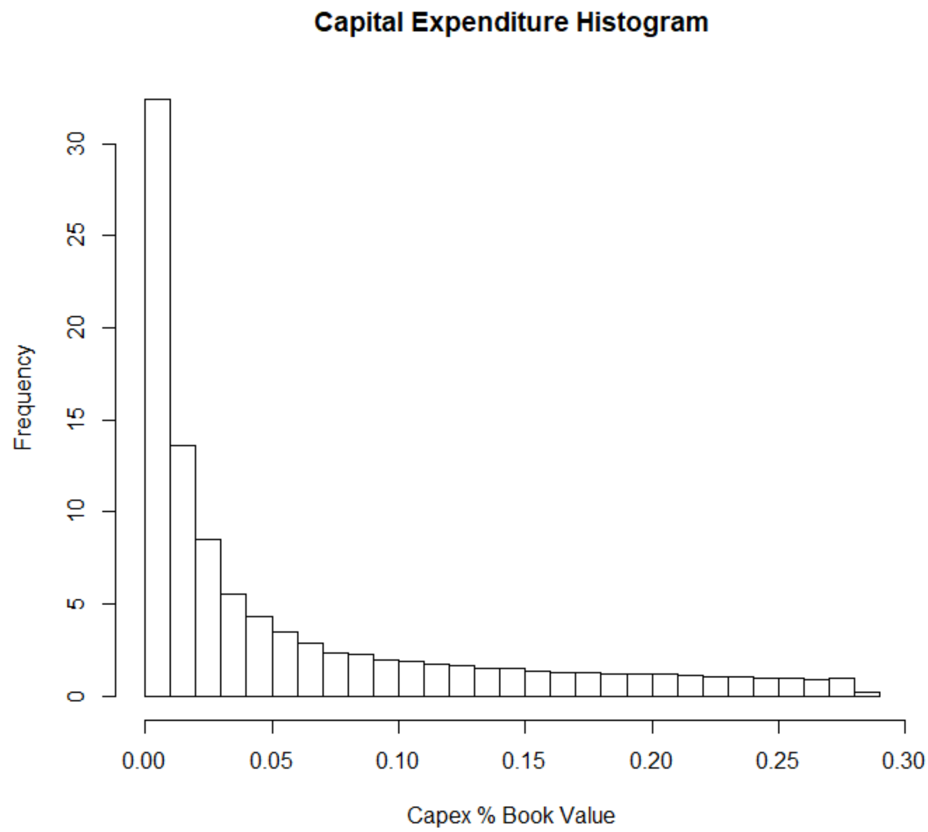

Figure 2.10. Histogram of Capital Expenditures

This histogram shows the distribution of the yearly capital expenditure as a percentage of the book value of the building that are greater than 0% and less than 30%. There are approximately 30% observation with 0 capital expenditure in a year. This are buildings owned by REITs that disclose information on the capita expenditures in the 10-k report section III.

year. I then estimate the same specifications using one year lagged values to test whether the impact of parameters takes place prior to the year of the investment in capital expenditure. I present estimates for these specifications in Table 2.5. The motivation behind this specification is that capital expenditure could be budgeted for and approved a year earlier; therefore, the variables affecting that decision should be from the previous year.

At first glance, it seems that parameters associated with the outside option have

Table 2.3. Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Cumulative Capex	101,356	9,485.34	8,649.19	24.00	25,640.00
Net Book Value	101,356	22,637.26	60,506.70	1.00	3,268,458.00
Capex $	101,356	978.33	2,535.01	0.00	21,391.00
Capex %	101,356	0.07	0.17	0.00	1.00
Age	101,356	20.63	14.11	0	211
V	101,356	1.01	0.71	0.10	23.98
O	101,356	1.26	1.23	0.37	10.84
σ_V	101,356	0.28	0.34	0.03	22.08
σ_O	101,356	0.33	0.54	0.04	3.25
ρ_{vo}	101,356	0.14	0.30	-0.80	0.88
δ_V	101,356	0.09	0.02	0.04	0.20
δ_O	101,356	0.08	0.02	0.01	0.19
Square footage	101,356	192,508.80	262,534.10	678	15,072,280
Year Observation	101,356	2,010.05	3.26	2,003	2,015

Note: This table shows the summary statistics for the data used in the empirical analysis. $Cumulative\quad Capex$ represents the declared cumulative subsequent improvements in the property in \$1000 reported in the 10-k for each property. $Net\quad Book\quad Value$ is the declared book value at the property level in the 10-k Section III. $Capex\quad \$$ is the dollar amount of capital expenditure for any given year $(Cumulative\quad Capex_t - Cumulative\quad Capex_{t-1})$. $Capex\quad \%$ is the capital expenditure as a % of the $Net\quad Book\quad Value$. Age is the age of the building. V is commercial property price index of the property type of the building at the region level. O is the property price index for all the alternative uses to that V.σ_V and σ_O are the volatility of the price index of V and O respectively. ρ_{vo} is the correlation between the two CPPIs. δ_V and δ_O are the capitalization rates for the property type of the building and the capitalization rates of the alternative use. Square footage is the size of the property in terms of the square feet.

a statistically significant role. For instance, σ_O is statistically significant independent of the specification used. The impact of the correlation between HBU and SHBU is also statistically significant. The impact of the price is statistically significant on the first specification but not in the second, and the sign is not consistent through different specifications. This could be due to the interaction terms and fixed effects in different specifications. The economical impact depends on other interaction terms which I will discuss using Specification 2 [10].

[10] I show in Table 2.4 all the estimated parameters, Specification 3 has more than 10 other interaction terms that I cannot show due to space constraints.

Table 2.4. Regression Analysis on Capital Expenditure

	Dependent variable:		
	Capex as % of Book Value		
	Specification 1	Specification 2	Specification 3
O	-0.002^{*}	0.0002	0.010
	(0.001)	(0.001)	(0.006)
ρ_{vo}	0.001	-0.015^{*}	-0.019^{***}
	(0.009)	(0.009)	(0.007)
δ_O	0.056	-0.147	0.016
	(0.067)	(0.090)	(0.110)
V/O	-0.019^{***}	-0.014^{***}	0.006
	(0.004)	(0.003)	(0.004)
$\rho_{vo}V/O$	-0.007^{*}	0.006	0.015^{*}
	(0.004)	(0.005)	(0.008)
σ_O	0.009^{***}	0.017^{***}	0.014^{***}
	(0.003)	(0.001)	(0.003)
$\rho_{vo}\sigma_V\sigma_O$	-0.085^{***}	-0.047^{***}	-0.052^{***}
	(0.005)	(0.006)	(0.007)
σ_V	0.023^{**}	0.010^{**}	0.003
	(0.009)	(0.004)	(0.006)
V	0.009^{**}	0.013^{***}	-0.010
	(0.004)	(0.002)	(0.009)
δ_V	-0.410^{***}	-0.059	-0.070
	(0.089)	(0.041)	(0.061)
Constant	0.112^{***}	0.008	0.003
	(0.013)	(0.012)	(0.021)
REIT F.E.	No	Yes	Yes
Year F.E.	No	No	Yes
Region F.E.	No	Yes	Yes
Property Type F.E.	No	Yes	Yes
Observations	101,356	101,356	101,356
Adjusted R^2	0.010	0.083	0.086

Note: This table shows the result from 3 different OLS specifications for the empirical model in Equation 2.29. The dependent variable is the capital expenditure as a percentage of the net book value. V and O are the property price indexes for the property type and the second highest best use. σ_O and σ_V are the volatilities of the indexes, while δ_O and δ_V are the capitalization rates. Specification 3 includes the following interaction terms not display here: $\delta_V * V$, $\delta_O * V$, $\delta_O * O$, $\sigma_v * V$, $\sigma_v * O$, $\sigma_o * V$, $\sigma_O * O$,$\rho_{vo} * V$, $\rho_{vo} * O$, $\rho_{vo}\sigma_V\sigma_O$. All specifications include Age and Sqft/1000, not shown here due to space constraints. Age is the age of the building and Sqft/1000 is the size of the building in 1000 square feet. All standard errors are clustered robust.*p<0.1; **p<0.05; ***p<0.01

Table 2.5. Regression Analysis on Capital Expenditure with lagged independent variables

	Dependent variable:		
	Capex as % of Book Value		
	Specification 1	Specification 2	Specification 3
O_l	0.001^{**}	0.002^{***}	0.0002
	(0.0004)	(0.001)	(0.004)
ρ_{vol}	-0.005	-0.017^{**}	-0.011^{*}
	(0.008)	(0.007)	(0.006)
δ_{Ol}	0.050	-0.157^{**}	-0.081
	(0.066)	(0.076)	(0.087)
V_l/O_l	-0.013^{***}	-0.006^{**}	0.007^{*}
	(0.004)	(0.003)	(0.004)
$\rho_{vol} V_l/O_l$	-0.005	0.005	0.003
	(0.005)	(0.005)	(0.006)
σ_{Ol}	0.003	0.012^{***}	0.011^{***}
	(0.003)	(0.001)	(0.003)
$\rho_{vol}\sigma_{Vl}\sigma_{Ol}$	-0.040^{***}	-0.018^{**}	-0.014
	(0.008)	(0.007)	(0.010)
σ_{Vl}	0.011^{**}	0.005^{**}	0.0001
	(0.005)	(0.002)	(0.005)
V_l	0.004^{*}	0.006^{***}	-0.005
	(0.002)	(0.002)	(0.009)
δ_{Vl}	-0.326^{***}	-0.014	-0.025
	(0.065)	(0.036)	(0.063)
Constant	0.104^{***}	0.010	0.012
	(0.011)	(0.009)	(0.019)
REIT F.E.	No	Yes	Yes
Year F.E.	No	No	Yes
Region F.E.	No	Yes	Yes
Property Type F.E.	No	Yes	Yes
Observations	101,356	101,356	101,356
Adjusted R^2	0.008	0.082	0.085

Note: This table shows the result from 3 different OLS specifications for the empirical model in Equation 2.29. The dependent variable is the capital expenditure as a percentage of the net book value. V and O are the property price indexes for the property type and the second highest best use. σ_O and σ_V are the volatilities of the indexes, while δ_O and δ_V are the capitalization rates. Specification 3 includes the following interaction terms not display here: $\delta_V * V$, $\delta_O * V$, $\delta_O * O$, $\sigma_v * V$, $\sigma_v * O$, $\sigma_o * V$, $\sigma_O * O$,$\rho_{vo} * V$, $\rho_{vo} * O$, $\rho_{vo}\sigma_V\sigma_O$. All specifications include Age and Sqft/1000, not shown here due to space constraints. Age is the age of the building and Sqft/1000 is the size of the building in 1000 square feet. All standard errors are clustered robust. *p<0.1; **p<0.05; ***p<0.01

2.6.1 Discussion on Correlation ρ_{vo}

The first set of results that I would like to discuss are the ones related to the correlation between the HBU and SHBU. If we consider that all the variables have their mean values and we alter only the ρ_{vo} one standard deviation, I want to see what the impact of that change has on the investment in capital expenditure . Table 2.4 provides the estimates for the empirical model generally described in equation 2.29. There are 3 estimates that include $\rho_v o$: the correlation itself (ρ_{vo}), the interaction between volatilities and correlation ($\rho_{vo}\sigma_O\sigma_V$), and the interaction between the correlation and the price ratio ($\rho_{vo}V/O$). The impact would depend on the combination of all three estimates.

I start by testing whether the coefficients are collectively different from 0. An ANOVA analysis suggests that all the coefficients associated ρ_{vo} are statistically different from zero[11]. The impact of the increase of one standard deviation of ρ_{vo} from the mean then needs to be considered as a joint impact of all three coefficients. The increase of 0.3 of ρ_{vo} translates into a decrease of 6.2% in capital expenditures[12]. This result, although statistically significant, has the opposite sign from that of Hypothesis 1.

When I use the estimates from Table 2.5, the result is practically the same: one standard deviation increase in ρ_{vol} decreases the investment in Capital Expenditure by 6.3%. Although these results are statistically significant, suggesting that the parameters of the model are relevant, they contradict the expected sign from Hypothesis 1. To further study this issue, I test whether or not the impact of ρ_{vo} depends on the subsample I use. I estimate Specification 2 with only observations where the ratio between V/O is closer to 1. This is relevant as areas where prices are relative close the correlation might be more important.

Table 2.6 provides the estimates for the subsample analysis. The result remains the same: a one standard deviation increase in the correlation translates into a

[11] I reject the null of no combined effect at the 1% level.

[12] I obtain the result by calculating the following: $-0.015 * \rho_{vo} + 0.006 * \rho_{vo}\sigma_O\sigma_V$ $-0.047\rho_{vo}V/O$ for ρ_{vo} equal to 0.14 and 0.44. With $\sigma_O = 0.33$, $\sigma_V = 0.28$ and $V/O = 0.8$. I divide the difference between the two estimated impacts by the average Capex %, therefore I get a 6.2 % increase.

decrease of approximately 7% of the amount invested in capital expenditure. Overall, the results suggest that there is evidence of the importance of the correlation with the outside option at the moment of making an investment decision but the sign is not the expected one.

Table 2.6. Regression Analysis on Capital Expenditure Subsample

	Dependent variable:
	Capex as % of Book Value
	Specification 2
O	0.259**
	(0.127)
ρ_{vo}	0.324**
	(0.137)
δ_O	-0.392 **
	(0.173)
V/O	0.269**
	(0.11)
$\rho_{vo}V/O$	-0.332**
	(0.138)
σ_O	0.0205
	(0.019)
$\rho_{vo}\sigma_V\sigma_O$	-0.087
	(0.068)
σ_V	0.049***
	(0.017)
V	-0.273**
	(0.127)
δ_V	0.017
	(0.132)
Constant	-0.253**
	(0.112)
REIT F.E.	Yes
Year F.E.	No
Region F.E.	Yes
Property Type F.E.	Yes
Observations	13,563
Adjusted R^2	0.096

Note: This table shows the result from 3 different OLS specifications for the empirical model in Equation 2.29. The dependent variable is the capital expenditure as a percentage of the net book value. V and O are the property price indexes for the property type and the second highest best use. σ_O and σ_V are the volatilities of the indexes, while δ_O and δ_V are the capitalization rates. Specification 3 includes the following interaction terms not display here: $\delta_V * V$, $\delta_O * V$, $\delta_O * O$, $\sigma_v * V$, $\sigma_v * O$, $\sigma_o * V$, $\sigma_O * O$,$\rho_{vo} * V$, $\rho_{vo} * O$, $\rho_{vo}\sigma_V\sigma_O$. All specifications include Age and Sqft/1000, not shown here due to space constraints. Age is the age of the building and Sqft/1000 is the size of the building in 1000 square feet. All standard errors are clustered robust. *p<0.1; **p<0.05; ***p<0.01

2.6.2 Discussion on Volatility σ_O

It is clear from Table 2.4 that the impact of σ_O depends on the interaction of two estimates, the estimates for σ_O and $\rho_{vo}\sigma_V\sigma_O$. Again, it is important to first test that these two estimates are jointly different from zero, and the ANOVA test suggests that the coefficients associated with σ_O are jointly different from 0[13]. The impact of σ_O then depends on the interaction of these coefficients.

Table 2.4 provides the estimates for the coefficients associated with σ_O and $\rho_{vo}\sigma_V\sigma_O$. The estimates suggest that an increase of one standard deviation of volatility of the outside option translates to an increase in capital expenditure by 11.9%. Again, the estimates for the volatility of the outside option are economical and statistically significant. This suggests that σ_O is an important parameter to estimate the value of the options that managers have at the moment of making a decision whether to invest in capital expenditures. Unfortunately, the sign of the impact is not what the model predicted.

2.6.3 Discussion on Cap Rate δ_O

The impact that the capitalization rate of the outside option, δ_O, has on the decision to invest in Capex depends solely on the coefficient associated with it, and there are no interaction terms to consider in the analysis. Table 2.4 shows that the coefficient associated with the capitalization rate of the outside option is negative but not statistically significant, as the p-value of the estimate is 0.102. This suggests that the capitalization rate has the sign expected by the model, but it is not statistically significant different from 0.

Overall, I find strong empirical evidence of the importance that the outside option has at the moment of making a decision to invest in capital expenditures. The

[13] I reject the null of coefficients equal to 0 at the 1% level

correlation between the HBU and SHBU, the cap rate, and price of the SHBU are relevant pieces of information for managers at the moment of making a decision to invest.

2.7 Conclusion

In this paper, I develop a real option model to estimate the impact that the option to invest in an alternative use has on a manager's decision of whether or not to invest in capital expenditures. I derive an analytical solution to the problem and then using a numerical representation, I describe 3 testable hypotheses. Using data on capital expenditures from more than 20,000 commercial properties owned by REITs, I test the hypotheses.

The findings suggest that the parameters of the model are statistically significant but the model predicts opposite signs. For instance, the correlation between the HBU and SHBU prices is relevant and statistically significant, but the model predicts the opposite sign from that of the empirical model. The same happens with the findings related to the volatility of the outside option. On the one hand, the theoretical model suggests that an increase in volatility translates into a reduction of capital expenditures. On the other hand, the empirical model suggests that an increase of one standard deviation translates into an increase of 11.9% in capital expenditures. Finally, the theoretical model predicts that an increase in the cap rate of the outside option has a negative effect on capital expenditures, and the empirical model appears to agree.

Chapter 3
Information Effect of Online Reviews on Investment in the Real Estate Industry

Online consumer review forums have the power to reduce information asymmetries between consumers and building managers, thereby generating incentives for managers to invest in quality. I use data on revenue, occupancy, and investment decisions for 1,815 hotels that are linked to a dataset of 895,768 online consumer reviews to explore this question. Using a panel fixed effect regression to estimate the impact of consumer reviews on hotel capital expenditures, I show that a drop of 0.5 star rating increases the probability of investing in Capex by 2.7%. However, the impact of changes in reviews is not homogeneous across locations or building characteristics. A subsample analysis shows that hotels in more competitive areas respond more aggressively to reviews. Subsequently, I show that hotels improve their reputation after investing in capital expenditures, leading to an increase in ratings of 0.35 stars. Taken together, I conclude that information flow among consumers and managers through these reputation systems makes buildings' quality salient, thereby providing incentives to invest.

3.1 Introduction

Lately in the real estate industry a number of websites provide information about the quality of buildings, services, or locations. These allow consumers to make more informed decisions, decreasing problems arising from information asymmetry (Akerlof, 1970). Examples of these reputation systems include AirBnB (Bed and Breakfast), TripAdvisor (Hotels/Restaurants), Lending Tree (Mortgages), and other websites that allow consumers to post a review of a product or service (i.e Amazon, and ebay). Many recent studies focus on how positive consumer reviews increase customers' patronage of stores. For example, Luca (2016) examines the impact of Yelp.com reviews on restaurant revenue, and Anderson and Magruder (2012) link consumer reviews to restaurant bookings. These findings suggest that reviews provide information to consumers, decreasing information asymmetry between store and customer, and therefore altering store patronage. Little is known, however, about the impact of these mechanisms on managers' decision.

This paper expands the empirical literature on the relation between information provision and firm quality. The literature on mandatory and voluntary disclosure looks at information effects on price and firm quality. For example, Lewis (2011) uses ebay online car auctions to argue that information provided for a car on a website influences the price offered in an auction due to a decrease in information asymmetry. Jin and Leslie (2003) study a restaurant policy about making hygiene grade cards visible to customers and argue that the information embedded in the grade cards impacts revenues and improve the hygiene quality of restaurants. The contribution of my paper is to focus on information provision from consumers and the subsequent effect on firms' decision to invest in capital improvements.

Another relevant strand of literature examines review manipulations (Mayzlin et al., 2014) which affect the information provision of reviews. In this paper, also I explore some implications of review manipulation on firm investment. Although it is not possible to tell whether a review is fake, I look at hotels that are more likely to suffer from these manipulations. From the literature on information asymmetries in real estate, we know that information considerations affect prices and how agents make transactions (Garmaise and Moskowitz, 2003; Levitt and Syverson, 2008). This paper explores how the reputation mechanism of online reviews can mitigate these problems. This paper contributes to the literature on capital expenditure in real estate by expanding the study of information asymmetries on capital expenditures after controlling for market conditions (Bond et al., 2014; Ambrose and Steiner, 2017).

I examine commercial properties from the hospitality industry to argue that online reputation systems impact investment decisions. In particular, I use hotel reviews from TripAdvisor.com to infer changes in the reputation of the hotel and how these affect the investment decisions of capital expenditures. The underlying assumption is that consumer reviews are a noisy signal of buildings' quality and they reveal information to consumers, thereby affecting their consumption pattern. As a result, landlords recognize potential future losses due to reputation deterioration and respond by increasing investment in capital improvements.

I propose a simple reputation model to explore the hypotheses of this paper. Consumers observe reviews that are a noisy signal of buildings' quality and decide whether or not to patronage a hotel the next period. The problem managers face is whether they allow quality to deteriorate or maintain the same quality level. This sets two rates at which consumers visit the hotel the next period: first, a rate that represents the case in which managers decide to maintain the reputation; second, a rate that represents the case in which managers decide to produce at a lower quality. With this information I estimate the price that guarantees quality depending on how consumers set the rates for the next period. Using this model, I discuss the potential impact that reviews have on setting the rates and derive my testable hypotheses and the empirical strategy to do so.

I use a dataset with more than 850,000 online reviews from TripAdvisor.com and a panel dataset of capital expenditures from more than 1,800 buildings to test the hypotheses. I analyze the type of information embedded in these online reviews and provide summary statistics, then show that messages posted by consumers on these websites affect the vacancy of corresponding hotels, suggesting that reviews are informative to consumers and that they alter their consumption pattern. Subsequently, I show that after controlling for market conditions, by using a combination of fixed effects and a time variant fixed effect, the probability of landlord investment in capital expenditures increases as facility reputation falls. Finally, by using an event study around the investment decision, I show that capital expenditures help restore facility reputation. Taken together, I conclude that information flow among consumers through message boards helps mitigate asymmetries of information between a consumer and the hotel suggesting an improvement in market efficiency. I also show that the impact of these mechanisms is not ubiquitous and depends on the characteristics of the competition and of the hotel itself.

The remainder of the paper proceeds as follows. In the next section, I provide an overview of the relevant literature and discuss the main findings of the optionality of capital expenditures, the impact of online reviews on hotel bookings, information disclosure and firms' quality, and the use of textual analysis to retrieve qualitative information from reviews. After describing the firm problem and deriving the main hypotheses of the paper, I present the main empirical model to test the hypotheses. Finally, I provide a detailed discussion of the results.

3.2 Prior Literature

The theoretical foundation of this paper builds on the literature in information asymmetry by Akerlof (1970), and the reputation and quality decision models by Shapiro (1983) and by Rogerson (1983). This paper is also informed by the empirical literature on verifiable disclosure information (Jin and Leslie, 2003; Lewis, 2011), the empirical analysis of online review manipulation (Mayzlin et al., 2014), and the studies on the influence reviews have on consumption patterns (Anderson and Magruder, 2012; Sparks and Browning, 2011). Finally, the paper fits within the literature on information asymmetries in real estate (Garmaise and Moskowitz, 2003; Levitt and Syverson, 2008) and the strand of literature that studies capital improvements in this industry (Bond et al., 2014; Ambrose and Steiner, 2017).

In his seminal work, Akerlof (1970) emphasizes the presence of adverse selection in markets with information asymmetries. If consumers are uncertain about the quality of the product and can only use some market-wide statistics to assess product grade, then there are incentives for firms to market poor quality products. Due to that benefits of producing high quality merchandise accrues to the entire market rather than the seller of higher grade. This phenomenon leads to a drop in the grade of products available and consequently a drop in welfare. Garmaise and Moskowitz (2003) and Levitt and Syverson (2008) study market distortions related to information asymmetries in the real estate industry. Using quality of property tax assessment as an exogenous variation of the level of information, Garmaise and Moskowitz (2003) study how agents mitigate asymmetries of information. They find that buyers tend to buy properties at closer distances, prefer buildings with a longer history of net operating income, and most importantly decrease transactions with counterparts with information advantages. On this same line of research, Levitt and Syverson (2008) study the market distortions that arise from informational advantages that real estate agents have on the markets in which they operate. Their results suggest that real estate agents sell their own houses at a higher price and wait longer to sell these properties. This empirical work points out the actions taken by agents in order to mitigate some of the market distortions and suggests that real estate markets suffer from information asymmetries. The model proposed by Akerlof (1970) rests on the assumption of a one-time game, with consumers assessing quality using only market wide statistics, and that firms do not signal the market that they are of high quality.

Shapiro (1983) relaxes the assumption of one-time transaction related to consumer sole dependence on a market statistic to assess the quality of a product. His model uses a reputation effect as an incentive for firms to market high quality products. His model postulates that firms recognize that future profits deteriorate if consumers flag the firm as poor in quality. Therefore, the firm produces high quality if the profit from doing so plus discounted future cash flow is greater than the current profit from producing low quality plus the lower future cash flow due to reputation deterioration. The most important insight from the model is the importance of discount rates on the incentives to produce high quality; a higher discount rate decreases the incentives to produce high quality as it decreases the premiums from future cash flows. However, the main weakness of the model is that it rests on the assumption that consumers can perfectly assess the quality of a product after consumption.

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

The internet is a growing source of information, and consumers use the content to their advantage. Web-based services allow customers to share information about products and services between each other, similar to what is traditionally known as word-of-mouth (WOM). With the goal of determining if the traditional theory for off-line WOM applies to online word of mouth, Brown et al. (2007) examined the purchasing pattern of consumers who were members of an online community. The authors investigated three characteristics of online networks: homophily of the community, tie strength among members, and credibility of the source. These three characteristics of the network affect the WOM's influence on individuals' decisions. Their findings suggest there are marked differences between the off-line and online network characteristics. The most relevant is the credibility of the source, where members in an online community weigh not only the credibility of other members but also the credibility of the website itself. These communities, though, still have the power to transmit information among members and alter the latter's perceptions of products.

There are several online services that allow customers to book hotels and read consumers' comments on a message board. The ability of those reviews to alter the decision of future guests to book a hotel is the research question studied by Sparks and Browning (2011). In an experimental setting, Sparks and Browning found that online reviews increase the willingness to book a hotel when the message is positive and decrease willingness to book when the message is negative. Sparks and Browning do not find support for a positive or negative impact of star ratings associated with the message, but the interaction of the star rating with the strength of the language used in the reviews does have an impact on willingness to book. The evidence found in the literature to date indicates that consumers base their decision to book a hotel on WOM, specifically, online reviews. Likewise, in the empirical literature, Anderson and Magruder (2012) use a regression discontinuity design to estimate the impact of half a star rating on restaurant booking. Their results suggest that changes of one half star are associated with a 49% increase in the times a restaurant sells out. The effect of a change in stars exacerbated when other sources of information are scarce. Along the same line and using the same methodology, Luca (2016) finds that 1 star translates into an increase of 5% to 9% in revenue.

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

One of the most cited weaknesses of working with consumer generated content is that given the impact it has on consumer decisions, firms have the incentive to manipulate reviews. For example, Mayzlin et al. (2014) examined review manipulation in the hotel industry. Their experimental setting consisted of comparing reviews of two different sites, Expidia.com and Tripadvisor.com, and further exploring the differences in how reviews are posted in each site. Expedia.com allows only verifiable customers to post reviews while Tripadvisor.com allows anyone to post comments[1]. Mayzlin et al. find that the presence of a near distance competitor, neighbor characteristics, and hotel characteristics affect review manipulation. Brand affiliated hotels have fewer incentives to manipulate reviews as these hotels manage their reputation through marketing campaigns that signal brand quality rather than any individual hotel. The findings of their study suggest that brand affiliated hotels receive less positive reviews and that independent hotels have a larger amount of positive reviews in TripAdvisor.com compared to Expedia.com where there is a cost to post, suggesting review manipulation by independent owners. The findings also suggest that brand-affiliated hotels neighboring independent hotels receive on average five percent more negative reviews than brand-affiliated hotels without independent neighbors. Luca and Zervas (2016) studied a similar phenomenon in the restaurant industry using Yelp reviews, and their conclusion was similar to that of Mayzlin et al. (2014). Overall, this literature suggests the existence of review manipulation and that this phenomenon is more pronounced for smaller businesses and businesses not affiliated with a brand. Review manipulation creates distortion in the market by altering the information provided to consumers and therefore decreases the impact of

[1]Although both sites have systems that filters out reviews most likely to be fake in order to reduce fraud. The details of how these filters operate is proprietary information and is not disclosed on the website. More details on the system can be found on this link: `https://www.tripadvisor.com/vpages/review_mod_fraud_detect.html`

reviews on firms' fundamentals, and probably in investment decisions.

Similar to the options available to consumers, managers have real options when it comes to capital investment decisions. Flexibility in capital improvements could have significant implications on various performance measures of the firm (Dixit and Pindyck, 1994) because the flexibility to invest or not in capital expenditure can skew cash flow as managers may invest differently during stronger economic conditions rather than weaker ones. For instance, Titman et al. (2004) developed a structural model to determine rate spread to price risky debt. Their findings suggest that capital investment flexibility alters volatility as well as skewness of the project's value. Most of these models were developed with the intention of determining the optimal policy of default on a commercial mortgage. The findings in this strand of literature suggest that capital improvements are linked to high lease rates.

Bond et al. (2014) extend the model proposed by Mueller and Reardon (1993), moving from examination of capital investment at the firm level to capital expenditures in real estate. Bond et al. follow the work of Dixit and Pindyck (1994), Titman et al. (2004) and Childs et al. (2004) to set up a model that depends on a long term and a noisy short-term lease rate. The short lease rate is a mean reverting process to the long run rate. In their model there is a quality component that depends on the stock of maintenance that depreciates at a constant rate, and that determines the revenue generated by combining lease rates with stock of maintenance. They test the model by using NCREIF data, and their findings suggest that capital expenditures lead to higher incomes. The authors also find that investments conducted during periods of low lease rate perform better than capital investment during high lease rate periods. On this same line Ambrose and Steiner (2017) study the disposition effect of properties after a capital expenditure takes place. Both disposition and capital expenditures are real options available to managers. Their findings suggest owners are encouraged to increase capital investment during periods of high rent in order to capture larger profits. Alternatively, during periods of lower rent or if owners foresee lower rent in the future, the incentive is to delay capital expenditures.

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

3.3 Model

In a world with uncertainty on quality, firms have little incentive to invest in improving their products or services (Akerlof, 1970). One solution to this problem is signals sent by the firms, such as advertising, to help consumers infer quality (Milgrom and Roberts, 1986) or to establish a reputation in the market. In this model, I consolidate the methodological approach from the reputation model proposed by Shapiro (1983) with the methodological approach introduced by Rogerson (1983). Rogerson (1983) proposes a reputation model with imperfect information initially and with customers making imperfect observations about the product's quality after its purchase, in part due to idiosyncratic observations of quality by consumers. For example, not every individual focuses on the same hotel's features to evaluate its quality which results in an imperfect observation.

The model outlined in this section consist of a three periods model. Figure 3.1 shows the time line. In period $t-1$ consumers provide information through reviews, where R is a function of the quality of the building (Q_M) and an error made by

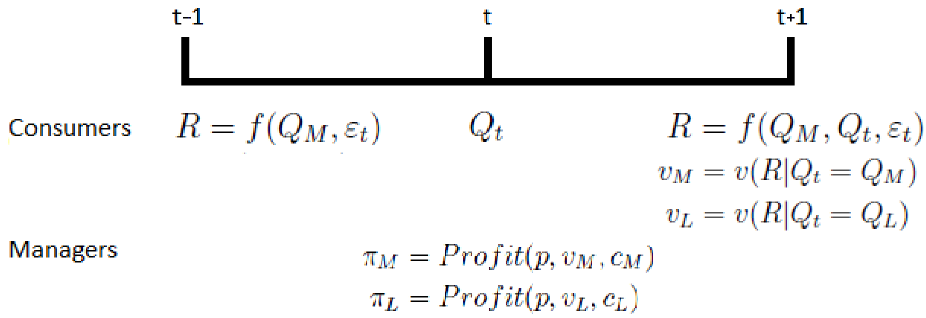

Figure 3.1. Model Set Up

This time line represents the decisions made by managers and consumers. At $t-1$, R represents the reviews that depend of the quality of the building (Q_M) and the error made by previous reviewers(ε_t). Consumers observe reviews and make the decision to purchase at time t, at this same time manager need to make the decision to either maintain quality or let the building deteriorate a produce at a lower quality. In order to maintain quality profits $\pi_M > \pi_L$. The vacancy rate of the next period depend on whether manager maintain quality or let building deteriorate.

reviewers (ε_{t-1}). In period t consumer make the decision to visit or not the hotel and at the same time manager need to decide whether to maintain quality or let the property deteriorate and produce at a lower quality (Q_L). This decision influence the vacancy of next period, that could be either v_L in the case managers decide to produce at low quality or v_M if they maintain the quality. A profit maximizer Manager maintains the property if profits from doing so are greater than letting the property deteriorate, with this condition I derive the price that guarantees the quality of the building. In the next subsections I go into more details on the implications of the decision made by the managers and what is the role of the reviews.

3.3.1 Reviews

Consumers seek information by reading reviews and make decisions about whether or not to purchase based on the probability that the product is of a certain quality. I assume that reviews given by consumer i in period t are correlated with the quality at t of the product, and that this review is a noisy signal of the quality.

$$R_{it} = \rho Q_t + \varepsilon_t \tag{3.1}$$

where R represents the rating given in the review (i.e., star rating from 1 to 5), Q_t is the quality of the product at period t. In the model, managers have the choice to either maintain production at the quality $Q_t = Q_M$ or extract profits from the reputation it has and produce at a lower quality $Q_t = Q_L$. In order to simplify the math and the analysis, I consider these two types of production and there is only one lower quality production (i.e., a firm with a reputation of 3.5 stars decides to maintain reputation or produce at the equivalent of 3.0 stars). The parameter ρ is the correlation coefficient and ε denotes the error made by the reviewer. The error is considered random and normally distributed.

Before making a purchase, a consumer observes the information provided by previous reviewers. For simplicity, I assume that the aggregation of reviews is the average of previous ratings given by other consumers rounded to the nearest 0.5 star. This rounding is the way consumers observe the aggregation of reviews on the website. As studied by Anderson and Magruder (2012) and Angrist and Pischke (2010), this methodology allows identification of information revelation on reviews [2].

With information on reviews, consumers can infer that a building is of a certain quality. For example, a consumer estimates probability P, that is a function of the history of reviews $P(\hat{R})$, where $\hat{R}$ is the average of reviews rounded to the nearest 0.5 star. Since the reviews are correlated with the quality of the building, the probability of giving a review according to Q_M quality is greater after visiting a building that produces at that level of quality. The rates at which guests stop visiting the hotel

[2] The rounding of reviews to the nearest 0.5 can make two hotels that are very similar in terms of average reviews look different in quality. For example, a hotel that has 3.74 average stars is very similar to one that has 3.76 but with the rounding and what consumers see is that the former is a 3.5 stars and the latter is a 4 stars hotel.

are $v_L(\hat{R})$ for lower quality and $v_M(\hat{R})$ for production when maintaining quality Q_M. These rates are a function of what consumers observe in the reviews.

As studied by Mayzlin et al. (2014) and by Luca and Zervas (2016), reviews suffer from manipulation. The problem of review manipulation arises from the incentives hotels or restaurants have to manipulate their own reviews or the ones of the competition in order to increase patronage to their own locations. For example, a small competitor may want to increase ratings by posting a false positive review to convince consumers the hotel is of greater quality than it actually is with the hope of increasing patronage. The problem with this phenomenon is that consumers may perceive the hotel differently than it really is, therefore affecting the departure rates v_M and v_L. With this information the firm then needs to optimize their quality investment decision.

3.3.2 The Firm Problem

The firm decides whether to maintain its quality or extract rents from its reputation by producing at a lower quality. If the firm picks the latter, then the next period will suffer the loss of v_L customers. Equally, if the firm decides to maintain its quality, the next period the firm losses v_M consumers[3]. The firm weighs immediate profits against the discounted future profits in order to make the quality investment decision.

There are two types of production; one is maintaining quality with a constant marginal cost c_M and the other is a lower quality product with a marginal cost of c_L, where $c_M > c_L$. Firms are price takers in this market and p represents the price of the product. I assume that the market price for quality Q_M is p, and firm makes the quality decision with this price. For simplicity, I assume the firm makes once-and-for-all quality decision and I leave out oscillations of quality over-time. The firm's problem then can be reduced to maximizing the present value of the profits of either option:

[3] v_M and v_L are departure rates, so they can be seen as negative growth of customers.

$$\pi_M = (p - c_M) + (p - c_M)\frac{(1 - v_M)}{(1 + r)} + (p - c_M)\frac{(1 - v_M)^2}{(1 + r)^2} + (p - c_M)\frac{(1 - v_M)^3}{(1 + r)^3} + ...$$
$$\pi_L = (p - c_L) + (p - c_L)\frac{(1 - v_L)}{(1 + r)} + (p - c_L)\frac{(1 - v_L)^2}{(1 + r)^2} + (p - c_L)\frac{(1 - v_L)^3}{(1 + r)^3} + ... \tag{3.2}$$

The profit maximizer firm will maintain quality of the product if and only if, $\pi_M > \pi_L$. Let's simplify the expression for π_M and π_L (see Appendix B.0.1.1).

$$\begin{aligned} \pi_M &= (p - c_M) + (p - c_M)\frac{(1 - v_M)}{(r + v_M)} \\ \pi_L &= (p - c_L) + (p - c_L)\frac{(1 - v_L)}{(r + v_L)} \end{aligned} \tag{3.3}$$

Therefore, by rearranging terms and using the profit maximization condition for firms to maintain quality, $\pi_M > \pi_L$, we obtain the price that guarantees quality (see Appendix B.0.1.2 for derivation):

$$\begin{aligned} &\pi_M > \pi_L \\ &p^* > \frac{(c_M - c_L) + \left[c_M\frac{(1 - v_M)}{(r + v_M)} - c_L\frac{(1 - v_L)}{(r + v_L)}\right]}{\left[\frac{(1 - v_M)}{(r + v_M)} - \frac{(1 - v_L)}{(r + v_L)}\right]} \end{aligned} \tag{3.4}$$

Immediately we observe that the larger difference of marginal cost would result in an increase of the price at which the firm maintains quality. Also, it is important to observe that as the discount rate r gets larger, the price at which a firm invests in maintaining quality also increases; as a result, more impatient firms will be less willing to maintain quality. Now let's focus on the consumer departure rules v_M and v_L, and the consequences these have on price. Basically, this is the number of customers the firm loses each period v_M and v_L and how the price responds to changes on those parameters. Propositions 1 and 2 summarize the effects of customer loss based on the assumption that consumers are identical in preference.

Proposition 1. *Lower departure rates of customers from firms maintaining quality, v_M, decreases the price required by the firm to invest in quality.*

Proof see Appendix B.0.1.3

Proposition 2. *Lower departure of consumers from lower quality firms results in higher prices required by firms to maintain its quality.*

Proof see Appendix B.0.1.4

From Propositions 1 and 2 we observe that the price would depend on how high departure rates from firms producing at lower quality are, or how loyal consumers are to firms that maintains quality. Thus, it is of interest to analyze what affects the departure rates v_M and v_L. For example, do consumers have a hard time recognizing firms that maintain their quality, confusing them for firms of lower quality? If so, then the departure rate v_M increases. In the previous example, consumers will depart from the firm as they observe that previous ratings are of lower quality, even though the firm maintains its quality. As a result of the higher departure v_M, we observe an increase in the price at which firms maintain the quality goods or services. These results from the model indicate that whenever consumers have a hard time recognizing the true quality of the product, the departure rules might not be set optimally and the consumers might depart from firms that maintain quality or stay with lower quality firms. Therefore, it is important to understand what may help consumers discern when a firm maintains quality, and what is the role of consumer generated content. The reviews may suffer from manipulation of competitors and in those cases a firm that invests in quality may end up been flagged as lower quality and therefore the departure rule v_M will be set up high by consumers. Markets where competitors alter each other's reviews results in a higher price required by firms in order to maintain quality, and consequently lower investment in the buildings.

3.4 Empirical Specification

The primary interest of this project is to examine the impact of information revelation on investment decisions of capital expenditure. Equation 3.4 provides the relation between the departure rate of consumers and the price at which a firm starts investing in quality. Hotels have various attributes in relation to quality, for example, service quality, hygiene quality, location quality, and property quality. In this project, I am interested in the quality of the real estate of the hotel measured by investment in capital improvements. The parameters required to determine the price at which the firms invest in quality are the marginal costs, the departure rates from the good and bad quality firms ex-post, and the discount rate. In order to estimate the relation between reviews and performance variables like occupancy rate, revenue per room and average daily rate, I estimate the following equation

$$Perf_{it} = \alpha_i + \beta_0 S_{it} + \beta_1 A_{it} + \beta_2 Y_t + \beta_3 L_i + \beta_4 PT_i + \beta_5 PT_i : Y_t + \varepsilon_{it} \tag{3.5}$$

where $Perf_{it}$ represents the log value of a performance measure (occupancy rate, revenue per room and average daily rate), α_i is the property fixed effect, S_{it} is the log value of the cumulative average of star rating rounded to the nearest 0.5 star, A_{it} is the age of the building, Y_t is the year fixed effect, L_i is the location fixed effect, PT_i property type fixed effect, $PT_i : Y_t$ is the interaction term between property type and year, and ϵ_{it} is the residual. The parameter of interest in the model is β_0 as this will be revealed if consumers respond to changes in reviews, which implies that the reviews inform the market of the quality of the building.

In order to further study the impact of a 0.5 star rating change on performance, I use a model similar to the one proposed by Anderson and Magruder (2012). The idea is that consumers observe only an approximation of the average star rating to the nearest 0.5 star. Therefore, I create eight thresholds that identify when a hotel increases to the next 0.5 star. For example, if the average rating of the hotel is 3.24

stars, the approximation shows the hotel rating as 3.00 stars, but for a 3.25 rating the approximation shows the hotel rating as 3.5 stars. In this example, 3.25 stars is the threshold. This discontinuity allows the identification of the impact of reviews on performance. The thresholds I use for this specification are 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, and 4.75. I use the following empirical model:

$$Perf_{it} = \alpha + \vec{\beta_1} f(\bar{R_{it}} - R_{thre}) + \beta_2 I(\bar{R_{it}} > R_{thre}) + \vec{\beta_1} 2 f(\bar{R_{it}} - R_{thre}) I(\bar{R_{it}} > R_{thre}) + \vec{\beta_3} \vec{F.E.} + \varepsilon_t \tag{3.6}$$

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

The second model of interest is the impact of reviews on investment decisions. A characteristic of capital expenditure is that its amount varies significantly from period to period, ranging from a relatively small percentage of the accounting value of the property, to percentages that exceed 20% of its value. Since there are periods where capital improvements are sufficiently small, I dichotomize the values by assigning a 1 to the top 20% of the capital expenditure measured as a percentage of the net book value of the property and zero otherwise[4]. This allows me to estimate a

[4]I divide the capital expenditure for a given year by the book value, then I rank these values and create an indicator variable that takes the value of 1 if the value was above the 80th percentile. In terms of revenue, these capital expenditures exceed 5% of the revenues of the hotel for a given year

linear probability model that associates the value of investment to consumer reviews, building fixed effects, time varying fixed effects at various locations, and market levels. The idea of the introduction of year fixed effect is to control for unobservables specific to a year. The location fixed effect allows me to control for invariant location or market characteristics, and finally the time varying location interaction captures changes over time at the location level or property type level in an effort to control for lease rates and changes in lease rates at the market level.

$$Pr(Capex_t = 1) = \alpha_i + \beta_0 S_{it} + \beta_1 A_{it} + \beta_2 Y_t + \beta_3 C_t + \beta_4 L_i + \beta_5 PT_i + \beta_6 PT_i : Y_t + \varepsilon_{it} \quad (3.7)$$

Here the dependent variable represents the probability that capital expenditure exceeds the 80^{th} percentile of the relation capital expenditure over net book value. C_t is the time since the last investment in capital expenditure. The rest of the variables are the same as in the model in Equation 3.5 and ϵ_{it} is the residual. I use some independent variables including the textual analysis of the consumer review (i.e., word count of messages in the review), building characteristics (i.e., chain affiliation) to subsample the data and run models 3.5 and 3.7. I also run the regression discontinuity design from Equation 3.6 to further study the impact of reviews on investment decision.

In this study, I am interested in the relation between investment decisions and online reviews. From previous literature, we know that reviews affect consumers' purchase patterns (Brown et al., 2007; Anderson and Magruder, 2012), and that hotels therefore have the incentive to manipulate those reviews. At the moment of manipulating reviews, hotels need to weigh the benefits of manipulation against the cost of getting caught (Mayzlin et al., 2014). The literature tells us that hotels affiliated with a chain brand have less incentive to manipulate reviews than independent

---

and on average are equal to 52% of the revenues of a hotel for a year. This is important since 5% of sales is the standard Capex required by hotel brands in the management contract or franchise agreement.

hotels since one brand property's reviews can damage the reputation of an entire chain. We also know that small competitors are associated with manipulation of negative reviews (Mayzlin et al., 2014; Luca and Zervas, 2016).

First, I test whether or not hotels are affected in terms of performance metrics (i.e., occupancy rate; average daily rate; and revenue per available room) by reviews. The importance of this hypothesis is that it directly relates to Propositions 1 and 2, as this influences the departure rates of customers after observing high or low ratings in the reviews.

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

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

Hypothesis 2: A change in the rating received by a hotel changes the probability of investing in the property.
Now, from the review manipulation literature, we know that not every hotel has truthful reviews. Independent hotels tend to have larger numbers of review manipulation (Mayzlin et al., 2014). We should expect that in cases of hotels with independent brands, the impact of reviews would be smaller. Thus, the next 4 hypotheses deal with the quality of the information embedded in the reviews. Review manipulation decreases the value of the information provided by reviews either due to a positive bias in the case of promotional review or negative bias in the case of a competitor writing a bad review.

Hypothesis 3: Reviews of hotels affiliated to a chain should reveal more realistic information of hotel quality and therefore have greater effects when changes in ratings occur.

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

Hypothesis 4: Reviews of hotels in areas with larger numbers of competitors have a stronger effect on investment decisions.

My next hypothesis tests whether or not the number of reviews that a hotel has influences the statistical significance and economic magnitude of the impact of reviews. If a Type II error occurs when a customer posts a review, then the departure rate from a high quality firm will be too high, or in the opposite case, in a bad quality firm the departure rate will be too low.

Hypothesis 5: The number of reviews received increases the economic and statistical significance of changes in ratings.

The last hypothesis relates to proposition 1 and 2, and looks at the implications of review manipulation overall. Hotels that have more accurate reviews should have lower cutoff prices at which they invest in good quality (see equation 3.4). If this is the case, hotels with less review manipulation should have more investment overall, as price at which starts investing in quality decreases.

Hypothesis 6: Hotels that are brand affiliated, receive a larger number of reviews, and face larger competitors invest more often in capital expenditure.

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

3.4.1 Treatment Effect Capex on Reviews

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

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

$$Rating_t - Rating_t^{Comp} = \beta_0 + \vec{\beta_1} f(T) + \beta_2 I(T > T_0) + \vec{\beta_1}2 f(T) I(T > T_0) + \vec{\beta_3} \vec{F.E.} + \varepsilon_t \quad (3.8)$$

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

Hypothesis 7: Investment in capital expenditures leads to rating improvements.

3.5 Data

Consumer generated Internet content in the travel industry is particularly important at the moment of booking a hotel. Prior to booking a room, independent of the site used to do so, the consumer has information on the star rating given by prior guests and a brief description of the experience during the stay. Although this study uses only hotels in the U.S.A, TripAdvisor.com has consumer reviews for more than 7 million accommodations, attractions and restaurants in 49 markets[5]. With 535 million reviews and more than 415 million visits per month, TripAdvisor.com is the leading provider of consumer generated content in the travel industry. The main data source for this project is a collection of 8,663,790 reviews for hotels within the U.S.A. from this website; Table 3.1 provides summary statistics for the full sample of reviews.

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

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

Table 3.1. Summary Statistics Reviews

Statistic	N	Mean	St. Dev.	Min	Max
Number of Reviews	10,338,729	5.75	16.91	0	584
Star Rating	9,744,305	40.36	11.66	10	50
Positive-Negative Words	9,109,599	4.15	3.98	−394.48	281.11
Location Words Count	9,109,599	6.17	5.61	0	473
Quality Words Count	9,109,599	2.48	2.69	0	252
Building Words Count	9,109,599	5.14	5.73	0	497
Words Count	9,109,599	116.74	111.67	1	12,682
Manager Response	10,339,562	0.46	0.50	0	1
Sleep Rating	5,525,009	4.18	1.11	1	5
Location Rating	6,147,327	4.39	0.89	0	5
Room Rating	6,183,886	4.08	1.13	1	5
Service Rating	7,872,697	4.23	1.13	1	5
Value Rating	6,352,449	4.02	1.17	1	5

Note: This table provides the summary statistics of all the reviews collected. N. of Helpful Reviews is the number of helpful votes reviewers received. Star Rating is the rating reviewers gave the hotel after their visit. Positive - Negative Words is the count of positive minus the count of negative words within the reviews. Location Words Count is the number of words that refer to location within a review. Quality Words Count is the number of words that refer to quality. Building Words Count is the number of words that refer to buildings' parts. Words Count is the number of words in a review. Manager Response is an indicator variable that takes the value of 1 if the manager responds to the review. Sleep, Location, Room, Service and value rating are the specific ratings reviewers gave to Sleep, Location, Room, Service or Value attributes.

minimum distance between the hotels of two samples are a match. I drop hotels for which the minimum distance was larger than 150 feet[6].

Investment and performance data comes from SNL, a company that collects information on publicly traded REITs. SNL is a subsidiary of S&P Global Market Intelligence. The information available is an unbalanced panel of hotels and other property types in the USA and around the world. I focus on hotels located in the

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

Table 3.2. Hotel Sample

City Hotel	Birmingham			
Hotel Name	EmbassySuites			
Address	2300 Woodcrest Place			
Property ID	17603			
State_Hotel	AL			
Number of Reviews	931			
Age as of 2015	28			
Number of Room	242			
Initial Value Land	2843			
Initial Value Building	29286			
Year	**Cumulative Capex**	**Net Book Value**	**Age**	**Invest Capex**
2005	95	25,564	18	0
2006	213	25,163	19	0
2007	2,447	26,094	20	1
2008	3,127	25,834	21	0
2009	3,171	25,002	22	0
2010	3,677	24,650	23	0
2011	3,690	23,735	24	0
2012	3,762	22,991	25	0
2013	3,796	22,122	26	0
2014	3,806	21,209	27	0
2015	3,862	20,377	28	0

Note: This table has an example of the type of information available for each hotel. Invest Capex is an indicator variable that takes the value of 1 whenever the investment is in the top 20% of the sample information. The hotel website in Tripadvisor is `https://www.tripadvisor.com/Hotel_Review-g30375-d72325-Reviews-Embassy_Suites_by_Hilton_Birmingham-Birmingham_Alabama.html` .

U.S.A. that have data available for subsequent building improvements. The capital expenditure information comes from the 10-k reports in section III, in which REITs provide a detail by property of the subsequent improvement after acquisition. In Table 3.2 in the appendix is a sample of the type of information available for each hotel. Figure 3.2 shows the distribution of capital expenditure as a percentage of the book value of the hotels in my sample. I create a binary variable for the capital expenditure, whenever investment exceed 3.5% of the book value of the building I assigned a 1. Figure 3.3 shows the distribution of this binary variables and approximately 32% of the time capital expenditures exceeds the 3.5% threshold. I use this binary variable to estimate the linear probability model of the decision of managers to investment larger amount in building improvements. Although SNL provides information for more than 4,900 hotels which represents 10% of the total number of hotels in the USA, I use a subsample of 1,815 buildings with information on Capital Expenditures and that have consumer reviews. My final sample has 6,416 property year observations.

Figure 3.4 shows the distribution across space of the hotels in the sample. The background of the map is a heat map that has the density of the economic activity for Real Estate and Rental and Leasing industry (NAICS 53). It seems by looking at the graph that hotels in the sample are not biased in terms of location and have similar spatial distribution as the economic activity of interest. To further review this claim, I use the list of competitors and perform a T-test on the sample matched, and the entire list of competitors; Table 3.3 shows this analysis. Although the buildings in the sample are spatially distributed according to the entire population of hotels, there is an over representation of buildings in the MidAtlantic (+4%), and an under representation Southeast (-3%) and West(-5%). Unsurprisingly given that I use REITs properties, other hotel characteristics differ in attributes that describe size, brand[7], and competitive environment.

Importantly, the goal of this paper is to study ratings and their impact on investment. While attributes that refer to size may differ, the attributes that refer to reviews are not different from the entire population of hotels. Hotels in my sample receive on average 5.29 reviews per room and the entire population receives 5.29 with no economical or statistical significance for the difference. Also the ratings in my sample receive on average 3.75 star rating while the population of hotels receives

[7]The brand variable is a dummy variable that takes the value of one if the hotel has one of the brands described in Table 3.4.

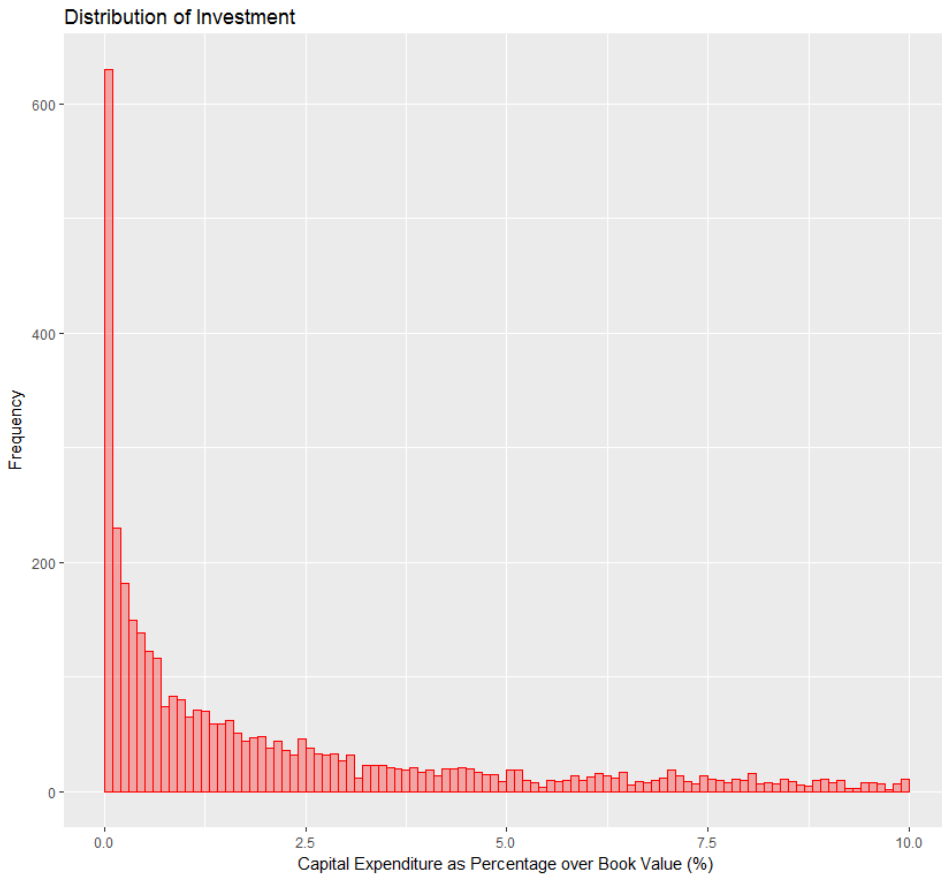

Figure 3.2. Distribution of Investment

This histogram shows the Capital Expenditure for the buildings. Capex is expressed as a % of the initial book value. The distribution goes beyond 10% of initial value but the density beyond that number is very low. I consider for the model any investment greater than 3% as a major capital expenditure and a renovation of the building.

a 3.73. Similarly, in Mayzlin et al. (2014) the average rating of their sample was 3.52 for TripAdvisor.com and 3.95 for Expedia.com.To construct my variable I use Tripadvisor.com on-line reviews. I read every consumer review posted on their website and construct average rating per year for each building on a scale from 1 to 5 stars. Figure 3.5 provides an example of the list of hotels in the system. Figures 3.6 and 3.7

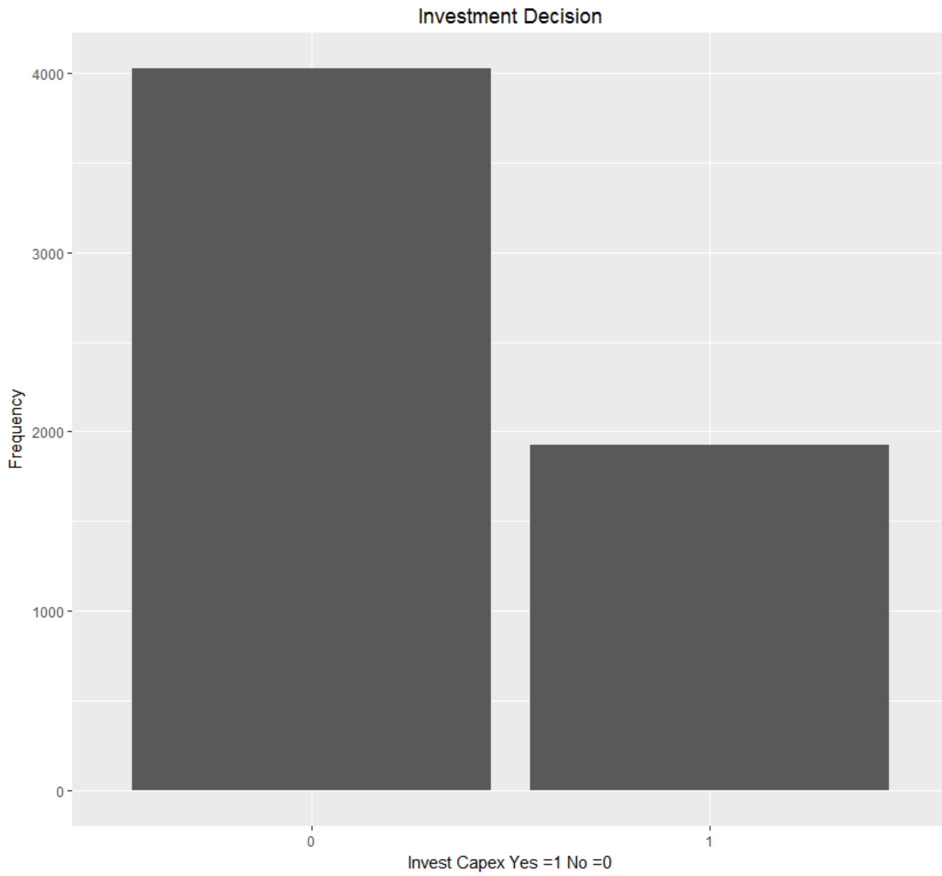

Figure 3.3. Distribution of Decision Investment for sample

This histogram shows the number of times managers of buildings choose to use capital expenditures above the threshold considered a renovation. The threshold I use in this study is 3% given the distribution of events. In the graph, 1 means that the investment was above the threshold and 0 means it was not.

show examples of 5 stars(good) or 1 star(bad) reviews.

TripAdvisor.com, as well as most sites that offer consumer reviews and ratings, rounds the star rating to the nearest 0.5 star. For example, if the average ratings of reviews received by the hotel up to the date the customer visits the site is 3.24,

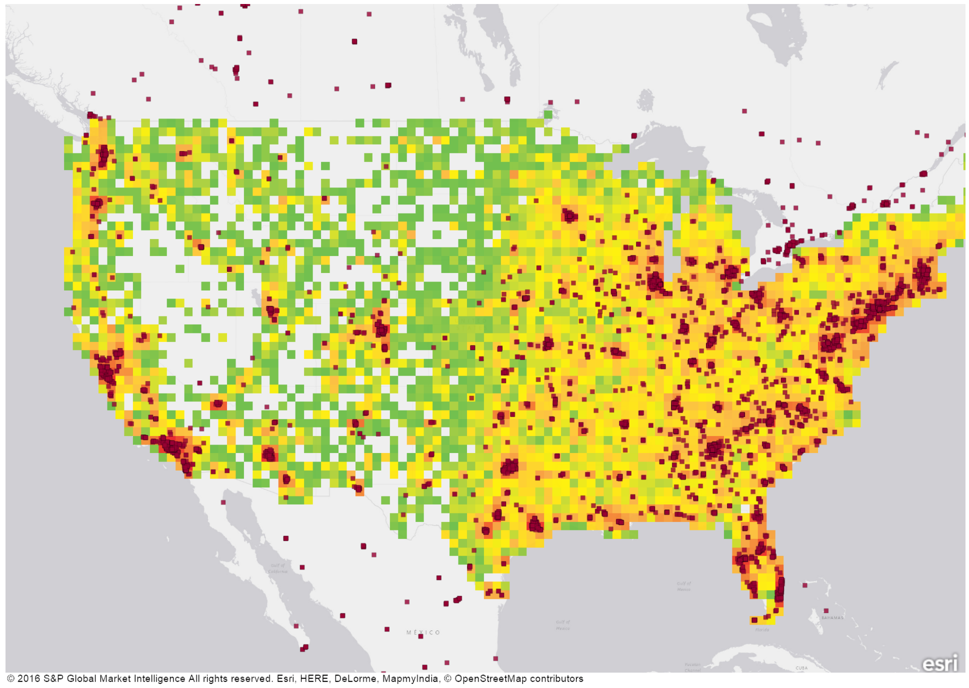

Figure 3.4. Hotels and Census

Each dot on the map represents a hotel from a publicly traded company. For the study, I use only those with information on capital expenditure and with on-line reviews. The map shows the hotels as dots and in the background the colors indicate the density of activity in the USA at the NAICS Industry Classification level for Real Estate and Rental and Leasing (NAICS:53). It seems there is no sampling bias, but I do show in Table 3.3 that North East and Mid Atlantic are slightly over-represented.

TripAdvisor.com will round and show consumers that the hotel is a 3 star. By the same token, if the hotel has an average of 3.25, it will show consumers the hotel is a 3.5 star. Following the discussion by Angrist and Pischke (2010) and by Anderson and Magruder (2012) this difference between the 3.24 hotel and the one that has an average of 3.25 creates exogenous variation that allows recognition of the impact of rating changes on performance data. I use this methodology to relate changes in ratings to performance and investment decisions. Therefore, for each year I create a cumulative average up to the date of interest and then round to the nearest 0.5 star.

Table 3.3. T-test of Sample and Population

Variable	Population	Matched	T-stat	P Value
Number of Rooms	102.48	154.60	-17.33	0
MidAtlantic	0.09	0.13	-5.87	0
Midwest	0.18	0.20	-2.52	0.01
Northeast	0.06	0.05	1.40	0.16
Southeast	0.29	0.26	3.19	0.01
Southwest	0.14	0.16	-2.68	0.01
West	0.24	0.19	4.17	0.00
Franchise	0.50	0.60	-7.89	0
Brand	0.58	0.72	-11.73	0
Year Established	1,997	2,004	-10.84	0
Employees Competitors 10 Miles	29.02	36.65	-15.62	0
Number of BnB 10 Miles	6.43	9.46	-11.59	0
Number of Reviews per Room	5.29	5.29	-0.01	1.00
Rating in Stars	3.73	3.75	-1.52	0.13

Note: This table shows the summary statistics of the hotels from the entire population I collected (49,247 hotels) and compares them to the subsample of hotels I use in the analysis that follows. Number of Rooms is the average number of rooms per hotel. Mid Atlantic is an Indicator variable that takes the value of 1 if the hotel is in that region. The same idea goes for variables Mid-west, Northeast, Southeast, Southwest and West. Brand and Franchise are indicator variables that take the value of 1 if the hotel is a Franchise or Brand, and 0 if not. Year Established is the year the Hotel open its doors. Employees Competitors 10 Miles is the average number of employees of all the competitors in a radius of 10 miles. Number of Competitors 10 Miles is the number of competitors within that radius. Number of BnB 10 Miles is the number of Bed and Breakfasts within 10 miles. Number of Reviews per Room is the number of reviews a hotel has received per room. Rating in Stars is the average rating measured between 10-50, where 10 represents 1 star and 50 presents 5 stars. %Manager Comments per Room is the number of times manager responded to reviews scaled by the number of rooms.

Figure 3.8 provides the density of ratings in my sample, and Figure 3.9 provides the average hotel ratings overtime. The graph at the top of the figure shows the average rating with a band of one standard deviation, and the bottom graph shows examples of how hotels ratings change overtime.

Table 3.4. Brands for Brand Dummy Construction

Parent Companies	Number of Hotels
Wyndham	5,656
Choice Hotels	5,335
Hilton Brands	4,615
IHG Brands	4,163
Marriot Brands	4,014
Bestwestern Hotels	1,492
La Quinta	1,188
Motel6	1,140
Extended Stay America	964
Red Lion	902
Radisson	489
Red Roof Inn	478

Note: This table shows the brands considered for the construction of the dummy brand. If a hotel is part of one the brands within those parent companies, the brand dummy takes the value of one. Next, I show you the brands of the top 5 chains. **Wyndham Brands**: Baymont Inn & Suites, Days Inn, Howard Johnson's, Knights Inn, Microtel, Ramada, Super 8, Travelodge, Wyndham, Wyndham Garden Hotels, Hawthorn Suites, and Wingate. **Choice Hotels** Brands includes: Ascend Hotel Collection, Cambria Hotels & Suites, Clarion Hotel, Comfort Inn, Comfort Suites, Econo Lodge, Mainstay Suites, Quality Inn, Rodeway Inn, Sleep Inn, Suburban Extended Stay Hotel, and WoodSpring Suites. **Hilton Brands**: Conrad Hotels & Resorts, Canopy, Curio - A Collection, Hilton Hotels & Resorts, DoubleTree, Embassy Suites Hotels, Hilton Garden Inn, Hampton, Homewood Suites, Home2 Suites, Hilton Grand Vacations and Waldorf Astoria Hotels & Resorts, Tru, and Tapestry Collection. **IHG Brands**: Avid Hotels, Candlewood Suites, Crowne Plaza, Even Hotels, Holiday Inn Hotels & Resorts, Holiday Inn Club Vacations, Holiday Inn Express, Holiday Inn Garden Court, Holiday Inn Resort, Hotel Indigo, Hualuxe Hotels & Resorts, InterContinental, Kimpton Hotels & Restaurants, and Staybridge Suites. **Marriott Brands:** JW Marriott, Ritz-Carlton, St. Regis, Bulgari Hotels & Resorts, Edition Hotels, W Hotels, Delta Hotels, Marriott, Sheraton, Autograph Collection, Design Hotels, Gaylord Hotels, Le Méridien, Renaissance Hotels, Tribute Portfolio, Westin Hotels & Resorts, Courtyard, Fairfield Inn, Four Points, Protea Hotels, SpringHill Suites, Moxy Wien-Schwechat, AC Hotels, Aloft Hotels, Moxy Hotels, Marriott Executive Apartments, Residence Inn, TownePlace Suites, and Element.

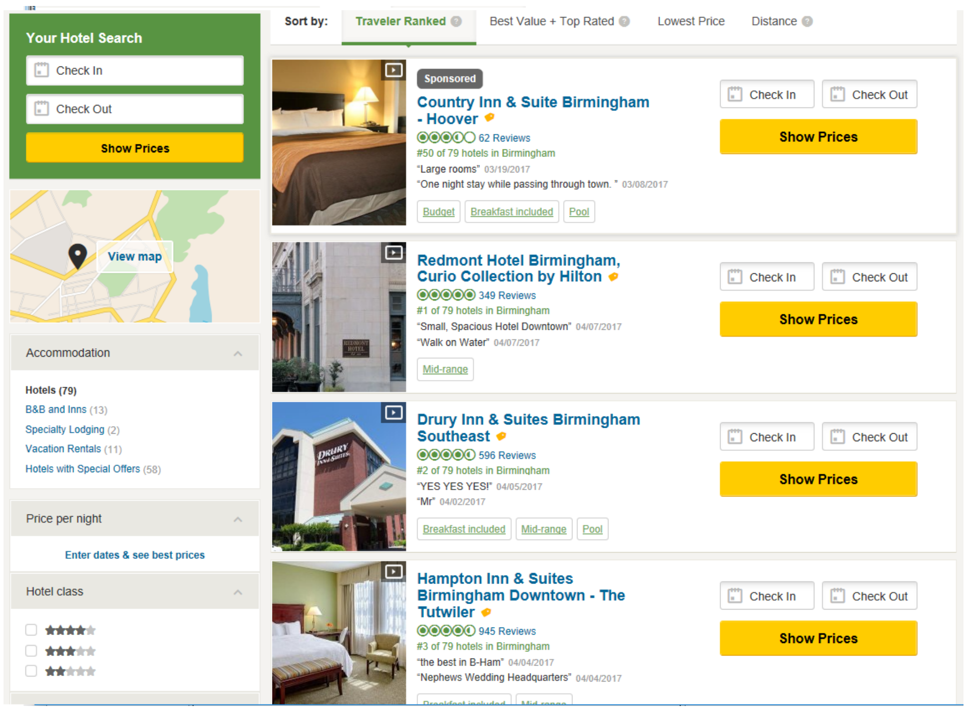

Figure 3.5. Hotels List within a City.

The site includes a list of hotels within each city. For each hotel I read the characteristics posted in the site as well as the address of the hotel. I then geocode that address and link the hotel information to the SNL data base.

As pointed out by Anderson and Magruder (2012), hotels may have incentives to manipulate reviews, given that a change of 0.5 stars may have repercussions on their occupancy and revenue. Although it is true that hotels may manipulate reviews to profit from this behavior in the short run, the benefits in the long run are not clear as most systems filter reviews and the risk of getting caught cheating can be quite significant[8]. Nevertheless, it is likely that hotels may manipulate reviews and therefore this needs to be controlled for. In order to do so, I use the AcNeilsen

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

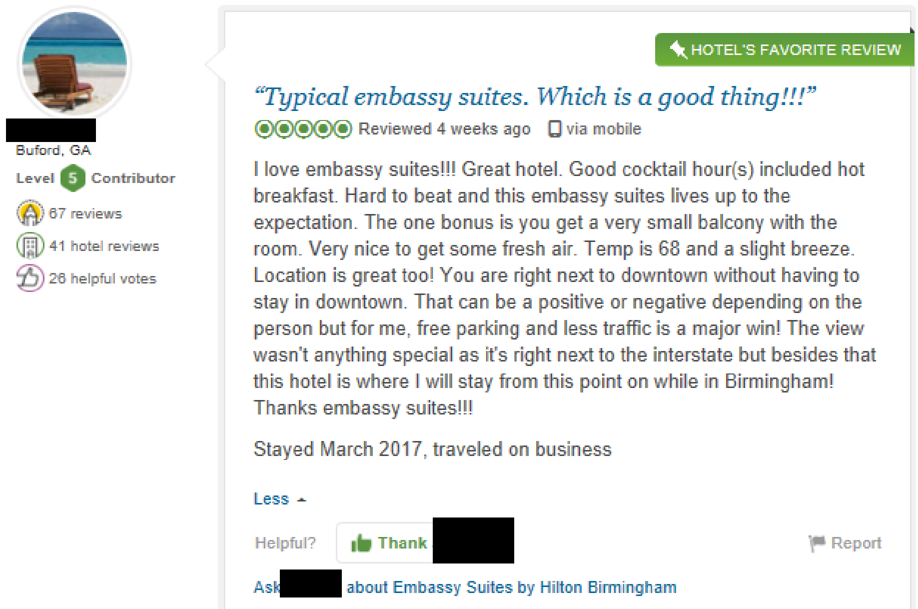

Figure 3.6. Example Positive Review

The picture provides an example of a good review for the hotel used as an example. I collect the information of the star rating (5 bubbles in this case). I collect the comments left on this review and count positive or negative words to asses the tone of the message. I also do the same thing for the response of the manager of the hotel.

Business listing, which contains information of brand, franchise status, estimated number of employees, address, and year established. This information allows me to estimate subsamples and run the analysis for subset of interest.

Although most of the properties provide latitude and longitude information, I use Google Maps API to geocode buildings with missing data or misleading data (i.e Hotel Latitude and Longitude reflects the centroid of the city rather than the centroid of the lot). The location allows me to build rings around each hotel and estimate the number of competitors and average size in terms of number of employees within the distance of analysis. Table 3.3 provides information of the average size of competitors for the matched sample. The number of employees of competitors in my sample is

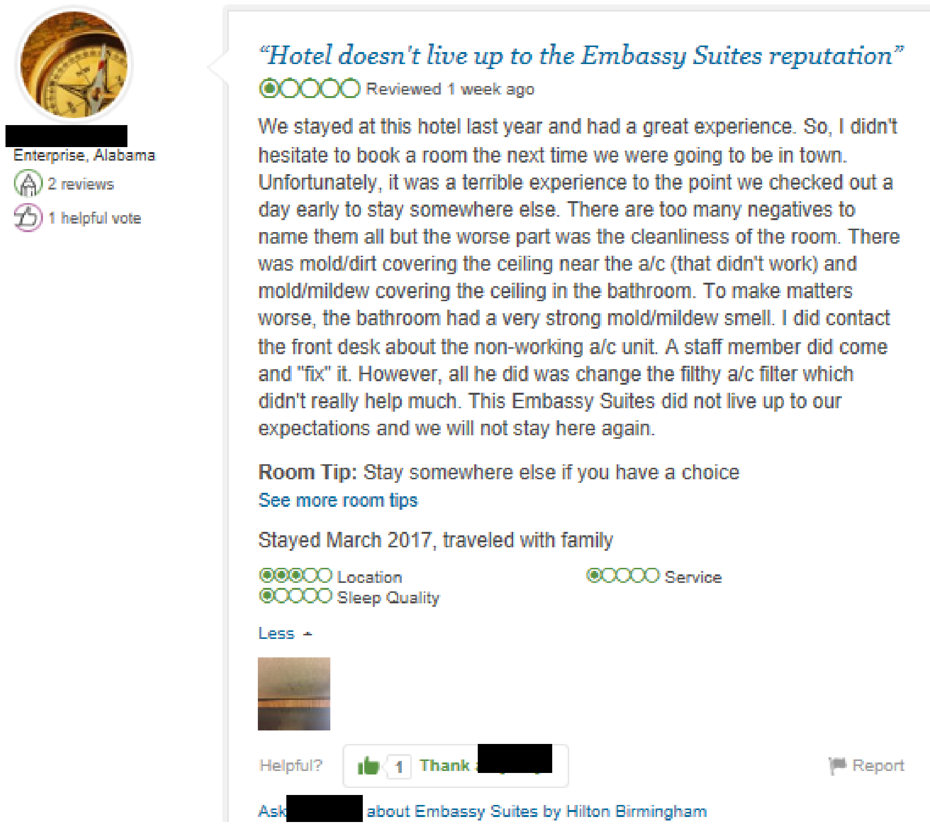

Figure 3.7. Example Negative Review

The picture provides an example of a bad review for the hotel used as an example. I collect the information of the star rating (1 bubble in this case). I collect the comments left on this review and count positive or negative words to asses the tone of the message. I also do the same thing for the response of the manager of the hotel.

larger than the number of employees of competitors for the entire population. This helps to recognize hotels that, given their competitive environment, may suffer from review manipulation and therefore see a decrease in the quality of the information provided on the reviews.

As result, I have a dataset with 6,416 property-year observations for 1,815 buildings,

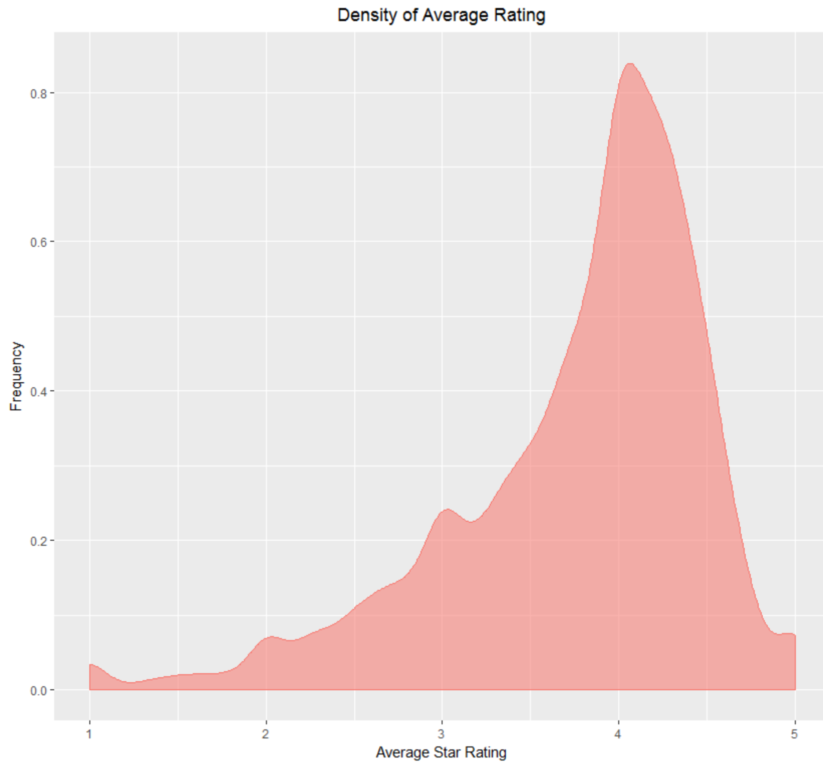

Figure 3.8. Density of Reviews for each building per year.

The graph shows the distribution of reviews for buildings in the sample for model 2. Each building has an average review per year. The reviews are on a scale from 1 to 5 stars.

with investment information for a time frame between 2006 and 2015. Reviews collected range from the year 2001 to 2015 and there is a total of 895,768 reviews with ratings information associated to those buildings.

3.6 Empirical Results

In this section, I present a set of results for the estimation of the model discussed in equations 3.5, 3.6 and 3.7. I provide some evidence for the impact of reviews on performance and investment decisions of hotels. The results, though, are not

Table 3.5. Summary Statistics

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

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

ubiquitous and there is indication that information effect may be heterogeneous across location characteristics as well as building characteristics. I use clustered-robust standard errors at the state level throughout the analysis to correct for standard error correlation within state.

3.6.1 The Effect of Reviews on Hotels' Performance

I start by exploring whether there is a relation between reviews and the performance of the hotels. I test whether reviews have any impact on occupancy rate, average daily rate, and/or revenue per available room. I begin by showing the results for the regression discontinuity analysis for the occupancy of the building. Table 3.6 shows the results for equation 3.6. The coefficient of interest is the one associated to Dummy Threshold; this indicates how occupancy changes with the increase of 0.5 stars. The coefficients are economically and statistically significant. The base case scenario suggests that an increase of 0.5 stars translate into a 5% increase in occupancy. The estimate seems consistent with the model that includes property type fixed effects, but becomes less statistically significant with the inclusion of other fixed effects. The last model of Table 3.6 includes Property Type, Region, and Building level fixed effects. Table 3.7 presents the results for the same model but in this case the dependent variable is revenue per room. The change in magnitude and statistical significance in both tables when including property fixed effect suggests that part of the results from the base case scenario are related to building specific characteristics not controlled for. To further analyze this possibility, I run the fixed effect model that includes Property Type, Region, Building level fixed effects, as well as interaction term with year fixed effect to control for market characteristics that change over time.

Table 3.8 presents the results of this analysis and reports the estimation of equation 3.5. The dependent variable for all columns are the log value of occupancy, the log value of the average daily rate, and the log of revenue per available room. The independent variable of interest is the log of rounded Star Rating to the closest 0.5

Table 3.6. Regression Discontinuity Model Occupancy

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

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

Table 3.7. Regression Discontinuity Model Revenue

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

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

star and lagged. The first three columns show the analysis for the full sample, while the last three columns show the regression for a subsample with only hotels above the 20th percentile ranked in terms of the number of competitors within 10 miles. The reason to use this subsample is that in areas where hotels do not have competition, consumers cannot select the hotel they go to (or at least have fewer choices); in a sense, consumers have no choice but to go to the hotel available independent of the information set. Therefore, the impact of reviews decreases.

Figure 3.9. Density of Reviews over time.

The graph shows the distribution of reviews for buildings overtime. Each building has an average review per year. The reviews are in a scale from 1 to 5 stars. The top graph shows reviews overtime for buildings in the sample. The bottom graph shows the reviews for a subsample of hotels, I round the star rating to the nearest 0.5 star for the analysis.

Table 3.8. Reviews on Performance and Competition Subsample

	All *Dependent variable:*			*Competitive Environment* *Dependent variable:*		
	Occupancy	Average Daily Rate	Revenue Per Room	Occupancy	Average Daily Rate	Revenue Per Room
	(1)	(2)	(3)	(4)	(5)	(6)
Star Rating Lag	0.119	−0.026	0.095	0.143	−0.024	0.122**
	(0.091)	(0.040)	(0.070)	(0.091)	(0.044)	(0.059)
Building Age	0.017***	−0.008***	0.009***	0.007***	0.032***	0.039***
	(0.002)	(0.002)	(0.003)	(0.000)	(0.000)	(0.000)
Constant	3.630***	5.319***	2.701***	3.572***	5.145***	4.098***
	(0.341)	(0.150)	(0.232)	(0.335)	(0.162)	(0.218)
Property F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Property Type F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time Variant Pro. Type F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Std. Errors	State Level	State Level	State Level	State Level	State Level	State Level
Observations	941	941	941	789	789	789
Adjusted R^2	0.799	0.973	0.960	0.844	0.977	0.972

Note: This table presents the OLS on variables that measure the performance of the hotels. Occupancy represents the log of the average occupancy rate of hotels within a year. Average Daily Rate represents the log of the average rate per room within a year. Rev Par is the log of the average revenue per room within a year. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. *Building Age* represents the age of the building. The columns under All use the entire sample of hotels that have performance variables available. The columns of **Competitive Environment** subsample of hotels that are in the top 80% of hotels in terms of number of competitors within 10 miles. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

The coefficients of interest are the ones associated with the Star Rating Lag. For the full sample, the estimated effect of star ratings on occupancy and revenue per room are positive but are not statistically significant. A 1% increase in the star rating increases occupancy by 0.119% and revenue per available room by 0.095%, although both coefficients are not statistically different from 0 but are of the expected sign. The impact of star rating on the average daily rate is smaller, negative, and again, not statistically significant. The effects in the sample that drops the least competitive areas in terms of number of hotels are statistically significant and of greater economic impact. In this case the effect of a 1% increase in the star rating translates into a 0.143% increase in occupancy rate and a 0.122% increase in revenue per available room; in both cases the effects are statistically and significantly different from zero. The impact of star rating on the average daily rate remains statistically insignificant and of a small magnitude. The economic impact of star rating on occupancy and revenue is quite significant; one standard deviation of star rating is 0.61 stars (or 16.02% change) which would translate into a 2.29% increase in occupancy and a 1.95% increase in revenue which is equivalent to $117,963[9] per year per hotel. These values are slightly smaller than the impact of star rating in the restaurant industry. Luca (2016) finds that a 0.5 star increase in ratings translates into an increase in revenues of 2.5% to 4.5%[10].

If reviews are manipulated, we should expect that they are less informative to consumers, and therefore, their impact on performance should decline. In order to test whether review manipulation affects the impact of the star rating on performance, I run a series of regressions on various subsamples that may have more truthful reviews. Following Mayzlin et al. (2014) and Luca and Zervas (2016), I subsample by brand affiliation and the size of competitors since both variables are associated with manipulation. Hotels that are brand affiliated have less incentives to create fake reviews as the reputation benefits come from brand rather than the reviews. Also, if caught cheating they risk the reputation of the chain. Another variable

[9] 1.95%*Revenue Per Room * Average Number of Room * 365 days = 1.95%* 106.96*154.60*365

[10] In my model the increase in revenue is 1.6%(13.13% increase in stars * 0.122).

that is associated with review manipulation is the presence of competitors and the size of these competitors. Hotels with smaller and independent competitors tend to receive abnormally negative reviews. Therefore, I subsample by brand and then I drop observation of hotels with competitors below the 20^{th} percentile in terms of average number of employees of competitors. I also consider a sample where I filter out areas with a larger number of bed and breakfasts to drop observations with small size competitors. One of the goals of this study is to identify the effects of information, so I subsample by the quantity of information. Specifically, I subsample by the number of reviews received and by the average word count in the reviews as proxies for quantity of information.

Tables 3.9 and 3.10 show the results from the subsample analysis previously described for occupancy rate and revenue per room available respectively. Table 3.9 indicates that hotels with high review counts, with fewer bed and breakfast in the area, with larger number of competitors experience a larger impact from changes in the star ratings and in all cases are statistically significant. I do not find a significant impact in the other subsamples. For example, regarding the impact of star rating in the brand sample, although the coefficient has the expected sign, it is not significant. One explanation for this is that reputation benefits in brand hotels do not come from reviews but rather the brand itself, so changes in reviews should not alter the occupancy as much as in independent brands. Now, Table 3.10 shows the results for the revenue per room equation. As with Table 3.9 all coefficients have the expected sign, but in this case the coefficient in the subsample with the larger number of review count is not significant.

As a robustness check, I run Equation 3.5 again, but in this case I include an indicator variable that takes the value of one if the hotel meets the criteria to be in all subsamples. I also include an interaction term between star rating and the indicator variable. I present the results for a regression using a full sample and the indicator variable in Table 3.11. Treatment indicates that a hotel meets the criteria of all subsamples and "Treatment:Star Rating Lag" is the interaction term between the indicator variable and star rating. The estimates for the coefficients suggest that the impact of star ratings on the hotels that do not meet the criteria to be in the subsample of truthful reviews is practically zero and insignificant from a statistical standpoint. Most importantly the sensitivity of the performance variable to changes in the star rating for the group of hotels that are likely to have more truthful reviews is large in terms of magnitude and statistical significance. These results suggest that the hotels more likely to have truthful reviews are more sensitive to changes in ratings.

Taken together, Table 3.8 shows evidence for Hypothesis 1 but is dependent on the level of competitions. Hotels with a lower number of competitors may not be as responsive to changes in ratings as hotels in highly competitive areas. There is no evidence for Hypothesis 3 in any of the estimates. One reason for this could be that reputation effects of online reviews for brand affiliated hotels might not be as important since reputation effects might come from the brand itself. Table 3.9 provides evidence that supports Hypothesis 5 for occupancy rates but not for revenue per available room in Table 3.10. Finally, Table 3.11 runs the analysis using an indicator variable that determines if a hotel is part of the group with more truthful reviews. This last analysis provides stronger evidence for Hypotheses 1,4 and 5.

Table 3.9. Reviews on Occupancy Subsample Analysis

	Dependent variable:					
	Occupancy					
	N. Review	Brand	Competition	Comp. Size	BnB	Words Count
	(1)	(2)	(3)	(4)	(5)	(6)
Star Rating Lag	0.179*	0.139	0.144*	0.143	0.140*	0.127
	(0.092)	(0.096)	(0.085)	(0.091)	(0.083)	(0.097)
Constant	3.660***	3.692***	3.752***	3.755***	3.715***	3.691***
	(0.320)	(0.368)	(0.318)	(0.364)	(0.369)	(0.370)
Property F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Property Type F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time Variant Pro. Type F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Std. Errors	State Level	State Level	State Level	State Level	State Level	State Level
Observations	676	724	783	789	684	793
Adjusted R^2	0.810	0.833	0.810	0.844	0.762	0.799

Note: This table examines the OLS for Occupancy Rate in various subsamples of hotels. The *dependent variable* for these models is the log of the average occupancy rate of hotels within a year. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. Column **Review Count**Count uses only hotels that have review counts in the top 80% of hotels in terms of review counts per room. Column **Brand** uses only hotels that are part of the top 30 hotel brands in terms of number of hotels by brand. Column **Competition** uses a subsample of hotels that are in the top 80% of hotels in terms of average number competitors within 10 miles. Column **Comp. Size** uses a subsample of hotels that are in the top 80% of hotels in terms of average number of employees of competitors within 10 miles. Column **BnB** uses a subsample of hotels that are in the bottom 80% of hotels in terms of the number of Bed and Breakfasts within 10 miles. Finally, column **Words Count** uses a subsample of hotels that are in the top 80% of hotels in terms of numbers of the average word count of the reviews received. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

Table 3.10. Reviews on Revenue per Room Subsample Analysis

	Dependent variable:					
	Revenue per Room					
	Review	Brand	Comp. Size	Competition	BnB	Words Count
	(1)	(2)	(3)	(4)	(5)	(6)
Star Rating Lag	0.112	0.120	0.116**	0.122**	0.125**	0.092
	(0.097)	(0.075)	(0.056)	(0.059)	(0.063)	(0.080)
Constant	3.552***	3.798***	4.312***	3.798***	4.873***	4.118***
	(0.385)	(0.310)	(0.223)	(0.267)	(0.366)	(0.292)
Property F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Property Type F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time Variant Pro. Type F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Cluster Std. Errors	State Level	State Level	State Level	State Level	State Level	State Level
Observations	724	724	783	789	684	793
Adjusted R^2	0.963	0.965	0.961	0.972	0.920	0.957

Note: This table examines the OLS for Revenue per Room in various subsamples of hotels. The *dependent variable* for these models is the log of the average revenue per room of hotels within a year. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. Column **Review Count** uses only hotels that have review counts in the top 80% of hotels in terms of review counts per room. Column **Brand** uses only hotels that are part of the top 30 hotel brands in terms of number of hotels by brand. Column **Competition** uses a subsample of hotels that are in the top 80% of hotels in terms of average number competitors within 10 miles. Column **Comp. Size** uses a subsample of hotels that are in the top 80% of hotels in terms of average number of employees of competitors within 10 miles. Column **BnB** uses a subsample of hotels that are in the bottom 80% of hotels in terms of the number of Bed and Breakfasts within 10 miles. Finally, column **Words Count** uses a subsample of hotels that are in the top 80% of hotels in terms of the average word count of the reviews received. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

Table 3.11. Ratings on Performance

	Dependent variable:	
	Occupancy Log Occupancy	Revenue Log Revenue
	(1)	(2)
Star Rating Lag	0.007	−0.033
	(0.082)	(0.059)
Treatment	−0.697	−0.787***
	(0.527)	(0.256)
Treatment:Star Rating Lag	0.193	0.212***
	(0.144)	(0.073)
Constant	4.202***	4.464***
	(0.281)	(0.249)
Property F.E.	Yes	Yes
Year F.E.	Yes	Yes
State F.E.	Yes	Yes
Property Type F.E.	Yes	Yes
Time Variant Pro. Type F.E.	Yes	Yes
Cluster Std. Errors	State Level	State Level
Observations	941	941
Adjusted R^2	0.802	0.960

Note: This table examines the OLS for Occupancy Rate and Revenue per Room. The *dependent variables* for these models are the log of the average occupancy rate and the log of the average revenue per room of hotels within a year. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. Variable *Treatment* takes the value of 0 if the hotel is not in or 1 if hotel in: takes the value of 0 if the hotel is not in, or 1 if the hotel is in:, the top 80% of hotels in terms of the number of reviews received; the top 80% of hotels in terms of the average number of employees of competitors within 10 miles; the bottom 80% of hotels in terms of the number of bed and breakfast within 10 miles; the top 30 brands of hotels in terms of the number of buildings; and, the top 80% in terms of words count within reviews. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

3.6.2 The Effect of Reviews on Hotels' Investment Decision

The main goal of this paper is to establish the relation between online consumer review and investment decisions in capital expenditure. Propositions 1 and 2 establish a theoretical relation between the departure rule of consumers and the decision to invest in quality. If this holds, then we should expect that changes in the departure rule would translate in more or less investment, depending on the direction of the change. From the literature on the optionality of capital expenditure, we know that managers would delay or accelerate investments depending on the market conditions, specifically rent and volatility of rent. I run various specifications of Equation 3.7 which take different definitions of market. Table 3.12 presents the estimates for linear probability models (LPM), and I show the estimates for the probit models as a robustness test. The dependent variable for all the models in this subsection is an indicator variable that takes the value of 1 if the capital expenditure exceeds the 3.5% of the book value of the building. Each column has a combination of other fixed effects to control for building, year and location characteristics. I also control for age of the building and time since last investment of significant capital expenditure to rule out that the results are driven from just pure obsolescence of the building or by recurring investment required by brands in order to keep building up to standard.

Table 3.12. Regression Effect Reviews on Investment

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

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

In the first column in Table 3.12, I use the hotel type as my market definition. I use fixed effect for Full Service, Limited Service, Extended Stay and Budget Hotel, and time varying fixed effects at the hotel type level. The results indicate that a 1% drop in the star rating increases the probability to invest in capital expenditure above 3.5% of the book value by approximately 0.200% of investment and is statistically significant at the 1% level. This gives strong evidence for Hypothesis 2; hotels change the probability of investing in capital expenditures depending on star ratings. Other LPM specifications are consistent with this finding and all of them imply that changes in star ratings affect the decision to invest, even after controlling for building characteristics, location fixed effect and time varying fixed effect. The results are also robust to specifications using the Probit model. However, for this last robustness check the inclusion of a property fixed effect does not allow the convergence of the Probit model; therefore, I drop the fixed effect at the property level. I also drop time since capex variable as it gives me the same problem. I do include location fixed effect at various levels, state, county and zip code, property type fixed effects and a time variant fixed effect to control for market conditions. The results still hold, estimates are statistically significant, and the take away from this table is that changes in star rating alter the decision to invest.

The existence of review manipulation creates some challenges for consumers to recognize which hotels are of good or bad quality, thus affecting the impact that changes in star ratings have on investment decisions. For example, areas where reviews are manipulated may see a Type II error. Propositions 1 and 2 establish the relation between the departure rule from good and bad quality firms with the price at which firms start investing in quality. In the presence of a Type II error, a hotel that is producing at high quality but whose reviews are being manipulated by a competitor to make it look bad in quality could be flagged by consumers as low quality. This Type of II error would lead to making the departure rule, v_M, from a good quality large and therefore making the price at which the hotel invests in capital expenditure higher(See Proposition 1). In the same way, poor quality hotels that write promotional reviews depicting the hotel as high quality may lead consumers to set a very low departure rule from low quality. Proposition 2 indicates that lower departure rates from poor quality hotels would lead to higher prices at which firms invest in capital expenditures. In the real estate scenario, if consumers do not punish low quality buildings with increases in the vacancy or do not reward good quality buildings with greater occupancy rates, then the rent at which investment in capital expenditure takes place is greater, leading to a drop in capital expenditures.

Following the manipulation literature, I take steps similar to the subsamples in Tables 3.9 and 3.10. I select samples of hotels that are less likely to have manipulated reviews like brand hotels and hotels with bigger competitors (Mayzlin et al., 2014; Luca and Zervas, 2016), and also filter by the number of reviews received. Table 3.13 shows the estimates for various subsamples and the evidence suggests that the hotels with larger competitors have a significant and larger effect of star rating on their decisions to invest. The evidence for other subsamples is significant but of lower magnitude and of the expected sign, the coefficients are less statistically significant.

Table 3.13. Regression Effect Various Sample of Information

	Dependent variable:					
	Invest in Capex Yes(1)/No(0)					
	Base	Brand	Reviews	Words	Competition	Competition Reviews
	(1)	(2)	(3)	(4)	(5)	(6)
Star Rating Lag	-0.200^{***}	-0.224^{***}	-0.284^{**}	-0.185^{**}	-0.202^{***}	-0.192
	(0.069)	(0.076)	(0.125)	(0.082)	(0.075)	(0.121)
Building Age	0.023	0.017	-0.005	0.038^{**}	0.016	0.125^{***}
	(0.017)	(0.020)	(0.009)	(0.019)	(0.020)	(0.022)
Constant	0.087	0.816^{***}	2.136^{***}	0.144	0.767	0.974^{*}
	(0.549)	(0.260)	(0.565)	(0.382)	(0.755)	(0.579)
Property F.E./Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State/County/Zip Code F.E.	No	No	No	No	No	No
Property Type F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time Variant F.E.	Pro. Type	Pro. Type	Pro. Type	Pro. Type	Pro. Type	Pro. Type
Cluster Std. Errors	State Level	State Level	State Level	State Level	State Level	State Level
N of Buildings	1815	1334	1635	1470	1445	1307
Observations	5,884	4,360	4,894	4,555	4,691	3,948
Adjusted R^2	0.389	0.378	0.262	0.395	0.400	0.422

Note: I run the model from Table 3.12 in column 1 and then subsample my data to a select group of hotels that have less odds of being misclassified due to reviews that mis-characterize the hotel. This is a Linear Probability Model with the dependent variable Invest in Capex, which is an indicator variable that takes the value of 1 if the capital expenditure is within the top 20% of capital expenditures. All standard errors are clustered robust at the State level. *Star Rating Lag* is the log of average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. *Building Age* represents the age of the building. I remove *Time Since Capex* as introduces multicollinarity with age and year fixed effect. Column **Brand** has the subsample of hotels affiliated to one of the top 30 brands in terms of the number of hotels. Column **Reviews** is the subsample of hotels within the top 80% of hotels in terms of the number of reviews received. Column **Words** is the subsample of hotels within the top 80% of hotels in terms of average word count of reviews. Column **Competition** is the subsample of hotels within the top 80% of hotels in terms of average number of employees of competitors within 10 miles. Column **Competition Review** is the subsample of the top 80% of hotels in terms of Number of Reviews and Average Number of Employees of Competitors. *p<0.1; **p<0.05; ***p<0.01

To further study these subsamples, I estimate the same model of Equation 3.7 but in this case I add an indicator variable that takes the value of one if the hotel is within a group likely to have fewer review manipulations and is located in a competitive environment. On one hand, Table 3.14 suggest that there is no statistical evidence that in competitive areas hotel respond to reviews. Hotels with larger review counts may suffer from less Type II error; in that case the departure rates will be set more efficiently (large for bad quality firms and small for good quality firms). Therefore, if that is the case, we should expect more investment in these hotels. Column 2 of Table 3.14, shows that Treatment:Star Rating Lag is positive suggesting that indeed they are more sensitive to changes. Also, the results in column 3 indicate that the hotels in the last subsample are more sensitive to changes in star ratings. A 1% increase in star rating generates a 0.224% drop in the probability to invest in Capex, but for the subsample of interest this probability decreases and extra 0.03%. Both coefficients are statistically significant and with a large economic impact.

Table 3.14. Regression Effect of Reviews on Investment

	Dependent variable:		
	Invest in Capex Yes(1)/No(0)		
	Competetive Env.	Reviews Count	Both
	(1)	(2)	(3)
Star Rating Lag	-0.250^{*}	-0.194^{***}	-0.199^{***}
	(0.129)	(0.069)	(0.069)
Treatment:Star Rating Lag	0.084	-0.019^{**}	-0.016
	(0.154)	(0.008)	(0.010)
Building Age	0.024	0.030^{*}	-0.0001^{***}
	(0.017)	(0.017)	(0.00003)
Constant	-0.034	-0.135	0.670^{**}
	(0.576)	(0.560)	(0.269)
Property F.E.	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes
Property Type F.E.	Yes	Yes	Yes
Time Variant F.E.	Pro. Type	Pro. Type	Pro. Type
Cluster Std. Errors	State Level	State Level	State Level
N of Buildings	1815	1812	1812
Observations	5,884	5,875	5,875
Adjusted R^2	0.389	0.390	0.390

Note: I run the model from Table 3.12in column 1, but in this case, I include and indicator variable *Treatment* as well as an interaction term between *Treatment* and the *Star Rating Lag*. *Treatment* is an indicator variable that takes the value of 1 if the hotel is part of the subsample of interest and takes the value 0 if not in the group. In column **Competetive Env.** (Short for Competitive Environment) the *Treatment* variable takes the value of 1 if the hotel is in an area of the top 80% of hotels in terms of the number of competitors within 10 miles and in terms of the size of competitors within 10 miles. In column **Reviews Count** the *Treatment* variable takes the value of 1 if the hotel is in an area of the top 80% of hotels in terms of the number of of reviews received. In both columns the Treatment variable takes the value of 1 in hotels that are in the top 80% of hotels in terms of number of competitors within 10 miles, the top 80% of hotels in terms of size of competitors within 10 miles(measured in average number of employees per competitor) and in the top 80% of hotels in terms of the number of reviews received. *Star Rating Lag* is the average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. *Building Age* represents the age of the building. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

My last robustness test examines the impact of a February 2012 ruling by the Advertising Standard Authority in the U.K. which affected the way TripAdvisor.com screens the reviews posted[11]. The ruling requested TripAdvisor.com to stop claiming that the reviews were "Reviews you can trust"after KwikChex Ltd and two hotels challenged the veracity of that claim and argued that the quote was misleading. This brought changes and an increase in the efforts by TripAdvisor.com to filter fraudulent reviews. For this reason, if there was a decline of fraudulent reviews posted on the site, we should expect that reviews to become more informative after this year. I run the same test in column 3 of Table 3.14 but split the sample into pre and post 2012. I also exclude the restriction of number of competitors as I want to test the quality of reviews. Table 3.15 indicates that the coefficients are larger for the period after 2012, suggesting that the impact of reviews on investment decision increased by approximately 33 %. The subsample of hotels is also more sensitive to changes in star ratings; a 1% increase in ratings translates into an extra 0.051% drop in the probability to invest.

These results provide evidence for Hypotheses 2, 4, 5 and 6. Investment decisions are sensitive to changes in star ratings of reviews; more reviews and competitive environments translate into larger and statistically significant impacts of star ratings on the probability to invest in capital expenditure. Again, I do not find evidence for Hypothesis 3, brand affiliated hotels have smaller but statistically significant effects.

[11] https://www.asa.org.uk/rulings/tripadvisor-llc-a11-166867.html

Table 3.15. Regression Effect Pre(<) and Post(≥) 2012

	Dependent variable:	
	Invest in Capex Yes(1)/No(0)	
	Pre (<) 2012	Post (≥) 2012
	(1)	(2)
Star Rating Lag	−0.222*	−0.280*
	(0.117)	(0.156)
Treatment:Star Rating Lag	−0.024	−0.037
	(0.017)	(0.027)
Building Age	−0.0003**	0.000
	(0.0001)	(0.000)
Constant	0.489	0.900*
	(0.402)	(0.501)
Property F.E.	Yes	Yes
Year F.E.	Yes	Yes
Property Type F.E.	Yes	Yes
Time Variant F.E.	Pro. Type	Pro. Type
Cluster Std. Errors	State Level	State Level
N of Buildings	770	1571
Observations	1,999	3,876
Adjusted R^2	0.219	0.278

Note: The ruling of the UK Advertising Standard Authority, in February of 2012, affected TripAdvisor.com, and since then, the firm has introduced new procedures in order to control fake reviews (https:www.tripadvisor.comvpagesreview_mod_fraud_detect.html). I run the model from Table 3.14in column labeled Both but in this case I run two separate regressions, pre 2012 from post 2012(post includes 2012). The variable *Treatment* is an indicator variable that takes the value of 1 if the hotel is part of the subsample of interest and takes the value 0 if not in the group. In both columns the Treatment variable takes the value of 1 in hotels that are in the top 80% of hotels in terms of number of competitors within 10 miles and in the top 80% of hotels in terms of the number of reviews received. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after their visit, rounded to the nearest 0.5. *Building Age* represents the age of the building. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

Table 3.16. Regression Falsification Test

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

Note: I run the model from Table 3.12in column 1, but in this case, I falsify the investment variable. I recreate a random investment decision (0 or 1). The *Treatment* takes the value of 1 in hotels that are in both the top 80% of hotels in terms of number of competitors within 10 miles and in the top 80% of hotels in terms of the number of reviews received. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after visit, rounded to the nearest 0.5. *Building Age* represent the age of the building. *Time Since Capex* is the number of years since the last capital expenditure exceed the top 20% of all observation, I also include a quadratic term and cubic term of this variable in the regression. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

3.6.3 The Impact of Investment on Reviews

My last analysis is an event study around the period during which the investment takes place. Whenever I observe an investment greater than 3.5%, I construct the average rating per quarter for the 8 quarters pre and post the year of the event. I estimate Equation 3.8 and test different trends in the pre and post period for various subsamples. I drop reviews from the year the investment takes place and focus only on the pre and post period, as observations of ratings within that year are affected by construction in the building due to capital expenditures.

Table 3.17 shows the estimated coefficients for Equation 3.8. Columns 1 and 2 show estimates of the regression using the full sample. The trend variable is positive and significant; the ratings difference between the hotel and the closest competitor increase 0.038 stars per year. The trend coefficients for the specifications using all competitors within 14 kilometers buffers are statically significant and of the same sign. Treatment represents the period after the investment in capital expenditure takes place. The coefficients in the full sample regressions suggest that after investment takes place, star ratings increase by approximately 0.18 star relative to the closest competitor. Although not statistically significant, the interaction term between trend and treatment suggest that after investment takes place the effects of capital expenditure on ratings start decreasing. The relation between treatment and interaction term suggest that the effects of capital expenditure disappear after approximately 18 quarters relative to the closest competitors. Figure 3.10 shows the fitted values of the regression in column 1 of Table 3.17, the bottom graph shows the detrended values and leaves only the treatment effect of investment and the change in trend after treatment. I run several robustness tests that use a subsample for different definitions of competitors to study the effect of potential review manipulation; columns 3 to 4 show this analysis. The impact of treatment for Hotels that are less likely to have manipulated reviews.

For the columns under the Filter group, I use the subset of hotels that are less

Table 3.17. Regression Reviews

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

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

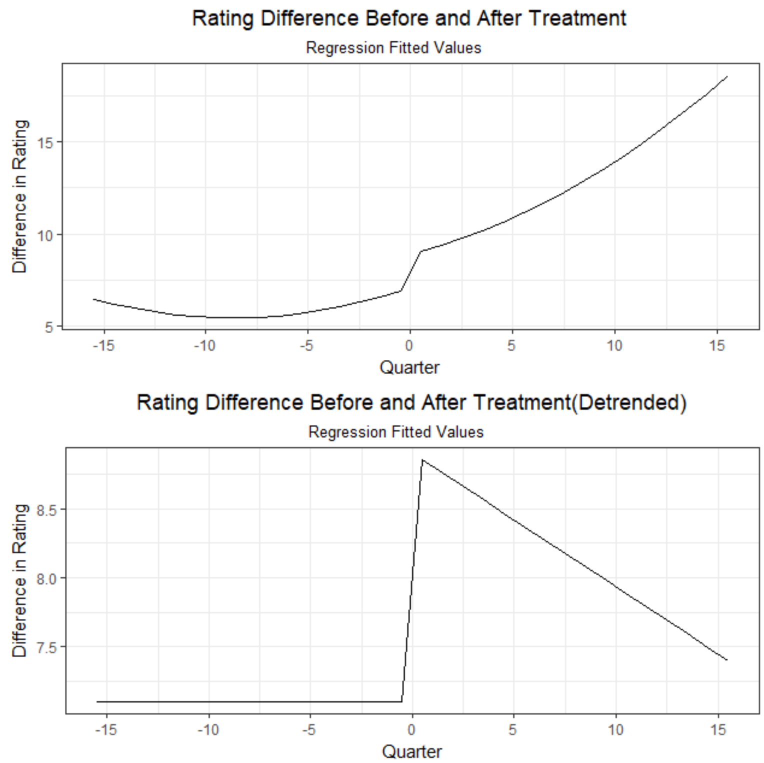

Figure 3.10. Regression Discontinuity Design.

The graph shows the regression model to estimate treatment effect of capital expenditure. The line represents the fitted values from regression in column 4 of Table 3.17. Difference in rating describes the difference in rating between the hotel that invest and its closest competitor. The variable has a range of -40 minimum to 40 maximum with a mean of -1.9. The grey line at Quarter 0 represents the treatment period, and indicates the year of the maximum capital expenditure observed for the building. The analysis looks at the 8 previous quarters and 8 quarters subsequent to the year of treatment.

likely to suffer from review manipulation. I filter out hotels not associated with a brand, I leave out hotels with small competitors, and, finally I drop hotels with more bed and breakfast accommodations within a 10 miles radius[12]. The results from the

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

subsample analysis suggest that investment affects the ratings in subsequent periods. Moreover, the results for treatment hold even after using various buffer analyses to retrieve the close competitors. In the case when I only use the closest hotel as control group, the treatment implies an increase of 0.3098 star.

I also test different definitions of competition. Table 3.18 shows the analysis using only competitors of the same property type. For example: if a building is a Full-Service hotel, I use only Full-Service hotels within 14 kilometers as competitors. The results suggest a rating improvement after capital expenditure takes place. The only test that yields that the improvement in ratings is not statistically significant is for the Limited-Service hotels.

As I mentioned in Section 3.6.1, an improvement in the reputation of the hotel leads to increases in revenue and occupancy. These results relate to the findings in Bond et al. (2014), where the authors conclude that capital improvement leads to higher income. In this paper, I relate revenue increases to and improvement in the hotel reputation online.

3.6.4 Alternative Hypothesis

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

Table 3.18. Regression Reviews Property Type

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

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

To test this alternative hypothesis, I use Equation 3.7 and add a dummy variable to control for the period right after the acquisition of the building. Table 3.19 shows this analysis where variable Brand Standard Dummy takes the value of 1 if the observation is within 2 years of the acquisition date. The results suggest that in fact new owners are 11.4% more likely to invest during this time period. Nevertheless, the negative relation between ratings and capital expenditures still holds and are statistically significant. In addition, the impact of ratings on investment increased by 7% with respect the base model in Table 3.12 column 1. In other words, this finding shows that the results on the negative relation is not due to capital expenditure required by contract to bring hotel to new brand standards.

Overall, the results from Subsections 3.6.1 , 3.6.2 and 3.6.3 indicate that the reputation mechanism of online consumer reviews leads to changes in the industry. Consumers alter their consumption patterns as reviews reveal the quality of hotels. I show that these changes in consumption patterns create the incentive for hotels to invest, and finally, that the investments lead to more positive consumer reviews in subsequent periods.

Table 3.19. Alternative Hypothesis

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

Note: I run the model from Table 3.12 in column 1, but in this case, I include a dummy to control for the first two year since acquisition. The *Brand Standard Dummy* takes the value of 1 in hotels that are in within two years since acquisition date. *Star Rating Lag* is the log of the average star rating reviewers gave the hotel after visit, rounded to the nearest 0.5. *Building Age* represent the age of the building. *Time Since Capex* is the number of years since the last capital expenditure exceed the top 20% of all observation, I also include a quadratic term and cubic term of this variable in the regression. All standard errors are clustered robust at the State level. *p<0.1; **p<0.05; ***p<0.01

3.7 Conclusion

In this paper I propose a novel question regarding online reviews in the real estate industry. In particular, I examine the impact of online reviews on firm investment decisions. I link online reviews of hotels from TripAdvisor.com to financial information on those hotels. I control for buildings and location characteristics using various fixed effect controls as well as time varying fixed effects. Overall, the results suggest there is a link between consumer generated content and a firm's decision to invest in capital improvements. I find that consumers alter their consumption based on information available in the form of online reviews. The findings suggest that an extra 0.5 star increases the occupancy rate by 1.9% and revenue by 1.6%. I provide a subsample analysis that filters hotels by variables that are correlated with review manipulation. I then filter out those hotels that may suffer from review manipulation and the results from the analysis are exacerbated, indicating that as the quality of the information provided by reviews improves, consumers' consumption patterns and hotels investment decisions become more sensitive to that information. I also show that hotels less likely to receive review manipulation are more sensitive to changes in ratings. Taken together, the information disclosed by previous consumers has a significant impact on the market by altering consumption as well as quality.

My research contributes to the literature of information asymmetry in real estate by studying the impact that reputation mechanism has on the quality of buildings. Prior literature on information disclosure looks at mandatory and voluntary disclosure and its impact on price and quality. In this paper, I look at consumer generated information and the impact it has on the quality of buildings. The contribution of this paper also expands to the literature of review manipulation. I use a set of variables previously described in this literature and find that these variables correlate with lower levels of investment as well as a smaller sensitivity of investment decisions to changes in the star rating. Following the literature on capital expenditure in real estate, this paper's contribution resides in the impact of information considerations at the moment of making investment decisions in quality. This assumes that capital expenditure is a good proxy for quality investment in a building.

Chapter 4
Consumer Evaluation Index and Mortgage Default Rate: An Empirical Investigation Using TripAdvisor.com

There is extensive empirical literature on determinants of mortgage termination that focuses on macro conditions and location characteristics. This literature finds that location fixed effects are economically and statistically significant, but these fixed effects fail to control for changes overtime at each location. In this paper I propose an index that uses location consumer reviews to asses changes in the economic activity at a granular level. I construct the index using Tripadvisor.com hotels' location ratings and then match this index to other local economic data as well as mortgages of commercial properties. I then estimate how the index relates to local economic activity and mortgage termination. The findings suggest that changes in the consumer location evaluation index are a leading indicator of employment, unemployment, real estate prices and mortgage termination.

4.1 Introduction

There is a robust list of empirical studies on default and prepayment of commercial mortgage loans that focus on characteristics such as Loan to value (LTV), Debt coverage ratio (DCR), and Interest rates. While aggregate economic and financial data are available to control for in risk models, these are usually available at long lags and in aggregates that do not capture the heterogeneity found at different locations. Estimates on location characteristics that change over time or economic conditions at small geographic areas are nonexistent. The scarcity of information on more refined geographic areas is striking given the role neighborhood characteristics have on rent, property values and ultimately on mortgage defaults. For example, Archer et al. (2002), Yildirim (2008) and Cho et al. (2013) find that location characteristics have an economic and statistically significant impact on default rates[1]. To control for location characteristics, these studies use location fixed effects and post origination economic variables to control for changes in economic conditions that occur over time. These economic drivers are at aggregate levels (i.e. MSA level) that are not able to capture heterogeneity at a more granular level and are usually available with a 1 to 2 month lag.

In this paper, I propose an index that captures consumer sentiment toward location. I seek to capture real time economic activity and default risk of commercial mortgage loans at the granular geographic level (i.e 5 digit zip code). To develop and test such an index, I use more than 6 million online consumer reviews of hotels from TripAdvisor.com, in which consumers evaluate each hotel and provide a rating to the location of the hotel. I look specifically at these ratings given to the location of the hotel. These reviews are posted daily, and the repetitive basis of the evaluation allows me to build a repetitive observation index which aims to capture location characteristics that are changing over time. The Consumer Location Evaluation Index (CLEI) can be tabulated at different frequencies and geographic levels, which results in an advantage relative to other potential local economic drivers of commercial

[1]Ambrose and Sanders (2003) find that location is statistically and economically significant for explaining prepayment risk but not for default risk.

mortgage default, such as consumer disposable income or unemployment data released with lags and at more aggregate levels.

The information provided in the reviews allows consumers to determine whether or not to grant patronage to a given hotel. While consumers also have other sources of information about hotels and their locations, usually expert opinions, that may influence their consumption patterns, the advantage of the reviews is that peer reviews may be closer to the prospective consumer's taste than that of experts. Therefore changes in these ratings may detract consumers from those areas. Also these ratings may capture other characteristics that may be changing over time at those locations, for example, crime, attractiveness of recreation activities, or just economic activity. If these location ratings affect consumption patterns, then the CLEI relates to the theory and empirics of commercial mortgage default. Changes in the patronage decisions of consumers to certain areas affects the net operating income of properties through changes in rental income as well as vacancy. These changes then translate to changes in the debt coverage ratio as well as the property values that are the main drivers of mortgage default[2].

The use of online reviews has the advantage of allowing the researcher to recognize the changes in preference of consumers for a given location. Relative to other sentiment based indexes, such as the one from the University of Michigan, that capture macro economic consumer sentiment, the reviews are geocoded, which allows the construction of the index for a more refined geographic area and for different frequencies. Although economic indicators are available at smaller geographic areas such as MSA or 3 digit zip codes, the aggregation makes it impossible to capture the heterogeneity within the geography of the indicator. For example, the 60600 zip code (606 three digit zip code) captures all the zip codes within the city of Chicago, where the distribution of industries is not ubiquitous across space. The financial industry is heavily skewed towards downtown (loop) and manufacturing towards the north

[2]For a complete discussion on the empirical literature on determinants of commercial mortgage default read Jones and Sirmans (2016)

west side of the city. A loss of employment in the manufacturing sector would have a different impact in the the northwest than in the loop. Another limitation of the economic data for more refined geographic areas is that data only becomes available with a lag. For example, unemployment rate data for MSAs becomes available with a one month lag or more. However, the CLEI its available on a daily basis. With these attributes, the CLEI is another source of information that researchers, policy makers and investors can use to evaluate commercial mortgage defaults .

Using the CLEI, I study the explanatory power the index has on employment, house prices, and commercial property prices at the MSA level. I use a panel fixed effect model with heteroscedastic and autocorrelated robust standard errors. The findings suggest that lag values of the CLEI are statistically significant, even after controlling for unemployment rates at the MSA level. The statistical significance is important as I treat unemployment data as if research can accurately predict the unemployment rate. In other words, even if researchers or forecasters can perfectly predict unemployment, the index has explanatory power. To further study the usefulness of the CLEI, I then estimate a competing risk model for mortgage termination. The results suggest that duration models that use the CLEI index are superior to models without it from the log likelihood and Akaike Information Criterion perspectives. Overall, the index leads economic data at the MSA level and improves duration models.

The remainder of the paper proceeds as follows. In the next section, I provide an overview of the relevant literature and discuss the main findings on the determinants of mortgage default and prepayment, and the evidence on the impact which online information has on real estate. I then present a simplified mortgage default model, and the relation the index has to defaults. In the third section I describe the data used and the methodology to construct the CLEI, as well as the TREPP data on mortgage defaults. I describe the empirical model in the fifth section and describe the main findings in section six. I provide a discussion of the results and conclude in the last section.

4.2 Prior Literature

This paper builds on empirical literature regarding the use of Internet consumers' information to asses mortgage risk (Chauvet et al., 2016) as well as asset pricing (Das et al., 2015). It contributes to this strand of literature by proposing a new index which captures consumers' evaluation of locations that directly affect commercial real estate fundamentals. The paper also makes a contribution to the empirical literature of determinants of commercial mortgage termination by complementing mortgage duration models with the index developed in this paper. Most of the literature in mortgage termination uses data that is only available with long lags (i.e. employment data at the MSA level) and is aggregated for large geographic areas, whereas this index allows the control of local economic activity for a more refined geographic area and in "real time".

Research on the link between online information and firms' fundamentals can be categorized into two broad strands: the first links consumer searches with mortgage defaults and firm fundamentals (Chauvet et al., 2016; Das et al., 2015); while the second links consumers' online reviews to real estate prices and businesses' operating performances (Anderson and Magruder, 2012; Kuang, 2017). On the one hand, a consumer's searches in search engines reveal relevant information about her concerns which precede an action. For example, if a house owner has a concern about potential job loss and sees that this could result in defaulting on her mortgage, she could do a Google search for "consequences for defaulting on my mortgage" and by doing so reveal her concern. Therefore, this information has the potential to be used for predicting a future default. On the other hand, consumers sharing reviews of local businesses and products reveal their perception of the quality of those businesses and therefore inform other consumers before they consume at that particular location. These reviews then carry information that could be used to predict a customer's patronage to stores and buildings.

An early paper by Wu and Brynjolfsson (2015)provides evidence of the link between a Google search and home sales volume as well as the house price index. The literature on search engine searches concludes that Google searches can be used to predict housing returns and mortgage delinquency, and that an increase in searches results in larger premiums to insure subprime mortgage bonds against default (Chauvet et al., 2016). Das et al. (2015) use Google trends to relate searches to fundamental variables in real estate such as vacancy rates, rental rates and real estate asset prices. Their findings suggest that after controlling for location, year fixed effect and macro variables changes in the search trends are negatively associated to vacancy rates and are not statistically significant for determining rental rates. Das et al. also present evidence that REIT stock returns are associated with trends even after controlling for the market CAPM factor, small - big Fama-French factor, high - low Fama-French factor and Carhart's momentum factor. From these studies it is clear that household searches reveal information to market participants that can be used for forecasting purposes.

Consumer reviews are another source of information that is associated with real estate fundamentals (Anderson and Magruder, 2012; Kuang, 2017). One potential explanation for why consumer reviews affect fundamentals, is that reviews make the quality of the business salient, therefore allowing consumers to differentiate businesses and finally alter their consumption pattern. In the real estate case, higher consumer ratings are associated with more restaurant visits and higher hotel occupancy (Anderson and Magruder, 2012; Gárate, 2018). This not only affects the fundamentals of the building itself but also property prices in the surrounding area due to externalities generated by these amenities (i.e. bars, restaurants, cafes, grocery stores, hotels, etc.). Pope and Pope (2015) find that a Walmart entry has a positive impact on prices of surrounding housing; more specifically, a store entry implies a 3% increase in house prices. Although the entry of new amenities may affect the prices, the quality of the new facility also affects the type of externalities generated. Kuang (2017) finds that consumer ratings of consumption amenities improve the measure of implicit market value of these amenities, and this increases property values in the surrounding areas. It is clear that consumer generated content has an impact on real estate fundamentals, and therefore it is of interest if this information is then used by property owners to make decisions on real options such as mortgage prepayments and defaults.

Empirical literature on determinants of commercial mortgage termination focuses on the impact of four variables: loan and underwriting characteristics; macroeconomic conditions; geographic region (usually by using fixed effect); and property type fixed effects. Loan and underwriting variables seek to control for underwriting loan characteristics that may affect the position a borrower has at the moment of origination, and this may affect the exercise of the prepayment or default option. Macro variables control changes in the value of the prepayment and default options due to market conditions. Finally, the inclusion of region and property fixed effect responds to the significant variation in prepayment and default that exists among property types and regions.

During the loan underwriting process, instead of presenting all participants with the same contract, originators adjust LTV and DSCR requirements. For example, lenders may request higher DSCR and lower LTV if borrowers are riskier Archer et al. (2002). This creates a challenge as the statistical significance of these variables diminishes due to the endogeneity to the origination process. This then explains why some contingents claim that model prediction relative to origination variables does not find support in the empirical literature.

Due to the substitution effect between prepayment and default, competing risk models are commonly used in the empirical literature. For example, Ciochetti et al. (2002) develop a competing risk proportional hazard model that estimates the probability of termination due to default and prepayment. Similar to this idea, Ambrose and Sanders (2003) also use a competing risk proportional hazard model and use mortgages underlying CMBS deals to estimate their hazard functions. Yildirim (2008) proposes a model that is consistent with data clustering and allows for data to be right censored. Most of the data in prepayment and default incidence is correlated within a geographical area or property type, making data clustering a central issue at the moment of estimating prepayment and default models.

Mortgages have an embedded put and call option that will be exercised depending on the prices of those options. To be consistent with the mortgage pricing literature, prepayment and default models require the inclusion of proxies for the underlying stochastic processes (Schwartz and Torous, 1992; Kau et al., 1992). Also lenders at origination assess the risk of the borrowers, and therefore offer borrowers contracts, such that the lender is indifferent to borrowers risk profiles (Ambrose and Sanders, 2003). Mortgage pricing using a contingent claim approach infers that prepayment and default are a function of DSCR and LTV. Since greater LTV and lower DSCR result in greater mortgage defaults, it is not a surprise then that lenders control future potential losses by adjusting these two variables at origination (Archer et al., 2002). Another set of variables adjusted at origination to control for potential losses and important for including as parameters are prepayment penalties and lock-out provisions in a loan. For example, lenders can deter borrowers from exercising the prepayment option using stricter prepayment penalties. These variables are usually observable at origination and through the life of the mortgage.

There are risks associated with a mortgage that are not observable at origination. To control for such risks, the use of fixed effect allows the researcher to control for the risks that are not changing though time and are correlated with the property type or location. For example, there is empirical evidence that default varies by property type (i.e. Episcopos et al. (1998); Ambrose and Sanders (2003); Grovenstein (2005); Yildirim (2008)) and region (i.e Yildirim (2008); Cho et al. (2013); Episcopos et al. (1998)). These studies control for location and property type using a combination of fixed effect; they also include local economic data at the MSA level such as unemployment, population and income. The problem with these local variables is that fixed effect fails to capture changes over time and local economic data are usually available for decision makers with a couple months of lag time.

In this article, I introduce an index to control for local market characteristics that change over time. I construct an index using the methodology proposed by

Bailey et al. (1963) and Case et al. (1989) to incorporate consumer ratings to default and prepayment hazard models. I estimate a proportional hazard model for the sub-distribution of competing risk, proposed by Fine and Gray (1999), which accounts for the substitution effect between prepayment and default options. In order to control for the clustering of default and prepayment incidence, I use Zhou et al. (2012) which extends Fine and Gray (1999) to allow for correlation within a cluster. In the next section, I give details on the construction of the index to include consumer ratings to the empirical estimation of the prepayment and default functions.

4.3 Construction of the Index

In order to construct the CLEI, I follow the methodology proposed by Bailey et al. (1963) (hereafter, BMN) to construct a home price index, and I use the Case et al. (1989)(hereafter, CS) methodology to correct for heteroscedasticity with respect to the holding period. The idea in the construction of the index is that I want to maintain the location characteristic of the buildings as constant, such as distance to the central business district, proximity to the beach, distance to parks and other amenities. In that sense the evaluation is a repeated evaluation of the location that allows me to capture changes in consumer preference towards those locations.

In a case of not using a repeated observation methodology but using a straight average of the ratings provided by consumers, the average will not be of constant location characteristics. For example, the average rating of one month could represent only hotels close to amenities while next month the predominant type of hotels on average could be hotels distant from those amenities. If consumers are not indifferent to being close or far from amenities, the average will vary, not because consumers change their preference, but because the hotels in the evaluations have different location characteristics.

The rating system evaluates not only the hotel service, building, rooms and cleanliness, but also the location. Location ratings relate to the proximity to amenities as well as characteristics of the neighborhood which change overtime, for example, crime or commercial activity around the location. I use location ratings to construct the index and capture the variability of the index in each location. To do so I average the ratings for each period and for each building and consider this an observation. I then create pairs of observations of period 1 and period 2, where period 1 represents the last observation before period 2. I then apply BMN. The description of the repeated observation is as follows:

$$RR_{itt'} = \frac{CLEI_{t'}}{CLEI_t} xE_{itt'} \tag{4.1}$$

where $RR_{itt'}$ represents the rating ratio of period t' over the rating received in t for the i-th pair location evaluations. $CLEI_{t'}$ and $CLEI_t$ are the indexes for period t' and t that are unknown and that I want to estimate. To estimate the model, I use the logarithm transformation of Equation 4.1. This results in the following:

$$rr_{itt'} = -clei_t + clei_{t'} + \varepsilon_{itt'} \tag{4.2}$$

where the lower case letter indicates the logarithm of the variables in Equation 4.1, where t in [0,1,2,...T-1] and t' in [1,2,...T]. The estimation of the unknown indexes $clei_t$ and $clei_{t'}$ can be treated as a regression problem in the following form:

$$rr_{itt'} = \sum_{j=1}^{T} clei_j x_j + \varepsilon_{itt'} \tag{4.3}$$

In Equation 4.3, x_j take the value of -1 in period t that represents the first evaluation of the pair, +1 in following evaluation period t', and 0 in the rest of the periods. In other words, Equation 4.3 could be expressed in matrix notation as:

$$rr = xclei + \varepsilon \tag{4.4}$$

where rr and ε are vectors of length equal to the number of pairs of evaluations,

n. Vector clei has the length of the number of periods and stores the indexes for each period. Matrix x contains -1s, +1s, and 0s; and, has n rows and T columns. For example, if the first evaluation for the i-th pair is in period t and the second evaluation for the pair is in t' then the i-th row in matrix x will have a -1 in column t, a +1 in column t' and 0 in all other columns. Therefore, the least square estimator for the *clei* index is the following:

$$\widehat{clei} = (x'x)^{-1}(x'rr) \tag{4.5}$$

Now to correct for heteroscedasticity, which may arise from the interval between evaluation periods, I use the CS approach. Case et al. (1989) introduce a weighted repeated sale index to control for heteroscedaticity, which increases with the holding period. The approach in my model uses ratings and the period in between evaluations to correct for heteroscedasticity. The procedure to estimate the weights for the weighted least square is as follows:

$$\varepsilon^2 = \beta_0 + EP \tag{4.6}$$

where ε^2 is the square of the error term from the estimation in Equation 4.5, β_0 is a constant term and EP is the period in between evaluations. Then I use the fitted values from Equation 4.6 as weights to reestimate the $\widehat{clei}$ vector. The weighted regression can be expressed as:

$$\widetilde{clei} = (x'W^{-1}x)^{-1}(x'W^{-1}rr) \tag{4.7}$$

The estimated $\widetilde{clei}$ has logarithmic estimates, in order to reconstruct the CLEI estimate the exponent of $\widetilde{clei}$. In the next section I describe the data and the results of the index.

4.4 Data

To examine the impact which the CLEI has on mortgage default and prepayment, I use data from 5 sources. I first collect consumer location reviews from TripAdvisor.com. In order to estimate mortgage default and prepayment models, I use TREPP data feed for CMBS deals. The Country and MSA level unemployment data comes from the Bureau of Labor Statistics. To build commercial property price indexes, I use transaction data from SNL and appraisal data from TREPP. I have also downloaded macro economic variables from FRED at the Federal Reserve Bank of St. Louis. I provide a brief description of the variables used in Table 4.1, and in the subsequent subsection I go over more details.

Table 4.1. Variables list and source

Name	Short description	Source
Rating information		
Rating of hotels	Rating of hotel given by consumer.	TripAdvisor
Location rating	Rating of the location of the hotel.	TripAdvisor
CLEI U.S.A.	Consumer location evaluation index U.S.A. level.	
CLEI zipcode	Consumer location evaluation index 5 digit zipcode.	
CLEI 3 digit zipcode	Consumer location evaluation index 3 digit zipcode.	
Macro Variables		
U. Mich Sent.	Consumer sentiment index from surveys of consumers, University of Michigan.	FRED
10 Year T. Rate	10-Year Treasury Constant Maturity Rate	FRED
1 Year T. Rate	1-Year Treasury Constant Maturity Rate	FRED
HPI U.S.A.	U.S. monthly Purchase-Only Indexes	FHFA
HPI MSA	S&P/Case-Shiller MSA Home Price Index	FRED
CPPI U.S.A.	Commercial property price index	SNL, TREPP
Yield Curve	10 Year T. Rate - 1 Year T. Rate	FRED
AAA	BofAML US Corporate AAA debt effective yield	FRED
BBB	BofAML US Corporate AAA debt effective yield	FRED
Spread AAA-BBB	BBB-AAA	FRED
MSA		
Employment	Number of civilian in labor force employed	BLS
Unemployment Rate	Unemployment rate of civilian labor force	BLS
Mortgage Data		
Since origination	Months since the origination of loan	TREPP
Outstanding balance	End of period outstanding scheduled principal balance	TREPP
Origination balance	Outstanding loan balance at origination	TREPP
Actual rate	Annualized gross rate used to calculate scheduled interest amount	TREPP
Origination term	Original term of the loan	TREPP
Amortization term	Amortization term of the loan	TREPP
Ppoption	(Actual rate -10 Year T. Rate)/10 Year T. Rate	
Lock-out period	Dummy indicating lock-out period	TREPP
Yield maintenance period	Dummy indicating yield maintenance period	TREPP
Negative equity	Dummy indicating that the loan has negative equity	
DSCR	Most recent debt service coverage ratio	TREPP
LTV Securitization	Loan to value at securitization.	TREPP

Note: Variables are from TripAdvisor.com(TripAdvisor), Bureau of Labor and Statistics (BLS), the FRED Economic Data from the Federal Reserve Bank of St. Louis(FRED), and CMBS database from TREPP (TREPP).

4.4.1 Location ratings from TripAdvisor.com

The popularity of TripAdvisor makes it the leading travel site on the Internet, and a must do for the customer looking for information about hotels and locations. TripAdvisor publishes consumer reviews of hotels and locations worldwide. Each review has a rating given to the hotel and, at the same time, consumers can rate more specific attributes of the hotel. For example, a review could have a rating given to the hotel plus a rating for Cleanliness, Sleep Quality and its Location. I focus on the Location rating to construct the index described in Section 4.3. The sample of more than 10 million reviews represents 24,841 hotels, located in 6,736 five digit zip codes and approximately 860 MSAs. From the 10 million reviews, only 6 million have a consumer evaluation of the location of the hotels. Table 4.2 provides the descriptive statistics of these reviews.

Table 4.2. Reviews Information

Variable	Count				
Number of Hotels	24,841				
Number of 5 Digit Zipcodes	6,736				
Number of 3 Digit Zipcodes	859				
Number of MSAs	866				
Variable	Count	Min.	Mean	Max.	St. Dev.
Ratings of Hotels	10,339,562	1	4.036	5	1.17
Location Rating	6,014,158	1	4.389	5	0.89

Note: This table provides information on the consumer reviews used in the analysis. Number of hotels represents the number of hotels used to collect reviews. Number of 5 and 3 Digit Zip codes as well as MSAs describes the number of geographies represented in reviews. Ratings of hotel represent the number of reviews available. Location Rating is the number of reviews that the location rating used in the construction of the Consumer Location Evaluation Index.

To build the index, I use the location reviews. It is crucial for the project to have reviews distributed across space in order to be able to construct the index for different geographies. Table 4.3 provides the distribution of reviews across region.

Another important feature required for the construction of the index is the distribution through time of reviews. The upper graph of Figure 4.1 presents the number of reviews received by quarter. The number of reviews received picked in mid 2013 with more than 300,000 reviews in a single quarter. The average rating of all reviews received is stable through the sample period with a small increase between 2008 and 2009. The lower graph from 4.1 reports examples of the average ratings at 5 digit zip codes. This bottom graph shows that at a more granular level, the ratings are not as stable as at the national level. These significant variations at a more granular level indicate that reviews at a more granular level may contain information for those locations that differ from those found at the national level.

Table 4.3. Review by Location

Region	Count Reviews	Percentage
EN	481,676	8.01%
ME	650,633	10.82%
MT	600,319	9.98%
NE	885,136	14.72%
PC	1,296,103	21.55%
SE	1,392,125	23.15%
SW	486,225	8.08%
WN	221,941	3.69%
Total Number of Reviews	6,014,158	100%

Note: This table provides the distribution of hotel by region. EN includes the following states: OH, IL, IN, MI and WI. ME includes SC, MD, DE, NC, DC, VA, KY and WV. MT contains AZ, UT, CO, MT, NV, WY and ID. The states NJ, PA, NY, CT, MA, RI, ME, VT and NH are part of the NE region. PC includes states in the Pacific Ocean such as CA, WA, OR, HI and AK. SE are the states located in the south east, including FL, AL, GA, TN and MS. SW includes TX, LA, A. Finally WN includes MO, KS, NE, MN, ND, IA and SD.

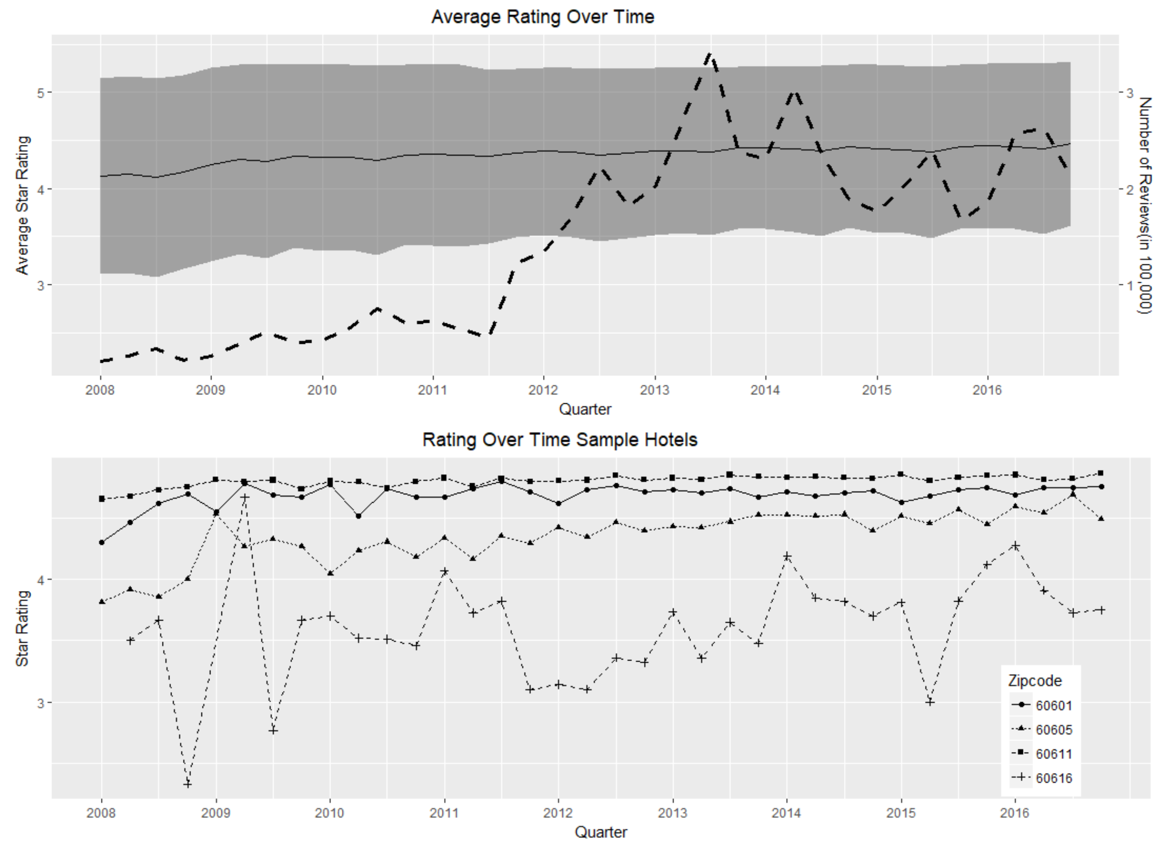

Figure 4.1. Reviews Overtine

The graph shows reviews information. The upper graph shows the average location rating given to hotels, the bands indicate one standard deviation. The dashed line in the upper graph indicates the number of reviews posted every month in 100,000s. The lower graph shows the average rating at a zipcode level. The examples are zipcodes within the city of Chicago, 60616 is Near South side, Douglas and China Town while 60601 is the New East side.

4.4.2 Commercial Mortgage Data

Mortgage performance data comes from the data feed provided by TREPP, which is the largest commercially available database of securitized mortgages. The data feed spans more than 100,000 CMBS loans across the USA and includes more than 1,500 CMBS deals. The dataset contains information on loan performance as well as collateral building operating performance. The dataset contains origination, securitization and time series information for each loan including: LTV, DSCR, origination balance, current balance, current rate, origination gross rate, originator, debt service, lock out period, yield maintenance information, mortgage type, underlying property type and location, among 300 hundred other variables.

4.4.3 Mortgage Data

Tables 4.4 through 4.8 provide descriptive statistics of the loan used in this analysis. The sample includes 15,988 loans whose origination ranges from prior to year 2000 to 2015, and Table 4.4 provides the distribution of monthly observations by year of origination. As one of the advantages of working with TREPP data is that it covers a wide range of originators (such as Wells Fargo, JPM, Wachovia, etc), 4.5 describes the distribution of loan observation by originator. Collateral information allows me to describe the loans by location as well as property type. Thus, Table 4.6 describes the distribution of loan observation by location of the underlying property. As can be seen, the region with the largest share of loan is the South East (SE)[3]. Table 4.7 shows the distribution of loan observation where Office, Retail and Multi-Family represent approximately 75% of the loan monthly observation. Finally, Table 4.8 presents the maturity type of the loan, with 97.9% of the observation having Balloon type amortization.

I show the Kaplan-Meier estimates of the survival probabilities for default and

[3]SE are the states located in the south east, including FL, AL, GA, TN and MS.

Table 4.4. Loans by year of origination

Year Origination	Frequency	Percentage
<=2000	21,479	2.71
2001	29,304	3.70
2002	34,849	4.40
2003	55,441	7.00
2004	72,073	9.10
2005	140,186	17.70
2006	175,035	22.10
2007	134,642	17.00
2008	1,584	0.20
2009	6,336	0.80
2010	17,424	2.20
2011	27,720	3.50
2012	30,096	3.80
2013	25,344	3.20
2014	15,840	2.00
2015	3,960	0.50

Note: This table provides the distribution of loan performance observations by the year in which the loan was originated.

prepayment events in Figure 4.2. From the sample we observe that after 10 years, approximately 10% of the loans have been prepaid and approximately 2% have defaulted. These graphs suggest that the majority of loans are performing after 10 years since origination.

In Table 4.9, I provide the summary statistics for the loans used to calculate the prepayment and default hazard functions. The average time since origination of the loans performance observation is 60.8 months with an average loan term of 123.725 months at origination. The mean DSCR is 1.729 and the mean LTV at securitization is 66.7%. The lock-out period variable indicates that 81.4% of the observations are from a period within the period established in the lock-out provision and 13.2%

Table 4.5. Source of data by originator

Source of Loan Data	Percentage	Source of Loan Data	Percentage
Wells Fargo	6.0	Greenwich	2.2
JPM	5.9	PNC	2.2
Wachovia	5.8	Bear Stearns	2.1
LaSalle	5.4	Artesia	2.0
Morgan Stanley	4.9	Goldman Sachs	2.0
Column	4.4	CIBC	1.9
BOA	3.5	Key	1.9
Lehman Brothers	3.4	Prudential	1.9
Principal	3.4	Merrill Lynch	1.8
GE	3.0	CBRE	1.7
NA	2.9	WAMU	1.7
CRF	2.5	NCCI	1.2
GACC	2.5	BARCLAYS	1.1
Citigroup	2.4	Berkadia	1.0
UBS	2.3	Others	17.0
Total	100		

Note: This table provides the distribution of loan performance observations by originator.

within the period of yield maintenance.

4.4.4 Macro Economic Data

As mentioned before, the mortgage pricing literature infers that borrowers would exercise the prepayment and default options if the options were "in the money". The "moneyness" of the option depends on the underlying stochastic risk processes associated with the mortgage (Kau et al., 1992; Schwartz and Torous, 1992).

Table 4.6. Loan data by region

Region	Percentage
EN	7.4
ME	14.9
MT	12.2
NE	10.7
PC	17.4
SE	22.8
SW	12.8
WN	1.8
Total	100

Note: This table provides the distribution of performance observation by region. EN includes the following states: OH, IL, IN, MI and WI. ME includes SC, MD, DE, NC, DC, VA, KY and WV. MT contains AZ, UT, CO, MT, NV, WY and ID. The states NJ, PA, NY, CT, MA, RI, ME, VT and NH are part of the NE region. PC includes states in the Pacific Ocean such as CA, WA, OR, HI and AK. SE are the states located in the south east, including FL, AL, GA, TN and MS. SW includes TX, LA, A. Finally, WN includes MO, KS, NE, MN, ND, IA and SD.

Table 4.7. Loan data by property type

Type	Percentage
Industrial	6.1
Lodging	11.5
Multi-Family	24.3
Office	24.5
Other	7.7
Retail	25.9
Total	100

Note: This table provides the distribution of performance observation by type of loan amortization.

Table 4.8. Loan data by maturity type

Type	Percentage
Balloon	97.9
Self Amortizing	2.1
Total	100

Note: This table provides the distribution of performance observation by type of loan amortization.

4.4.5 Macroeconomic Data

The prepayment option dynamics depend on the relation between the contract interest rate and the actual market. The relative position that a borrower has can be measured by the relation between the spread of the contract rate relative to the market. Following Ambrose and Sanders (2003), I define this relationship as the contract rate $r_c(t)$minus the government treasury rate $r_G(t)$ all divided by the treasury rate.

$$PPmtOption = \frac{r_c(t) - r_G(t)}{r_G(t)} \tag{4.8}$$

This relationship is established if the prepayment option is becoming more valuable or not. An increase in the $PPmtOption$ indicates that prepayment is becoming more valuable and therefore we should expect it to have a positive impact on the prepayment function. The $PPmtOption$ controls for the "moneyness" of the option using current information, but it does not control for the expectation of future market interest rates. The interest rate yield curve allows us to do exactly that; it establishes the spread between long term debt and short term debt. Ambrose and Sanders (2003) and Grovenstein (2005) use the spread between the 10 year treasury bond rate and the one year treasury bond rate to control for future expectations of interest rates[4]. Finally, a portion of the mortgage rate represents a market

[4]Cho et al. (2013) use 10 year minus 3 year treasury rate, both 1 year and 3 year treasuries help

Table 4.9. Summary statitics loan data

Statistic	Mean	St. Dev.	Min	Max
Since origination	60.882	33.772	0	295
Outstanding balance	17,958,205	37,666,852	0	1,450,000,000
Origination balance	18,289,254	37,123,804	150,000	1,450,000,000
Actual rate	5.781	0.961	0.0001	114.500
Origination term	123.725	42.158	12	427
Amortization term	322.828	98.354	12	999
Ppoption	1.164	0.658	−1.000	44.609
Volatility 10 year treasury	0.456	0.130	0.219	0.737
Yield curve	2.190	0.695	0.200	3.400
Spread AAA-BBB bonds	1.788	0.746	0.881	4.420
Volatility spread AAA-BBB bonds yield	0.482	0.327	0.083	1.166
Lock out period	0.814	0.389	0	1
Yield maintenance period	0.132	0.339	0	1
Negative equity	0.314	0.464	0	1
DSCR	1.729	1.105	−8.720	70.748
LTV securitization	66.694	13.186	0.900	128.600
Unemployment rate MSA	7.325	2.517	2.600	30.000
Change CLEI zipcode	0.030	0.307	−0.923	23.138
Volatility CLEI zipcode	0.244	2.058	0.005	602.089
Mean change CLEI zipcode	0.041	0.606	−0.515	173.883
Change CLEI 3 digit zipcode	0.009	0.131	−0.850	6.458
Volatility CLEI 3 digit zipcode	0.108	0.104	0.009	2.738
Mean change CLEI 3 digit zipcode	0.011	0.028	−0.100	0.760
Change CLEI USA	0.001	0.009	−0.017	0.029
Volatility CLEI USA	0.009	0.002	0.005	0.015
Mean change CLEI USA	0.001	0.002	−0.002	0.006

Note: This table provides the summary statistics for the data use for the competing risk model of commercial mortgage termination by default and prepayment. Origination represents the number of months since the origination of loan. Outstanding balance is outstanding scheduled principal balance at the end of the period. Origination balance is the loan amount at origination. Actual rate is the annualized gross rate used to calculate the current scheduled interest amount. Origination term is the original term of the loan. Amortization term is the amortization term of the loan. Ppoption is the relative position of the contract interest rate with respect to the market rate. Lock-out period and Yield maintenance are dummy variables that take the value of one if the loan is within a lock-out period and yield maintenance period respectively. Negative equity is a dummy variable that indicates whether or not a loan has negative equity in a given month. DSCR is the most recent debt service coverage ratio reported for the loan. LTV securitization is the value of the loan to value ratio at securitization of loan. Change CLEI is the month by month change of the CLEI index. Volatility is the standard deviation of the prior 12 months Change CLEI. The Mean change CLEI is the average Change CLEI of the prior 12 months. Zip code indicates that the index is calculated at the 5 digit zip code level. The 3 digit zip code indicates that the index is at the 3 digit level. Finally USA indicates that the index is calculated at the country level.

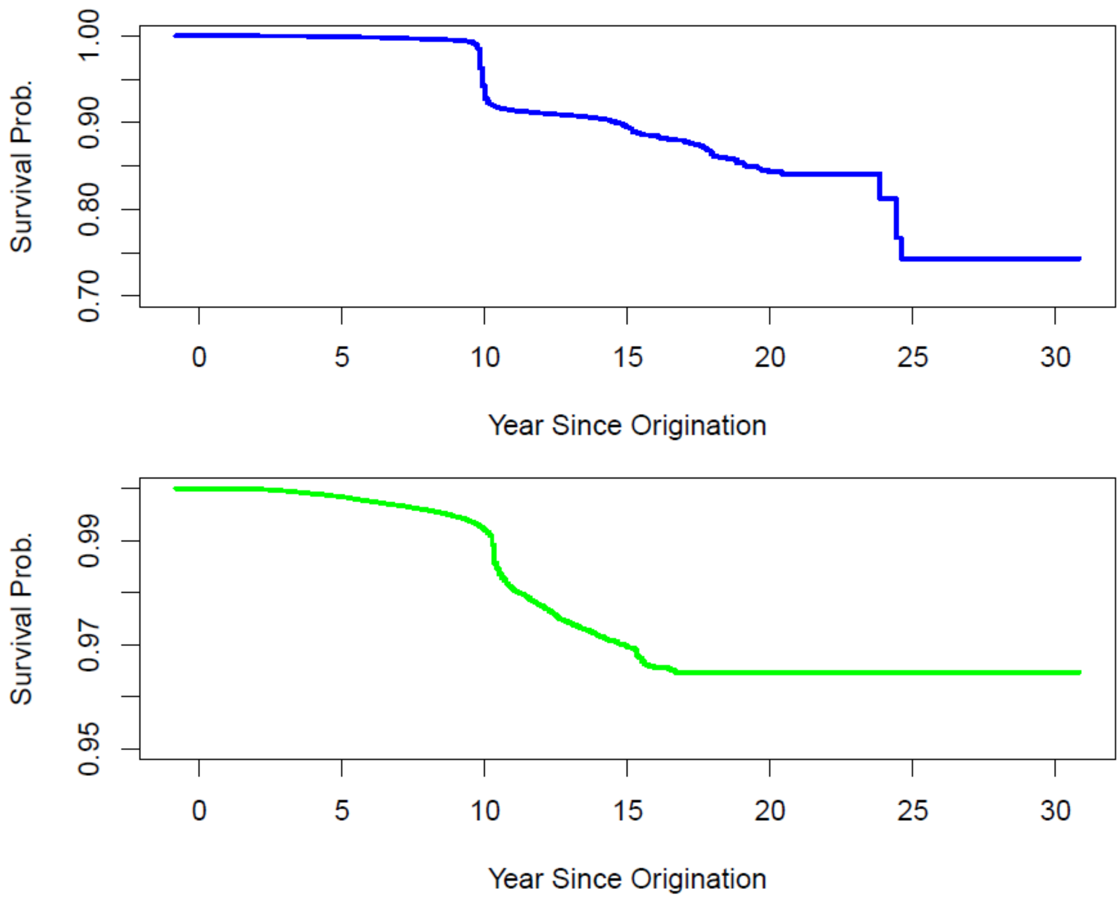

Figure 4.2. Survival Probability CMBS Mortgages

Kaplan-Meier estimates of survival defined as time to mortgage termination by default or prepayment. Default is a loan that is 90 or more days delinquent and is an absorbing state.

risk premium. To control for this premium, I use a spread between the AAA and BBB corporate bonds rates. So far, the variables of interest refer to levels of the contract rate relative to the treasury rates or spreads, but this is only part of the risks.

control for expectation of interest rates.

Table 4.10. Summary Statistic Macroeconomic Variables

Statistic	N	Mean	St. Dev.	Min	Max
CLEI	114	1.08	0.04	0.98	1.14
U. Mich Sent.	114	77.33	11.02	55.30	98.20
10 Year T. Rate	114	2.72	0.82	1.50	5.00
1 Year T. Rate	114	0.65	0.98	0.10	4.96
HPI U.S.A.	114	202.56	17.20	178.80	241.24
CPPI U.S.A.	114	0.72	0.14	0.56	1.19
Yield curve	114	2.07	0.74	0.04	3.40
Spread AAA-BBB	114	1.69	0.72	0.69	4.42
Vol. 10 Year T. R.	114	0.43	0.14	0.22	0.74
Vol. HPI	114	6.03	2.31	2.16	11.04
Vol. CPPI	114	0.06	0.05	0.02	0.18

Note: This table provides the summary statistics for the macroeconomic variables. CLEI represents the Consumer Location Evaluation Index estimated at the country level. U. Mich Sent. represents the consumer sentiment index calculated by University of Michigan. 10 Year T. Rate and 1 Year T. Rate represent the 10-Year and 1-Year Treasury Constant Maturity Rates. HPI U.S.A. is the monthly house price index estimated by FHFA at the country level. CPPI is the Commercial Property Price Index estimated using Case & Shiller methodology and by using TREPP collateral appraisal data, as well as, SNL REIT property transaction data. Yield curve is the 10 Year T. Rate minus 1 Year T. Rate. Spread AAA-BBB represents the difference between BofAML US Corporate AAA and BBB debt effective yield. Vol. represents the previous 12 month moving standards deviation of the variable.

Theoretical pricing mortgage literature argues that interest rate volatility decreases the value of the prepayment option. The empirical literature uses a 10 year treasury rate volatility as control. I follow Seslen and Wheaton (2010) and use the standard deviation of the 24 month rolling window of the 10 year treasury rate to measure the market volatility. I also include the volatility of the spread between the AAA and BBB corporate bond rates using the same 24 month rolling window.

Table 4.10 provides summary statistics for the macro variables during the time frame of this study. CLEI represents the monthly index at the country level, where the U. Mich Sent. is the consumer sentiment from University of Michigan. The HPI is the national housing price index estimated by the FHFA, with vol. HPI being the rolling previous 24 month standard deviation of the index.

I provide the correlation among variables in Table 4.11. The correlation of CLEI estimated at the national level reveals that the index has a positive correlation with the University of Michigan Index, and this correlation coefficient is statistically significant at the 1% level. The correlation of the CLEI with treasury interest rates is negative and statistically significant at the 1% level, while the unconditional correlation with the commercial property index is negative and statistically significant.

Table 4.11. Correlation CLEI with Macro Variables

	CLEI	U. Mich. Sent.	10 Yr. T. Rate	1 Yr. T. Rate	HPI U.S.A.	CPPI U.S.A.	Yield Curve	Spread AAA - BBB	Vol. 10 Y T. Rate	Vol. HPI
U. mich sent.	0.46***									
10 year T. Rate	-0.81***	-0.41***								
1 year T. Rate	-0.76***	-0.07	0.68***							
HPI U.S.A.	0.00	0.70***	-0.12	0.34***						
CPPI U.S.A.	-0.45***	0.30**	0.33***	0.80***	0.69***					
Yield curve	0.10	-0.38***	0.22*	-0.57***	-0.58***	-0.70***				
Spread AAA-BBB	-0.26**	-0.69***	0.10	-0.14	-0.55***	-0.35***	0.30**			
Vol. 10 year T. R.	0.08	-0.60***	-0.09	-0.30**	-0.68***	-0.49***	0.30**	0.67***		
Vol. HPI	0.05	-0.06	0.01	-0.23*	0.23*	-0.11	0.32***	0.30***	0.10	
Vol. CPPI	-0.55***	-0.51***	0.50***	0.21*	-0.07	0.05	0.27**	0.64***	0.35***	0.62***

Note: This table shows the lower half of the monthly correlation matrix. CLEI represents the Consumer Location Evaluation Index estimated at the country level. U. Mich Sent. rep-resents the consumer sentiment index calculated by University of Michigan. 10 Year T. Rate and 1 Year T. Rate represent the 10-Year and 1-Year Treasury Constant Maturity Rates. HPI U.S.A. is the monthly house price index estimated by FHFA at the country level. CPPI is the Commercial Property Price Index estimated using Case & Shiller methodology and by using TREPP collateral appraisal data, as well as, SNL REIT property transaction data. Yield curve is the 10 Year T. Rate minus 1 Year T. Rate. Spread AAA-BBB represents the difference between BofAML US Corporate AAA and BBB debt effective yield. Vol. represents the previous 12 month moving standards deviation of the variable. The asterisk represents the significance levels, *p<0.1; **p<0.05; ***p<0.01

4.4.6 Commercial Property Price Index

Option pricing literature not only includes the risk of interest rates, but also includes the price of the underlying asset as a determinant of the option price. A decrease in the price of the property increases the price of the option by default. Unfortunately I do not observe the LTV every period as the market price of the property is not observable. In order to proxy the property price, I inflate the last appraisal value of the underlying asset by a commercial price property index (CPPI) calculated by region and property type. This way I am able to proxy the equity position of the borrower.

To construct the CPPI, I use REITs' property transaction data from SNL and appraisal data from the underlying properties of the CMBS loans. I follow the methodology proposed by Bailey et al. (1963) and Case et al. (1989) to construct the repeated sales index and use more than 25,000 repeated observations including appraisals and transactions of commercial properties. I estimate the index property type, NCREIF region and month. The index includes Multi-Family, Retail, Office, Industrial, Lodging and Other property type buildings. The regions cover all of the United States of America, and are divided into East North Central (EN), Mideast (ME), Mountain (MT), Northeast (NE), Pacific (PC), Southeast (SE), Southwest (SW), and West North Central (WN). Table 4.6 provides a description of the states included in each region.

Estimating the index for various regions and property types is important for capturing commercial property market conditions at a more local level. Graphs 4.4 and 4.3 show that prices for various property types may differ significantly by region. For example, in Multi-Family indexes (see Graph 4.3) the Northeastern properties did not experience a significant and steady decline in property prices as did properties of this type in the West North Central between 2009 and 2012. Graph 4.4 shows the same phenomenon for the retail properties; in this case, East North Central prices have not recovered as much as prices in the Southeast. This heterogeneity of the indexes indicates that in order to capture the equity position, the appraisal values need to be inflated by more local price indexes.

To test the methodology used, I estimated the index for the USA using the same set of repeated observations and compared this to a commercial property index published by the Federal Reserve Bank of St. Louis with the authorization of the International Monetary Fund. Graph 4.5 presents this comparison between the changes of the two indexes. From the graph we observe that the changes of both indexes track each other closely.

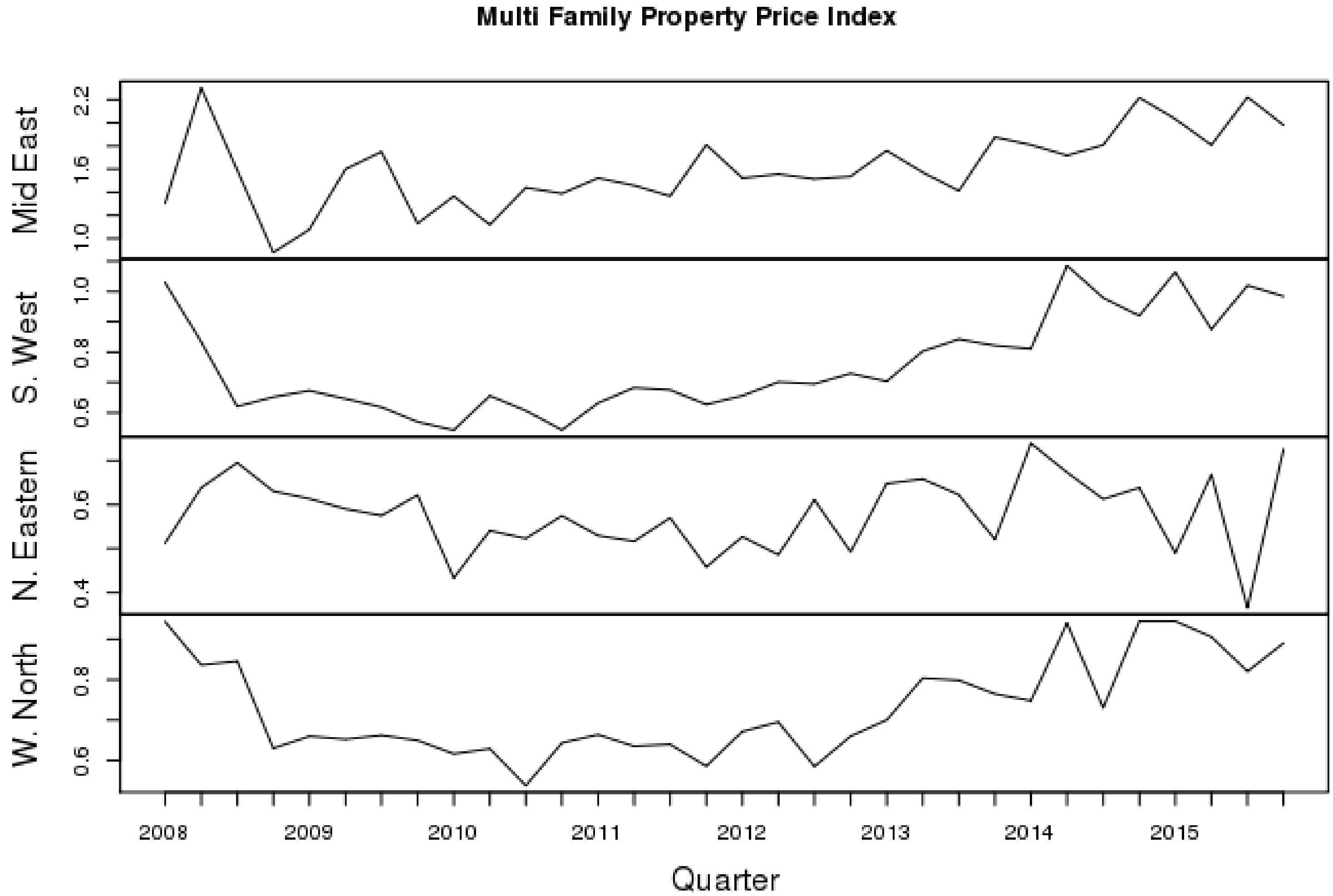

Figure 4.3. Quarterly Multi-Family Property Price Index

The graph provides the information on the estimation of the quarterly commercial property price index by region and property type. In this case the property type is Multi-Family and the data sources are TREPP collateral appraisal data and REIT property transaction data from SNL.

Figure 4.4. Quarterly Retail Property Price Index

The graph provides the information on the estimation of the quarterly commercial property price index by region and property type. In this case the property type is Retail and the data sources are TREPP collateral appraisal data and REIT property transaction data from SNL.

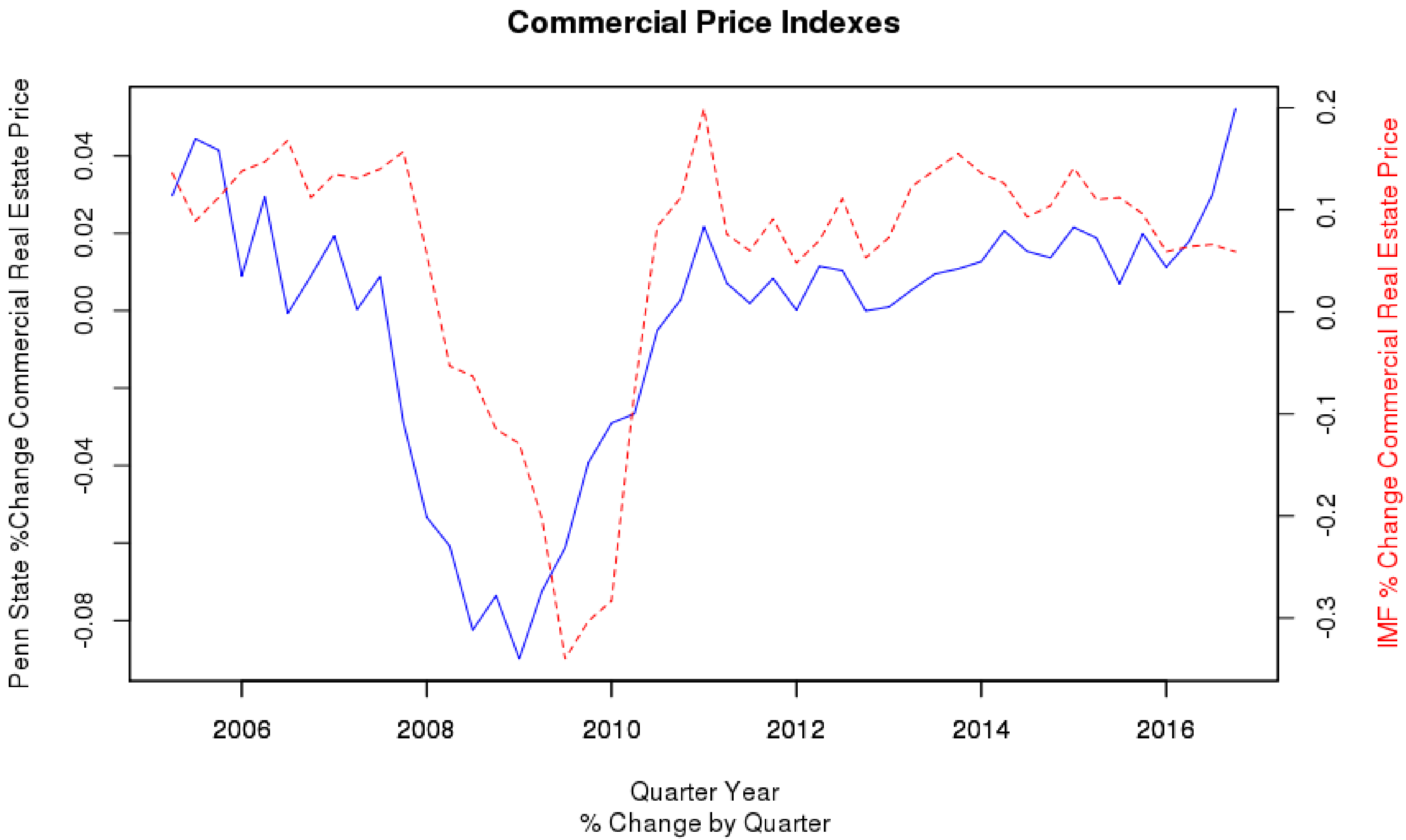

Figure 4.5. Comparison IMF Index vs Penn State Database

This graph compares the change of the Commercial Property Price Index estimated using TREPP and SNL(Penn State Index) against the IMF Commercial Propety Index available in FRED at St. Louis FED. The correlation of the two changes of index is 0.71 and the R-squared when regressed one onto the other is 0.51.

I use this commercial property index at the regional - property type level to construct an indicator of the equity position of borrowers. Table 4.9 indicates that 31.4% of the observation had negative equity during the time frame of analysis. Mortgage pricing literature recognizes that property price decline increases the probability of defaults (Kau et al., 1992) due to an increase in the probability of negative equity. The negative equity indicator in this paper captures precisely that.

4.4.7 Metropolitan Statistical Area Data

In an effort to estimate the relation of the CLEI to more granular level economic data, I use MSA level data from the Bureau of Labor Statistics, the Case et al. (1989) HPI at the MSA published by Federal Reserve of St. Louis, and the CPPI estimated at MSA level[5]. I complement the MSA data for the analysis with the macro level data from Section 4.4.4.

Table 4.12 provides the summary statistics of the data at the MSA level. The sample includes New York City, Chicago, Dallas, Los Angeles, Phoenix, San Diego, Detroit, Seattle, Las Vegas, Miami, Portland, Cleveland, Charlotte, Tampa, Minneapolis, Atlanta, Washington, Denver and San Francisco. These MSAs represent almost one third of the USA population, with an estimated population of 107,791,390 living in these MSAs as of 2016 [6].

4.5 Empirical Results

In this section, I provide the empirical results to asses the explanatory power of the Consumer Location Evaluation Index (CLEI). First, I start by establishing the relation of the CLEI to macro economic variables. Second, I asses the explanatory power of the index on metropolitan statistical area data. Finally, I use commercial mortgage

[5] I do not estimate the index by property type due to insufficient number of observation

[6] `https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP_2017_PEPANNRES&src=pt`

Table 4.12. Summary statitics MSA data

Statistic	N	Mean	St. Dev.	Min	Max
Employment	22,848	407,710.200	854,502.800	34,917	9,516,703
Unemployment Rate	22,848	7.859	2.926	2.300	30.000
CLEI	22,848	1.103	0.728	0.021	15.302
HPIFED	1,596	140.108	33.760	66.920	241.590
U. Mich Sent.	22,848	76.895	9.809	55.800	98.100
CPPI	1,623	0.806	0.370	0.072	5.083
10 year T. Rate	22,848	2.584	0.625	1.530	3.850
Yield curve	22,848	2.338	0.571	1.340	3.400
Spread AAA-BBB	22,848	1.739	0.677	0.896	4.364
Vol. 10 year T. R.	22,848	0.468	0.130	0.247	0.737
Vol. Spread AAA-BBB	22,848	0.478	0.332	0.131	1.166

Note: This table provides the summary statistics for the MSA data. CLEI represents the Consumer Location Evaluation Index estimated at the country level. U. Mich Sent. represents the consumer sentiment index calculated by University of Michigan. 10 Year T. Rate and 1 Year T. Rate represent the 10-Year and 1-Year Treasury Constant Maturity Rates. HPI is the S&P Case & Shiller house price index at the MSA level. CPPI is the Commercial Property Price Index estimated using Case & Shiller methodology and by using TREPP collateral appraisal data, as well as, SNL REIT property transaction data. Yield curve is the 10 Year T. Rate minus 1 Year T. Rate. Spread AAA-BBB represents the difference between BofAML US Corporate AAA and BBB debt effective yield. Vol. represents the previous 12 month moving standards deviation of the variable. Employment and Unemployment Rate represent the civilian labor force employed and the fraction of unemployed over labor force.

data to asses if the information embedded in the consumer evaluation index improves the prepayment and default hazard rate models for commercial mortgages. In the subsequent subsections, I describe the empirical findings for each aforementioned test.

4.5.1 Macroeconomic Data and CLEI

Table 4.11 presents the correlation between the CLEI at the national level and other macroeconomic variables. First, I show that the CLEI is associated with the

University of Michigan consumer sentiment index (UMCSI). This is relevant as the residual from regressing UMSCI onto macroeconomic variables predicts small stocks and stocks with low institutional ownership (Lemmon and Portniaguina, 2006)[7]. The CLEI has a strong relation with the spread between AAA and BBB corporate bond rates, suggesting that high levels of the index can be observed during periods of lower risk premium. The analysis also suggests a negative correlation between the CLEI and treasury bonds rates and with commercial property prices and volatility. I will test this last relation at the more granular level, controlling for economic conditions.

Figure 4.6 shows the plot of monthly the CLEI versus other key monthly macroeconomic indicators. It is clear that the CLEI and the UMCSI track each other in the long run (2008-2016). Both graphs of the indexes show a drop in early 2011, probably related to negative growth during the first quarter of 2011 when the economy experienced the first negative growth since 2009. The graph after the CLEI and the UMCSI is the volatility of the CLEI calculated as the standard deviation of the previous 12 months. There is a decline in the long run and the trend is comparable to that of the 10 year treasury bond.

[7] I do not provide an analysis of the predictive power of the CLEI and stock returns, but there is no doubt that this is an interesting extension of the project.

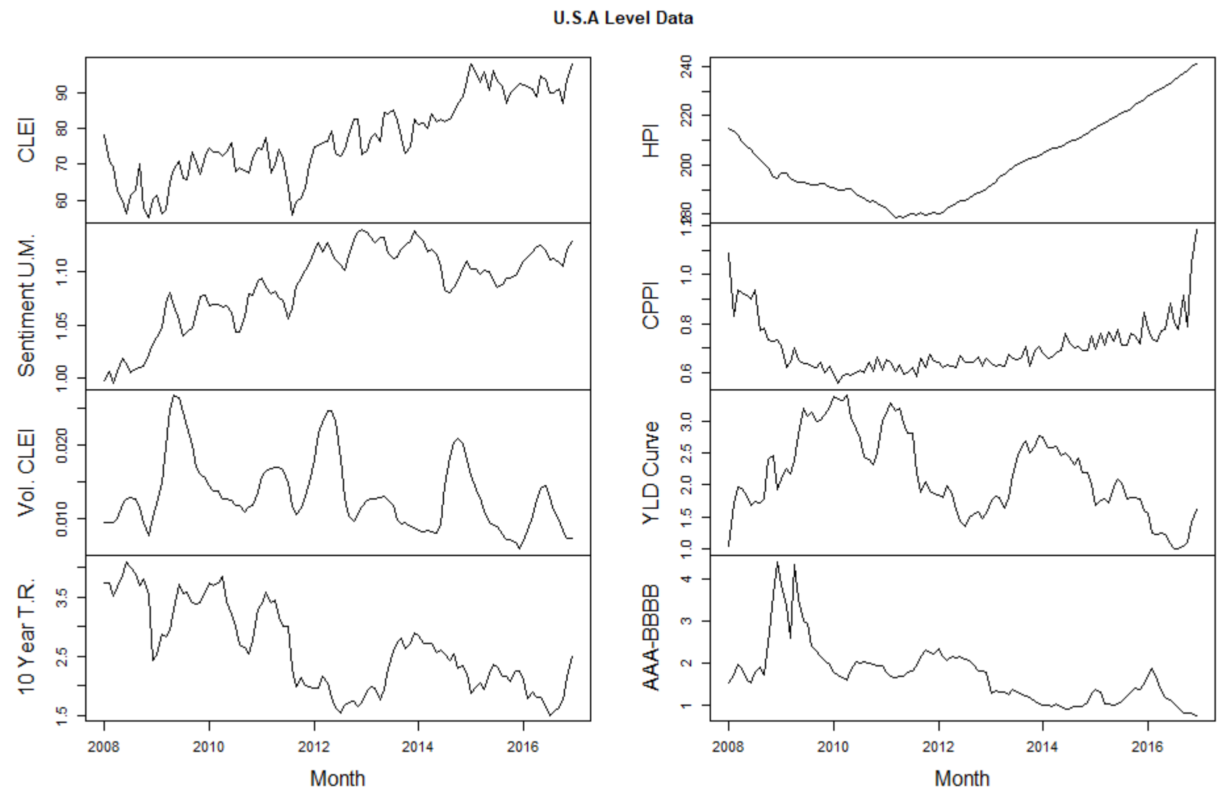

Figure 4.6. CLEI and Other Macro Economic Variables

Each graph is the monthly information of different variables. CLEI represents the Consumer Location Evaluation Index estimated at the country level. U. Mich Sent. represent the consumer sentiment index calculated by University of Michigan. 10 Year T. Rate and 1 Year T. Rate represent the 10-Year and 1-Year Treasury Constant Maturity Rates. HPI U.S.A. is the monthly house price index estimated by FHFA at the country level. CPPI is the commercial property price index estimated using Case & Shiller methodology and by using TREPP collateral appraisal data, as well as, SNL REIT property transaction data. Yield curve is the 10 Year T. Rate minus 1 Year T. Rate. AAA-BBB represents the Spread between BofAML US Corporate AAA and BBB debt effective yield. Vol. represents the previous 12 month moving standards deviation of the variable.

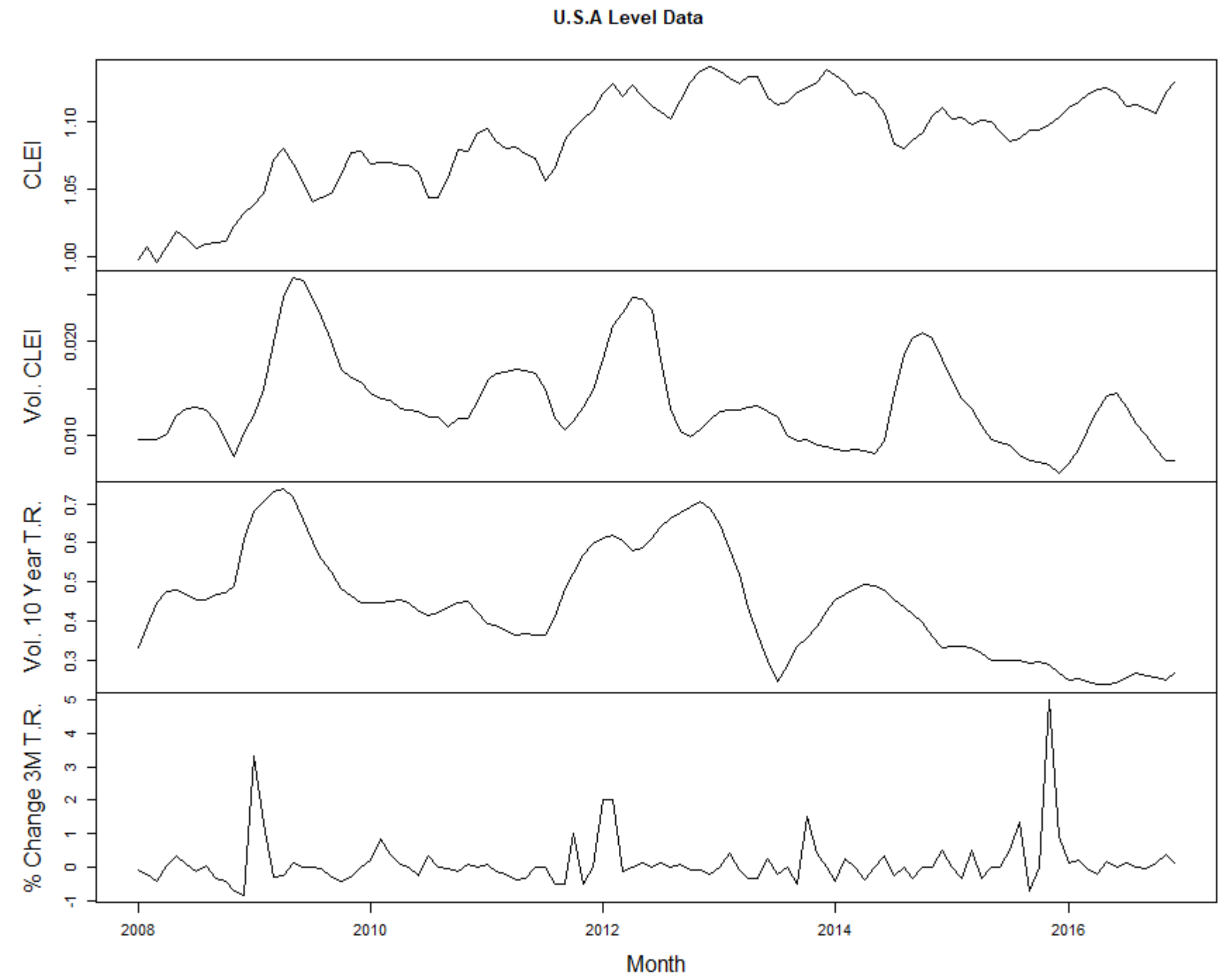

Figure 4.7. CLEI and Interest Rates

Each graph is the monthly information of different variables. CLEI represents the Consumer Location Evaluation Index estimated at the country level. 10 Year T. Rate and 3 month T. Rate represent the 10-Year and 3-Month Treasury Constant Maturity Rates. Vol. represents the previous 12 month moving standards deviation of the variable. % Change 3 T.R. represents the monthly percentage change of the 3-Month treasury rate.

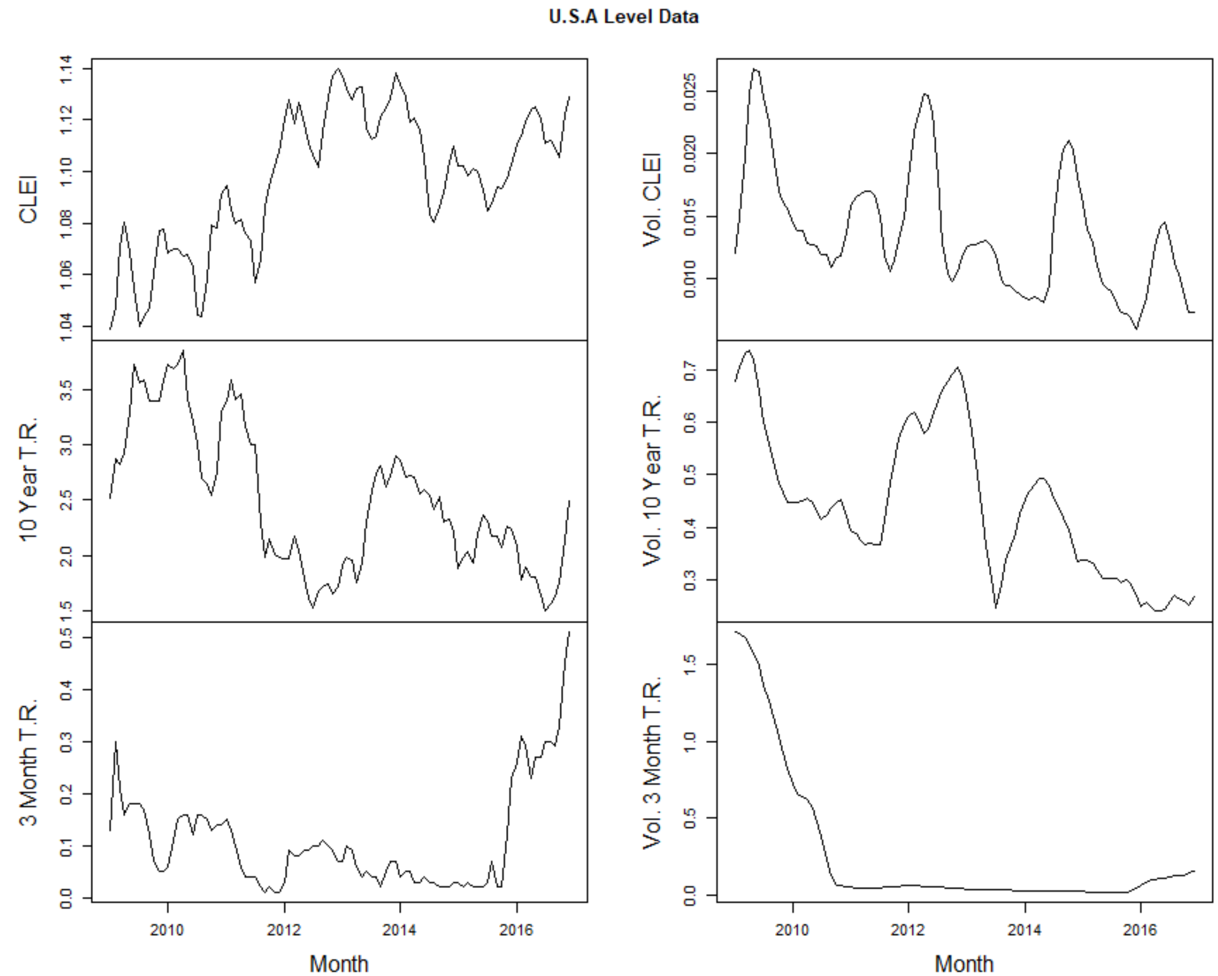

Figure 4.8. CLEI and Interest Rates

Each graph is the monthly information of different variables. CLEI represents the Consumer Location Evaluation Index estimated at the country level. 10 Year T. Rate and 3 month T. Rate represent the 10-Year and 3-Month Treasury Constant Maturity Rates. Vol. represents the previous 12 month moving standards deviation of the variable.

The panel on the right of Figure 4.6 shows the graph for the monthly national HPI from the FHFA and the CPPI calculated using the data described in Section 4.4.6. It is clear that commercial properties suffered a rapid decline in prices and the recovery has been slower than the recovery of house prices. The next two graphs after the price indexes reflect the market yield curve (YLD Curve) and the risk premium (AAA-BBB). The YLD Curve indicator tracks the Vol. CLEI indicator. While both pick similar time periods, the intensity of the picks are not as big and in some cases it lags and in others it leads this indicator.

To further analyze the relation between the CLEI and treasury bonds, Figures 4.7 and 4.8 show the monthly observations for 10 year and 3 month treasury bonds rates, and CLEI. The volatility for the different variables represents the standard deviation of the previous 12 month observations. The percentage change of the 3 month treasury bond represent the percentage change with respect to the previous month. It is clear that volatility of CLEI follows increases of the 3 month interest rates; the relation to the CLEI itself is less clear.

Overall it seems that the CLEI at the national level is correlated with information at the macro level. The consumer location evaluation index tracks some key indicators, such as the University of Michigan Consumer sentiment index, quite well. Although this is relevant for forecasting purposes, the advantage of the index is that it allows for more granular spatial economic activity.

4.5.2 Metropolitan Statistical Area Data Analysis

The main advantage of constructing the CLEI is that it can capture economic activity at a more granular level. In this section I provide evidence that the lag value of the CLEI is statistically significant when explaining economic activity. I use employment data from the Bureau of Labor Statistics, the HPI for the MSAs described in Section 4.4.7, and the CPPI at the MSA level.

Figure 4.10 plots a subsample of the CLEI at the MSA level for Chicago, New York, Seattle and Miami. The Chicago index has some variation between 2007 and 2012, with a significant decline after 2012. New York, on the other hand, after 2012 experienced a significant increase. I merge these indexes with the employment data and estimate the following fixed effect empirical model:

$$y_{it} = \alpha_i + \beta_0 CLEI_{it} + \beta_1 x_{1it} + \beta_2 x_{2it} + \varepsilon_{it}. \tag{4.9}$$

where y_{it} represents the dependent variable of interest (i.e. employment, unemployment, HPI or CPPI); α_i is the MSA fixed effect; $CLEI_{it}$ is the index of interest; x_{2it} is any local economic indicator (i.e. unemployment or employment); and x_{2it} represents a set of macroeconomic variables. The estimate of interest is β_0, and I am interested in whether this estimate is statistically significant.

Table 4.13 provides the results from the estimated regressions, and all the standard errors are heteroscedastic, autocorrelated and cross-sectionally correlated consistent following Driscoll and Kraay (1998). The first column is the estimate of the log transformation of employment. The regression includes not only both of the sentiment indexes, but also macroeconomic variables. I include both the CLEI and the UMCSI with a lag as I am interested in the predictive power of the CLEI. It is clear from column 1 that there is a strong relationship between the CLEI and Employment; a 1% change in the previous month translates into an increase of the population of approximately 0.005% which represents 20.4 new jobs . Column 2 shows that an increase of 1% in the CLEI translates into a 0.012% decline in the unemployment rate. Both estimates for the CLEI in column 1 and 2 are statistically significant at the 1% level.

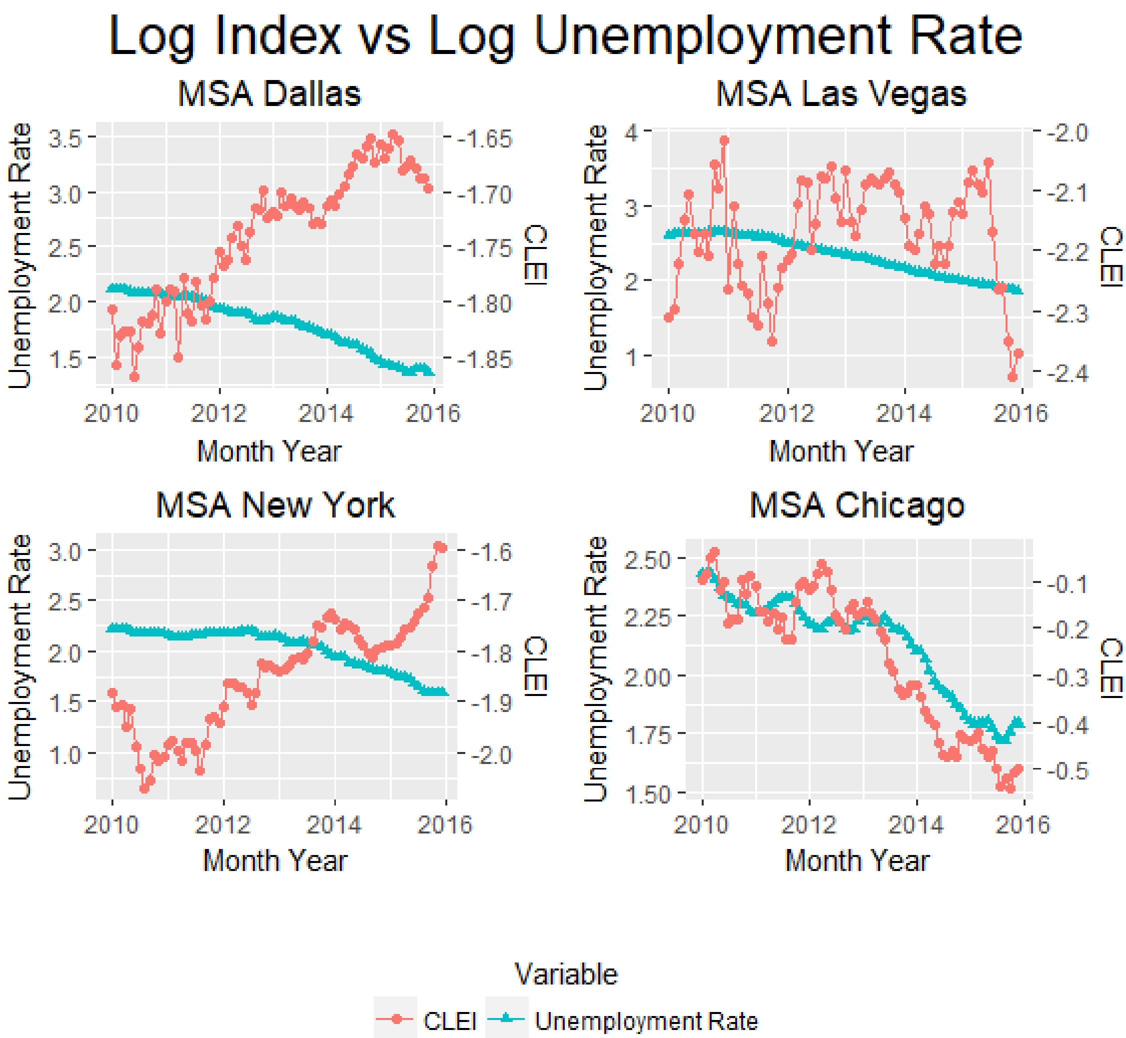

Figure 4.9. CLEI index and unemployment at the MSA level

The graph shows the comparison between the CLEI index and the unemployment rate in each MSA.

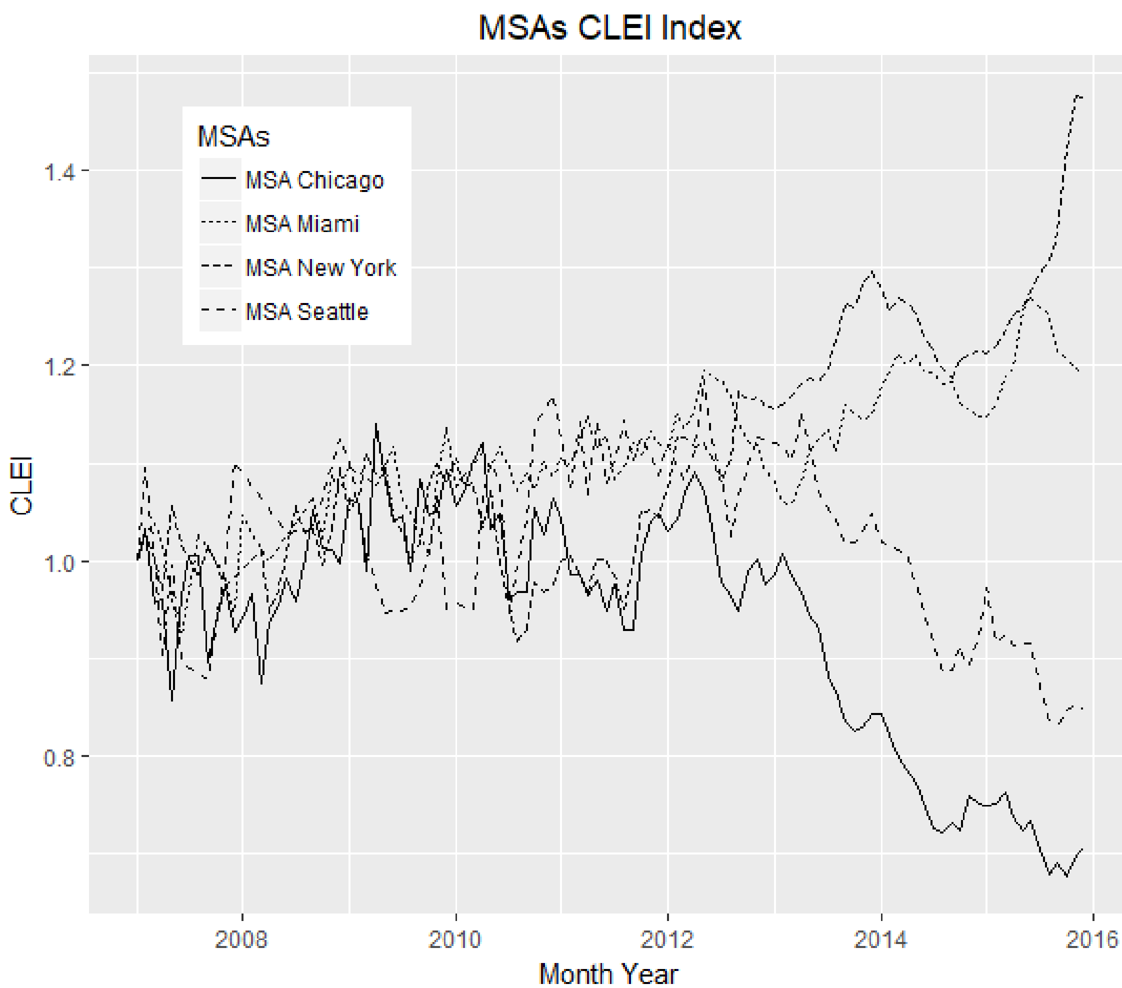

Figure 4.10. CLEI index at the MSA level

The graphs shows the CLEI indexes for select MSAs.

Table 4.13. Regressions Economic Variables on CLEI MSAs Data

	Dependent variable:				
	Log Employment	Log Unemployment Rate	Log HPI	Log CPPI	
	(1)	(2)	(3)	(4)	(5)
Lag Log CLEI	0.005***	−0.012*	0.072***	0.089*	0.038
	(0.002)	(0.007)	(0.014)	(0.050)	(0.049)
Lag Log UM Index	0.052***	−0.595***	0.043	0.180*	−0.124*
	(0.015)	(0.158)	(0.049)	(0.092)	(0.066)
Log Unemployment Rate			−0.323***		−0.220***
			(0.019)		(0.056)
Log Employment			0.498**		1.910***
			(0.196)		(0.302)
Standard Error	H.A.C.	H.A.C.	H.A.C.	H.A.C.	H.A.C.
Panel Estimation	within	within	within	within	within
Observations	22,576	22,576	1,577	1,608	1,608
Adjusted R^2	0.249	0.657	0.778	0.053	0.111

Note: CLEI represents the Consumer Location Evaluation Index estimated at the country level. U. Mich Sent. represents the consumer sentiment index calculated by University of Michigan. HPI is the monthly Case & Shiller house price index estimated at MSA level. CPPI is the Commercial Property Price Index at the MSA level estimated using Case & Shiller methodology and by using TREPP collateral appraisal data, as well as, SNL REIT property transaction data. Employment is the number of civilians in labor force employed at the MSA level. Unemployment Rate is the fraction of unemployed over civilian labor force at the MSA level. Other controls: 10 year T. Rate, Yield curve, Spread AAA-BBB, Vol. Spread AAA-BBB and Vol. 10 year T. Rate. The standard errors are heteroskedastic, auto-correlated and cross-sectional correlated consistent using Driscoll and Kraay (1998) methodology. The use of Newey and West (1987) standard errors does not change the significance level. The asterisks represent the significance levels, *p<0.1; **p<0.05; ***p<0.01

Next, columns 3 through 5 show the estimates for the impact of the CLEI on housing prices and commercial property prices. The HPI regression provides evidence that even after controlling for unemployment at the MSA, MSA fixed effects, and macro economic conditions (i.e. 10 year T. Rate, Yield curve, Spread AAA-BBB, Vol. Spread AAA-BBB and Vol. 10 year T. Rate), the lag CLEI is statistically significant and positive. This suggests that a 10% increase in the CLEI translates into a 0.7% increase in house price or a 1 standard deviation increase translates into an increase in house price of 4.9%. For the commercial properties index, the impact is also positive; the significance, though, depends on how good the unemployment rate forecast is. If the unemployment rate has a perfect forecast (column 5), the impact of the CLEI is not statistically different from 0. If the unemployment rate is not available (column 4) the impact of the CLEI becomes significant.

Overall, the findings so far suggest that the CLEI is positively correlated with employment and provides even more information than just unemployment. An Akaike-Information-Criterion analysis suggests that the models from column 1 through 5 in Table 4.13, including the CLEI variable, reduce the information loss. One important question now is whether this index can be used to improve models for mortgage prepayment and default.

4.5.3 Commercial Mortgage Data Analysis

In this section, I use Zhou et al. (2012) competing risk model to estimate the hazard rate for clustered data. My assumption here is that the data of prepayment and default are correlated within a state, given that within a state, mortgages are governed with the same set of rules. Therefore, standard errors need to be consistent with clustered data; thus, the Zhou et al. (2012) model.

Table 4.14 presents the results of the empirical competing risk model. Although not shown in the table, the empirical model includes location and property type fixed effects. First, the impact of the CLEI in the probability of termination is statistically

Table 4.14. Competing risk model of commercial mortgages

	Default	Prepayment
Lag Change CLEI	0.117	-0.005
	(0.237)	(0.234)
Lag Mean change	-2.188	3.463***
	(1.521)	(1.213)
Lag vol. change CLEI	0.533	-2.973***
	(0.487)	(0.555)
Unemployment Rate	0.116***	-0.025
	(0.019)	(0.019)
PPmtOption	0.372***	0.066***
	(0.030)	(0.012)
Vol. 10 year T.R.	0.163	-0.052
	(0.369)	(0.248)
Yield curve	0.531***	0.297***
	(0.110)	(0.039)
Spread AAA - BBB	0.051	-0.284***
	(0.075)	(0.068)
Vol. spread AAA-BBB	0.436***	-0.876***
	(0.152)	(0.132)
Lock-out indicator	-1.485 ***	-5.271***
	(0.105)	(0.097)
Yield maintenance indicator	-1.458***	-3.166***
	(0.122)	(0.085)
Negative Equity	0.707***	-0.155***
	(0.091)	(0.050)
DSCR	-1.191***	0.018
	(0.069)	(0.023)
LTV securitization	0.027***	0.007***
	(0.003)	(0.002)

Note: This table provides the estimates for model 5 from Tables 4.15 and 4.16. The CLEI variables are estimated at the 3 digits zip code. The unemployment rate is at the MSA level. The model also includes NCREIF region and Property type fixed effects.

significant for prepayment. Increases in the index in the previous 12 months suggest a larger number of prepayments. On the contrary, the impact of the volatility of the index is negative and statistically significant at the 1% level. The unemployment data is available at the MSA level and the estimate for default suggests that unemployment in the MSA increases the probability of default. To put this in perspective, an increase of one percentage point in unemployment increases the odds of default by 12%.

The model selection analysis in Tables 4.15 and 4.16 suggests that models that include MSA level data are preferred to those without local information. More interesting, the best models from an Akaike Information Criterion (AIC) perspective are the ones that include both the CLEI index at the National and 3 digit zip code levels. A Chi-square analysis suggests that the coefficients of MSA level data are statistically different from 0. The base model includes macro variables that are standard in the empirical literature on Mortgage termination.

The other variables in the model are standard in the literature. The prepayment option position (PPmtOption) has a direct impact on the termination of a mortgage due to either prepayment or default. An extra one standard deviation in the PPmtOption index translates into an increase in the odds of prepayment by 29.7% and the odds of default by 4.5%. Yield curve is also significant and positive. A one hundred basis points increase in the difference between the 10 years and 1 year treasury bond rates translates into an increase of the odds of prepayment by 70% and of default by 34.5%. The spread between AAA and BBB rated corporate bonds is statistically significant only for the prepayment option, which suggests that increases in the risk premium reduce the prepayment of mortgages. Volatility of the AAA-BBB spread has a negative impact on the odds of prepayment and an increase in the odds of defaults.

Table 4.15. Akaike information criterion test default

	Default Models					
	Model	**loglik**	**df**	**AIC**	**AIC diff**	**Better than Base Model**
0	Null Model	-29063	0	58127	4247.8	
1	Base Model (B.M.)	-26985	22	54014	135.3	
2	B.M.+ Index 3 digits zipcode Set 1 Variables	-26982	25	54015	136.2	No
3	B.M.+ Index 5 digits zipcode variables Set 1 Variables (Set1 Var.)	-26983	25	54017	137.9	No
4	B.M.+ Index USA Set1 Var.	-26975	25	54001	121.9	Yes
5	B.M.+ Index 3 digits zipcode Set1 Var.+ Unemployment Rate (U.R)	-26920	26	53893	13.6	Yes
6	B.M.+ Index 5 digits zipcode Set1 Var.+ U.R	-26921	26	53894	15.2	Yes
7	B.M.+ Index USA Set1 Var.+ Unemployment Rate	-26914	26	53881	2	Yes
8	B.M.+ Index 3 digits zipcode +Index USA Set2 Var.+ U.R	-26914	26	53880	1.6	Yes
9	B.M.+ Index USA+ Unemployment Rate	-26921	24	53890	11.3	Yes
10	B.M.+ Index 3 digits zipcode Set2 Var. +Index USA Set2 Var.+ U.R	-26913	27	53881	1.6	Yes
11	B.M.+ Index 5 digits zipcode Set2 Var.+Index USA Set2 Var.+ U.R	-26914	27	53883	3.6	Yes
12	B.M.+ Index USA Set2 Var.+U.R	-26914	25	53879	0	Yes

Note: This table provides the model selection analysis. Column loglik shows the log-likelihood of the model and column df the degrees of freedom. Column AIC shows the Akaike-Information-Criterion and AIC diff is the difference between the model's AIC and the lowest AIC. The null model considers all the coefficients of available explanatory variables as being equal to 0. The base model includes the following variables: Ppoption, Volatility 10 year treasury, NCREIF region fixed effects, Yield curve, Spread AAA-BBB bonds yield, Volatility spread AAA-BBB, Lock-out and Yield maintenance period indicators, Property type fixed effects, Negative equity indicator, most recent DSCR and LTV at securitization. Index 3 and 5 digit zipcode, and USA set 1 of variables include the one month lag change of the CLEI, the one month lag of the 12 month moving average of the change in the CLEI and the one month lag of the 12 month standard deviation of the change in the CLEI at the respective geographical level. Unemployment rate variable is the unemployment rate at the MSA level. The set 2 of CLEI variables is the same as the set one but in this case I drop the lag one month change of CLEI.

Table 4.16. Akaike information criterion test prepayment

	Prepayment Models					
	Model	**loglik**	**df**	**AIC**	**AIC diff**	**Better than Base Model**
0	Null Model	-46934	0	93867	15768.4	
1	Base Model (B.M.)	-39113	22	78270	171.7	
2	B.M.+ Index 3 digits zip code Set 1 Variables	-39080	25	78210	111.1	Yes
3	B.M.+ Index 5 digits zip code variables Set 1 Variables (Set1 Var.)	-39111	25	78273	173.9	No
4	B.M.+ Index USA Set1 Var.	-39032	25	78115	16.2	Yes
5	B.M.+ Index 3 digits zip code Set1 Var.+ Unemployment Rate (U.R)	-39077	26	78206	107.4	Yes
6	B.M.+ Index 5 digits zip code Set1 Var.+ U.R	-39106	26	78264	164.9	Yes
7	B.M.+ Index USA Set1 Var.+ Unemployment Rate	-39030	26	78111	12.6	Yes
8	B.M.+ Index 3 digits zip code +Index USA Set2 Var.+ U.R	-39030	26	78113	14.3	Yes
9	B.M.+ Index USA+ Unemployment Rate	-39032	24	78111	12.5	Yes
10	B.M.+ Index 3 digits zip code Set2 Var. +Index USA Set2 Var.+ U.R	-39022	27	78099	0	Yes
11	B.M.+ Index 5 digits zip code Set2 Var.+Index USA Set2 Var.+ U.R	-39029	27	78113	14.3	Yes
12	B.M.+ Index USA Set 2 Variables+U.R	-39030	25	78111	12.3	Yes

Note: This table provides the model selection analysis. Column loglik shows the log-likelihood of the model and column df the degrees of freedom. Column AIC shows the Akaike-Information-Criterion and AIC diff is the difference between the model's AIC and the lowest AIC. The null model considers all the coefficients of available explanatory variables as being equal to 0. The base model includes the following variables: Ppoption, Volatility 10 year treasury, NCREIF region fixed effects, Yield curve, Spread AAA-BBB bonds yield, Volatility spread AAA-BBB, Lock-out and Yield maintenance period indicators, Property type fixed effects, Negative equity indicator, most recent DSCR and LTV at securitization. Index 3 and 5 digit zipcode, and USA set 1 of variables include the one month lag change of the CLEI, the one month lag of the 12 month moving average of the change in the CLEI and the one month lag of the 12 month standard deviation of the change in the CLEI at the respective geographical level. Unemployment rate variable is the unemployment rate at the MSA level. The set 2 of CLEI variables is the same as the set one but in this case I drop the lag one month change of CLEI.

Lock-out and Yield maintenance are time-varying indicators that take the value of 1 if the loan is under a lock-out provision or under a yield maintenance and 0 otherwise. The impact of a lock-out clause is economically and statistically significant at the 1% level. For example, when a loan is under a lock-out period (i.e. Lock-out indicator equal to 1) the odds of prepayment decrease by 99.5% and for the yield maintenance indicator the odds decrease by 95.7%. Negative equity increases the probability of default and decreases the odds of prepayment. Increases in the most recent DSCR reduce the odds of default but the impact on prepayment is null. One standard deviation of the DSCR reduces the odds of defaults by 76%. Finally, although the LTV at securitization is statistically significant the economic impact is small, which relates to the idea that LTV at origination is endogenous to the origination process; therefore, its impact should be insignificant (Archer et al., 2002).

The main take away of the commercial loan analysis is that models which include location consumer reviews to proxy for localized economic activity reduce the information loss from the Akaike Information Criterion perspective. A Chi-square analysis on the log likelihood of the models suggests that the coefficients for variables involving the CLEI are statistically different from 0.

4.5.3.1 CLEI by Hotel Class

The model proposed in Section 4.3 makes the assumption that CLEI is independent to the type of costumer visiting the hotel. In this section I relax this assumption and estimate the CLEI for different hotel classes. More specifically, I divide hotels into two groups[8]. The first group includes hotels with a hotel class less or equal to 2.5 stars, while the second group includes hotels with a hotel class greater than 2.5 stars. If customers visiting higher class hotels value changes in local amenities differently or focus on a completely different set of amenities than those of customers visiting lower class hotels, then the index could be improved by separating the two groups

[8]Hotel class is independent to consumer reviews and describes the general level of features and amenities to expect in a hotel. This information is disclosed for each hotel by TripAdvisor and comes either from a third party vendor or GIATA.

and creating two indexes.

Figure 4.11 shows the CLEI at the USA level estimated for different hotel classes. The graph shows a difference between the CLEI estimated using the hotels with a class of 2.5 stars or lower and the CLEI using hotels with a class higher than 2.5 stars. It is clear that the two CLEIs track each other closely, specially before 2012, but then they diverge. This might suggest that consumers from high end hotels value amenities differently or focus on a different set of amenities to rate a location than those consumers visiting lower class hotels. Therefore, analyzing whether these different indexes provide more information than just using one index I run the same analysis as in Tables 4.15 and 4.16.

Tables 4.17 and 4.18 show the Akaike Information Criterion (AIC) analysis for the default and prepayment models. The results suggest that separating the indexes between lower or equal (LE) and greater than (GT) 2.5 stars improve the models. The best default model from the AIC perspective is the one that includes the indexes at the USA level separated in the two groups and that also includes the indexes at the zip 3 level for the two groups (model 6). The prepayment model also improves when including the CLEI by hotel type. The best models from the AIC are models 5 and 6 from Table 4.18.

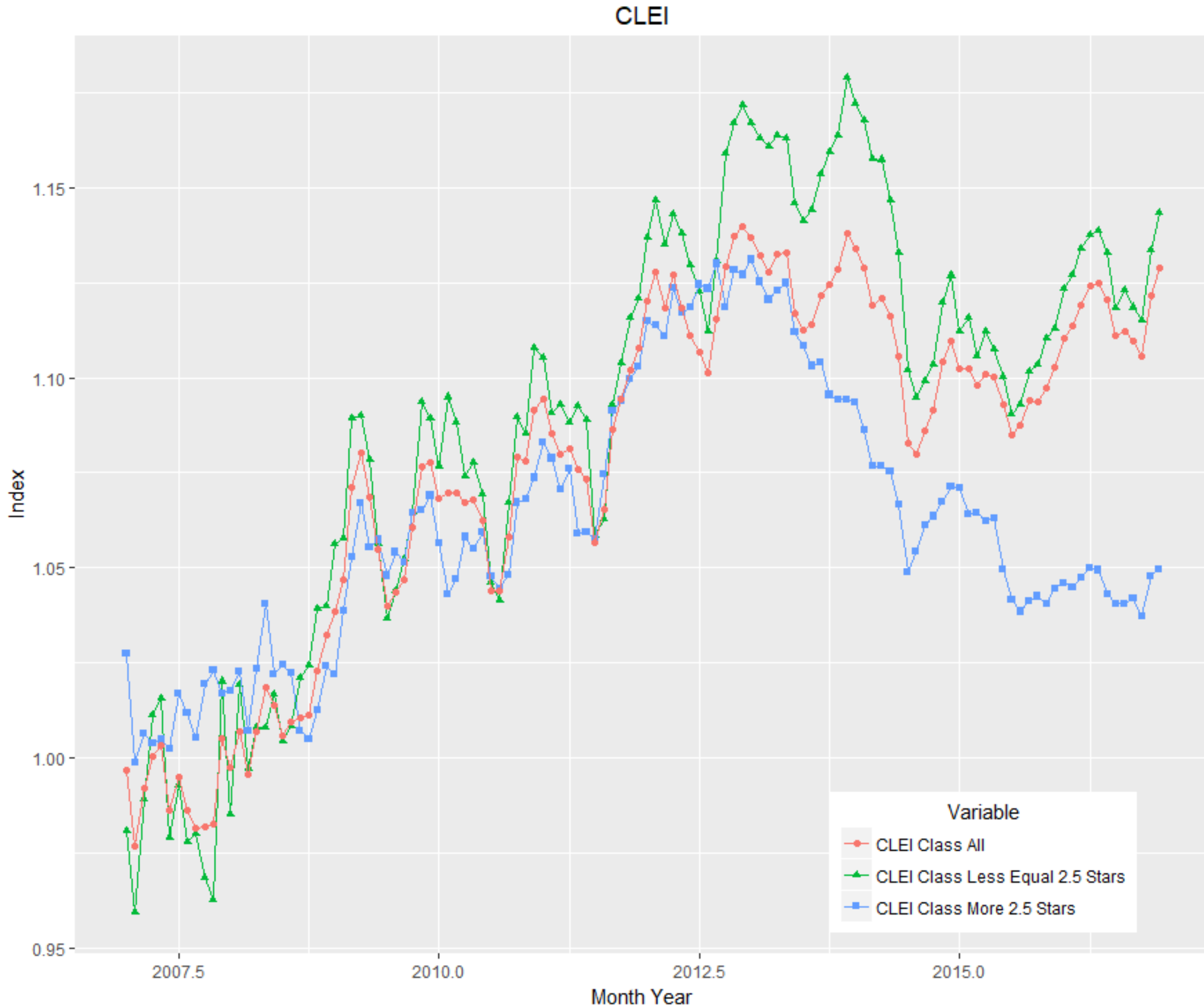

Figure 4.11. CLEI by Hotel Class

Each graph line represents the CLEI estimated for different subsamples of hotels by their class. CLEI Class All is the index estimated at the USA level using all hotels' location reviews. CLEI Less Equal 2.5 Stars represents the index at the USA level using only location reviews of hotels that have a class lower or equal to 2.5 stars. Finally, CLEI Class More 2.5 Stars is the index estimated at the USA level using only location reviews of hotels with a class greater than 2.5 stars.

Table 4.17. Akaike information criterion test default by hotel class

	Default Models					
	Model	**loglik**	**df**	**AIC**	**AIC diff**	**Better than Base Model**
0	Null Model	-28117	0	56234		
1	Base Model (B.M.)	-26139	21	52320	130	
2	B.M.+ Unemployment Rate (U.R)	-26088	22	52220	30	Yes
3	B.M.+ Index USA Set 2 Variables+U.R	-26081	24	52210	20	Yes
4	B.M.+ Index Zip 3 Set 2 Variables GT and LE + U.R	-26081	26	52214	24	Yes
5	B.M.+ Index USA Set 2 Variables GT and LE + U.R	-26072	26	52196	6	Yes
6	B.M.+ Index USA and Zip 3 Set 2 Variables GT and LE + U.R	-26065	30	52190	0	Yes

Note: This table provides the default model selection analysis. Column loglik shows the log-likelihood of the model and column df the degrees of freedom. Column AIC shows the Akaike-Information-Criterion and AIC diff is the difference between the model's AIC and the lowest AIC. The null model considers all the coefficients of available explanatory variables as being equal to 0. The base model includes the following variables: Ppoption, Volatility 10 year treasury, NCREIF region fixed effects, Yield curve, Spread AAA-BBB bonds yield, Volatility spread AAA-BBB, Lock-out and Yield maintenance period indicators, Property type fixed effects, Negative equity indicator, most recent DSCR and LTV at securitization. Index Zip 3 and USA set 2 variables include the one month lag of the 12 month moving average of the change in the CLEI and the one month lag of the 12 month standard deviation of the change in the CLEI at the respective geographical level. LE indicates that the CLEI set 2 variables were estimated for hotels with a hotel class less than or equal to 2.5 stars and GT is the index set 2 variables for hotels with a class greater than 2.5 stars. Unemployment rate variable is the unemployment rate at the MSA level.

Table 4.18. Akaike information criterion test prepayment by hotel class

	Prepayment Models					
	Model	**loglik**	**df**	**AIC**	**AIC diff**	**Better than Base Model**
0	Null Model	-45369	0	90738		
1	Base Model (B.M.)	-37800	21	75642	164	
2	B.M.+ Unemployment Rate (U.R)	-37798	22	75640	162	Yes
3	B.M.+ Index USA Set 2 Variables+U.R	-37735	24	75518	40	Yes
4	B.M.+ Index Zip 3 Set 2 Variables GT and LE + U.R	-37771	26	75594	116	Yes
5	B.M.+ Index USA Set 2 Variables GT and LE + U.R	-37713	26	75478	0	Yes
6	B.M.+ Index USA and Zip 3 Set 2 Variables GT and LE + U.R	-37709	30	75478	0	Yes

Note: This table provides the prepayment model selection analysis. Column loglik shows the log-likelihood of the model and column df the degrees of freedom. Column AIC shows the Akaike-Information-Criterion and AIC diff is the difference between the model's AIC and the lowest AIC. The null model considers all the coefficients of available explanatory variables as being equal to 0. The base model includes the following variables: Ppoption, Volatility 10 year treasury, NCREIF region fixed effects, Yield curve, Spread AAA-BBB bonds yield, Volatility spread AAA-BBB, Lock-out and Yield maintenance period indicators, Property type fixed effects, Negative equity indicator, most recent DSCR and LTV at securitization. Index Zip 3 and USA set 2 variables include the one month lag of the 12 month moving average of the change in the CLEI and the one month lag of the 12 month standard deviation of the change in the CLEI at the respective geographical level. LE indicates that the CLEI set 2 variables were estimated for hotels with a hotel class less than or equal to 2.5 stars and GT is the index set 2 variables for hotels with a class greater than 2.5 stars. Unemployment rate variable is the unemployment rate at the MSA level.

4.6 Conclusion

I find evidence that the consumer location evaluation index (CLEI) has explanatory power on employment, unemployment rate, house price indexes, and commercial price indexes at the MSA level. Also, the index at the national level exhibits significant correlation with macro economic data like the 10 and 1 year Treasury Bonds, the market risk premium measured by the difference between the AAA - BBB bond rates, and the commercial property price index. The results from a panel regression model using MSA level data suggest that lag value of the CLEI can explain values of the MSA level employment, unemployment rate, house prices and commercial property prices. The findings also suggest that the index created using consumers' location reviews reveals economic activity at a granular level. The advantage of this index is that it allows researchers to control for a localized economic environment that may affect an underlying real estate asset.

I then use the index to further examine the explanatory power on commercial mortgage default and prepayment by using a competing risk hazard rate model for mortgage termination. The estimates from this model suggest that the inclusion of the CLEI reduce the information loss relative to a base model with variables typically found in the empirical literature of determinants of mortgage termination Jones and Sirmans (2016). The coefficients associated with the CLEI in the competing risk model are statistically significant. For example, a 1 standard deviation increase on the 12 months moving average of the change in the index results in an increase on the odds of prepayment by 4 times, and a 1 standard deviation on the volatility of the monthly change in the index translates into a drop of prepayments of about 10%. Overall, the findings shown here suggest that consumer reviews reveal information about the economic activity at a granular level allowing researchers, practitioners and policy makers to construct timely indexes in order to control for location characteristics changing over time.

Appendix A
Derivations Option Model

A.0.0.1 The Outside Option

The value of the option to sell the building to another developer depends on the sunk cost D paid to acquire a building and the value of the building for an alternative use S, which follows the geometric Brownian motion described by Equation 2.1. Since there is no cash flow associated with holding this investment opportunity the only return from holding this option is the capital gains.

Now in order to price this investment opportunity consider a portfolio that has a long position on the option to invest and a short position on the underlying asset S equal to $\frac{\partial O}{\partial S}$.

$$\begin{aligned} \pi_O &= +O - \frac{\partial O}{\partial S} S \\ d\pi_{sell} &= dO - \frac{\partial O}{\partial S} dS \end{aligned} \tag{A.1}$$

The short position requires a payment of the revenue stream equal to δS per unit invested. Given that there are $\frac{\partial O}{\partial S}$ invested, the total payment equals $\delta S \frac{\partial O}{\partial S}$. Taking into account the payment, the portfolio return then can be expressed as:

$$dO - \frac{\partial O}{\partial S}dS - \delta S\frac{\partial O}{\partial S}dt. \tag{A.2}$$

By using Ito's lemma dO can be expressed as:

$$dO = \frac{\partial O}{\partial S}dS + \frac{1}{2}\frac{\partial^2 O}{\partial S^2}dS^2. \tag{A.3}$$

Replacing A.3 into A.2 I get the total return of the portfolio in A.1:

$$\begin{aligned}
&\frac{\partial O}{\partial S}dS + \frac{1}{2}\frac{\partial^2 O}{\partial S^2}dS^2 - \frac{\partial O}{\partial S}dS - \delta S\frac{\partial O}{\partial S}dt. \\
&\frac{1}{2}\frac{\partial^2 O}{\partial S^2}dS^2 - \delta S\frac{\partial O}{\partial S}dt.
\end{aligned} \tag{A.4}$$

Moreover I know that dS^2 is given by $\sigma^2 S^2 dt$, therefore the total return of the portfolio is given by:

$$\frac{1}{2}\frac{\partial^2 O}{\partial S^2}\sigma^2 S^2 dt - \delta S\frac{\partial O}{\partial S}dt. \tag{A.5}$$

and since the return of the portfolio is free of risk I get the following equality:

$$\begin{aligned}
r_f \cdot \pi_{sell} dt &= \frac{1}{2}\frac{\partial^2 O}{\partial S^2}\sigma^2 S^2 dt - \delta S\frac{\partial O}{\partial S}dt. \\
r_f \cdot [O - \frac{\partial O}{\partial S}S]dt &= \frac{1}{2}\frac{\partial^2 O}{\partial S^2}\sigma^2 S^2 dt - \delta S\frac{\partial O}{\partial S}dt. \\
r_f \cdot [O - \frac{\partial O}{\partial S}S] &= \frac{1}{2}\frac{\partial^2 O}{\partial S^2}\sigma^2 S^2 - \delta S\frac{\partial O}{\partial S}. \\
0 &= \frac{1}{2}\frac{\partial^2 O}{\partial S^2}\sigma^2 S^2 + (r_f - \delta)S\frac{\partial O}{\partial S} - r_f O
\end{aligned} \tag{A.6}$$

Where the solution to the last differential equation in A.6 represents the option value of the alternative project or sell option (O). In order to solve this equation I establish some boundary conditions:

$$\begin{aligned} O(S=0) &= 0 \\ O(S^*) &= S^* - D \\ \frac{\partial O(S^*)}{\partial S} &= 1 \end{aligned} \tag{A.7}$$

The first boundary condition states that the option to invest would be of no value when S goes to zero. S^* is the free boundary condition and represents the value of S when it is optimal to invest, where D is the investment required to develop. The second boundary condition is the value matching condition. Finally, the third condition is the smooth pasting condition.

In order to satisfy the first boundary condition, the option O must be of the following form:

$$O(S) = AS^{\beta_1} \tag{A.8}$$

where, A is a constant that needs to be identified. Using the second and third boundary conditions, I can write the the optimal value of S in order to invest in such project as:

$$S^* = \frac{\beta_1}{\beta_1 - 1} D, \tag{A.9}$$

and the value of A as:

$$
\begin{aligned}
A &= \frac{S^* - D}{(S^*)^{\beta_1}} \\
&= \frac{\frac{\beta_1}{\beta_1 - 1} D - D}{(\frac{\beta_1}{\beta_1 - 1} D)^{\beta_1}} \\
&= \frac{(\beta_1 - 1)^{(\beta_1 - 1)}}{(\beta_1^{\beta_1}) D^{\beta_1}}
\end{aligned}
\tag{A.10}
$$

Equation A.6 is a second order homogeneous differential equation that is linear in variable S and derivatives, therefore the solution can be expressed as a linear combination of two independent solutions of the form AS^{β_1}. A general solution requires that β_1 is a solution to the fundamental quadratic equation given by:

$$
0 = \frac{1}{2}\sigma^2 \beta_1(\beta_1 - 1) + (r_f - \delta)\beta_1 - r_f \tag{A.11}
$$

with this I have two roots:

$$
\begin{aligned}
\beta_{11} &= \frac{1}{2} - \frac{r_f - \delta}{\sigma_S^2} + \sqrt[2]{\left[\left(\frac{r_f - \delta}{\sigma_S^2}\right) - \frac{1}{2}\right]^2 + 2\frac{r_f}{\sigma_S^2}} \\
\beta_{12} &= \frac{1}{2} - \frac{r_f - \delta}{\sigma_S^2} - \sqrt[2]{\left[\left(\frac{r_f - \delta}{\sigma_S^2}\right) - \frac{1}{2}\right]^2 + 2\frac{r_f}{\sigma_S^2}}
\end{aligned}
\tag{A.12}
$$

The resulting solution could be any combination of two independent solutions, like for example the sum of them:

$$
O(S) = A_1 S^{\beta_{11}} + A_2 S^{\beta_{12}} \tag{A.13}
$$

The condition $O(S = 0) = 0$ requires that $A_2 = 0$ due to the negative root β_{12}. The resulting solution then equals the solution in Equation A.9 with the value of A given by equation A.8 and the value of β_1 given by β_{11}.

The estimated value of the alternative investment opportunity $O(S)$ follows a geometric Brownian motion. The process $O(S)$ follows, can be derived by using Ito's lemma:

$$dO = \frac{\partial O}{\partial t}dt + \frac{\partial O}{\partial S}dS + \frac{1}{2}\frac{\partial^2 O}{\partial S^2}dS^2 \tag{A.14}$$

By replacing the conditions that in this case $\dfrac{\partial O}{\partial t} = 0$, that $dS^2 = \sigma^2 S^2 dt$, and by using the solution derived in Equations A.8 through A.13 equation can be re-written as follows:

$$\begin{aligned}
dO &= \beta_{11}A_1S^{\beta_{11}}\gamma dt + \beta_{11}A_1S^{\beta_{11}}\sigma_S dZ_S + \frac{1}{2}\beta_{11}(\beta_{11}-1)A_1S^{\beta_{11}}\sigma_S^2 dt \\
dO &= [\beta_{11}\gamma + \frac{1}{2}\beta_{11}(\beta_{11}-1)\sigma_S^2]Odt + \beta_{11}\sigma_S O dZ_S \\
\frac{dO}{O} &= [\beta_{11}\gamma + \frac{1}{2}\beta_{11}(\beta_{11}-1)\sigma_S^2]dt + \beta_{11}\sigma_S dZ_S \\
\frac{dO}{O} &= \gamma_O dt + \sigma_O dZ_S
\end{aligned} \tag{A.15}$$

For the real estate manager the value of the option to sell is important as this represents the termination value of the project and higher values of that option decreases the range of rents at which the manager would maintain a depreciated property or even increase the rent at which it is optimal to invest in CAPEX. Geltner et al. (1996) examine this option in the presence of two potential stochastic processes with correlation factor ρ between the two. Their numerical solution findings suggest lower correlation of stochastic processes of alternative rent increases the value of the sell option. Moreover, even if both processes are highly correlated, the option to sell is still higher in value than when there is only one other use for the property.

In this paper, I'm interested in reviewing what is the impact that an alternative land uses has on the value of an option to invest in either type of building use. Therefore, I will set up the option as a free boundary problem in order to find an analytical solution.

A.0.0.2 The Depreciating Property

For the development of this section I follow the approach used by Dixit and Pindyck (1994) and Marcus and Modest (1984). In their model a depreciating asset has the option to be replaced. Once the asset is substituted, the asset starts depreciating again and this option gets renewed every time the asset fully depreciates. In our case due to either physical or economical depreciation the building needs to be remodeled in order to keep it up to date and operational. Now if the price of the highest best use is sufficiently low not to immediately remodel but high enough to keep the option alive to invest in capex, lets say above R_L, then the REIT maintains the depreciated or "dead" property. We can imagine that the depreciated property has a decay such that eventually the maintenance cost increases enough to make the property cash flows zero. Let's assume the correlation between dz_R and dz_C is ρ_{rc} and is constant.

Call the option to maintain a depreciated property and keep the option to invest alive V(R,C). Let's build a portfolio that is long in one unit of V(R,C) and short in $\partial V/\partial R$ units of the asset R and short in $\partial V/\partial C$ units of C and call that portfolio π_{RC}. Then by using Ito's lemma I obtain:

$$\begin{aligned} d(V - \frac{\partial V}{\partial R}R - \frac{\partial V}{\partial C}C) &= \frac{1}{2}[\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 + \frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 + 2\frac{\partial^2 V}{\partial C \partial R}CR\sigma_C\sigma_R\rho_{rc}]dt + \\ &+ \frac{\partial V}{\partial R}R + \frac{\partial V}{\partial C}C - \frac{\partial V}{\partial R}R - \frac{\partial V}{\partial C}C \\ &= \frac{1}{2}[\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 + \frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 + 2\frac{\partial^2 V}{\partial C \partial R}CR\sigma_C\sigma_R\rho_{rc}]dt \end{aligned} \tag{A.16}$$

The resulting equation A.16 is the capital gains associated with holding such a portfolio. The short position requires a payment to investors equal to the convenience yields of the asset, δ_C and δ_R. Convenience yields are the difference between a risk adjusted discount rate μ_R and μ_C for R and C respectively, and their corresponding drifts parameters α and θ. Therefore the total return on the portfolio π_{RC} is:

$$\frac{1}{2}[\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 + \frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 + 2\frac{\partial^2 V}{\partial C \partial R}CR\sigma_C\sigma_R\rho_{rc}]dt - \delta_R\frac{\partial V}{\partial R}Rdt - \delta_C\frac{\partial V}{\partial C}Cdt \tag{A.17}$$

Portfolio π_{RC} is a risk neutral portfolio thus the expected return on the portfolio would be the risk free rate.

$$\begin{aligned} r_f\pi_{RC}dt &= \frac{1}{2}[\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 + \frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 + 2\frac{\partial^2 V}{\partial C \partial R}CR\sigma_C\sigma_R\rho_{rc}]dt - \delta_R\frac{\partial V}{\partial R}Rdt- \\ &- \delta_C\frac{\partial V}{\partial C}Cdt \\ 0 &= \frac{1}{2}[\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 + \frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 + 2\frac{\partial^2 V}{\partial C \partial R}CR\sigma_C\sigma_R\rho_{rc}] + (r_f - \delta_R)\frac{\partial V}{\partial R}R+ \\ &+ (r_f - \delta_C)\frac{\partial V}{\partial C}C - r_fV \end{aligned} \tag{A.18}$$

This last differential equation applies over the area where it is optimal to hold the depreciated property and not exercise the option to invest in CAPEX. When the rent exceeds an optimal threshold then the option is exercised, giving us the following condition:

$$V(R, C) = B(R, \lambda) - C \tag{A.19}$$

where $B(R, \lambda)$ is the value of a building with revenue flow R and depreciation parameter λ. This last equality is the value matching condition. I will discuss what the value of the building is in the next subsection, but first let's state the smooth pasting conditions:

$$\begin{aligned}\frac{\partial V(R,C)}{\partial R} &= \frac{\partial B(R,\lambda)}{\partial R} \\ \frac{\partial V(R,C)}{\partial C} &= -1\end{aligned} \tag{A.20}$$

Basically what this last condition tells us is that at the boundary, the functions need to meet tangentially.

The Expected Value of the Depreciating Property

Let's describe the depreciation process of the building and its implication on the option values. The building's lifetime is random and follows a Poisson process, where the probability of death of the property during the next short period, dT, conditional on living up until T is λdT. The probability density is given by:

$$\lambda e^{-\lambda T} \tag{A.21}$$

and the cumulative by:

$$1 - e^{-\lambda T} \tag{A.22}$$

Given the rental stochastic process the percentage change in HBU rental rate can be represented by:

$$\begin{aligned}dR &= \alpha R dt + \sigma R dz \\ \frac{dR}{R} &= \alpha dt + \sigma dz \qquad take \qquad expectation \\ \varepsilon[\frac{dR}{R}] &= \alpha dt \\ \frac{1}{dt}\varepsilon[\frac{dR}{R}] &= \alpha\end{aligned} \tag{A.23}$$

The last equation gives us the expected change in rent for the HBU during a short period of time dt. Now I can construct the expected rent of the project and therefore the value of the project with these characteristics represented by B(R). Let's define the discount factor for this project as μ_R and let's use the expected change of R described in A.23 to derive the value of the functioning building, B(R).

$$
\begin{aligned}
B(R) &= \int_0^T e^{-\mu_R t} R e^{\alpha t} dt \\
&= R \int_0^T e^{-(\mu_R - \alpha)t} dt \\
&= R \frac{1 - e^{-(\mu_R - \alpha)T}}{\mu_R - \alpha} \\
&= R \frac{1 - e^{-\delta_R T}}{\delta_R}
\end{aligned}
\tag{A.24}
$$

where $\delta_R = \mu_R - \alpha$ represents the shortfall. Now in order to incorporate the impact that an exponential depreciation has on the value of the building, let's consider the probability density function of a Poisson process for the lifetime.

$$
\begin{aligned}
B(R, \lambda) &= \int_0^\infty \lambda e^{-\lambda t} \frac{R}{\delta_R} [1 - e^{-\delta_R t}] dt \\
&= \frac{R\lambda}{\delta_R} \int_0^\infty e^{-\lambda t} [1 - e^{-\delta_R t}] dt \\
&= \frac{R\lambda}{\delta_R} [-\frac{1}{\lambda} e^{-\lambda t} - \frac{-1}{\lambda + \delta_R} e^{-(\lambda + \delta_R)t}]\Big|_0^\infty \\
&= \frac{R}{\lambda + \delta_R}
\end{aligned}
\tag{A.25}
$$

This last result is the value of a building in operation receiving a revenue stream R with an exponential depreciation and depreciating parameter λ. With this, our boundary conditions then are given by:

$$
\begin{aligned}
V(R,C) &= \frac{R}{\lambda+\delta_R} - C \\
\frac{\partial V(R,C)}{\partial R} &= \frac{\partial B(R,\lambda)}{\partial R} = \frac{1}{\lambda+\delta_R} \\
\frac{\partial V(R,C)}{\partial C} &= -1
\end{aligned}
\tag{A.26}
$$

The partial differential equation A.18 along with the boundary conditions in equation A.26 yield a solution for $V(R,C)$. Given the homogeneity of the problem, if I double the values of R and I the decision is not affected by the scaling. Therefore the optimal decision depends on the ratio between R and C, $r = R/C$.

$$
V(R,C) = Cv(\frac{R}{C}) = Cv(r) \tag{A.27}
$$

where v is the function to be determined. Using the function described in A.27 and differentiation I obtain:

$$
\begin{aligned}
\frac{\partial V(R,C)}{\partial R} &= v'(r) \\
\frac{\partial V(R,C)}{\partial C} &= v(r) - rv'(r) \\
\frac{\partial^2 V(R,C)}{\partial R^2} &= \frac{v''(r)}{C} \\
\frac{\partial^2 V(R,C)}{\partial C^2} &= r^2\frac{v''(r)}{C} \\
\frac{\partial^2 V(R,C)}{\partial R \partial C} &= -r\frac{v''(r)}{C}
\end{aligned}
\tag{A.28}
$$

Now if I replace equations A.28 into A.18 I get the following differential equation.

$$
\begin{aligned}
0 &= \frac{1}{2}[\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 + \frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 + 2\frac{\partial^2 V}{\partial C \partial R}CR\sigma_C\sigma_R\rho_{rc}] + (r_f - \delta_R)\frac{\partial V}{\partial R}R + \\
&+ (r_f - \delta_C)\frac{\partial V}{\partial C}C - r_f V \\
0 &= \frac{1}{2}[\frac{v''(r)}{C}R^2\sigma_R^2 + r^2\frac{v''(r)}{C}C^2\sigma_C^2 - 2r\frac{v''(r)}{C}CR\sigma_C\sigma_R\rho_{rc}] + \\
&+ (r_f - \delta_R)v'(r)R + (r_f - \delta_C)(v(r) - rv'(r))C - r_f Cv(r) \quad divide \quad by \quad C \\
0 &= \frac{1}{2}[v''(r)r^2\sigma_R^2 + r^2v''(r)\sigma_C^2 - 2r^2v''(r)\sigma_C\sigma_R\rho_{rc}] + (r_f - \delta_R)v'(r)r + \\
&+ (r_f - \delta_C)(v(r) - rv'(r)) - r_f v(r) \\
0 &= \frac{1}{2}[\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}]v''(r)r^2 + (r_f - \delta_R)v'(r)r + (r_f - \delta_C)(v(r) - rv'(r)) - \\
&- r_f v(r) \\
0 &= \frac{1}{2}[\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}]v''(r)r^2 + (\delta_C - \delta_R)v'(r)r - \delta_C v(r)
\end{aligned}
\tag{A.29}
$$

Now by replacing A.28 into the boundary conditions described in equations A.19 and A.20 I get the following conditions:

Value Matching condition

$$
\begin{aligned}
V(R, C) &= B(R, \lambda) - C \\
Cv(r) &= \frac{R}{\lambda + \delta_R} - C \\
v(r) &= \frac{r}{\lambda + \delta_R} - 1
\end{aligned}
\tag{A.30}
$$

Smooth pasting with respect R

$$
\begin{aligned}
\frac{\partial V(R, C)}{\partial R} &= \frac{\partial B(R, \lambda)}{\partial R} \\
v'(r) &= \frac{1}{\lambda + \delta_R}
\end{aligned}
\tag{A.31}
$$

Smooth pasting with respect C

$$
\begin{aligned}
\frac{\partial V(R,C)}{\partial C} &= -1 \\
v'(r) &= -1
\end{aligned}
\tag{A.32}
$$

To solve the differential equation, I use the boundary conditions plus the condition that as $r \to 0$ the option $v(r) \to 0$. Equation A.29 is a homogeneous linear equation of second order, so its solution is a combination of linearly independent solutions:

$$
v(r) = A_1 r^{\beta_1} + A_2 r^{\beta_2} \tag{A.33}
$$

With this the quadratic fundamental equation is then given by:

$$
\begin{aligned}
Q &= \frac{1}{2}[\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}]v''(r)r^2 + (\delta_C - \delta_R)v'(r)r - \delta_C v(r) \\
Q &= \frac{1}{2}[\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}]A_1(\beta_1 - 1)\beta_1 r^{\beta_1 - 2}r^2 + (\delta_C - \delta_R)\beta_1 A_1 r^{\beta_1 - 1}r - \delta_C A_1 r^{\beta_1} \\
Q &= \frac{1}{2}[\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}](\beta_1 - 1)\beta_1 + (\delta_C - \delta_R)\beta_1 - \delta_C = 0 \\
Q &= \frac{1}{2}(\beta_1 - 1)\beta_1 + \frac{\delta_C - \delta_R}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}}\beta_1 - \frac{\delta_C}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}} = 0
\end{aligned}
\tag{A.34}
$$

The roots to the quadratic equation in A.34 are given by:

$$
\begin{aligned}
\beta_{11} = \frac{1}{2} - \frac{\delta_C - \delta_R}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}} + \\
+ \sqrt[2]{\left[\left(\frac{\delta_C - \delta_R}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}}\right) - \frac{1}{2}\right]^2 + 2\frac{\delta_C}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}}} \\
\beta_{12} = \frac{1}{2} - \frac{\delta_C - \delta_R}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}} - \\
- \sqrt[2]{\left[\left(\frac{\delta_C - \delta_R}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}}\right) - \frac{1}{2}\right]^2 + 2\frac{\delta_C}{\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc}}}
\end{aligned}
\tag{A.35}
$$

with this the only thing remaining to estimate for $v(r)$ is A_1[1], which can be derived by using the boundary conditions:

$$
\begin{aligned}
v'(r) &= \frac{1}{\lambda + \delta_R} \\
A_1\beta_1 r^{\beta_1 - 1} &= \frac{1}{\lambda + \delta_R} \\
A_1 r^{\beta_1} &= \frac{r}{(\lambda + \delta_R)\beta_1} \\
v(r) &= \frac{r}{(\lambda + \delta_R)\beta_1} \\
\frac{r^*}{\lambda + \delta_R} - 1 &= \frac{r^*}{(\lambda + \delta_R)\beta_1} \\
\frac{r^*}{\lambda + \delta_R}(1 - \frac{1}{\beta_1}) &= 1 \\
r^* &= (\lambda + \delta_R)(\frac{\beta_1}{\beta_1 - 1})
\end{aligned}
\tag{A.36}
$$

Now let's replace the optimal r^* to obtain the value of A_1.

[1]Given the condition that as r approaches zero the value of v also approaches zero and since the second root is negative the only plausible value for A_2 is zero.

$$
\begin{aligned}
v(r^*) &= \frac{r^*}{\lambda+\delta_R} - 1 \\
A_1 r^{*\beta_1} &= \frac{r^*}{\lambda+\delta_R} - 1 \\
A_1[(\lambda+\delta_R)(\frac{\beta_1}{\beta_1-1})]^{\beta_1} &= \frac{(\lambda+\delta_R)(\frac{\beta_1}{\beta_1-1})}{\lambda+\delta_R} - 1 \\
A_1 &= [\frac{\beta_1}{\beta_1-1} - 1][\frac{\beta_1-1}{\beta_1(\lambda+\delta_R)}]^{\beta_1} \\
A_1 &= [\frac{1}{\beta_1-1}][\frac{\beta_1-1}{\beta_1(\lambda+\delta_R)}]^{\beta_1} \\
A_1 &= \frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}
\end{aligned}
\tag{A.37}
$$

by replacing equations A.37 and A.35 into equation A.33 I obtain the solution such that:

$$
\begin{aligned}
V(R,C) &= Cv(r) \\
V(R,C) &= CA_1 r^{\beta_1} \\
V(R,C) &= C\frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}[\frac{R}{C}]^{\beta_1}
\end{aligned}
\tag{A.38}
$$

So far I have three values that are of interest: the sell option, O, which depends on the second highest best use rent; the option to keep a depreciated building and maintain the option to invest in CAPEX depicted by $V(R,C)$, and finally, I have the value of the building in operation given a depreciation rate represented by B(R).

The value of $V(R,C)$ also follows a geometric Brownian motion that I will use to estimate the price of the option to hold both investment opportunities. That is, the investment in a depreciating building or the investment in the outside option.

To estimate the stochastic process $V(R,C)$ follows, I use Ito's lemma.

$$dV = [\frac{\partial V}{\partial t} + \frac{\partial V}{\partial R}\alpha R + \frac{\partial V}{\partial C}\theta C + \frac{1}{2}\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 + \frac{1}{2}\frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 + \\ + 2\frac{\partial^2 V}{\partial C \partial R}CR\sigma_C\sigma_R\rho_{rc}]dt + \sigma_C C\frac{\partial V}{\partial C}dz_C + \sigma_R R\frac{\partial V}{\partial R}dz_R \tag{A.39}$$

Using the value of the option estimated by equation A.38, $V(R,C)$, I estimate the value of each derivative.

$$
\begin{aligned}
\frac{\partial V}{\partial t} &= 0 \\
\frac{\partial V}{\partial R}\alpha R &= \frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}\beta_1[\frac{R}{C}]^{\beta_1-1}\alpha R \\
&= V\beta_1\alpha \\
\frac{\partial V}{\partial C}\theta C &= \frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}(1-\beta_1)[\frac{R}{C}]^{\beta_1}\theta C \\
&= V(1-\beta_1)\theta \\
\frac{\partial^2 V}{\partial R^2}R^2\sigma_R^2 &= C\frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}\beta_1(\beta_1-1)[\frac{R}{C}]^{\beta_1}\sigma_R^2 \\
&= V\beta_1(\beta_1-1)\sigma_R^2 \\
\frac{\partial^2 V}{\partial C^2}C^2\sigma_C^2 &= C\frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}\beta_1(\beta_1-1)[\frac{R}{C}]^{\beta_1}\sigma_C^2 \\
&= V\beta_1(\beta_1-1)\sigma_C^2 \\
\frac{\partial^2 V}{\partial C\partial R}CR\sigma_C\sigma_R\rho_{rc} &= \frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}\beta_1(1-\beta_1)R^{\beta_1-1}C^{-\beta_1}CR\sigma_C\sigma_R\rho_{rc} \\
&= V\beta_1(1-\beta_1)\sigma_C\sigma_R\rho_{rc} \\
\frac{\partial V}{\partial C}\sigma_C C dz_C &= \frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}C^{-\beta_1}R^{\beta_1}(1-\beta_1)\sigma_C C dz_C \\
&= V(1-\beta_1)\sigma_C dz_C \\
\frac{\partial V}{\partial R}\sigma_R R dz_R &= \frac{(\beta_1-1)^{\beta_1-1}}{\beta_1(\lambda+\delta_R)^{\beta_1}}C^{1-\beta_1}R^{\beta_1-1}\beta_1\sigma_R R dz_R \\
&= V\beta_1\sigma_R dz_R
\end{aligned}
\tag{A.40}
$$

Now by replacing the value of the derivatives described in equations A.40 into equation A.39 that described the value of dV we get the following:

$$
\begin{aligned}
dV &= V\beta_1\alpha + V(1-\beta_1)\theta + \frac{1}{2}(V\beta_1(\beta_1-1)\sigma_R^2 + V\beta_1(\beta_1-1)\sigma_C^2 - \\
&\quad - 2V\beta_1(\beta_1-1)\sigma_C\sigma_R\rho_{rc})dt + V\beta_1\sigma_R dz_R + V(1-\beta_1)\sigma_C dz_C \\
\frac{dV}{V} &= [\beta_1\alpha + (1-\beta_1)\theta + \frac{1}{2}\beta_1(\beta_1-1)(\sigma_R^2 + \sigma_C^2 - 2\sigma_C\sigma_R\rho_{rc})]dt + \\
&\quad + \beta_1\sigma_R dz_R + (1-\beta_1)\sigma_C dz_C \\
\frac{dV}{V} &= \gamma_V dt + \sigma_V dz_V
\end{aligned}
\tag{A.41}
$$

The result from equation A.41 indicates that the value of the option to invest in the building follows a geometric Brownian motion.

Appendix B
Derivations for Reputation Model

B.0.1 Proofs and Derivations

B.0.1.1 Profits

Let's define π_M as the present value of future cash flows for a firm that maintains quality.

$$\begin{aligned}
\pi_M &= (p - c_M)\frac{(1 - v_M)}{(1 + r)} + (p - c_M)\frac{(1 - v_M)^2}{(1 + r)^2} + \text{...multiply both sides by } \frac{(1 + r)}{(1 - v_M)} \\
\pi_M\frac{(1 + r)}{(1 - v_M)} &= (p - c_M) + (p - c_M)\frac{(1 - v_M)}{(1 + r)} + \text{...where second term and ... represent } PV \\
\pi_M(1 + r) &= +PV_M(1 - v_M) \\
\pi_M &= (p - c_M)\frac{(1 - v_M)}{(r + v_M)}
\end{aligned} \tag{B.1}$$

The same could be done for the lower quality firms but with subscript L. Now π_M and π_L could be expressed as:

$$\begin{aligned}
\pi_M &= (p - c_M) + (p - c_M)\frac{(1 - v_M)}{(r + v_M)} \\
\pi_L &= (p - c_L) + (p - c_L)\frac{(1 - v_L)}{(r + v_L)}
\end{aligned} \tag{B.2}$$

B.0.1.2 Price for high quality

The price that guaranties the firm will maintain its quality, p^*, is the one that makes the condition $\pi_M > \pi_L$ possible. By rearranging terms we get the following:

$$\begin{aligned}
\pi_M &> \pi_L \\
(p - c_M) + (p - c_M)\frac{(1 - v_M)}{(r + v_M)} &> (p - c_L) + (p - c_L)\frac{(1 - v_L)}{(r + v_L)} \\
(p - c_M)\frac{(1 - v_M)}{(r + v_M)} - (p - c_L)\frac{(1 - v_L)}{(r + v_L)} &> (c_M - c_L) \\
p\left[\frac{(1 - v_M)}{(r + v_M)} - \frac{(1 - v_L)}{(r + v_L)}\right] &> (c_M - c_L) + \left[c_M\frac{(1 - v_M)}{(r + v_M)} - c_L\frac{(1 - v_L)}{(r + v_L)}\right] \\
p^* &> \frac{(c_M - c_L) + \left[c_M\frac{(1 - v_M)}{(r + v_M)} - c_L\frac{(1 - v_L)}{(r + v_L)}\right]}{\left[\frac{(1 - v_M)}{(r + v_M)} - \frac{(1 - v_L)}{(r + v_L)}\right]}
\end{aligned} \tag{B.3}$$

B.0.1.3 Proposition 1

To prove proposition 1, I estimate the limit of v_M when it tends to $-\infty$ (increase in the amount of customers to infinity) and compare this to $v_M > 0$.

$$\lim_{v_M \longrightarrow -\infty} \frac{(c_M - c_L) + \left[c_M\frac{(1 - v_M)}{(r + v_M)} - c_L\frac{(1 - v_L)}{(r + v_L)}\right]}{\left[\frac{(1 - v_M)}{(r + v_M)} - \frac{(1 - v_L)}{(r + v_L)}\right]} \tag{B.4}$$

From equation B.4 we observe the need to estimate the limit only to the part involving v_M as the rest remains constant. The case with v_M tending to 0 (no consumer leaving the firm) is straightforward and can be solve just by replacing 0 in the equation. The limit of interest is what happens if the amount of consumers increases to ∞. This last result could be interpreted as if the firm that maintain quality captures the entire market and the resulting price is a bound.

$$
\begin{aligned}
\lim_{v_M \longrightarrow 0} p^* &= \frac{(c_M - c_L) + \left[c_M \frac{1}{r} - c_L \frac{(1-v_L)}{(r+v_L)} \right]}{\left[\frac{1}{r} - \frac{(1-v_L)}{(r+v_L)} \right]} \\
\lim_{v_M \longrightarrow -\infty} p^* &= \lim_{v_M \longrightarrow -\infty} \frac{\left(c_M - c_L - c_L \frac{(1-v_L)}{r+v_L} \right)}{\left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)} \right]} + \frac{c_M}{\left[1 - \frac{\frac{(1-v_L)}{(r+v_L)}}{\frac{(1-v_M)}{(r+v_M)}} \right]}
\end{aligned} \tag{B.5}
$$

In order to solve the equation B.5 limit for $-\infty$ let's focus only on the fraction that contains v_M.

$$
\begin{aligned}
\lim_{v_M \longrightarrow -\infty} \frac{(1-v_M)}{(r+v_M)} &= \lim_{v_M \longrightarrow -\infty} \frac{(1-v_M)}{(r+1-(1-v_M))} \\
&= \lim_{v_M \longrightarrow -\infty} \frac{1}{(\frac{(r+1)}{(1-v_M)} - 1)} \\
&= -1
\end{aligned} \tag{B.6}
$$

Replacing B.6 into B.5 gives the price at the limit:

$$
\begin{aligned}
\lim_{v_M \longrightarrow -\infty} p^* &= -\frac{c_M}{\left[1+\dfrac{(1-v_L)}{(r+v_L)}\right]} + \frac{\left(-c_L(1+(\dfrac{(1-v_L)}{r+v_L}))\right)}{\left[-1-\dfrac{(1-v_L)}{(r+v_L)}\right]} + \frac{c_M}{\left[1+\dfrac{(1-v_L)}{(r+v_L)}\right]} \\
&= \frac{-c_L\left[1+\dfrac{(1-v_L)}{(r+v_L)}\right]}{-\left[1+\dfrac{(1-v_L)}{(r+v_L)}\right]} \\
&= c_L
\end{aligned}
\tag{B.7}
$$

Now the lower bound of the price is c_L. Let's look at the case where there is no loss of customers, $v_M = 0$, when the firm maintain its quality. To prove that the price is lower when v_M is lower, we prove it by contradiction, and let's suppose that the price when v_M is greater than zero is lower than when $v_M = 0$.

$$\frac{(c_M - c_L) + \left[c_M \frac{1}{r} - c_L \frac{(1-v_L)}{(r+v_L)}\right]}{\left[\frac{1}{r} - \frac{(1-v_L)}{(r+v_L)}\right]} > \frac{(c_M - c_L) + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{(1-v_L)}{(r+v_L)}\right]}{\left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right]}$$

$$\begin{aligned}
[(c_M - c_L)] \left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right] + &> [(c_M - c_L)] \left[\frac{1}{r} - \frac{(1-v_L)}{(r+v_L)}\right] + \\
+ \left[c_M \frac{1}{r} - c_L \frac{(1-v_L)}{(r+v_L)}\right] \cdot &\quad + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{(1-v_L)}{(r+v_L)}\right] \cdot \\
\cdot \left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right] &\quad \cdot \left[\frac{1}{r} - \frac{(1-v_L)}{(r+v_L)}\right]
\end{aligned}$$

$$\begin{aligned}
\left[c_M + c_M \frac{1}{r}\right] \left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right] - &> \left[c_M + c_M \frac{(1-v_M)}{(r+v_M)}\right] \left[\frac{1}{r} - \frac{(1-v_L)}{(r+v_L)}\right] - \\
- \left[c_L + c_L \frac{(1-v_L)}{(r+v_L)}\right] \left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right] &\quad - \left[c_L + c_L \frac{(1-v_L)}{(r+v_L)}\right] \left[\frac{1}{r} - \frac{(1-v_L)}{(r+v_L)}\right]
\end{aligned}$$

$$\begin{aligned}
c_M \frac{1-v_M}{r+vg} - c_M \frac{(1-v_L)}{r(r+vb)} - &> c_M \frac{1}{r} - c_M \frac{(1-v_M)(1-v_L)}{(r+v_M)(r+v_L)} - \\
- \left[c_L \frac{1-v_M}{r+v_M} + c_L \frac{(1-v_M)(1-v_L)}{(r+v_M)(r+v_L)}\right] &\quad - \left[c_L \frac{1}{r} + c_L \frac{(1-v_L)}{r(r+v_L)}\right]
\end{aligned}$$

$$\begin{aligned}
c_L \left[+\frac{1}{r} - \frac{1-v_M}{r+v_M} + \frac{(1-v_L)}{r(r+v_L)}\right] - &> c_M \left[+\frac{1}{r} - \frac{1-v_M}{r+vg} + \frac{(1-v_L)}{r(r+vb)}\right] - \\
-c_L \left[\frac{(1-v_M)(1-v_L)}{(r+v_M)(r+v_L)}\right] &\quad - c_M \left[\frac{(1-v_M)(1-v_L)}{(r+v_M)(r+v_L)}\right]
\end{aligned}$$

$$c_L > c_M \tag{B.8}$$

This last result is a contradiction as c_L by definition is smaller than c_M. The marginal cost of maintaining quality is higher than that of lower quality. Also v_M and v_L are positive numbers therefore the equation multiplying c_M and c_L result in a positive number, so there is no change in sign when simplifying. v_M came from the assumption that in this case the loss of clients was positive and I compared that to no loss of customer. The contradiction therefore shows that the price decreases as the customers set low v_M for firms that maintain quality.

B.0.1.4 Proposition 2

Proposition two introduces the question of what happens if the customers set low V_L. Although the steps are fairly similar to the proof in section B.0.1.3 in this case the question of interest is what happens to price when v_L goes to zero, or even worse, becomes a negative number increasing in firms with low quality. Having negative numbers in v_M and v_L implies that the customer base is increasing rather than decreasing. To prove that lower rates of v_L (meaning close to zero or even negative) result in higher prices, first I show what happen to the price when v_L is equal to 0 and then estimate the limit when v_L goes to $-\infty$.

$$\begin{aligned}
\lim_{v_L \longrightarrow 0} p^* &= \lim_{v_L \longrightarrow 0} \frac{(c_M - c_L) + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{1}{r}\right]}{\left[\frac{(1-v_M)}{(r+v_M)} - \frac{1}{r)}\right]} \\
\lim_{v_L \longrightarrow -\infty} p^* &= \lim_{v_L \longrightarrow -\infty} \frac{(c_M - c_L) + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{(1-v_L)}{(r+v_L)}\right]}{\left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right]}
\end{aligned} \tag{B.9}$$

In order to solve the equation B.9 limit for $-\infty$ let's focus only in the fraction that contains v_L.

$$\begin{aligned}
\lim_{v_L \longrightarrow -\infty} \frac{(1-v_L)}{(r+v_L)} &= \lim_{v_L \longrightarrow -\infty} \frac{(1-v_L)}{(r+1-(1-v_L))} \\
&= \lim_{v_L \longrightarrow -\infty} \frac{1}{(\frac{(r+1)}{(1-v_L)}-1)} \\
&= -1
\end{aligned} \tag{B.10}$$

Replacing B.10 into B.9 gives the price at the limit:

$$\begin{aligned}
\lim_{v_L \longrightarrow -\infty} p^* &= \lim_{v_L \longrightarrow -\infty} \frac{(c_M - c_L) + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L(-1)\right]}{\left[\frac{(1-v_M)}{(r+v_M)} + 1\right]} \\
&= \frac{-c_M \left[1 + \frac{(1-v_M)}{(r+v_M)}\right]}{-\left[1 + \frac{(1-v_M)}{(r+v_M)}\right]} \\
&= c_M
\end{aligned} \tag{B.11}$$

The price in this case increases to marginal cost of production c_M. To show that lower values of v_L result in higher prices at which the firm maintain quality, let's prove by contradiction and assume it actually decreases prices.

$$\frac{(c_M - c_L) + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{1}{r}\right]}{\left[\frac{(1-v_M)}{(r+v_M)} - \frac{1}{r}\right]} < \frac{(c_M - c_L) + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{(1-v_L)}{(r+v_L)}\right]}{\left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right]}$$

$$\begin{aligned}
[(c_M - c_L)] \left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right] + &< (c_M - c_L) \left[\frac{(1-v_M)}{(r+v_M)} - \frac{1}{r}\right] + \\
+ \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{1}{r}\right] \cdot \quad & \quad + \left[c_M \frac{(1-v_M)}{(r+v_M)} - c_L \frac{(1-v_L)}{(r+v_L)}\right] \cdot \\
\cdot \left[\frac{(1-v_M)}{(r+v_M)} - \frac{(1-v_L)}{(r+v_L)}\right] \quad & \quad \cdot \left[\frac{(1-v_M)}{(r+v_M)} - \frac{1}{r}\right]
\end{aligned}$$

$$\begin{aligned}
-c_M \frac{1-v_L}{r+vb} + c_L \frac{(1-v_L)}{(r+vb)} - &< -c_M \frac{1}{r} - c_M \frac{(1-v_M)}{(r+v_M)r} + \\
-\left[c_L \frac{1-v_M}{r(r+v_M)} + c_M \frac{(1-v_M)(1-v_L)}{(r+v_M)(r+v_L)}\right] \quad & \quad + \left[c_L \frac{1}{r} - c_L \frac{(1-v_L)(1-v_M)}{(r+v_L)(r+v_M)}\right]
\end{aligned}$$

$$\begin{aligned}
c_M \left[\frac{1}{r} - \frac{1-v_L}{r+v_L} + \frac{(1-v_M)}{r(r+v_M)}\right] - &< c_L \left[\frac{1}{r} - \frac{1-v_L}{r+vb} + \frac{(1-v_M)}{r(r+v_M)}\right] - \\
-c_M \left[\frac{(1-v_M)(1-v_L)}{(r+v_M)(r+v_L)}\right] \quad & \quad - c_L \left[\frac{(1-v_M)(1-v_L)}{(r+v_M)(r+v_L)}\right]
\end{aligned}$$

$$c_M < c_L \tag{B.12}$$

Again the assumption that prices are lower when there is no departure of customers from lower quality firms, $v_L = 0$, results in a contradiction. Therefore, prices at which firms start investing in quality decreases when customers treat firms producing at a lower quality.

Bibliography

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Ambrose, B. and Sanders, A. (2003). Commercial mortgage-backed securities: Prepayment and default. *Journal of Real Estate Finance and Economics*, 26:179 – 196.

Ambrose, B. and Steiner, E. (2017). Capital expenditures, asset dispositions, and the real estate cycle. *Working Paper*.

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

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30.

Archer, W., Elmer, P., Harrison, D., and Ling, D. (2002). Determinants of multifamily mortgage default. *Real Estate Economics*, 30(3).

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304):933–942.

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

Brown, J., Broderick, A. J., and Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, 21(3):2–20.

Capozza, D. R. and Helsley, R. W. (1990). The stochastic city. *Journal of Urban Economics*, (2):187 – 203.

Case, K., Shiller, R. J., et al. (1989). The efficiency of the market for single-family homes. *The American Economic Review*, 79(1):125–137.

Chauvet, M., Gabriel, S., and Lutz, C. (2016). Mortgage default risk: New evidence from internet search queries. *Journal of Urban Economics*, 96:91 – 111.

Chetty, R. (2007). Interest rates, irreversibility, and backward-bending investment. *The Review of Economic Studies*, 74(1):67–91.

Childs, P. D., Ott, S. H., and Riddiough, T. J. (2004). Effects of noise on optimal exercise decisions: The case of risky debt secured by renewable lease income. *The Journal of Real Estate Finance and Economics*, 28(2):109–121.

Cho, H., Ciochetti, B., and Shilling, J. (2013). Are commercial mortgage defaults affected by tax considerations?. *Journal of Real Estate Finance and Economics*, 46(1):1 – 23.

Ciochetti, B., Deng, Y., Gao, B., and Yao, R. (2002). The termination of commercial mortgage contracts through prepayment and default: A proportional hazard approach with competing risks. *Real Estate Economics*, 30(4).

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

Das, P., Ziobrowski, A., and Coulson, N. E. (2015). Online information search, market fundamentals and apartment real estate. *The Journal of Real Estate Finance and Economics*, 51(4):480–502.

Dixit, A. K. and Pindyck, R. S. (1994). *Investment under Uncertainty*. Princeton University Press.

Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics*, 80(4):549–560.

Episcopos, A., Pericli, A., and Hu, J. (1998). Commercial mortgage default: a comparison of logit with radial basis function networks. *J. Real Estate Finance and Econ*, 17:163 – 178.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509.

Gárate, S. (2018). Information effect of online reviews on investment in the real estate industry. *Working Paper*.

Garmaise, M. J. and Moskowitz, T. J. (2003). Confronting information asymmetries: Evidence from real estate markets. *The Review of Financial Studies*, 17(2):405–437.

Geltner, D., Riddiough, T., and Stojanovic, S. (1996). Insights on the effect of land use choice: The perpetual option on the best of two underlying assets. *Journal of Urban Economics*, (1):20–50.

Grovenstein, R. (2005). Commercial mortgage underwriting: how well do lenders manage the risks?. *Journal of Housing Economics*, 14:355 – 383.

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

Jones, T. and Sirmans, G. S. (2016). Commercial mortgage termination and pricing: Theory and evidence from 1970 to 2015. *Journal of Real Estate Literature*, 24(2):213–249.

Kau, J. B., Keenan, D. C., Muller, W. J., and Epperson, J. F. (1992). A generalized valuation model for fixed-rate residential mortgages. *Journal of money, credit and banking*, 24(3):279–299.

Kuang, C. (2017). Does quality matter in local consumption amenities? an empirical investigation with yelp. *Journal of Urban Economics*, 100:1 – 18.

Lemmon, M. and Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, 19(4):1499–1529.

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

Lewis, G. (2011). Asymmetric information, adverse selection and online disclosure: The case of ebay motors. *The American Economic Review*, 101(4):1535–1546.

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

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

Magill, M. and Quinzii, M. (2002). *Theory of incomplete markets*, volume 1. Mit press.

Marcus, A. J. and Modest, D. M. (1984). Futures markets and production decisions. *Journal of Political Economy*, (3):409–426.

Mauer, D. C. and Ott, S. H. (1995). Investment under uncertainty: The case of replacement investment decisions. *Journal of Financial and Quantitative Analysis*, (04):581–605.

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

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94(4):796–821.

Mueller, D. C. and Reardon, E. A. (1993). Rates of return on corporate investment. *Southern Economic Journal*, pages 430–453.

Peng, L. and Thibodeau, T. G. (2018). Interest rate and investment under uncertainty: Evidence from commercial real estate capital expenditures. *Working Paper*.

Pope, D. G. and Pope, J. C. (2015). When walmart comes to town: Always low housing prices? always? *Journal of Urban Economics*, 87:1 – 13.

Rogerson, W. P. (1983). Reputation and product quality. *The Bell Journal of Economics*, 14(2):508–516.

Samuelson, P. A. and McKean, H. P. (1965). Rational theory of warrant pricing. *IMR; Industrial Management Review (pre-1986)*, (2):13.

Schwartz, E. S. and Torous, W. N. (1992). Prepayment, default, and the valuation of mortgage pass-through securities. *Journal of Business*, pages 221–239.

Seslen, T. and Wheaton, W. (2010). Contemporaneous loan stress and termination risk in the cmbs pool: how 'ruthless' is default?. *Real Estate Economics*, 38(2):225 – 255.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations*. *The Quarterly Journal of Economics*, 98(4):659–679.

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

Titman, S. (1985). Urban land prices under uncertainty. *The American Economic Review*, 75(3):505–514.

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

Williams, J. T. (1991). Real estate development as an option. *The Journal of Real Estate Finance and Economics*, (2):191–208.

Wu, L. and Brynjolfsson, E. (2015). The future of prediction: How google searches foreshadow housing prices and sales. In *Economic analysis of the digital economy*, pages 89–118. University of Chicago Press.

Yildirim, Y. (2008). Estimating default probabilities of cmbs loans with clustering and heavy censoring. *The Journal of Real Estate Finance and Economics*, 37(2):93 – 111.

Zhou, B., Fine, J., Latouche, A., and Labopin, M. (2012). Competing risks regression for clustered data. *Biostatistics*, 13(3):371–383.

Vita

Sergio Rodolfo Gárate Alvarez

Education

Pennsylvania State University	University Park, PA
Smeal College of Business	
Ph.D. Business Administration	2018
Marquette University	Milwaukee, WI
College of Business	
Master Applied Economics	2013
Universidad de Chile	Santiago, Chile
School of Physics and Mathematics	
Industrial Engineer Degree	2003

Professional Activities

Academic Appointments:

Pennsylvania State University	2014 - 2018
Instructor	2017
Teaching Assistant	2014 - 2018
Research Assistant	2015 - 2017

Referee:

Real Estate Economics	2015 -Present

Industry Professional Positions:

Freddie Mac • Consultant for the Chief Economist Office	2016
Havi Global Solution • Senior Analyst	2013 - 2014
SA Gestiona • Real Estate Manager	2009 - 2012
Walmart-Chile • Real Estate Manager	2006 - 2009
Repsol-YPF • Financial Analyst	2003 - 2006
Citibank • Financial Planner	2002

Publication

Measuring the Impact of Agglomeration on Productivity: Evidence from Chilean Retailers, with Anthony Pennington-Cross. Urban Studies, 51(8), pp.1653-1671, (2014).

Awards

Best manuscript in housing at the *American Real Estate Society* meeting, (2017)

The Jeanne and Charles Rider Graduate Fellowship, *Smeal College of Business at PSU*, (2017)

Best manuscript in commercial real estate at the *American Real Estate Society* meeting, (2015)