Hedonic pricing and the sharing economy: How profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville

Master's Thesis submitted in fulfillment of the Degree

Master of Business Administration (MBA)

Submitted to Dr. Ulrich Gunter

Baldwin Tong

1702505

Vienna, 22.04.2019

Affidavit

I hereby affirm that this Master's Thesis represents my own written work and that I have used no sources and aids other than those indicated. All passages quoted from publications or paraphrased from these sources are properly cited and attributed.

The thesis was not submitted in the same or in a substantially similar version, not even partially, to another examination board and was not published elsewhere.

22.04.2019

Date

ABSTRACT

The sharing economy has allowed people from all over the world to more effectively utilize their assets if they choose to do so. Owners or controllers of assets in the sharing economy are free to set any price they want. However, from an economics perspective, they should only set sensible prices for which there is still demand.

This paper examines how various characteristics of an Airbnb listing (size, number of photos, ratings, host responsiveness, superhost status, distance from city center, etc.) affect the prices of accommodation and determines which factors strongly affect price using a quantitative approach.

A hedonic pricing model was developed and applied to data from the cities of Barcelona, Madrid, and Seville to determine how the different characteristics of an Airbnb listing affect the price of accommodation in these three Spanish cities. The results indicate that characteristics which are indicative of the size of the accommodation have the strongest positive influence on price while the number of reviews and distance from the city center have the strongest negative influence on price.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

List of Abbreviations

BLUE	Best Linear Unbiased Estimator
B2B	Business-to-business
B2C	Business-to-consumer
C2C	Consumer-to-consumer
EU	European Union
GLM	General linear model
GWR	Geographically weighted regression
ID	Identification
OLS	Ordinary least squares
WLS	Weighted least squares
URL	Uniform Resource Locator
USD	United States Dollar

1 THE SHARING ECONOMY

The sharing economy is the broad term given to the activity of individuals sharing, renting, or exchanging underutilized assets or resources in a peer-to-peer system for a fee (Zervas, Proserpio, & Byers, 2017). While this kind of activity is represented by other terms, this will be the term used throughout the thesis.

What is known as the sharing economy is also known as "collaborative consumption" (Felson & Spaeth, 1978; Belk, 2014) or "access-based consumption" (Bardhi & Eckhardt, 2012), and there is a considerable amount of debate as to whether the term "sharing economy" is in fact appropriate to describe the typical activities that fall under its umbrella (Murillo, Buckland, & Val, 2017). Nevertheless, regardless of what term is used, this type of economic activity has grown more and more popular over time due to technological advancements such as the rapid development of mobile technology.

The development of the internet has facilitated the ability of individuals to connect with other individuals. All forms of communication and activity have migrated online, including people's natural inclination to share, buy, and sell. As a result, internet-based platform businesses have been establiblished in a variety of different industries to facilitate these transactions. In addition to technological advancement, the popularity of sharing economy platforms are attributed to supply-side flexibility (Zervas et al., 2017). Individuals can choose when and where they would like to participate and are in no way obligated to offer a required minimum or maximum amount of services, allowing regular people to become freelancers or entrepreneurs.

In the sharing economy, it is typical for each company offering an online marketplace platform to be focused on one particular service. For example, Taskrabbit allows individuals to hire other individuals to help them with tasks such as furniture assembly, mounting and installation, or heavy lifting (TaskRabbit, Inc., 2018). Turo, previously known as RelayRides, offers peer-to-peer car rental (Turo Inc., 2018). Fon offers residential WiFi sharing (Fon Wireless Limited, 2018). It should be noted that the the sharing economy term is not exclusively confined to peer-to-peer or C2C activity but also includes business-to-consumer or B2C services such as Zipcar, a car sharing service, car2go, another car sharing service (Cho, et al., 2013), and B2B services (Demary, 2015). While the C2C and B2C sharing economy sectors are fairly well developed, business-to-business or B2B offerings such as FLEXE, an online marketplace offering warehouse space and services (FLEXE, Inc., 2018), are growing in popularity.

While there are many companies operating in many industries in the sharing economy space, the most successful and well-known sharing economy companies are the ones that operate in the transportation and accommodation sectors. Uber, a taxi service where individuals use their cars to provide rides to customers, is considered to be the leader of the sharing economy transportation companies. The company is targeting a 100 billion USD valuation for a planned IPO some time in 2019 and had revenues of 11.27 billion USD in 2018 but posted a loss of 1.85 billion USD in the same period (Rushe & Helmore, 2019). Airbnb, a platform where people can rent a room, an apartment, or a home, on a short-term basis, is the leader in the sharing economy accommodation sector and is profitable. The company has a valuation of 31 billion USD, and a 5.5 billion USD balance sheet according to Airbnb CEO Brian Chesky (Somerville, 2018).

1.1 History of Airbnb

Airbnb was founded in 2008 in San Francisco, California. The core idea for the company began when co-founders Brain Chesky and Joe Gebbia decided to rent air mattreses in their living room during a design conference when all local hotels were booked, turning their apartment into a bed and breakfast (Airbnb, Inc., 2019).

The company offers an online marketplace where individuals can become hosts and rent out their rooms, homes, or properties to other individuals using an internet-based platform. The platform gives hosts the ability to post and price their offerings and control the dates of when their accommodation is made available. Potential guests have the ability to search for accommodation in a city or country of their choice by date. The Airbnb platform then handles the booking and payment functions for both the host and the guest. It also has a messaging function in the event both parties need to communicate. The Airbnb platform can be accessed via internet browsers or mobile applications (iOS and Android).

According to the Airbnb website, there are currently over 6 million Airbnb listings worldwide covering 81,000 cities and over 191 countries. They have served 500 million guests since their founding in 2008 and have 31 offices globally (Airbnb, Inc., 2019). Airbnb earns revenue by charging a service fee to individuals on both sides of the platform: the host, and the guest. This service fee ranges from 9% to 15% of the total transaction. As mentioned in the previous section, Airbnb is valued at 31 billion USD (Somerville, 2018) and is estimated to have earned 3.5 billion USD in revenue in 2017 (Hook, 2018). Airbnb has since expanded its services to include the booking of tours, vacation packages, and restaurants.

1.2 Hedonic Pricing Model

The basic theory behind the hedonic pricing model was first mentioned in 1966 by Kelvin J. Lancaster, a mathematical economist, in his article entitled: "A New Approach to Consumer Theory". Lancaster (1966) proposed that the utility of a good was derived from its properties or characteristics and not directly from the good itself which was the accepted belief at the time.

The hedonic pricing model is a regression model typically used in the real estate industry to estimate the value of homes by relating its price to its characteristics (Wooldridge, 2012). It was first presented in 1974 by Sherwin Rosen, a labor economist and professor of economics at the University of Rochester. Hedonic pricing theory states that the total price of a product is the sum of the prices of the individual characteristics (hedonic prices) of that product (Rosen, 1974).

For example, the price of a house could be broken down into the prices of the characteristics of the house such as its size, number of bedrooms, number of bathrooms, etc.. While the application of the hedonic pricing model has been predominantly in the real estate sector, its use is not limited to pricing real estate. It has been also applied in the tourism industry to price accommodations such as hotel rooms (see Section 2.4), in the retail industry for things such as consumer packaged goods (Martínez-Garmendia, 2010), fish (McConnell & Strand, 2000), wine (Brentari, Levaggi, & Zuccolotto, 2011), and to assess the price of pollution (Bajari, Fruehwirth, Kim, & Timmins, 2012).

A more detailed description of the hedonic pricing model used in this thesis can be found in Section 4.2.

1.3 Monopolistic competition

As explained in Section 1.1, Airbnb hosts are free to set their own price and control the dates of their listings meaning that there are no entry or exit barriers. From a microeconomics perspective, all of these things combined point to a monopolistically competitive market.

Monopolistic competition is similar to a perfectly competitive market in that there are no entry or exit barriers and that there are many producers. What differentiates monopolistic competition from perfect competition is the fact that the products themselves are imperfect substitutes for one another and can be differentiated by the producers (Pindyck & Rubinfeld, 2018). Applying this theory to Airbnb, hosts are essentially producers and their ability to charge different prices on their listings and ultimate success is down to their ability to distinguish their accommodation from all the others since every host on Airbnb offers a product that is highly substitutable but not perfectly substitutable.

According to Pindyck & Rubinfeld (2018), the two key characteristics of a monopolistically competitive market are:

1. Firms compete by selling differentiated products that are highly substitutable but are not perfect substitutes.
2. There is free entry and exit of the market for firms.

An Airbnb host is a price setter in the market and due to the monopolistically competitive structure of the market they are operating in, they have a certain level of market power. In both the short and long-term, an Airbnb host faces a downward sloping demand curve since it is the sole producer of its own brand. And as a sole producer of its brand, an Airbnb host has a certain level of monopoly power. In the short-term, the host can charge a price above its average cost but over the long-term, as a result of an increase in competitors, price will equal average cost meaning a host will not earn a profit even though they have monopoly power (Pindyck & Rubinfeld, 2018)

1.4 Research aims and objectives

The aim of this paper is to highlight how profile characteristics in an Airbnb listing contribute to the final price of the listing and to which extent a relevant characteristic marginally contributes to the final price. To determine which characteristics contribute to price, a hedonic pricing model using multivariable linear regression will be applied to Airbnb listing data from three Spanish cities: Barcelona, Madrid, and Seville.

These cities were chosen due to their size and status as being one of the top five metropolitan areas in Spain, their distinct geographic locations within the country, and their popularity as tourist destinations. Barcelona and Madrid are both listed in Euromonitor International's list of Top 100 City Destinations of 2018 (Geerts, 2018) and Seville was named by Lonely Planet as the best city in the world to visit in 2018 (Roden, 2017). The country of Spain in general was chosen due to its importance as a tourist destination in the EU and worldwide. In 2017, Spain was the most common tourism destination in the EU according to Eurostat. The 305.9 million nights spent in tourist accommodation by non-residents represented 20% of the EU total and net receipts from travel amounted to 40.6 billion euros (Eurostat, 2018). Worldwide, Spain is the third most visited country with 81.8 million visitors in 2018 (Smith, 2018). The objective of performing an analysis of data from three different cities in Spain is to obtain results which could be considered applicable to the entire country.

1.5 Structure of thesis

Chapter 2 is a broad overview of the current literature concerning the sharing economy, pricing in the sharing economy, Airbnb, the effect of the sharing economy on the traditional accommodation industry, pricing in the traditional accommodation industry, Airbnb pricing, and hedonic pricing model applications in the sharing economy.

Chapter 3 provides a detailed description of the data set including descriptive statistic tables of the data used, a description of the preparation and treatment of the data as well as the procedures followed, and a description of the results of the preliminary regression analysis and statistical pre-testing.

Chapter 4 describes the methodology used for the thesis. It includes a detailed description of how variables were selected to be included in the hedonic pricing model, the building of the theoretical hedonic pricing model used to analyze the data set, and how the model is applied to the data using multivariable linear regression.

Chapter 5 outlines the results of the multivariable linear regression analysis using ordinary least squares (OLS) and weighted least squares (WLS) and includes a detailed discussion of the findings as well as a description of the robustness checks performed.

Chapter 6 marks the conclusion of the thesis and includes an overall summary, the thesis' contribution to knowledge, the implications of the findings for relevant stakeholders, the limitations of the thesis, and provides an outlook on potential future research using the methodology and findings from the thesis.

2 LITERATURE REVIEW

The literature review revolves around the topics introduced in the introduction section namely: the sharing economy, pricing in the sharing economy, Airbnb, pricing in the traditional accommodation industry, Airbnb pricing, and hedonic pricing model applications in the sharing economy.

2.1 The Sharing Economy

The sharing economy is an umbrella term that includes all forms of peer-to-peer, business-to-consumer, and business-to-business transactions involving the shared use of underutilized assets. The term was first used in 2008 by Lawrence Lessig, an attorney and professor of law, in his book *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. Lessig (2008) proposed the existence of two economies: commercial and sharing. The commercial economy is mandated by a market where services or goods have tangible economic value. The sharing economy exists outside of the exchange of money so is not measured by price.

Current literature about the sharing economy focuses on a variety of subjects including the use of the term “sharing economy” itself. One main area of focus is the analysis of the globalization of sharing economy platforms. Parente, Geleilate, & Rong (2018) developed a framework to guide future research of the sharing economy from a business ecosystem perspective while acknowledging that the success of these platforms are contingent on ecosystem engagement from all players.

Ravenelle (2017) looked at the socioeconomic effects of sharing economy economies and concluded that participation in the sharing economy left workers vulnerable to platform changes making them feel insecure as opposed to independent. Most workers also described themselves as simply seeking money as opposed to being entrepreneurial, counter to the message being communicated by sharing economy companies.

Gobble (2015), Erickson & Sørensen (2016), Munkøe (2017), Hou (2018), and Ganapati & Reddick (2018) all addressed the potential regulatory and economic implications of the sharing economy. Gobble (2015) wrote about how sharing economy companies have been operating without governmental permission and regulations are needed in order to create increased trust between users. Erickson & Sørensen (2016) discussed how understanding the social and economic motivations of participants is important for the development of regulations. Munkøe (2017) wrote about the challenges facing the regulation of sharing economy companies in the EU such as the status of workers, insurance issues, and tax avoidance. The author concluded that regulations are needed and the EU needs to strike a difficult balance in its regulation of the sharing economy. Vague rules would have no effect while strict regulations could impede innovation and the resulting economic rewards. Hou (2018) concluded that the sharing economy blurs the previously distinct line between businesses and consumers, and employers and employees making it difficult to incorporate into current regulations. Ganapati & Reddick (2018) concluded that the sharing economy may increase inequality since its benefits go primarily to those who own assets and that sharing economy companies should be subject to data transparency regulations for research and regulation purposes.

There is also a considerable amount of literature which is critical towards the sharing economy. Martin (2016) analyzed the discourse surrounding the sharing economy and found two distinct groups: one positive, where the sharing economy is an economic opportunity which is more sustainable, and one negative, where the sharing economy is the operation of unregulated marketplaces which allow for the reinforcement of the neoliberal paradigm. Oskam & Boswijk (2016) stated that sharing economy companies are digital platforms which coordinate supply and demand for products and services and concluded that this is not sharing—it is a "digitally enabled expansion of the market economy." They added that the term "sharing economy" itself should be eliminated since it is contradictory in nature and confuses the academic discourse. Ravenelle (2017) simply stated: "Sharing is usually free–a view shared by many workers interviewed for this research." While this statement seems obvious, what it reveals is that the term "sharing economy" seems to hold an element of cognitive dissonance among those who hear it as it does not accurately encapsulate and properly define the activities of the participants. Although it may be implied that sharing is usually free, the definition of the word "sharing" does not explicitly state that (Merriam-Webster, 2019). It is also of interest to note that Ravenelle also found that the sharing economy marketplaces which adhered to the commonly understood ethos of sharing were either defunct or had been superceded by for-profit companies.

While these papers provide interesting insight and raise interesting questions concerning the sharing economy and its implications for societies on the local, national, and international level, these areas of interest lie outside the scope of this thesis. As the objective of this thesis is an analysis of Airbnb prices in Spain, specifically the cities of Barcelona, Madrid, and Seville, it is important to narrow the focus of the literature review of the sharing economy to pricing in the sharing economy.

2.2 Pricing in the sharing economy

Pricing is an important topic, especially in the tourism and accommodation industry. Accurate pricing influences consumer purchases so it is important for the success of any sharing economy company that users price their services properly as it allows for a higher number of transactions and more revenue for the company. Taking this idea one dimension further, optimal pricing is also extremely relevant for revenue and profit maximization purposes for both the platform and the person sharing their resources.

Most of the current literature concerning pricing in the sharing economy deals predominantly with the two most successful and highly valued companies in the sharing economy space: Airbnb and Uber. The literature review concerning Airbnb pricing will be covered in Section 2.6.

The literature concerning Uber primarily concerns the analysis of Uber's dynamic pricing of rides, a practice also known as surge pricing. Dynamic pricing is a pricing strategy that sets flexible prices depending on the demand for a product or a service. Sharing economy companies such as Uber use a real-time dynamic pricing algorithm, a strategy that has become easier to implement due to advancements in technology, specifically the increased speed of data processing (Chen, Mislove, & Wilson, 2015; Brodeur & Nield, 2018).

A dynamic pricing strategy can be simple, such as in the tourism and accommodation industry where there is often a two-tier system, low season and high season, and the price changes based on time (Oses, Gerrikagoitia, & Alzua, 2016). Dynamic pricing can also be much more complex such as in the airline industry where prices of flights change depending on a variety of factors which may include time of day, day of the week, consumer, type, and number of seats available (Martin & Koo, 2009). This kind of flexible pricing should bring the market structure to something close to perfect competition by optimizing for the equilibrium of price and supply. Perfect competition relies on three basic assumptions: price taking, product homogeneity, and free entry and exit (Pindyck & Rubinfeld, 2018). However, given that the sharing economy is dominated by only a few large platforms, these companies could also become quasi-monopolists and control their respective markets. In the case of Uber, the company controls the market price of a ride on its platform giving it a certain level of monopoly power. If Uber chooses to exercise this monopoly power, they could charge "a price that exceeds marginal cost but by an amount that depends inversely on the elasticity of demand" (Pindyck & Rubinfeld, 2018). In the case of Airbnb, prices are set by the hosts but if an increasing number of hosts accept Airbnb's dynamic pricing feature, then the company would also have a certain amount of monopoly power from a pricing point of view.

Other literature focusing on pricing in the sharing economy includes Kung & Zhong (2017), who compared three pricing strategies for an online grocery delivery service provider, Weber (2016), who investigated how the price of products are affected with and without a sharing market, and Zhang, Jahromi, & Kizildag (2018), who examined consumers' willingness to pay a premium price during three stages of consumption: pre-consumption, mid-consumption, and post-consumption.

2.3 Airbnb

Due to its position alongside Uber as one of the most well-known, visible, and most valuable sharing economy companies in the world (Somerville, 2018; Rushe & Helmore, 2019), there is a fair amount of literature concerning Airbnb. As mentioned in Section 1, Airbnb is the leader in the sharing economy accommodation sector. According to the most recent company disclosures by CEO Brain Chesky, the company is profitable and has a valuation of 31 billion USD (Somerville, 2018). Airbnb has over 6 million Airbnb listings worldwide covering 81,000 cities in over 191 countries, and they have served 500 million guests since their founding in 2008 (Airbnb, Inc., 2019), a large indication of their global prevalence and successful penetration of the accommodation sector.

The focus of the literature concerning Airbnb involves a wide range of topics such as consumer segmentation, discrimination, and trust and reputation.

Lutz & Newlands (2018) analyzed the potential of consumer segmentation within a single sharing economy platform, in their case, Airbnb. By analyzing and comparing the characteristics of guests who booked shared rooms and those who booked entire homes, they concluded that there is often a mismatch between host expectations and guest intentions which indicates a lower than optimal matching efficiency. They also raised the question concerning the fine line between consumer segmentation and discrimination given hosts ability to decline service to specific consumer segments.

Discrimination of Airbnb guests is one of the prominent topics in the Airbnb literature and the one that has received the most widespread media attention. Edelman, Luca, & Svirsky (2017) found that guests with distinctively African American names were 16% less likely to be accepted as a guest compared to relatively identical guests with caucasian names. From the perspective of hosts who belong to minority groups, Kakar, Voelz, Wu, & Franco (2018) found that Asian and Hispanic hosts in San Francisco typically charge 8% to 10% less than their comparable caucasian counterparts which may be a result of a strategy to avoid possible discrimination by potential guests who would not have booked the accommodation due to the ethnicity of the host.

Given that the creation of trust between strangers is essential to Airbnb's success, this topic has also received a fair amount of attention. Abrahao, Parigi, Gupta, & Cook (2017) found that an artificially designed reputation system such as the review system Airbnb uses for its website is effective in counteracting people's natural tendency to base trustworthiness on social biases. In contrast, Ert, Fleischer, & Magen (2016) found that a host's reputation had no effect on listing price and did not increase the likelihood of a guest booking. They found that a host's photos and the perceived trustworthiness of the host from the photos increased the price of the listing as well as increased the likelihood of it being booked by a guest. Fagerstrøm, Pawar, Sigurdsson, Foxall, & Yani-de-Soriano (2017) found that the facial expressions of hosts increased or decreased the likelihood of guest booking. Neutral and positive facial expressions compared to negative expressions and absence of a facial image increased the likelihood of a guest booking. High reputation and low price were found to be insufficient in compensating for a negative image or lack of a facial image.

While the Airbnb literature reviewed above is interesting, many of its findings and conclusions are only tangentially related to the aims of this thesis by providing useful background information. A more relevant topic to this thesis in Airbnb literature is Airbnb's effect on the accommodation industry.

2.4 Effect on traditional accommodation industry

Airbnb's business model has been referred to as disruptive innovation (Guttentag, 2013) since it can have an impact on the traditional accommodation industry such as hotels, motels, and traditional bed and breakfasts. In addition to the literature mentiond in Section 2.3, there is a fair amount of literature concerning Airbnb's growing popularity and its effect on the traditional accommodation industry.

Research has shown that Airbnb's services can have an overall small negative effect on the revenues of traditional accommodation providers, but its effect on smaller operators in the lower-budget tiers of the accommodation industry is more pronounced as Airbnb listings often compete for this market segment (Zervas et al., 2017). The authors found that a 10% increase in Airbnb supply in Texas resulted in a 0.4% reduction in hotel revenue which translated to an 8% to 10% loss for budget and economy hotels in Austin, Texas. They also analyzed how Airbnb supply scaled during peak-demand periods such as the South by South West (SXSW) festival in Austin, Texas, and the Texas State Fair in Dallas, Texas. Zervas et al. found that Airbnb supply scaled up instantaneously during these peak-demand periods which significantly limited the pricing power of hotels during these peak-demand periods.

Xie & Kwok (2017) explored similar topical as well as geographical territory to Zervas et al. (2017). They concluded that while guests used Airbnb as an alternative to hotels, the pricing of Airbnb listings, which are set by hosts, reduced Airbnb's impact on hotels. This is primarily due to the large price differential that existed between the higher Airbnb listings prices and the lower hotel prices within the metropolitan Austin, Texas area as well as the large dispersion of prices among the Airbnb listings themselves.

Blal, Singal, & Templin (2018) found that in San Francisco, overall revenue per available room of a hotel was unrelated to Airbnb supply. Revenue was positively affected by higher Airbnb prices, but negatively affected by an increase of an Airbnb review score.

2.5 Pricing in the traditional accommodation industry

Accurate pricing is an extremely important issue in the traditional accommodation industry but the ability to price accurately remains a challenge. Generally, hotels use cost-based pricing, competition-driven pricing, and customer-driven pricing to price their rooms which may not be sufficient in establishing optimal prices in a dynamic environment (Hung, Shang, & Wang, 2010).

The authors used ordinary least squares and quantile regression to find the determinants of hotel room prices in Taiwan and found that hotel age, market conditions, and the proportion of foreign travelers positively influenced room prices in the high-price category.

The importance of location in the traditional accommodation industry has inspired a wide range of literature. In a study using a hedonic pricing model, Bull (1994) found that location is a significant determinant of room price of motels in Ballina, Australia. For every kilometer away from the town center a motel was located, the room rate fell an average of 5.17 AUD. Sainaghi (2011) confirmed the importance of location and also found that number of rooms, number of employees per room, and years since last refurbishment are also significant determinants of the price of hotel rooms in Milan. Balaguer & Pernías (2013) focused their research on the effect of spatial agglomeration on hotel prices in Madrid, Spain. They found that a greater density of competitors resulted in a lower average price as well as lower price variation. Along similar lines, Lee (2015), found that hotels in Texas of similar quality competed more distantly and that hotels of similar quality that are closely located showed signs of possible cooperation.

2.6 Airbnb pricing

The literature concerning Airbnb pricing is relatively new and much of it is concerned with identifying the determinants of the price of an Airbnb listing.

Wang & Nicolau (2017) analyzed 180,533 Airbnb listings in 33 cities using an ordinary least squares and quantile regression analysis to find the price determinants of Airbnb accommodation. With regards to the characteristics of the host, Superhost status, number of listings, and verified identities were associated with higher prices. Concerning the characteristics of the accommodation, location, size, and accommodation type affected prices. Higher prices were associated with moderate and strict cancellation rules, and high customer ratings. Number of ratings and the availability of breakfast negatively affected prices.

Aznar, Sayeras, Segarra, & Claveria (2018) compared the prices of hotels and Airbnb listings in Barcelona and found that Airbnb hosts did not dynamically price the same way as hotels did. While both hotels and Airbnb hosts priced based on seasonality, Airbnb hosts did not vary their prices significantly based on the day of the week, counter to common hotel room pricing where rooms are more expensive on weekends.

In 2015, Airbnb introduced a dynamic pricing feature known as “Smart pricing” within the Airbnb application. If a host turns on the feature, Airbnb suggests a price for a host’s offered accommodation and a host can choose to accept the suggestion or alter it (Hill, 2015). Gibbs, Guttentag, Gretzel, Yao, & Morton (2018) analyzed 39,837 Airbnb listings in five cities in Canada over a 12-month period and found that dynamic pricing was not uniformly used by hosts. Over half (52.2%) of listings essentially did not change in price.

Zhang, Chen, Han, & Yang (2017) employed a general linear model (GLM) and a geographically weighted regression (GWR) to determine which characteristics affected Airbnb listing prices in Nashville, Tennessee. Their sample set included 794 listings and they found that the GWR model had better explanatory power and identified distance from the convention center, number of reviews, and customer rating as being determinants of price. The GWR model was able to explain how the the effect of location had a variable effect on price having had a strong effect on listings in central Nashville and having had a diminishing effect in remote areas further away. An odd result was that an increase in customer rating was found to negatively affect prices.

Price behavior is another interesting topic within the Airbnb pricing literature. Gunter & Önder (2018) used ordinary least squares (OLS) with cluster-robust standard errors and found that Airbnb listings in Vienna were price inelastic suggesting that there was room for revenue increases from price hikes. Benítez-Aurioles (2018) used a two-stage least squares regression model and found that prices in Barcelona and Madrid were elastic and fairly similar (2.2, and 2.4 respectively).

As it is directly related to the research aims of this thesis, literature concerning the application of the hedonic pricing model in the sharing economy is addressed in Section 2.7.

2.7 Hedonic pricing model applications in the sharing economy

Current literature concerning the hedonic pricing model includes extensive research of the model's applications in various real estate markets throughout the world (Goodman, 1978; Witte, Sumka, & Erekson, 1979; Mok, Chan, & Cho, 1995; Yang, 2001; Liao & Wang, 2012), and in the traditional accommodation industry (Thrane, 2007; Abrate, Capriello, & Fraquelli, 2011; Zhang, Ye, & Law, 2011; Bull, 1994; Schamel, 2012). While the hedonic pricing model has been used extensively in real estate valuation, there are issues which can negatively affect the performance of the model such as: model specification procedures, multicollinearity, independent variable interactions, heteroskedasticity, non-linearity and outlier data points (Linsombunchai, Gan, & Lee, 2004). Advancements in technology have created new methods of valuation such as using an artificial neural network. Abidoye & Chan (2018) compared whether the hedonic pricing model or an artificial neural network model made more accurate price predictions of real estate in Lagos, Nigeria. They found that the artificial neural network performed better in predicting real estate prices which supports the findings of Linsombunchai, Gan, & Lee (2004) who first conducted this research using real estate prices in Christchurch, New Zealand.

Research concerning the application of a hedonic pricing model to the sharing economy is relatively new. As such, there is little literature on this topic with most of the literature being written and published within the last two years. Teubner, Hawlitschek, & Dann (2017) applied a hedonic pricing model to an Airbnb data set of 86 German cities and found that hosts' rating scores, duration of membership, and number of photographs positively affected prices. Number of ratings was found to have negatively affected prices while ID verification and Superhost status had no effect contrary to the findings of Wang & Nicolau (2017).

Gibbs, Guttentag, Gretzel, Morton, & Goodwill (2017) applied a hedonic pricing model to 15,716 Airbnb listings from 5 Canadian cities: Montreal, Calgary, Ottawa, Toronto, and Vancouver. Their results confirmed previous studies analyzing characteristics that affect the prices of accommodation. They found that traditional characteristics of accommodation such as size and location matter and are taken into account by Airbnb hosts when they set their prices.

Chen & Xie (2017) developed a hedonic pricing model and tested it using 5,779 Airbnb listings in Austin, Texas. They concluded that the basic functionality of an Airbnb listing were the main determinants of price with accommodation type having the largest effect. Speed of host response was also seen to have an effect on price whereas Superhost status did not. Customer reviews had an impact but only accounted for a 2% variance in price.

Magno, Cassia, & Ugolini (2018) used a hedonic pricing model to analyze 1,056 Airbnb listing to determine whether hosts in Verona, Italy dynamically adjusted the prices of their accommodation as a result of their prior experience with price management and the market demand. They found that price and dynamic pricing are significantly related to a host's experience measured by the number of months a host had been a part of the Airbnb platform, and the level of market demand. They also found that accommodation type and size positively affected price while number of reviews was negatively correlated to price.

Önder, Weismayer, & Gunter (2018) employed a hedonic pricing model to quantify the marginal contributions of an Airbnb listing characteristic to the prices of Airbnb listings in Tallinn, Estonia. The authors found after 381 regression iterations of their linear hedonic pricing model that the optimal radius surrounding an Airbnb listing was 650 meters with number of bedrooms, bathrooms, average competitors' prices, and the number of points of interest within the optimal radius having had a significantly positive effect on price. A significant negative effect on price was found the further away an Airbnb listing was from the city center. Concerning competitors' prices, it was discovered that there are price dependencies between Airbnb listings and the traditional accommodation sector as well as the fact that Airbnb accommodation is the more affordable of the two.

2.8 Summary

There is a wide range of literature concerning the sharing economy. However, given the specific focus of this thesis, most of the literature falls outside of its scope and concerns the popularity, socioeconomic consequences, regulatory and economic implications of the sharing economy as well as the sharing economy term itself.

It is clear that Airbnb, given its success and size, does have an effect on the traditional accommodation industry. Despite its potential as a hotel substitute, research indicates that Airbnb supply acts in more complementary ways. From the literature available, it is difficult to make broad generalizations since the effects of Airbnb on hotel demand and revenue vary depending on the location and time period analyzed.

Pricing is an important area of research in the traditional accommodation industry as well as in Airbnb. Accurate pricing is an essential component of a well-functioning economy and optimal pricing allows for revenue and profit maximization. Research concerning pricing in the traditional accommodation industry emphasized the importance of location. Literature concerning Airbnb pricing identified the determinants of price of an Airbnb listing and found that basic functionality of accommodation such as size, location, and type were found to be significant, similar to traditional accommodation.

The application of a hedonic pricing model on Airbnb listings is quite new and as such, there is very little literature on the topic. Therefore, it is hoped that the analysis provided by this thesis will contribute to this rapidly growing area of research.

3 DATA

The section includes a detailed description of the data set, the treatment and preparation of the data set for further analysis, preliminary regression analysis and statistical pre-testing, and the procedures followed to prepare the data along with descriptive statistics of the variables for each city.

3.1 Data set

The data set used for analysis was purchased from AirDNA, a company which tracks the daily performance of over 10 million short-term rental listings on websites such as Airbnb and HomeAway. The data and products AirDNA create are targeted towards hosts, property managers, academics, tourism boards, and investors. The methodology AirDNA employs to produce its data is to consider only active listings, use more precise geolocations of listings, and calculate revenue based on rates and fees. They also use machine learning technology to distinguish booked vs. blocked dates using statistical pattern-recognition techniques similar to those found in recommender systems for Netflix and Amazon (AirDNA, LLC, 2018).

The total amount of listing data in the data set from AirDNA for the three cities in Spain to be analyzed contains 59,567 separate listings: Barcelona (36,087), Madrid (18,026), and Seville (5,454). It is a cross-sectional data set where each listing includes 45 different characteristics such as: unique IDs for property and host, geographic information, price information, etc.. A list of the characteristics included in each listing can be found in Table A-1 in Appendix A. Limitations of the data set are the temporal nature of the data set itself as it only captures Airbnb listing data from Barcelona, Madrid, and Seville between 2015-2016.

3.2 Data preparation

While the data for all 3 cities is from 2015-2016, in order to have a consistent time frame across all cities, all listings with a last scraped date prior to 17.08.2015 have been dropped. This date was chosen as it is the earliest scrape date for all Seville Airbnb listings. Therefore, the period of analysis for all Airbnb listings is between 17.08.2015 and 02.08.2016. This reduces the number of eligible listings in Barcelona and Madrid to 31,512 and 16,438 respectively while the number of eligible listings in Seville remains unchanged. Therefore, the total number of listings across all three cities to be considered for further analysis is 53,404 aggregated for the time period from September 17, 2015 to August 2, 2016.

In accordance with the research question of the thesis, only variables which could affect price or are deemed necessary for data preparation have been considered. Therefore, certain variables included in each Airbnb listing (e.g 'Listing URL') have been excluded. In total, 23 out of 45 characteristics have been immediately excluded. A list of excluded variables can be found in Table A-2 in Appendix A.

In addition to excluding various variables which would not impact price and excluding certain listings due to time period, categorical variables have been recoded into dummy variables for the purposes of interpretation and analysis.

Similar to Gunter (2018), 'Cancellation Policy' has been recoded into a binary variable where 0 indicates a flexible policy and all other options (strict and moderate) are 1. The decision to turn what would normally be a categorical variable into a binary variable was made in order to create a clear distinction between the most guest friendly cancellation policy, 'flexible', and the other stricter cancellation policies, namely, 'moderate' and 'strict'. Given the research question of the thesis, this was determined to be the most appropriate approach to analyze the data.

For 'Listing Type', accommodation offering the entire apartment or home has been recoded to 1 while all other listing types (private room and shared room) have been recoded to 0. This was done to make a clear distinction between shared accommodation and unshared accommodation. Other variables such as 'Superhost Status', and 'Instantbook Enabled' have been recoded into 0 for "No" and 1 for "Yes".

A new variable called 'Professional Status' was created in order to gauge whether a host's professionalism, defined as a host that manages two or more properties, affects Airbnb prices which is similar to what what was explored by Gibbs et al. (2017) which built on similar work done by Li, Moreno, & Zhang (2016). Gibbs et al. (2017) found that professional status was only associated with a significant increase in price in one Canadian city (Montreal) whereas Li et al. (2016) found that professional hosts earn more in general. With the goal of building off of the research that has come before, it was determined that it would be interesting to see what effect this variable would have on the Airbnb prices in Spain.

In order to determine how distance would affect the prices of Airbnb listings, it was necessary to determine the city centers of Barcelona, Madrid, and Seville. The centers of Barcelona, Madrid, and Seville were chosen to be Plaça de Catalunya, Puerta del Sol, and Plaza Nueva respectively, in line with what is commonly recognized as the city center for each city. The geographic coordinates of each city center were taken from the website www.gps-coordinates.net which obtains the longitude and latitude of a given place from Google Maps.

To calculate the distance between each property and its respective city center, the distGeo function from the geosphere package was used in the software R. The distGeo function calculates the shortest distance between two points on an ellipsoid (called the geodesic), improving on the previous algorithms created by T. Vincenty (Vincenty, 1975), and is considered to be a highly accurate calculation of the shortest distance between two points on the Earth's surface (Karney, 2013). A descriptive summary of the variables to be analyzed for each city can be found in Table 3-1 including the ex-ante expected algebraic signs (+/-) of the explanatory variables if applicable. "±" is used if the expected sign is ambiguous. Untransformed descriptive statistics of the data set for each city can be found in Tables 3-2, 3-3, and 3-4.

TABLE 3-1. DESCRIPTION OF STUDY VARIABLES

Variable	Description
Price ($) (dependent)	Average price per night in USD including user defined cleaning fee
Bedrooms (+)	Number of bedrooms in the accommodation
Bathrooms (+)	Number of bathrooms in the accommodation
Number of Reviews (-)	Number of user reviews of the accommodation
Overall Rating (+)	Overall rating of the accommodation on a scale of 1-5
Max Guests (+)	Max number of guests allowed in the accommodation
Response Rate (+)	Percentage of users the host has responded to
Response Time (-)	Average amount of time a host takes to respond to a user in minutes
Number of Photos (+)	Number of photos of the accommodation in the Airbnb listing
Distance From City Center (km) (-)	Distance of the accommodation from the city center
Instantbook Enabled (0 no; 1 yes) (±)	Whether Instantbook is enabled on the accommodation listing
Professional Status (0 no; 1 yes) (+)	Whether the host manages more than 1 accommodation listing
Superhost Status (0 no; 1 yes) (+)	Whether the host is a Superhost
Cancellation Policy (0 flexible; 1 non-flexible) (+)	Whether the cancellation policy is flexible or non-flexible
Listing Type (0 shared; 1 entire) (+)	Whether the accommodation is shared or entire

TABLE 3-2. DESCRIPTIVE STATISTICS (BARCELONA)

	Arithmetic Mean	Std. Deviation	Minimum	Maximum
Price	104.05	118.32	10.00	8154.33
Bedrooms	1.45	.916	0	10
Bathrooms	1.270	.5861	0	8.0
Number of Reviews	14.92	26.966	0	335
Overall Rating	4.42	0.54	1	5
Max Guests	3.31	2.17	1	16
Response Rate	91.63	15.956	2	100
Response Time	240.92	359.13	0.02	1440.00
Number of Photos	15.70	11.60	1	358
Distance From City Center	1.91	1.25	0.006	39.70
Instantbook Enabled	0.20	0.40	0	1
Professional Status	0.64	0.48	0	1
Superhost Status	0.01	0.11	0	1
Cancellation Policy	0.66	0.47	0	1
Listing Type	0.50	0.50	0	1

TABLE 3-3. DESCRIPTIVE STATISTICS (MADRID)

	Arithmetic Mean	Std. Deviation	Minimum	Maximum
Price	81.33	65.63	10.00	928.93
Bedrooms	1.27	.808	0	10
Bathrooms	1.249	.5968	0	8.0
Number of Reviews	14.32	27.83	0	429
Overall Rating	4.50	0.53	1	5
Max Guests	3.11	2.00	1	16
Response Rate	93.30	14.032	10	100
Response Time	212.56	331.19	0.01	1440
Number of Photos	15.91	12.16	1	148
Distance From City Center	2.14	2.13	0.004	29.80
Instantbook Enabled	0.22	0.42	0	1
Professional Status	0.56	0.50	0	1
Superhost Status	0.07	0.26	0	1
Cancellation Policy	0.60	0.49	0	1
Listing Type	0.43	0.50	0	1

TABLE 3-4. DESCRIPTIVE STATISTICS (SEVILLE)

	Mean	Std. Deviation	Minimum	Maximum
Price	83.33	75.21	10.00	2082.00
Bedrooms	1.40	0.89	0	10
Bathrooms	1.254	0.63	0	8
Number of Reviews	18.44	36.14	0	406
Overall Rating	4.49	0.50	1	5
Max Guests	3.39	1.98	1	16
Response Rate	93.61	14.45	10	100
Response Time	195.15	337.13	0.01	1440
Number of Photos	17.14	11.97	1	116
Distance from City Center	1.37	1.64	0.03	89.55
Instantbook Enabled	0.26	0.44	0	1
Professional Status	0.58	0.49	0	1
Superhost Status	0.07	0.25	0	1
Cancellation Policy	0.57	0.50	0	1
Listing Type	0.37	0.48	0	1

In order to ensure a linear relationship between the dependent and explanatory variables, the natural logarithms of all continuous variables were taken. In the context of multivariable linear regression, this is called the log-log model and is defined by Wooldridge (2012) as: “a regression model where the dependent variable and (at least some of) the explanatory variables are in logarithmic form” (p.852). In a log-log model, the values of the parameter estimates of the explanatory variables are interpreted as elasticities. The advantages of taking the natural logarithm of variables are to lessen the impact of outliers on the model by making the distribution of values approximately normal.

All subsequent data analysis was conducted with IBM SPSS Statistics Version 25, a software package used for statistical analysis, using the transformed variables. Variables names have been kept the same.

3.3 Preliminary Regression Analysis and Statistical Pre-Testing

In order to ensure that the data has been appropriately transformed in order to be used in the hedonic pricing model (see Section 4.2) and that the results from the analysis are indeed the desired results of the chosen methodology, it is important to test for common issues that can arise during a regression analysis, specifically that none of the assumptions of multivariable linear regression have been violated (see Section 4.3). This is done to ensure that there are no problems when the results are interpreted. A model suffering from issues such as multicollinearity and heteroskedasticity can lead to inaccurate conclusions.

Multicollinearity occurs in a regression analysis when there is a high correlation between explanatory variables in a model. While the issue itself does not violate the assumptions of multivariable linear regression (see Section 4.3) and is not clearly defined, meaning that there is no specific level of correlation between explanatory variables that has been deemed problematic, it is generally accepted that in order to properly estimate the slope parameters of a model, it is better to have less correlation between the explanatory variables (Wooldridge, 2012). The presence of multicollinearity in a model also reduces the precision and statistical power of the model making it difficult to justify and interpret *p*-values and conclude that the model specification is correct.

To determine whether multicollinearity exists in the model, the variance inflation factor (VIF) for all considered explanatory variables was calculated with SPSS. VIF quantifies the severity of multicollinearity existing in the model by measuring how much the variance of an estimated coefficient is inflated by the correlation with other explanatory variables. While VIF is a statistic that allows for the interpretation of multicollinearity, there is no specific number which researchers have agreed is too high although it is commonly accepted that a VIF under 10 for certain explanatory variables indicates that these variables are not a source of multicollinearity and should be kept in the model (Kutner, Nachtsheim, & Neter, 2004).

TABLE 3-5. VARIANCE INFLATION FACTORS (VIF) OF EXPLANATORY VARIABLES (BARCELONA)

Coefficients[a]

Model		t	Sig.	Collinearity Statistics	
				Tolerance	VIF
	(Constant)	38.324	.000		
	Cancellation Policy	8.062	.000	.897	1.115
	Listing Type	84.780	.000	.434	2.304
	Number of Reviews	-39.066	.000	.822	1.216
	Overall Rating	16.396	.000	.932	1.073
	Bedrooms	15.697	.000	.356	2.811
	Bathrooms	18.547	.000	.777	1.288
	Max Guests	43.673	.000	.325	3.075
	Response Rate	7.026	.000	.813	1.231
	Response Time	1.080	.280	.614	1.629
	Superhost Status	5.791	.000	.970	1.031
	Number of Photos	18.582	.000	.749	1.336
	Instantbook Enabled	4.772	.000	.736	1.359
	Professional Status	-5.360	.000	.915	1.093
	Distance from City Center	-35.590	.000	.959	1.042

a. Dependent Variable: Price

TABLE 3-6. VARIANCE INFLATION FACTORS (VIF) OF EXPLANATORY VARIABLES (MADRID)

Coefficients[a]

Model		t	Sig.	Collinearity Statistics	
				Tolerance	VIF
	(Constant)	23.076	.000		
	Cancellation Policy	1.526	.127	.876	1.142
	Listing Type	62.665	.000	.488	2.048
	Number of Reviews	-22.027	.000	.782	1.278
	Overall Rating	11.151	.000	.917	1.091
	Bedrooms_	25.956	.000	.501	1.995
	Bathrooms	11.304	.000	.750	1.333
	Max Guests	19.611	.000	.386	2.593
	Response Rate	1.044	.297	.848	1.179
	Response Time	3.063	.002	.605	1.654
	Superhost Status	8.946	.000	.924	1.083
	Number of Photos	12.590	.000	.704	1.420
	Instantbook Enabled	-.686	.493	.695	1.440
	Professional Status	.253	.800	.884	1.131
	Distance from City Center	-26.627	.000	.876	1.141

a. Dependent Variable: Price

TABLE 3-7. VARIANCE INFLATION FACTORS (VIF) OF EXPLANATORY VARIABLES (SEVILLE)

Coefficients[a]

Model		t	Sig.	Collinearity Statistics	
				Tolerance	VIF
	(Constant)	13.474	.000		
	Cancellation Policy	6.096	.000	.883	1.133
	Listing Type	29.767	.000	.467	2.144
	Number of Reviews	-16.664	.000	.749	1.335
	Overall Rating	6.299	.000	.924	1.082
	Bedrooms	10.848	.000	.462	2.166
	Bathrooms	6.992	.000	.752	1.331
	Max Guests	15.296	.000	.366	2.731
	Response Rate	1.379	.168	.831	1.204
	Response Time	1.258	.208	.510	1.959
	Superhost Status	2.781	.005	.919	1.088
	Number of Photos	5.590	.000	.703	1.423
	Instantbook Enabled	1.120	.263	.624	1.602
	Professional Status	-2.669	.008	.904	1.106
	Distance from City Center	-15.474	.000	.830	1.206

a. Dependent Variable: Price

As seen in Tables 3-5, 3-6, and 3-7, the variance inflation factors are all well under 10 indicating that there are no multicollinearity issues that may affect the interpretation of the results.

Depending on the city, certain explanatory variables were found to be not stastically significant at the 5% level ($p > 0.05$). In Barcelona, the 'Response Time' explanatory variable was found to be not statistically significant. The same result was found for 'Response Rate', 'Instantbook Enabled', 'Professional Status', and 'Cancellation Policy' in Madrid. In Seville, 'Response Rate', 'Response Time', and 'Instantbook Enabled' were found to be not statistically significant. While these variables were found to be not statistically significant, they will nonetheless remain in the model as no variable was uniformly found to be not statistically significant for each city. In addition, these variables will not be left out of the model given their potential substantive significance. A fast 'Response Time' could reflect a host's willingness to deliver superior customer service which could translate into higher prices. 'Instantbook Enabled' could also translate into higher prices given that a host has given up the ability to review the profile of a potential guest thereby increasing the level of risk they are taking on. Typically, higher risk equals higher returns. These variables were also found to be significant in previous Airbnb research (Gunter, 2018; Gibbs et al., 2017) and this thesis seeks to determine whether similar or vastly different findings will be found for Barcelona, Madrid, and Seville. However, caution will be taken when interpreting the results for these variables and their respective cities.

Another common issue that needs to be checked for in the model is heteroskedasticity—it exists when the variance of the error term in the model is not constant which is a violation of one of the assumptions of multivariable linear regression. This causes a problem when conducting significance tests and calculating confidence intervals as the error term is biased and therefore unreliable (Wooldridge, 2012). Similar to multicollinearity, the presence of heteroskedasticity can lead to false conclusions about the model. A method to deal with heteroskedasticity is to perform a logarithmic transformation of the data. This has already been done as explained in Section 3.2. Therefore it is anticipated that there will be no heteroskedasticity in the model. However, if heteroskedasticity is found to be present, the regression will be performed using weighted least squares (WLS) in addition to ordinary least squares (OLS) and the results will be compared.

In order to check for heteroskedasticity, a Breusch-Pagan (Breusch & Pagan, 1979) test was performed on the data set. The Breusch-Pagan test is a "test for heteroskedasticity where the squared OLS residuals are regressed on the explanatory variables in the model" (Wooldridge, 2012, p.845). The test is done to see whether the variance of the error depends upon the explanatory variables in the model which would be a violation of the homoskedasticity assumption which is integral to multivariable linear regression. Since this functionality is not a part of SPSS, an SPSS extension using the ncv.test function from the R car package was used to conduct the test (Fox, 2007).

Given the assumptions of multivariable linear regression and the logarithmic transformation of the data, the null hypothesis is that the model is homoskedastic and thus exhibits no heteroskedasticity. This is determined by the Chi-Square (χ^2) value of the Breusch-Pagan test. Also, depending on the critical value, if the p-value < 0.05, we would reject the null hypothesis in favor of the alternative hypothesis at least at this conventional level. It should be mentioned that there is a considerable debate among researchers concerning the usage of this level as a means of validating statistical significance. However, according to Antonakis (2017), there is nothing wrong with using p-values in general if they are used correctly.

The null hypothesis for the Breusch-Pagan test is as follows:

$$H_0: \chi^2 \leq 3.84 \tag{1}$$

The alternative hypothesis is as follows:

$$H_1: \chi^2 > 3.84 \tag{2}$$

TABLE 3-8. BREUSCH-PAGAN TEST RESULTS (BARCELONA)

Non-constant Variance Score Test

	ChiSquare	df	Sig.
Test Result	120.248	1	.000

Variance model: fitted values

Computed by R ncvTest function

TABLE 3-9. BREUSCH-PAGAN TEST RESULTS (MADRID)

Non-constant Variance Score Test

	ChiSquare	df	Sig.
Test Result	4.738	1	.030

Variance model: fitted values

Computed by R ncvTest function

TABLE 3-10. BREUSCH-PAGAN TEST RESULTS (SEVILLE)

Non-constant Variance Score Test

	ChiSquare	df	Sig.
Test Result	.939	1	.333

Variance model: fitted values

Computed by R ncvTest function

Results of the Breusch-Pagan test in Tables 3-8, 3-9, and 3-10 for each of the three cities indicate that we fail to reject H_0 at the 5% level in favor of H_1 for Seville indicating that there is no statistically significant amount of heteroskedasticity in the model. However, for Barcelona and Madrid, we reject H_0 in favor of H_1 indicating that there is heteroskedasticity in the model. However, it is important to note that data sets with a large number of observations suffer from size distortions when the Breusch-Pagan test is applied making it easier to reject the null hypothesis (Pesaran, 2004).

To confirm the results of the Breusch-Pagan test, it was decided that a visual check for heteroskedasticity would be performed. The predicted values and residuals of the regression were plotted on a scatterplot. If the points are equally distributed above and below the x-axis and to the left and right of the y-axis, it is a sign that the homoskedasticity assumption has not been violated and that there is therefore no heteroskedasticity present.

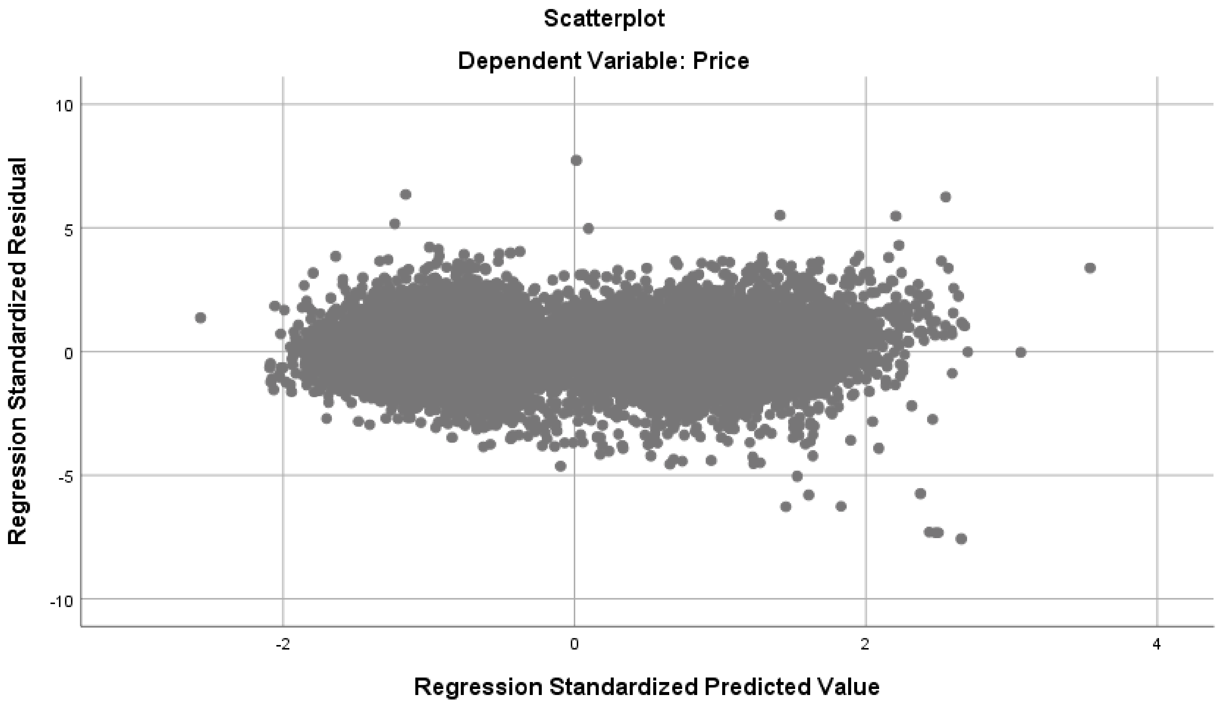

FIGURE 3-1. SCATTERPLOT OF STANDARDIZED PREDICTED VALUE VS. STANDARDIZED RESIDUAL VALUES (BARCELONA)

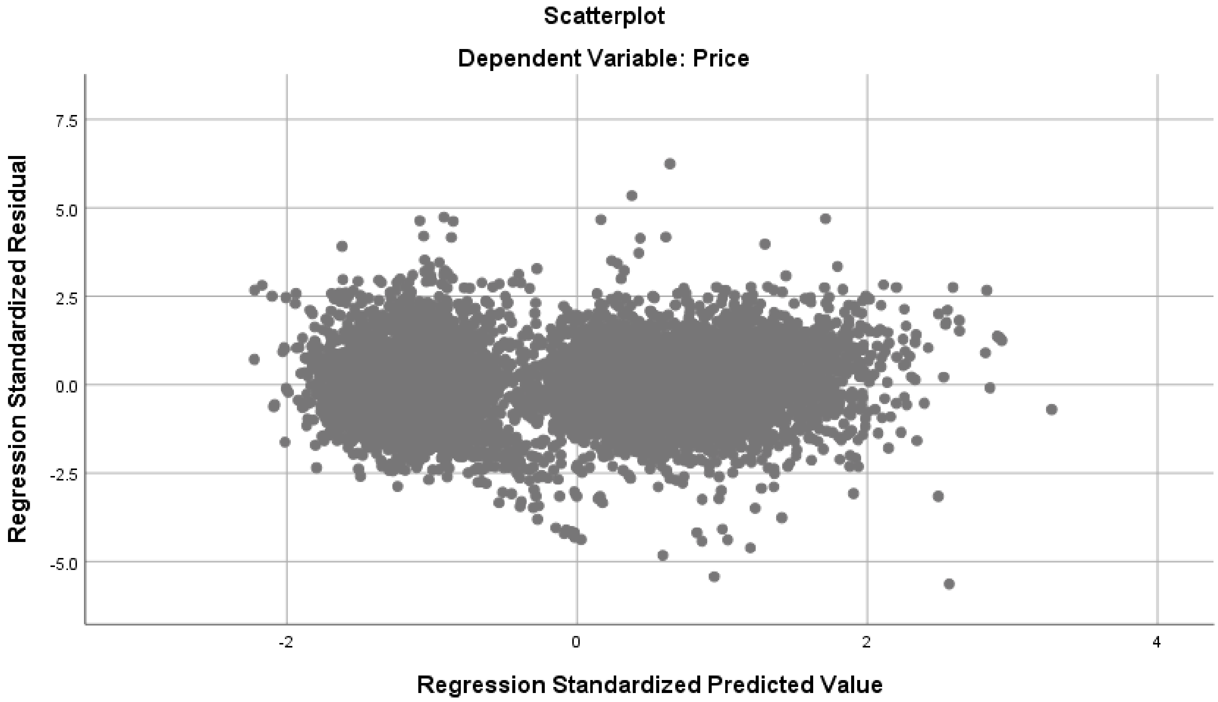

FIGURE 3-2. SCATTERPLOT OF STANDARDIZED PREDICTED VALUE VS. STANDARDIZED RESIDUAL VALUES (MADRID)

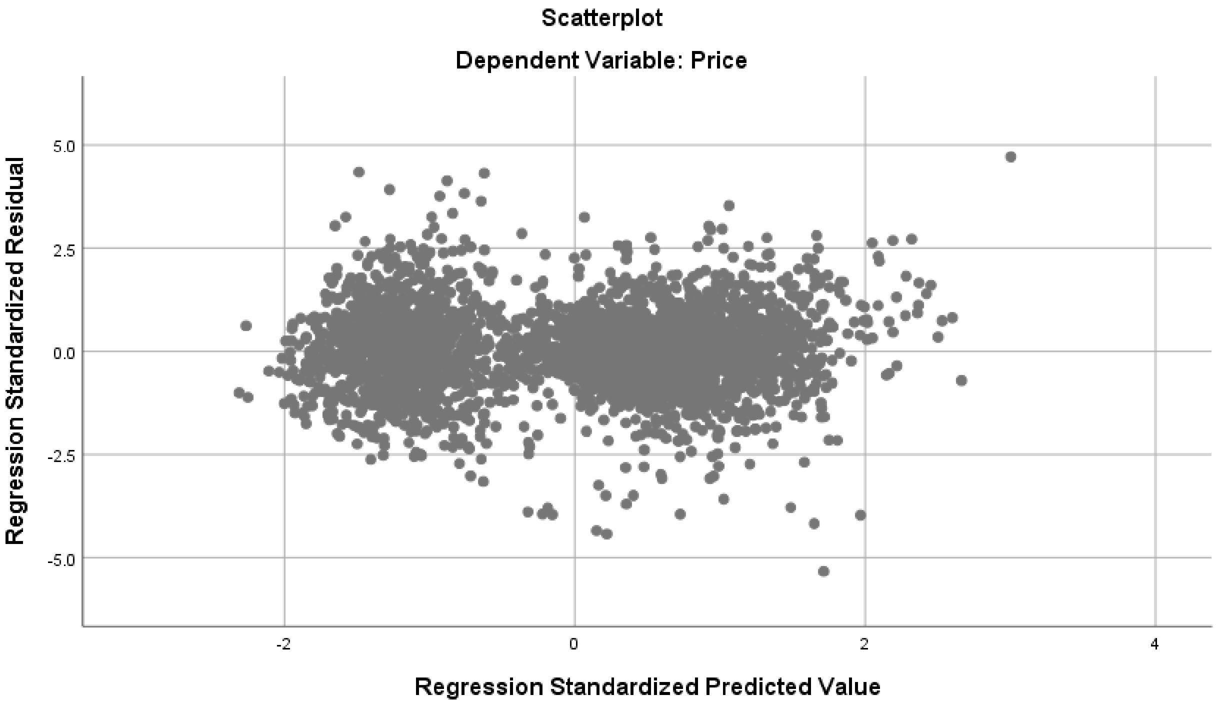

FIGURE 3-3. SCATTERPLOT OF STANDARDIZED PREDICTED VALUE VS. STANDARDIZED RESIDUAL VALUES (SEVILLE)

As seen in Figures 3-1, 3-2, and 3-3, the homoskedasticity assumption does not seem to have been violated as the observations for all three cities are fairly evenly distributed above and below the x-axis and to the left and to the right of the y-axis. As mentioned previously, if hetereoskedasticity was found to be present in the model, the regression would be performed using WLS in addition to OLS and the results would be compared. Since the results of the Breusch-Pagan test and the scatterplots point to two different conclusions, it was determined that this approach would be followed and the results of the regressions would be compared to determine whether there are any differences between using OLS and WLS.

Preliminary regression analysis also included the production of histograms and normal P-P Plots for the residuals of the regression for each city. The charts indicated that the data is normally distributed and linear, fulfilling the normality and linearity assumptions required for multivariable linear regression (see Section 4.3). Given that the natural logarithm was taken of all continuous variables, this was to be expected. These charts can be found in Figures B-1, B-2, B-3, B,4, B-5, and B-6 in Appendix B.

3.4 Summary

The data set which was used for analysis in this thesis was purchased from AirDNA, a company that tracks the performance of short-term rental listings. Approximately 10.34% of listings (6,163) were excluded from further analysis in order to keep the time period under observation consistent across all three Spanish cities. Several variables which were unrelated to price were also immediately excluded as they did not pertain to the research question of this thesis.

The remaining variables were then recoded into dummy variables as deemed necessary while the natural logarithm of all continuous variables were taken. New variables such as ‘Professional Status’ and ‘Distance to City Center’ were created in accordance with data preparation procedures followed by various researchers in the existing literature, specifically to answer whether a host’s professional ability (managing two or more Airbnb listings) and whether the distance from an Airbnb listing to its respective city center would affect the price of the listing.

Preliminary regression analysis and statistical pre-testing revealed that there was no multicollinearity issues in the model while the Breusch-Pagan test along with the plotting of the predicted values and the residuals of the regression pointed to different conclusions concerning heteroskedasticity. Therefore, it was decided, in order to ensure that the results can be interpreted correctly, that two regressions would be conducted—one with OLS and the other with WLS. The results of these regressions are compared and discussed in Chapter 5, specifically Section 5.3.

4 METHODOLOGY

The methodology section details the procedures followed to select the appropriate variables to be included in the hedonic pricing model, a detailed description of the hedonic pricing model to be used for the analysis of the data, and a detailed description of multivariable linear regression, the assumptions that need to be fulfilled in order for the analysis of the results to be valid, and a description of OLS and WLS.

4.1 Variable selection

Before the analysis can begin, the variables to be included in the model need to be selected. While all variables that were determined to have no impact on price or were not relevant for data preparation have been excluded from the original data set (see Table A-2 in Appendix A), there are certain variables that remain in the data set which may describe the same characteristic of an Airbnb accommodation. For example, both 'Bedrooms' and 'Bathrooms' can be considered a proxy for the overall size of the accommodation. In these cases, when incorporating variables into the hedonic pricing model, it is recommended that only one of these two variables is used. The inclusion of both may result in multicollinearity (see Section 3.3) and would be a violation of one of the assumptions (see Section 4.3) underpinning the classical linear model of cross-sectional regression. However, as demonstrated in Section 3.3, there are no multicollinearity issues and therefore these explanatory variables remain in the model. Also, excluding these variables completely from the data set would be unadvisable as these variables could still be used to create alternate hedonic pricing models which would serve as robustness checks (see Section 5.4) for the regression results from the original model (Gunter & Önder, 2018).

The current literature concerning hedonic pricing model applications in the sharing economy (see Section 2.4) provides significant guidance in the selection of variables from the data set to include in the hedonic pricing model of this thesis. 'Overall Rating', 'Number of Photos' (Teubner et al., 2017), 'Number of Reviews', 'Superhost', 'Cancellation Policy' (Wang & Nicolau, 2017), 'Distance to City Center', 'Number of Bathrooms', 'Professional Status' (Gibbs et al., 2017), 'Listing Type' and 'Response Time' (Chen & Xie, 2017) were all demonstrated to have affected the prices of Airbnb accommodation. Therefore, these variables will be included in the hedonic pricing model. In addition, 'Max Guests', 'Response Rate' and 'Instantbook Enabled' will also be included in the model.

4.2 Selection of methodology: hedonic pricing model

As mentioned in Section 1.2, the hedonic pricing model is a regression model typically used in the real estate industry to estimate the value of homes by relating its price to its characteristics (Wooldridge, 2012). Since its first use in 1974, it has been applied in various situations, including to determine which factors affect the price of accommodation (Bull, 1994). Given the fact that

Airbnb accommodation is considered to be a competitor to traditional accommodation such as hotels, it is reasonable to assume that the same analysis could be applied to Airbnb listings.

Therefore, to answer the research question of the thesis, how profile characteristics of an Airbnb listing affect accommodation prices, it was determined that a hedonic pricing model would be an appropriate vehicle to use due to its frequent use and effectiveness in determining the factors that influence hotel room prices. As mentioned in Section 2.5, the application of a hedonic pricing model to analyze Airbnb prices is still relatively new and it is desired that this thesis will add to this rapidly growing body of literature.

Due to the presence of more than one explanatory variable in a hedonic pricing model, a multivariable linear regression analysis will be performed on the data set. The general formula for the hedonic pricing model that will be applied to the Airbnb data for this thesis is:

$$Price = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i + u \tag{3}$$

where 'Price' is the natural logarithm of the average daily prices of Airbnb listings, β_0 is the intercept, β_i are the parameters of the model measuring the change in 'Price' with respect to x_i, where x_i are the explanatory variables, and u is the error term.

The hedonic pricing model specification to be used for this analysis is:

$$\begin{aligned} Price = \beta_0 + \ & \beta_1 Bedrooms + \beta_2 Bathrooms + \beta_3 Number\ of\ Reviews \\ & + \beta_4 Overall\ Rating + \beta_5 Max\ Guests + \beta_6 Response\ Rate \\ & + \beta_7 Response\ Time + \beta_8 Number\ of\ Photos \\ & + \beta_9 Distance\ from\ City\ Center + \beta_{10} Instanbook\ Enabled \\ & + \beta_{11} Professional\ Status + \beta_{12} Superhost\ Status \\ & + \beta_{13} Cancellation\ Policy + \beta_{14} Listing\ Type + u \end{aligned} \tag{4}$$

4.3 Multivariable Linear Regression

Multivariable linear regression is the most widely used method for empirical analysis in economics and other social sciences. It allows for the explicit control of the explanatory variables that simultaneously affect the dependent variable to determine the relationship between the two (Wooldridge, 2012). It allows for more complex analysis compared to a simple linear regression.

Multivariable linear regression, as opposed to simple linear regression, allows researchers to see how more than one explanatory variable affects a dependent variable. The dependent variable is the variable a researcher is interested in analyzing while the explanatory variables are the various factors that may influence the dependent variable.

According to Mooi & Sarstedt (2010), the key benefits of regression analysis are:

1. Indicate if explanatory variables have a significant relationship with a dependent variable.
2. Indicate the relative strength of different explanatory variables' effects on a dependent variable.
3. Make predictions. (p.162)

By performing a regression, a line of best fit is drawn through the observations. This line is an estimation of the parameters where the sum of squared residuals between the observations and the line of best fit are as small as possible. This is known as the ordinary least squares (OLS) method (Wooldridge, 2012).

While regression analysis is an effective method to determine relationships between variables, according to Wooldridge (2012), there are several classical linear model assumptions that need to be made for cross-sectional regression:

1. Linear in parameters: the parameters, β_i, of the model must be linear (i.e. $\neq \beta^x$).
2. Random sampling: there is a random sample of n observations which is representative of the population.
3. No perfect collinearity: none of the explanatory variables are constant and there are no exact linear relationships between the explanatory variables.
4. Zero conditional mean: the error term u has an expected value of 0 given any value of the explanatory variables.
5. Homoskedasticity: the variance of error term u is constant.
6. Normality: the population of error term u is not dependent on the explanatory variables and is normally distributed.

If any of these assumptions are violated, the results of the regression need to be called into question. For example, if any of the first four assumptions are violated, then the unbiasedness of the estimator for the population parameters becomes suspect. An unbiased estimator is integral to the validity of the regression results as it is considered to be an accurate estimate of the population. All the assumptions listed are important. However, when certain assumptions are violated, there are different consequences.

The reason for these assumptions and for the popularity of the use of OLS in econometrics is that according to the Gauss-Markov Theorem, under the first five assumptions, the OLS estimators are the Best Linear Unbiased Estimators (BLUE). This means that the OLS estimators have the smallest variance of all linear and unbiased estimators (Wooldridge, 2012) making them the most effective and accurate of all estimators.

Correct model specification in multivariable linear regression is also extremely important. While including irrelevant explanatory variables does not affect the assumed unbiasedness of the OLS estimators, it can have undesirable effects on the variances of the OLS estimators. Omitting a relevant variable is even more of a problem as it causes OLS estimators to become biased. (Wooldridge, 2012).

Due to the inconclusive existence of heteroskedasticity discovered during the preliminary regression analysis (see Section 3.3), it was decided that two regressions, one using OLS and the other using WLS would be conducted and the results will be compared. WLS allows for less weight to be given to observations with a higher error variance and more weight to be given to observations with a lower error variance. It is used to correct for heteroskedasticity. The difference between WLS and OLS is that for OLS, the same weight is given to all observations (Wooldridge, 2012). It is anticipated that the estimates and the standard errors for the regression using WLS will be slightly different from the regression using OLS.

4.4 Summary

Variable selection is a key component of constructing the appropriate model to address the research question of this thesis. Including irrelevant and omitting relevant variables each come with their own consequences. Most of the variables chosen were guided by previous research.

The hedonic pricing model was selected as an appropriate vehicle to answer the research question of this thesis due to its effectiveness and popularity in determing how various factors influence the price of a certain object under observation. Previous research has shown how various characteristics affect the price of a hotel room which can be considered analogous to the research question in this thesis since Airbnb is seen as a competitor to the traditional accommodation industry.

The use of a hedonic pricing model necessitates a multivariable linear regression due to the existence of more than one explanatory variable. Multivariable linear regression allows researchers to determine the relationships between explanatory variables and a dependent variable such as whether a relationship even exists and if so, how strong these relationships are. Certain assumptions are made in a regression analysis. If any of these assumptions are violated, the results of the regression are likely to be invalid. OLS estimators are considered to be BLUE under the Gauss-Markov assumptions. WLS corrects for heteroskedasticity by giving less weight to observations with a higher variance and more weight to observations with a lower variance. OLS assigns the same weight to all variances.

5 RESULTS AND DISCUSSION

After building the hedonic pricing model and conducting a multivariable linear regression on the data using both OLS and WLS, this section details whether the Airbnb listings analyzed fulfill the definition of a monopolistically competitive market (see Section 1.3). It also provides an analysis and comparison of the results of the regression using OLS and WLS, and a discussion concerning the implications of the findings.

5.1 Ordinary least squares (OLS)

The multivariable linear regression was run using SPSS on the hedonic pricing model specified in Section 4.2.

TABLE 5-1. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL (BARCELONA).

	Coefficients	Std. Error	t	Sig.
(Constant)	2.623	.068	38.324	.000
Cancellation Policy	.054	.007	8.062	.000
Listing Type	.698	.008	84.780	.000
Number of Reviews	-.088	.002	-39.066	.000
Overall Rating	.306	.019	16.396	.000
Bedrooms	.154	.010	15.697	.000
Bathrooms	.171	.009	18.547	.000
Max Guests	.356	.008	43.673	.000
Response Rate	.094	.013	7.026	.000
Response Time	.002	.001	1.080	.280
Superhost Status	.123	.021	5.791	.000
Number of Photos	.091	.005	18.582	.000
Instantbook Enabled	.036	.007	4.772	.000
Professional Status	-.033	.006	-5.360	.000
Distance from City Center	-.146	.004	-35.590	.000

TABLE 5-2. MODEL SUMMARY (BARCELONA).

Number of observations	17819
R	.866
R^2	.750
Adjusted R^2	.750
Std. Error of the Estimate	.3614
R^2 Change	.750
F Statistic	3824.286
Degrees of Freedom	14
Sig. *F* Change	.000

Table 5-2 indicates that the hedonic pricing model used to analyze the data for Barcelona is a good fit for the model. The R value of .866 and specifically the R^2 and Adjusted R^2 values of .750 indicate that the data are fitted very well to the regression line. In other words, the model explains approximately 75% of the variability of the data around the regression line.

Concerning the *F* Statistic, the p-value indicates that the value is significant at least at the conventional 5% level ($p < 0.05$).

The F-test is another measure indicating the overall significance of the model and tests whether the hedonic pricing model is a better fit for the data compared to a model which contains no explanatory variables. The null hypothesis is that the model with no explanatory variables using only the constant is as good a fit for the data as the hedonic pricing model. To determine whether we fail to reject the null hypothesis, the *F* Statistic must be lower than the critical F-Value for the model which depends on degrees of freedom and number of observations.

$$H_0: F \leq 1.692 \tag{5}$$

$$H_1: F > 1.692 \tag{6}$$

Given the *F* Statistic of 3824.276 for the hedonic pricing model, we reject H_0 in favor of H_1 and conclude that the hedonic pricing model is a better fit for the data than the intercept-only model.

TABLE 5-3. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL (MADRID).

	Coefficients	Std. Error	t	Sig.
(Constant)	2.795	.121	23.076	.000
Cancellation Policy	.014	.009	1.526	.127
Listing Type	.699	.011	62.665	.000
Number of Reviews	-.072	.003	-22.027	.000
Overall Rating	.301	.027	11.151	.000
Bedrooms	.345	.013	25.956	.000
Bathrooms	.152	.013	11.304	.000
Max Guests	.205	.010	19.611	.000
Response Rate	.026	.025	1.044	.297
Response Time	.006	.002	3.063	.002
Superhost Status	.119	.013	8.946	.000
Number of Photos	.080	.006	12.590	.000
Instantbook Enabled	-.007	.010	-.686	.493
Professional Status	.002	.008	.253	.800
Distance from City Center	-.115	.004	-26.627	.000

TABLE 5-4. MODEL SUMMARY (MADRID).

Number of observations	8565
R	.858
R^2	.736
Adjusted R^2	.736
Std. Error of the Estimate	.3514
R^2 Change	.736
F Statistic	1702.319
Degrees of Freedom	14
Sig. *F* Change	.000

Table 5-4 indicates that the hedonic pricing model used to analyze the data for Madrid is a good fit for the model. The R value of .858 and specifically the R^2 and Adjusted R^2 values of .736

indicate that the data are fitted very well to the regression line. In other words, the model explains approximately 73.6% of the variability of the data around the regression line.

Concerning the *F* Statistic, the p-value indicates that the value is significant at least at the conventional 5% level ($p < 0.05$). To determine whether we fail to reject the null hypothesis, the *F* Statistic must be lower than the critical F-Value for the model which depends on degrees of freedom and number of observations.

$$H_0: F \leq 1.693 \tag{7}$$

$$H_1: F > 1.693 \tag{8}$$

Given the *F* Statistic of 1702.319 for the hedonic pricing model, we reject H_0 in favor of H_1 and conclude that the hedonic pricing model is a better fit for the data than the intercept-only model.

TABLE 5-5. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL (SEVILLE).

	Coefficients	Std. Error	t	Sig.
(Constant)	2.588	.192	13.474	.000
Cancellation Policy	.091	.015	6.096	.000
Listing Type	.592	.020	29.767	.000
Number of Reviews	-.088	.005	-16.664	.000
Overall Rating	.331	.053	6.299	.000
Bedrooms	.243	.022	10.848	.000
Bathrooms	.149	.021	6.992	.000
Max Guests	.294	.019	15.296	.000
Response Rate	.054	.039	1.379	.168
Response Time	.004	.003	1.258	.208
Superhost Status	.069	.025	2.781	.005
Number of Photos	.066	.012	5.590	.000
Instantbook Enabled	.020	.018	1.120	.263
Professional Status	-.038	.014	-2.669	.008
Distance from City Center	-.149	.010	-15.474	.000

TABLE 5-6. MODEL SUMMARY (SEVILLE).

Number of observations	3291
R	.837
R^2	.700
Adjusted R^2	.699
Std. Error of the Estimate	.3732
R^2 Change	.700
F Statistic	545.840
Degrees of Freedom	14
Sig. *F* Change	.000

Table 5-6 indicates that the hedonic pricing model used to analyze the data for Seville is a good fit for the model. The R value of .837 and specifically the R^2 and Adjusted R^2 values of .700 and .699 respectively indicate that the data are fitted very well to the regression line. In other words, the model explains approximately 69.9% of the variability of the data around the regression line.

Concerning the *F* Statistic, the p-value indicates that the value is significant ($p < 0.05$) at the conventional 5% level. To determine whether we fail to reject the null hypothesis, the *F* Statistic must be lower than the critical F-Value for the model which depends on degrees of freedom and number of observations.

$$H_0: F \leq 1.695 \tag{9}$$

$$H_1: F > 1.695 \tag{10}$$

Given the *F* Statistic of 545.840 for the hedonic pricing model, we reject H_0 in favor of H_1 and conclude that the hedonic pricing model is a better fit for the data than the intercept-only model.

Overall, with regards to goodness of fit for the data, the model summaries as outlined in Tables 5-2, 5-4, and 5-6 for the hedonic pricing model in all three cities were similar. R, R^2 and Adjusted R^2 values were all relatively similar as were the values of the standard error of the estimates. Concerning the F-test, the null hypothesis was rejected in all three cities in favor of the alternative hypothesis, strengthening the model validity and robustness of the model as currently specified.

Interpretation of the regression coefficients for each individual city can be found in Section 5.3.

5.2 Weighted least squares (WLS)

Due to the inconclusive results of the heteroskedasticity tests of the model (see Section 3.3), it was decided that a WLS regression would be performed on the data and the results would be compared with OLS regression results from Section 5.1. In order to perform a WLS regression on the data set in SPSS, the WLS weight for the regression was calculated. According to Wooldridge (2012), "it is best to specify weights that are proportional to the inverse of the variance as it allows for the interpretation of weighted least squares estimates in the original model".

To calculate the WLS weight, first, the absolute value of the unstandardized residuals from the OLS were calculated. Then, the regression was rerun with the absolute value of the unstandardized residuals as the dependent variable and the unstandardized predicted values were calculated in order to capture the effects of the explanatory variables on the unstandardized residuals. Finally, the weights for WLS were calculated by dividing 1 by the squared values of the unstandardized predicted values. They are squared to minimize the effect of potential outliers by assigning less weight to these values.

TABLE 5-7. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING WLS (BARCELONA).

Coefficients[a,b]

	Coefficients	Std. Error	t	Sig.
(Constant)	2.478	.065	38.340	.000
Cancellation Policy	.048	.006	7.350	.000
Listing Type	.704	.007	99.000	.000
Number of Reviews	-.087	.002	-41.342	.000
Overall Rating	.427	.021	20.818	.000
Bedrooms	.090	.009	9.996	.000
Bathrooms	.187	.010	19.243	.000
Max Guests	.367	.008	48.136	.000
Response Rate	.083	.012	6.699	.000
Response Time	.001	.001	.854	.393
Superhost Status	.102	.018	5.846	.000
Number of Photos	.097	.005	21.170	.000
Instantbook Enabled	.029	.007	4.422	.000
Professional Status	-.022	.006	-4.007	.000
Distance from City Center	-.144	.004	-38.039	.000

a. Dependent Variable: Price

b. Weighted Least Squares Regression - Weighted by WLS_weight

TABLE 5-8. MODEL SUMMARY WLS (BARCELONA).

Number of observations	17819
R	.870
R^2	.757
Adjusted R^2	.757
Std. Error of the Estimate	1.3045
R^2 Change	.757
F Statistic	3958.126
Degrees of Freedom	14
Sig. *F* Change	.000

Compared to Table 5-2, Table 5-8 indicates that the hedonic pricing model using WLS to analyze the data for Barcelona is also a good fit for the model and that the values compared with OLS are similar. The R value of .870 and specifically the R^2 and Adjusted R^2 values of .757 indicate that the data is a slightly better fit using WLS than OLS. The model using WLS explains approximately 75.7% of the variability of the data around the regression line compared to 75% using OLS.

Concerning the *F* Statistic, the p-value indicates that the value is significant ($p < 0.05$) at the conventional 5% level and the *F* Statistic (3958.126) is higher than the critical *F*-value for Barcelona (1.692). Therefore, we reject H_0 in favor of H_1 (see Section 5.1) and conclude that the hedonic pricing model with WLS is a better fit for the data than the intercept-only model.

Comparing the coefficients for the individual explanatory variables, WLS produces similar values compared to OLS with the exception of “Overall Rating” and “Bedrooms”. WLS estimates produce a somewhat higher value for “Overall Rating” than OLS (0.427 vs. 0.306). For “Bedrooms”, it is the opposite (0.09 vs. 0.154).

TABLE 5-9. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING WLS (MADRID).

	Coefficients	Std. Error	t	Sig.
(Constant)	2.577	.129	20.017	.000
Cancellation Policy	.028	.008	3.386	.001
Listing Type	.668	.010	66.387	.000
Number of Reviews	-.071	.003	-24.566	.000
Overall Rating	.444	.032	13.943	.000
Bedrooms	.273	.013	21.648	.000
Bathrooms	.201	.015	13.647	.000
Max Guests	.221	.010	22.570	.000
Response Rate	.023	.027	.849	.396
Response Time	.005	.002	3.383	.001
Superhost Status	.112	.010	10.695	.000
Number of Photos	.086	.006	14.372	.000
Instantbook Enabled	-.004	.009	-.454	.650
Professional Status	.008	.007	1.085	.278
Distance from City Center	-.109	.004	-27.894	.000

TABLE 5-10. MODEL SUMMARY WLS (MADRID).

Number of observations	8565
R	.847
R^2	.718
Adjusted R^2	.717
Std. Error of the Estimate	1.2895
R^2 Change	.718
F Statistic	1553.465
Degrees of Freedom	14
Sig. *F* Change	.000

Compared to Table 5-4, Table 5-10 indicates that the hedonic pricing model using WLS to analyze the data for Madrid is also a good fit for the model and that the values compared with OLS are similar. In contrast to Barcelona, the R value of .847 and specifically the R^2 and Adjusted R^2

values of .718 and .717 respectively indicate that the data does not fit as well using WLS compared to OLS. The model using WLS explains approximately 71.7% of the variability of the data around the regression line compared to 73.6% using OLS.

Concerning the *F* Statistic, the p-value indicates that the value is significant ($p < 0.05$) at the conventional 5% level and the *F* Statistic (1553.465) is higher than the critical *F*-value for Madrid (1.693). Therefore, we reject H_0 in favor of H_1 (see Section 5.1) and conclude that the hedonic pricing model with WLS is a better fit for the data than the intercept-only model.

Comparing the coefficients for the individual explanatory variables, WLS produces similar values compared to OLS with the exception of "Cancellation Policy", "Bathrooms", "Overall Rating", and "Bedrooms". Similar to Barcelona, WLS estimates produce a somewhat higher value for "Overall Rating" than OLS (0.444 vs. 0.301). The same issue found in Barcelona for "Bedrooms" (lower value compared to OLS) also applies to Madrid (0.273 vs. 0.345). Using WLS, "Cancellation Policy" is found to be statistically significant and the value to be twice that of the OLS estimate (0.028 vs 0.014) and "Bathrooms" also exhibits a difference in value (0.201 vs. 0.152).

TABLE 5-11. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING WLS (SEVILLE).

	Unstandardized Coefficients			
	B	Std. Error	t	Sig.
(Constant)	2.397	.191	12.526	.000
Cancellation Policy	.088	.012	7.334	.000
Listing Type	.607	.017	35.645	.000
Number of Reviews	-.078	.004	-17.861	.000
Overall Rating	.432	.058	7.496	.000
Bedrooms	.183	.020	9.331	.000
Bathrooms	.215	.023	9.440	.000
Max Guests	.264	.017	15.940	.000
Response Rate	.059	.039	1.515	.130
Response Time	.005	.002	1.859	.063
Superhost Status	.069	.016	4.288	.000
Number of Photos	.071	.010	6.831	.000
Instantbook Enabled	.013	.015	.914	.361
Professional Status	-.025	.011	-2.246	.025
Distance from City Center	-.143	.008	-18.087	.000

TABLE 5-12. MODEL SUMMARY WLS (SEVILLE).

Number of observations	3291
R	.824
R^2	.678
Adjusted R^2	.677
Std. Error of the Estimate	1.2898
R^2 Change	.678
F Statistic	493.708
Degrees of Freedom	14
Sig. *F* Change	.000

Compared to Table 5-6, Table 5-12 indicates that the hedonic pricing model using WLS to analyze the data for Seville is also a good fit for the model and that the values compared with OLS are similar. Similar to Madrid, the R value of .824 and specifically the R^2 and Adjusted R^2 values of .678 and .677 respectively indicate that the data does not fit as well using WLS compared to OLS. The model using WLS explains approximately 67.7% of the variability of the data around the regression line compared to 69.9% using OLS.

Concerning the *F* Statistic, the p-value indicates that the value is significant ($p < 0.05$) at the conventional 5% level and the *F* Statistic (493.708) is higher than the critical *F*-value for Seville (1.695). Therefore, we reject H_0 in favor of H_1 (see Section 5.1) and conclude that the hedonic pricing model with WLS is a better fit for the data than the intercept-only model.

Comparing the coefficients for the individual explanatory variables, WLS produces similar values compared to OLS with the exception of “Bathrooms”, “Overall Rating”, and “Bedrooms”. Similar to Barcelona and Madrid, WLS estimates produce a somewhat higher value for “Overall Rating” than OLS (0.432 vs. 0.331). The same issue found in Barcelona and Madrid for “Bedrooms” (lower value compared to OLS) also applies to Seville (0.183 vs. 0.243). Similar to Madrid, “Bathrooms” also exhibits a noticeable difference in value when estimated with WLS compared to OLS (0.215 vs. 0.149).

Overall, from a model summary perspective, the hedonic pricing model using WLS has a similar goodness of fit in all three cities compared to the hedonic pricing model using OLS. The one main difference between WLS and OLS is the values for the standard error of the estimates across all three cities. While they are similar across all three cities in both cases, for WLS, they are higher than for OLS (approximately 1.3 vs. approximately 0.35) however this is to be expected.

From an overall coefficient estimates perspective, it is clear that the hedonic pricing model using WLS determines that “Overall Rating” has a larger influence on price and that “Bedrooms” has a smaller influence on price compared to OLS estimates across all three cities. “Bathrooms” using WLS has a higher influence on price in all three cities compared to OLS but the influence in Barcelona is more muted compared to Madrid and Seville. All other coefficient estimates are fairly similar. Given that the estimates of the hedonic pricing model using WLS produces very similar results compared to OLS with the exception of the previously mentioned explanatory variables, the interpretation of the results and the findings to be drawn from both methods would be similar. The only variation would be the degree to which the previously mentioned explanatory variables affect price.

5.3 Discussion

It is clear from a cursory analysis of the prices of the Airbnb listings in the data set that there is no equilibrium market price for Airbnb accommodation in Barcelona, Madrid, or Seville given the wide range of prices (see Tables 3-2, 3-3, and 3-4). This suggest that Airbnb accommodations are differentiated and thus heterogeneous products indicating that the market for Airbnb guests and hosts in these three cities fulfills the criteria of a monopolistically competitive market (see Section 1.3).

As explained in Section 3.2, the hedonic pricing model is a log-log model. Therefore, the results should be interpreted as elasticities. An exception which should be mentioned is the interpretation of the coefficients for dummy explanatory variables. When interpreting dummy variables, the results should be considered using a log-level model where an increase in x should be interpreted as a regime shift as opposed to a percentage increase as is the standard for a log-log model. Ex-ante algebraic signs are as expected with a few minor exceptions.

As Sections 5.1 and 5.2 demonstrated, the hedonic pricing model used for this paper is a good fit for the data. Results using either OLS or WLS are fairly similar except for certain exceptions where WLS differs from OLS with regards to the degree in which a certain explanatory variable has an effect on price. Other than that, the results from the hedonic pricing model across all three cities are fairly consistent. No city has substantially different results from the other cities. While certain generalizations could be made that would be applicable to all three cities, it is important to note that each city has its own individual particularities that are worth exploring in more detail.

From Tables 5-1, 5-3, and 5-5, as expected from the current literature, it is clear that explanatory variables such as ‘Bedrooms’, ‘Bathrooms’, ‘Listing Type’ and ‘Max Guests’ have a strong positive influence on price given that they are clear indicators of the overall size of the accommodation (Chen & Xie, 2017; Gibbs et al., 2017; Magno et al., 2018; Önder et al., 2018). The larger an accommodation is, the higher that price will be as is normally the case in the traditional accommodation industry as well.

The other explanatory variables to have a positive effect on price are the 'Overall Rating' of the accommodation as well as the 'Superhost Status' of the host. For every percentage of increase in 'Overall Rating', the price increases by a minimum of 0.3% across all three cities which supports the findings of Teubner et al. (2017). As was explained in Section 5.2, this effect is more pronounced using WLS. The relationship between 'Overall Rating' and price should be quite clear as it is a reflection of the objective overall quality of the accommodation as determined by previous guests so it is not a surprise that accommodations with higher ratings are more expensive. A potential guest is normally willing to pay more for the higher likelihood of being satisfied. Also, paying a higher price for an accommodation that has a higher rating is potentially done to mitigate risk. This may also be reflected in the explanatory variables 'Superhost Status' and 'Number of Photos'. If a host is a Superhost, it can be seen in the results that there is an increase in price although the effect is smaller than 'Overall Rating'. This could be interpreted as the host pricing in their Airbnb recognized high level of customer service, which guests are willing to pay for as it is also a signal of quality and reliability which would mitigate their own risk. Regarding the 'Number of Photos' explanatory variable, which across all three cities has a value of less than 0.1, while assessing the objective quality of the photos of an Airbnb listing is beyond the scope of this thesis (see Section 2.3 for previous literature addressing this topic), it is assumed that a high number of photos would allow a host the opportunity to differentiate their accommodation from their competitors and allow them to charge higher prices. From the guest's point of view, a higher number of photos may provide more extensive visual documentation regarding the state and quality of the accommodation, assuring that at the very least when a guest arrives at the accommodation, their expectations are met given the large amount of information they had access to.

Looking at explanatory variables that have a negative effect on price, there are mainly two: 'Number of Reviews' and 'Distance from City Center'. Similar to Wang & Nicolau (2017) but contrary to Teubner et al. (2017), an increase in 'Number of Reviews' was found to have a negative effect on price of at least 0.071. As explained by the authors, this may be due to the fact that overall, Airbnb accommodations are booked because they are often cheaper than alternatives in the traditional accommodation industry. In addition, cheaper Airbnb listings are booked more often than more expensive ones resulting in more reviews. This would explain why an increase in the amount of reviews would result in a negative effect on price in the model. While the authors provide a more detailed analysis of this explanatory variable, the replication of which is beyond the scope of this thesis, it is quite likely that their findings, which are that there is a structural and unequal distribution of bookings across the range of Airbnb prices, are likely the cause of this negative effect on price. This would explain the fairly consistent results for 'Number of Reviews' found across all three cities.

The explanatory variable 'Distance from the City Center' was found to have an even stronger negative effect on price than 'Number of Reviews'. This negative effect on price supports the findings of Önder et al. (2018), Wang & Nicolau (2017), and Zhang et al. (2017). The further an Airbnb accommodation is from the city center in Barcelona, Madrid, and Seville (see Section 3.2), the lower the overall price of the accommodation. The results from Tables 5-1, 5-3, 5-5, as well as Tables 5-7, 5-9, and 5-11, indicate that the negative effect on price is at least 0.109. This negative effect on price can also be seen in the traditional accommodation industry (Bull, 1994) so it is no surprise that this relationship was found in the three cities under analysis or in previous Airbnb pricing literature. Proximity to a city center is a source of convenience for guests who as a result of staying more centrally, will have an easier ability to access points of interests. It is clear that hosts can charge a higher price for this convenience and that they are likely factoring this into their pricing.

The only explanatory variable to exhibit no effect on price is 'Response Time'. The highest value is for Madrid (0.006) and it is the only value considered to be statistically significant for all three cities. It was expected that a longer response time would indicate a lower level of customer service which would result in a lower price but this does not seem to be the case.

The previous explanatory variables mentioned were the ones that were found to have the strongest effect on price, whether positive or negative. All the other explanatory variables, except the previously mentioned 'Response Time', in the hedonic pricing model have a value of less than 0.1 indicating relatively weak effects on price compared to the other explanatory variables. But, however negligible their effects, these explanatory variables should be further analyzed as even their lack of relationship or effect on price may provide insights of their own. Notable city specific particularities in results will also be detailed when they are deemed insightful. For the purposes of simplicity and given the similar results of OLS and WLS, going forward, the following interpretation of the results will only consider the results from OLS.

For Barcelona, 'Bathrooms', 'Max Guests', 'Response Rate', 'Superhost Status', 'Number of Photos' 'Instantbook Enabled' have the highest values compared to the other two cities. While 'Bathrooms' has the highest value (0.172), it is only 0.019 higher than the next closest value, Madrid, and 0.021 higher than Seville. However, 'Max Guests' has a much larger variation between the three cities. The coefficient indicates a 0.356 positive effect on price in Barcelona compared to 0.205 for Madrid and 0.294 for Seville. It could be that there is a higher demand for group accommodation in Barcelona and hosts are pricing this need into their listings.

Regarding the explanatory variables that address a host's behavior, 'Response Rate' was found to have a positive influence on price across all three cities with —the more frequent a host responded to a potential guest, the higher the price of the accommodation. The coefficient for Barcelona (0.094) is the highest and substantially higher than Madrid (0.026) and Seville (0.054) although it should be mentioned that for Madrid and Seville, 'Response Rate' is found to be not statistically significant. The reason for the positive effect of 'Response Rate' could be that hosts value their perceived level of customer service and are incorporating these costs into the price.

Concerning 'Superhost Status', the coefficient is highest in Barcelona (0.123) and similar to Madrid (0.119). For Seville, it is (0.069). For 'Number of Photos', the effect on price is fairly consistent across all 3 cities with the coefficient for Barcelona being the highest (0.091). For 'Instantbook Enabled', while the coefficient for Barcelona is highest (0.036), the overall effect on price is minimal and not statistically significant for Madrid (-0.007) and Seville (0.02). It is clear from the results that the Instantbook feature, while it may slightly increase the risk to the host by not allowing them to vet potential guests, is not considered special and does not greatly influence price.

However, 'Professional Status' was found to have a small negative effect on price. The possible explanation for this result is that hosts who manage more than one listing are more likely to be using Airbnb to make a living and are motivated to price their listings to maximize occupancy in order to maximize revenues. Therefore, while they may incorporate professionalism into their prices, they cannot unilaterally raise their prices to such a degree where their listings are deemed unattractive to potential guests. It is also important to note that 'Professional Status' unlike 'Superhost' is an explanatory variable taken from previous literature (see Section 3.2) and is not in effect something that is communicated to a guest or can be perceived as having value.

For Madrid, the highest coefficient values across all three cities are 'Listing Type' (0.699) and 'Bedrooms'. The value for 'Listing Type' is almost identical in Barcelona (0.698) so no city-specific particularity can be explained for this explanatory variable. 'Bedrooms' on the other hand, has a noticeably higher value in Madrid (0.345) than in Barcelona (0.154) and Seville (0.243). It could be that since Madrid is the capital of Spain, there are more business guests using Airbnb which increases the value of separate bedrooms since it is fairly standard to expect to stay in an individual room on a business trip.

For Seville, the highest coefficient values are 'Cancellation Policy' and 'Professional Status'. A stricter cancellation policy has the highest positive effect on price in Seville (0.091) which is higher than Barcelona (0.054) and Madrid (0.014). Essentially, the stricter the cancellation policy is, the higher the price of the accommodation. Gunter (2018) makes the assumption that a host who selects a stricter cancellation policy for their accommodation applies a similar level of strictness to themselves and implicitly communicates reciprocity with a guest to build trust. Along the same lines, this could also be interpreted as the level of professionalism a host is approaching their listing on Airbnb and therefore, the self-perception of the level of their services drives a host to charge higher prices.

For 'Professional Status', those managing more than one Airbnb listing, the effect on price is slightly negative in Seville (-0.038) which is roughly similar to Barcelona (-0.033) and more negative than Madrid (0.002) although for Madrid it should be noted it is not statistically significant. This essentially supports the findings of Gibbs et al. (2017) who found only a slight positive effect on price in Montreal (3.5%). Simply stated, the professional status of a host does not affect price to any noticeable degree. This is understandable considering that if hosts are using Airbnb to make a living, their main preoccupation would be to have a high occupancy rate and not necessarily to charge the highest price. In fact, the slightly negative values for Barcelona and Seville are indicative of what would be expected from a host in this situation, which is that they would potentially lower their prices in order to increase the level of occupancy. However, the values from this analysis are too low to make any definitive conclusion concerning this issue.

While the regression results are fairly similar across all three cities and that one of the objectives of this thesis was to obtain results which may be applicable to the entire country of Spain, it is in the opinion of the author of this thesis that it is unadvisable to generalize to such a degree given the temporal and geographical limitations of the data set and the results. It is also important to note that major metropolitan areas are likely to exhibit certain characteristics which are different compared to rural areas. It is possible that the findings of this thesis may be applicable to other metropolitan areas in Spain, however, prior to making these conclusions, it is preferable that data be obtained from these other metropolitan areas and the methodology as outlined be applied to it to determine whether in fact the results are similar.

5.4 Robustness checks

In order to check whether the model is robustness, two robustness checks using OLS were performed: one with an alternative dependent variable measuring the price of an Airbnb accommodation (‘Published Nightly Rate’), and another using only one explanatory variable indicative of the size of the accommodation (‘Bedrooms’) leaving out ‘Listing Type’, ‘Bathrooms’, and ‘Max Guests’ from the original hedonic pricing model. As was done with ‘Price’ in the original model, the natural logarithm of ‘Published Nightly Rate’ was taken prior to running the robustness check.

The results of the robustness checks are fairly similar indicating that the original model is correctly specified and is valid and robust. Certain algebraic signs for various explanatory variables whose values were near 0 in the original model switched in the robustness check models.

For the robustness check using only ‘Bedrooms’, all the coefficients for the explanatory variables are higher compared to the original model, especially ‘Bedrooms’, although this is to be expected since the previously mentioned explanatory variables were removed. Robustness check model summaries and estimation results can be found in Tables C-1, C-2, C-3, C-4, C-5, C-6, C-7, C-8, C-9, C-10, C-11, and C-12 in Appendix C.

5.5 Summary

Overall, the hedonic pricing model as outlined in Section 4.2 is a good fit for the data. Regression results using OLS and WLS are fairly similar and consistent across all three cities. While using WLS should have corrected the original hedonic pricing model for heteroskedasticity, the results did not indicate any substantial differences with regards to interpretation.

Similar to the findings from previous literature, the explanatory variables which indicate the size or functionality of the Airbnb accommodation, specifically 'Bedrooms', 'Bathrooms', 'Listing Type', and 'Max Guests', have the strongest positive effect of price in all three cities and are the explanatory variables which have the strongest influence on price. In addition, 'Overall Rating' also has a strong positive effect on price. 'Number of Reviews' and 'Distance from City Center' have a slight to strong negative effect on price while 'Response Time' has essentially no effect on price.

As for city-specific particularities, there are very few since the regression results across all three cities are fairly consistent. Certain coefficients in specific cities may have higher values than others but not to such a great or consistent degree where specific conclusions can be made. As for whether the results could be applicable to the entire country, it is in the opinion of the author that more data and analysis are required prior to making this claim.

The robustness checks, one using a different dependent variable and another leaving out specific explanatory variables, yielded similar results which in turn validate the robustness of the original hedonic pricing model used to analyze the data in this thesis and to answer the specific research question: how profile characteristics affect Airbnb accommodation prices.

6 CONCLUSION

This chapter outlines the overall summary of the thesis as well as its contribution to knowledge. Details are provided concerning the implications for relevant stakeholders as well as the limitations of the research conducted in this thesis, and the potential for future research.

6.1 Thesis summary

It is clear that the sharing economy has dramatically affected the way how economic agents interact in a market by providing them numerous uses for an asset that were previously not possible. Airbnb has significantly disrupted the accommodation industry by providing individuals with a platform to act as accommodation providers. It is also one of the most important companies operating in the sharing economy given its valuation, market size, and market power. Previous research about the sharing economy have explored Airbnb from various points of view whether it be socio-economical, political, or regulatory.

The microeconomic theory underpinning the Airbnb market is the concept of monopolistic competition. It was decided that an interesting way of analyzing the scraped Airbnb data from AirDNA would be to apply a hedonic pricing model to see which profile characteristics of an Airbnb listing affect price. This hedonic pricing model using multivariable linear regression was applied to data from Barcelona, Madrid, and Seville, with the aim of finding insights which may be applicable to the entire country. Hedonic pricing models have been used for decades to analyze and determine specific elements which make up the price of real estate but have also been applied in various other areas including the traditional accommodation industry. Given the similarities between Airbnb and the accommodation industry, this was deemed an appropriate methodology to use.

The data set was prepared and analyzed to make sure the most appropriate explanatory variables were selected for inclusion in the hedonic pricing model while also conforming to the assumptions required for a valid interpretation of multivariable linear regression results. Various variables were found to be not statistically significant in certain cities but were nonetheless included in the hedonic pricing model due to their importance in previous literature and their potential economic and substantive significance. The natural logarithms of all continuous variables was taken and certain explanatory variables were recoded into dummy variables as outlined in previous research and for ease of interpretation.

Preliminary regression analysis and statistical pre-testing uncovered certain issues which required the need to perform the regression twice: once using OLS and another using WLS due to the inconclusive tests which could not definitively fail to reject the null hypothesis of homoskedasticity.

The results of the regression analysis indicate that explanatory variables which are indicative of the size and functionality of the accommodation have the strongest positive influence on price while the number of reviews and distance from the city center have the strongest negative influence on price. City-specific results were not variable enough to form specific conclusions and more data and analysis are required prior to making generalizations which could be applicable to the country of Spain.

6.2 Contribution to knowledge

As the application of a hedonic pricing model in the sharing economy, specifically on Airbnb accommodation prices, is relatively new, it is the desire of the author that this thesis will add to the rapidly growing body of research on this topic and be a contribution to the accommodation pricing literature as well as Airbnb pricing research. The quantitative analysis provides a thorough framework in which to view Airbnb hosts and their pricing strategies. It is also one of the first papers to apply spatial distances as an explanatory variable in a hedonic pricing model for the sharing economy. The framework also held up after several robustness checks indicating that the model specifications are correct and could be used for future research on other cities or countries. It is also hoped that the specific focus on major metropolitan as well as major tourist destinations in Spain will contribute to the tourism research on Spain, specifically from the sharing economy perspective.

6.3 Implications for relevant stakeholders

Relevant stakeholders for this research are hospitality researchers and professionals, Airbnb hosts, and Airbnb. Previous research on various Airbnb markets have indicated that host pricing is not necessarily optimal on the Airbnb platform and that most hosts do not use the “Smart pricing” feature as explained in Section 2.6. The findings from this thesis could help Airbnb hosts optimize the pricing of their listings which could increase their occupancy rates and maximize their revenues. With regards to Airbnb, the findings of this thesis could help the company further improve and optimize their “Smart pricing” feature which would allow the company to maximize transactions thereby maximizing their revenues. As for hospitality researchers, it is hoped that this thesis provides some insight concerning the sharing economy dynamics in Spain.

6.4 Limitations

The research conducted in this thesis has various limitations. The first is geographical. While generalizations of findings have been made for Barcelona, Madrid, and Seville which may be applicable to the entire country of Spain, it is difficult to claim the applicability of the findings across the entire country without looking at the data from other metropolitan areas as well as rural areas.

The second limitation is temporal since the prices of Airbnb listings in the analysis were captured between 17.08.2015 and 02.08.2016. It should be noted that the market environment has likely changed since the capture of this data and is constantly evolving. For example, Airbnb hosts are now under stricter regulations in many regions of Spain including Barcelona, Madrid, and Seville. Hosts in these three cities are now required to register their Airbnb listing as "tourist accommodation" with the respective authority. The registration number must be displayed on an Airbnb listing (Airbnb, Inc., 2019). Hosts operating without a registration number may be fined or shut down. In addition, the city of Madrid recently approved a plan where hosts who rent out accommodation for a period of longer than 90 days per year need to obtain a special licence. In order to be granted a licence, the accommodation must have its own separate entrance. This is estimated to affect 95% of existing "professional" apartments (Urra, 2019).

These regulations may have altered the structure of the market and the way it operates, potentially changing the way each explanatory variable interacts with price. However, the fundamental methodology used to perform the analysis could be easily replicated and conducted on newly scraped data which would be an interesting topic for future research, more specifically, to see how the Airbnb market and prices have developed in the three Spanish cities.

Additional limitations include the lack of data on various other explanatory variables which may provide a more detailed analysis of market dynamics, such as competition between Airbnb listings and market segmentation of guests. Also, this thesis, given its quantitative focus and its reliance on economic modeling, does not take into account a host's behavior in setting their own prices, thus ignoring social and psychological factors affecting price which may be important.

6.5 Future research

While this thesis builds on the research that came before it, it also supports the relatively new area of research which is sharing economy pricing, more specifically, Airbnb pricing. The methodology of this thesis could be easily replicated and applied to different cities as well as entire countries. Applying the methodology on newer Airbnb data, as mentioned in Section 6.4, could also reveal specific insights concerning how increased regulations in certain cities, including the ones analyzed in this thesis, have changed the Airbnb market and the prices of Airbnb accommodations. It could also reveal how the pricing of Airbnb accommodations in Barcelona, Madrid, and Seville have developed.

Considering some of the limitations pointed out in Section 6.4, conducting a survey among Airbnb guests and hosts to obtain data concerning customer segmentation, preferences, and social and psychological factors affecting price, and combining this data with the quantitative approach as outlined in this thesis would allow for a more complete analysis of pricing as well as demand and supply-side sharing economy market dynamics.

Bibliography

Abidoye, R. B., & Chan, A. P. (2018). Improving property valuation accuracy: a comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal, 24*(1), 71-83.

Abrahao, B., Parigi, P., Gupta, A., & Cook, K. S. (2017). Reputation offsets trust judgments based on social biases among Airbnb users. *Proceedings of the National Academy of Sciences, 114*(37), 9848-9853.

Abrate, G., Capriello, A., & Fraquelli, G. (2011). When quality signals talk: Evidence from the Turin hotel industry. *Tourism Management, 32*(4), 912-921.

Airbnb, Inc. (2019, April 13). *Fast Facts*. Retrieved April 13, 2019, from Airbnb: https://press.airbnb.com/fast-facts/

Airbnb, Inc. (2019). *Responsible hosting in Spain*. Retrieved April 20, 2019, from Airbnb: https://www.airbnb.com/help/article/1388/responsible-hosting-in-spain#city

AirDNA, LLC. (2018, December 6). *About Us: Data Science meets Real Estate Investing in Vacation Rentals*. Retrieved from AirDNA: https://www.airdna.co/about

AirDNA, LLC. (2018, December 6). *Short-Term Rental Data Methodology - The AI Behind AirDNA*. Retrieved from AirDNA: https://www.airdna.co/blog/short-term-rental-data-methodology

Antonakis, J. (2017, February). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly, 28*(1), 5-21.

Aznar, J. P., Sayeras, J. M., Segarra, G., & Claveria, J. (2018). Airbnb landlords and price strategy: Have they learnt price discrimination from the hotel industry? Evidence from Barcelona. *International Journal of Tourism Sciences, 18*(1), 16-28.

Bajari, P., Fruehwirth, J. C., Kim, K., & Timmins, C. (2012). A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution. *American Economic Review, 102*(5), 1898-1926.

Balaguer, J., & Pernías, J. C. (2013). Relationship between spatial agglomeration and hotel prices. Evidence from business and tourism consumers. *Tourism Management, 36*, 391-400.

Bardhi, F., & Eckhardt, G. M. (2012, December 1). Access-Based Consumption: The Case of Car Sharing. *Journal of Consumer Research, 39*(4), 881-898.

Belk, R. (2014, August). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research, 67*(8), 1595-1600.

Benítez-Aurioles, B. (2018). The role of distance in the peer-to-peer accommodation market for tourist accommodation. *Tourism Economics, 24*(3), 237-250.

Blal, I., Singal, M., & Templin, J. (2018, July). Airbnb's effect on hotel sales growth. *International Journal of Hospitality Management, 73*, 85-92.

Brentari, E., Levaggi, R., & Zuccolotto, P. (2011). Pricing strategies for Italian red wine. *Food Quality and Preference, 22*(8), 725-732.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*(5), 1287-1294.

Brodeur, A., & Nield, K. (2018). An empirical analysis of taxi, Lyft and Uber rides: Evidence from weather shocks in NYC. *Journal of Economic Behavior and Organization, 152*, 1-16.

Bull, A. O. (1994). Pricing a Motel's Location. *International Journal of Contemporary Hospitality Management, 6*(6), 10-15.

Chen, L., Mislove, A., & Wilson, C. (2015). Peeking Beneath the Hood of Uber. *IMC '15 Proceedings of the 2015 Internet Measurement Conference* (pp. 495-508). Tokyo: ACM.

Chen, Y., & Xie, K. (2017). Consumer valuation of Airbnb listings: a hedonic pricing approach. *International Journal of Contemporary Hospitality Management, 29*(9), 2405-2424.

Cho, M., Lee, K., Choi, H., Hong, S., Park, C.-h., & Kim, H.-J. (2013). Introduction Strategy of Inter-corporate Sharing Economy for Small Businesses Competitiveness Reinforcement. *International Conference on Business Administration* (pp. 109-118). Tenerife: WSEAS.

Demary, V. (2015). *Competition in the sharing economy, IW Policy Paper, No. 19/2015.* Köln: Institut der deutschen Wirtschaft Köln.

Edelman, B., Luca, M., & Svirsky, D. (2017). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics, 9*(2), 1-22.

Erickson, K., & Sørensen, I. (2016). Regulating the sharing economy. *Internet Policy Review, 5*(2). doi:10.14763/2016.2.414

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management, 55*, 62-73.

Eurostat. (2018, December). *Tourism Statistics*. Retrieved April 13, 2019, from Statistics Explained: https://ec.europa.eu/eurostat/statistics-explained/index.php/Tourism_statistics

Fagerstrøm, A., Pawar, S., Sigurdsson, V., Foxall, G. R., & Yani-de-Soriano, M. (2017). That personal profile image might jeopardize your rental opportunity! On the relative impact of the seller's facial expressions upon buying behaviour on Airbnb. *Computers in Human Behavior, 72*, 123-131.

Felson, M., & Spaeth, J. L. (1978, March 1). Community Structure and Collaborative Consumption: A Routine Activity Approach. *American Behavioral Scientist, 21*(4), 614-624.

FLEXE, Inc. (2018, November 9). *About Us | FLEXE: On-Demand Warehouse Space*. Retrieved from FLEXE, Inc. Web site: https://www.flexe.com/company

Fon Wireless Limited. (2018, November 9). *Fon is the global WiFi network with millions of hotspots*. Retrieved from Fon Wireless Limited Web site: https://fon.com/

Fox, J. (2007, October 26). *ncv.test function*. Retrieved April 20, 2019, from R Documentation: https://www.rdocumentation.org/packages/car/versions/1.2-6/topics/ncv.test

Ganapati, S., & Reddick, C. G. (2018). Prospects and challenges of sharing economy for the public sector. *Government Information Quarterly, 35*(1), 77-87.

Geerts, W. (2018). *Top 100 City Destinations 2018.* Euromonitor International.

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2017, April 10). Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing, 35*(1), 46-56.

Gibbs, C., Guttentag, D., Gretzel, U., Yao, L., & Morton, J. (2018). Use of dynamic pricing strategies by Airbnb hosts. *International Journal of Contemporary Hospitality Management, 30*(1), 2-20.

Gobble, M. M. (2015). Regulating innovation in the new economy. *Research-Technology Management, 58*(2), 62-64.

Goodman, A. C. (1978, October). Hedonic prices, price indices and housing markets. *Journal of Urban Economics, 5*(4), 471-484.

Gunter, U. (2018). What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area. *Tourism Management, 66*, 26-37.

Gunter, U., & Önder, I. (2018). Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. *Tourism Economics, 24*(3), 270-293.

Guttentag, D. (2013, July 14). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism, 18*(12), 1192-1217.

Hill, D. (2015). How much is your spare room worth? *IEEE Spectrum, 52*(9), 32-58.

Hook, L. (2018, January 25). *Airbnb marks first full year of profitability in 2017 | Financial Times*. Retrieved August 15, 2018, from Financial Times: https://www.ft.com/content/96215e16-0201-11e8-9650-9c0ad2d7c5b5

Hou, L. (2018). Destructive sharing economy: A passage from status to contract. *Computer Law & Security Review, 34*(4), 965-976.

Hung, W.-T., Shang, J.-K., & Wang, F.-C. (2010). Pricing determinants in the hotel industry: Quantile regression analysis. *International Journal of Hospitality Management, 29*, 378-384.

Kakar, V., Voelz, J., Wu, J., & Franco, J. (2018, June). The Visible Host: Does race guide Airbnb rental rates in San Francisco? *Journal of Housing Economics, 40*, 25-40.

Karney, C. F. (2013). Algorithms for geodesics. *Journal of Geodesy, 87*(1), 43-55.

Kung, L.-C., & Zhong, G.-Y. (2017). The optimal pricing strategy for two-sided platform delivery in the sharing economy. *Transportation Research Part E, 101*, 1-12.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models* (4th ed.). Irwin: McGraw-Hill.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 132-157.

Lee, S. (2015). Quality differentiation and conditional spatial price competition among hotels. *Tourism Management, 46*, 114-122.

Lessig, L. (2008). *Remix: Making Art and Commerce Thrive in the Hybrid Economy.* Penguin.

Li, J., Moreno, A., & Zhang, D. J. (2016). *Pros vs Joes: Agent Pricing Behavior in the Sharing Economy.* University of Michigan. Ann Arbor: Ross School of Business. doi:10.2139/ssrn.2708279

Liao, W.-C., & Wang, X. (2012). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics, 21*(1), 16-27.

Linsombunchai, V., Gan, C., & Lee, M. (2004). House Price Prediction: Hedonic Price Model vs. Artificial Neural Network. *American Journal of Applied Sciences, 1*(3), 193-201.

Lutz, C., & Newlands, G. (2018). Consumer segmentation within the sharing economy: The case of Airbnb. *Journal of Business Research, 88*, 187-196.

Magno, F., Cassia, F., & Ugolini, M. M. (2018). Accommodation prices on Airbnb: effects of host experience and market demand. *The TQM Journal,, 30*(5), 608-620.

Martin, B., & Koo, B. (2009). Dynamic price dispersion in airline markets. *Transportation Research Part E: Logistics and Transportation Review, 45*(6), 1020-1029.

Martin, C. J. (2016). The sharing economy: A pathway to sustainability or a nightmarish form of neoliberal capitalism? *Ecological Economics, 121*, 149-159.

Martínez-Garmendia, J. (2010). Application of hedonic price modeling to consumer packaged goods using store scanner data. *Journal of Business Research, 63*(7), 690-696.

McConnell, K. E., & Strand, I. E. (2000). Hedonic Prices for Fish: Tuna Prices in Hawaii. *American Journal of Agricultural Economics, 82*(1), 133-144.

Merriam-Webster. (2019). *sharing*. Retrieved February 17, 2019, from Merriam-Webster: https://www.merriam-webster.com/dictionary/sharing

Mok, H. M., Chan, P. P., & Cho, Y.-S. (1995, January). A hedonic price model for private properties in Hong Kong. *The Journal of Real Estate Finance and Economics, 10*(1), 37–48.

Mooi, E., & Sarstedt, M. (2010). *A Concise Guide to Market Research.* Berlin: Springer.

Munkøe, M. (2017, January). Regulating the European Sharing Economy: State of Play and Challenges. *Intereconomics, 52*(1), 38-44.

Murillo, D., Buckland, H., & Val, E. (2017). When the sharing economy becomes neoliberalism on steroids: Unravelling the controversies. *Technological Forecasting & Social Change, 125*, 66-76.

Önder, I., Weismayer, C., & Gunter, U. (2018). Spatial price dependencies between the traditional accommodation sector and the sharing economy. *Tourism Economics*, 1-17.

Oses, N., Gerrikagoitia, J. K., & Alzua, A. (2016). Evidence of hotels' dynamic pricing patterns on an Internet distribution channel: the case study of the Basque Country's hotels in 2013–2014. *Information Technology & Tourism, 15*(4), 365-394.

Oskam, J., & Boswijk, A. (2016). Airbnb: the future of networked hospitality businesses. *Journal of Tourism Futures, 2*(1), 22-42.

Parente, R. C., Geleilate, J.-M. G., & Rong, K. (2018). The Sharing Economy Globalization Phenomenon: A Research Agenda. *Journal of International Management, 24*(1), 52-64.

Pesaran, M. H. (2004). General Diagnostic Tests for Cross Section Dependence in Panels. *Faculty of Economics, University of Cambridge* .

Pindyck, R. S., & Rubinfeld, D. L. (2018). *Microeconomics, Ninth Edition.* London: Pearson Education Limited.

Ravenelle, A. J. (2017). Sharing economy workers: selling, not sharing. *Cambridge Journal of Regions, Economy and Society, 10*, 281–295.

Roden, L. (2017, October 24). *Seville named best city to travel to in 2018 by Lonely Planet*. Retrieved April 13, 2019, from The Local: https://www.thelocal.es/20171024/seville-named-best-city-to-travel-to-in-2018-by-lonely-planet

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy, 82*(1), 34-55.

Rushe, D., & Helmore, E. (2019, April 11). *Uber warns it 'may not achieve profitability' as it aims for $100bn valuation*. Retrieved April 13, 2019, from The Guardian: https://www.theguardian.com/technology/2019/apr/11/uber-to-share-intimate-details-about-company-ahead-of-going-public

Sainaghi, R. (2011). Price determinants of individual hotels: evidence from Milan. *Tourism Review, 66*(4), 18-29.

Schamel, G. (2012). Weekend vs. midweek stays: Modelling hotel room rates in a small market. *International Journal of Hospitality Management, 31*, 1113-1118.

Smith, O. (2018, December 24). *Which was the world's most visited country in 2018?* Retrieved April 13, 2019, from The Telegraph: https://www.telegraph.co.uk/travel/lists/the-fastest-growing-travel-destinations-of-2018/

Somerville, H. (2018, February 1). *Airbnb says no listing in 2018, CFO to leave, names first COO*. Retrieved from Reuters: https://www.reuters.com/article/us-airbnb-ipo/airbnb-says-no-listing-in-2018-cfo-to-leave-names-first-coo-idUSKBN1FL6BI

TaskRabbit, Inc. (2018, November 9). *TaskRabbit connects you to safe and reliable help in your neighborhood*. Retrieved from TaskRabbit: https://www.taskrabbit.com/

Teubner, T., Hawlitschek, F., & Dann, D. (2017). Price determinants on Airbnb: How reputation pays off in the sharing economy. *Journal of Self-Governance and Management Economics, 5*(4), 53-80.

Thrane, C. (2007). Examining the determinants of room rates for hotels in capital cities: The Oslo experience. *Journal of Revenue and Pricing Management, 5*(4), 315-323.

Turo Inc. (2018, November 9). *Turo car sharing marketplace*. Retrieved from Turo: https://turo.com/

Urra, S. (2019, April 1). *Airbnb in Spain: The battle against vacation rentals spreads beyond Madrid*. Retrieved April 20, 2019, from El País: https://elpais.com/elpais/2019/03/28/inenglish/1553767032_107018.html

Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review, 23*(176), 88-93.

Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management, 62*, 120-131.

Weber, T. A. (2016). Product Pricing in a Peer-to-Peer Economy. *Journal of Management Information Systems, 33*(2), 573-596.

Witte, A. D., Sumka, H. J., & Erekson, H. (1979, September). An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets. *Econometrica, 47*(5), 1151-1173.

Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach* (5th ed.). Mason: Cengage Learning.

Xie, K. L., & Kwok, L. (2017). The effects of Airbnb's price positioning on hotel performance. *International Journal of Hospitality Management, 67*, 174-184.

Yang, Z. (2001). An application of the hedonic price model with uncertain attribute - The case of the People's Republic of China. *Property Management, 19*(1), 50-63.

Zervas, G., Proserpio, D., & Byers, J. W. (2017, October). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research, 54*, 687–705.

Zhang, T. C., Jahromi, M. F., & Kizildag, M. (2018, April). Value co-creation in a sharing economy: The end of price wars? *International Journal of Hospitality Management, 71*, 51-58.

Zhang, Z., Chen, R. J., Han, L. D., & Yang, L. (2017). Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. *Sustainability, 9*(9), 1635. doi:10.3390/su9091635

Zhang, Z., Ye, Q., & Law, R. (2011). Determinants of hotel room price: An exploration of travelers' hierarchy of accommodation needs. *International Journal of Contemporary Hospitality Management, 23*(7), 972-981.

APPENDIX A

TABLE A-1. LIST OF CHARACTERISTICS OF EACH PROPERTY IN AIRDNA DATA SET (IN US DOLLARS WHERE APPLICABLE).

Characteristic	Description	Data type
Property ID	A unique ID number of the property	Number
Host ID	A unique ID of the host of the property	Number
Property Type	The type of property (e.g. apartment, house, B&B)	String
Listing Title	The title of the listing on Airbnb	String
Listing Type	The type of the listing (e.g. private room, shared room, or entire home/apt)	String
Created Date	Date the listing was created	Date
Last Scraped Date	Date the listing information was last captured by AirDNA	Date
Country	Country where the property is located	String
State	State where the property is located (US Only)	String
City	City where the property is located	String
Zipcode	Zipcode where the property is located (US Only)	String
Neighborhood	Name of the neighborhood where the property is located	String
Metropolitan Statistical Area	MSA of the area (US only)	String
Average Daily Rate	Average daily price of the property including fees	Number
Annual Revenue LTM	Annual revenue of the property for the last 12 months	Number

Occupancy Rate LTM	Occupancy rate of the property in the last 12 months	Number
Number of Bookings LTM	Number of bookings of the property in the last 12 months	Number
Number of Reviews	Total number of reviews of the property	Number
Overall Rating	Overall rating of the property on a scale of 1 to 5	Number
Bedrooms	Number of bedrooms of the property	Number
Bathrooms	Number of bathrooms of the property	Number
Max Guests	Number of maximum guests for the property	Number
Calendar Last Updated	Date when the calendar was last updated	Date
Response Rate	Rate of response of the host to a message from guests on a scale from 0 to 100	Number
Response Time	Amount of time it took a host to respond to a message from a guest in minutes	Number
Superhost	Indication of whether a host has superhost status indicated by ‘t’ for True and ‘f’ for False	String
Cancellation Policy	Cancellation policy of the property (e.g. strict, flexible, or moderate)	String
Security Deposit	Amount of security deposit required by the host	Number
Cleaning Fee	Amount of cleaning fee in addition to the rental price	Number
Extra People Fee	Amount charged for each additional guest	Number
Published Nightly Rate	Price charged to guest per night	Number

Published Monthly Rate	Price charged to guest per month	Number
Published Weekly Rate	Price charged to guest per week	Number
Check-in Time	Time of day when a guest can check-in to the property	String
Checkout Time	Time of day when a guest must checkout of a property	String
Minimum Stay	Minimum amount of nights a guest must stay in order to book the property	Number
Count Reservation Days LTM	Number of days a guest has booked the property in the last 12 months	Number
Count Available Days LTM	Number of days where a guest can book the property for the last 12 months	Number
Count Blocked Days LTM	Number of days blocked by the host where a guest is unable to book the property in the last 12 months	Number
Number of Photos	Number of photos of the property in the Airbnb listing	Number
Instantbook Enabled	Whether instantbook is enabled on the property listing by the host which is indicated by a ‘No’ or a ‘Yes’	String
Listing URL	URL of the property listing	String
Listing Main Image URL	Main image of the property listing	String
Latitude	Latitude coordinates of the property	Number
Longitude	Longitude coordinates of the property	Number

SOURCE: AIRDNA (HTTPS://WWW.AIRDNA.CO/)

TABLE A-2. LIST OF EXCLUDED VARIABLES FROM AIRDNA DATA SET (IN US DOLLARS WHERE APPLICABLE).

Characteristic	Description	Data type
Property Type	The type of property (e.g. apartment, house, B&B)	String
Listing Title	The title of the listing on Airbnb	String
Created Date	Date the listing was created	Date
State	State where the property is located (US Only)	String
Zipcode	Zipcode where the property is located (US Only)	String
Neighborhood	Name of the neighborhood where the property is located	String
Metropolitan Statistical Area	MSA of the area (US only)	String
Annual Revenue LTM	Annual revenue of the property for the last 12 months	Number
Occupancy Rate LTM	Occupancy rate of the property in the last 12 months	Number
Number of Bookings LTM	Number of bookings of the property in the last 12 months	Number
Calendar Last Updated	Date when the calendar was last updated	Date
Security Deposit	Amount of security deposit required by the host	Number
Extra People Fee	Amount charged for each additional guest	Number
Published Monthly Rate	Price charged to guest per month	Number
Published Weekly Rate	Price charged to guest per week	Number

Check-in Time	Time of day when a guest can check-in to the property	String
Checkout Time	Time of day when a guest must checkout of a property	String
Minimum Stay	Minimum amount of nights a guest must stay in order to book the property	Number
Count Reservation Days LTM	Number of days a guest has booked the property in the last 12 months	Number
Count Available Days LTM	Number of days where a guest can book the property for the last 12 months	Number
Count Blocked Days LTM	Number of days blocked by the host where a guest is unable to book the property in the last 12 months	Number
Listing URL	URL of the property listing	String
Listing Main Image URL	Main image of the property listing	String

SOURCE: AIRDNA (HTTPS://WWW.AIRDNA.CO/)

APPENDIX B

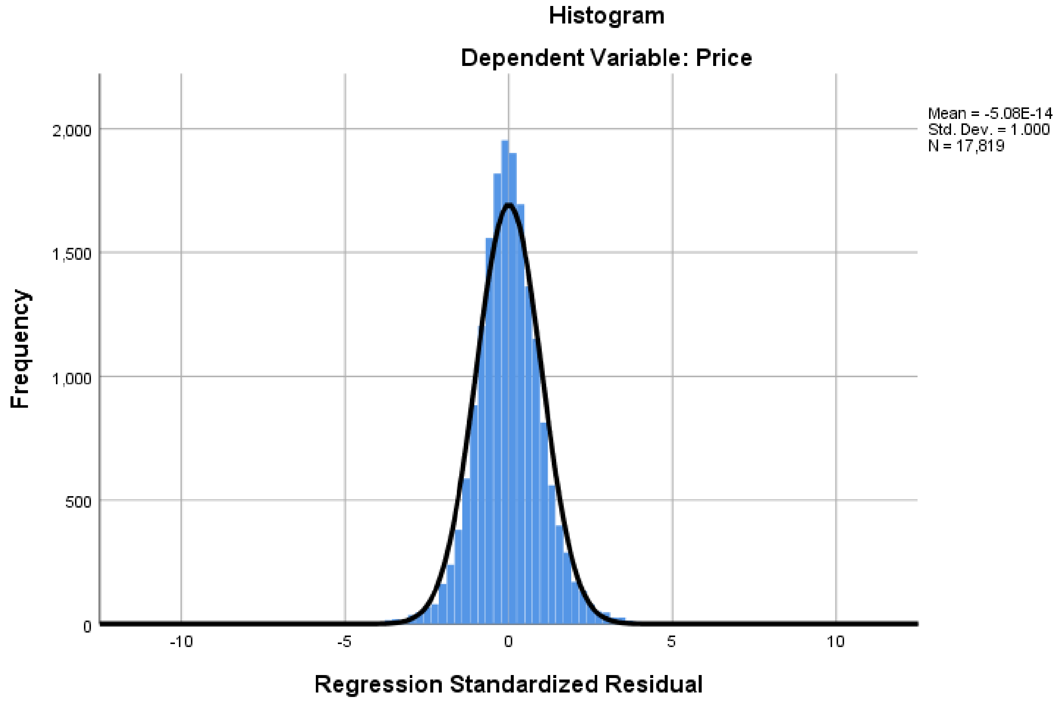

FIGURE B-1. HISTOGRAM OF STANDARDIZED RESIDUALS (BARCELONA).

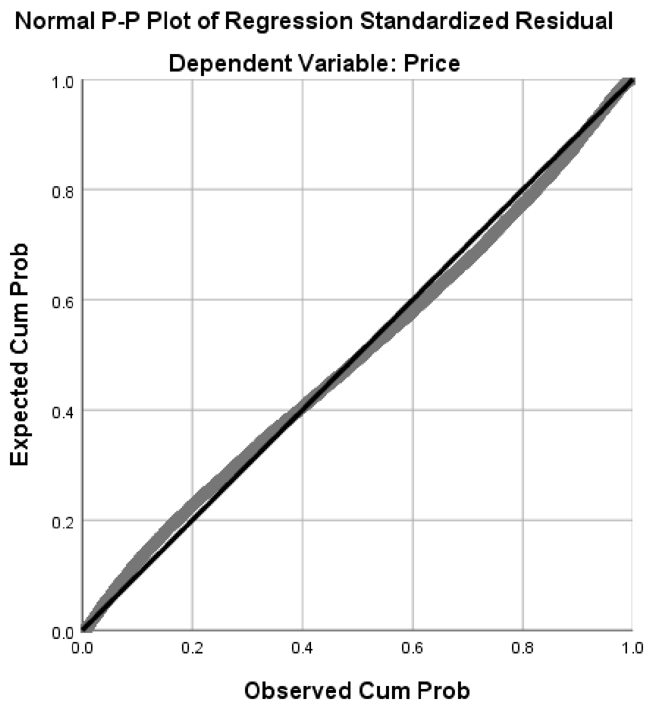

FIGURE B-2. NORMAL P-P PLOT OF STANDARDIZED RESIDUALS (BARCELONA).

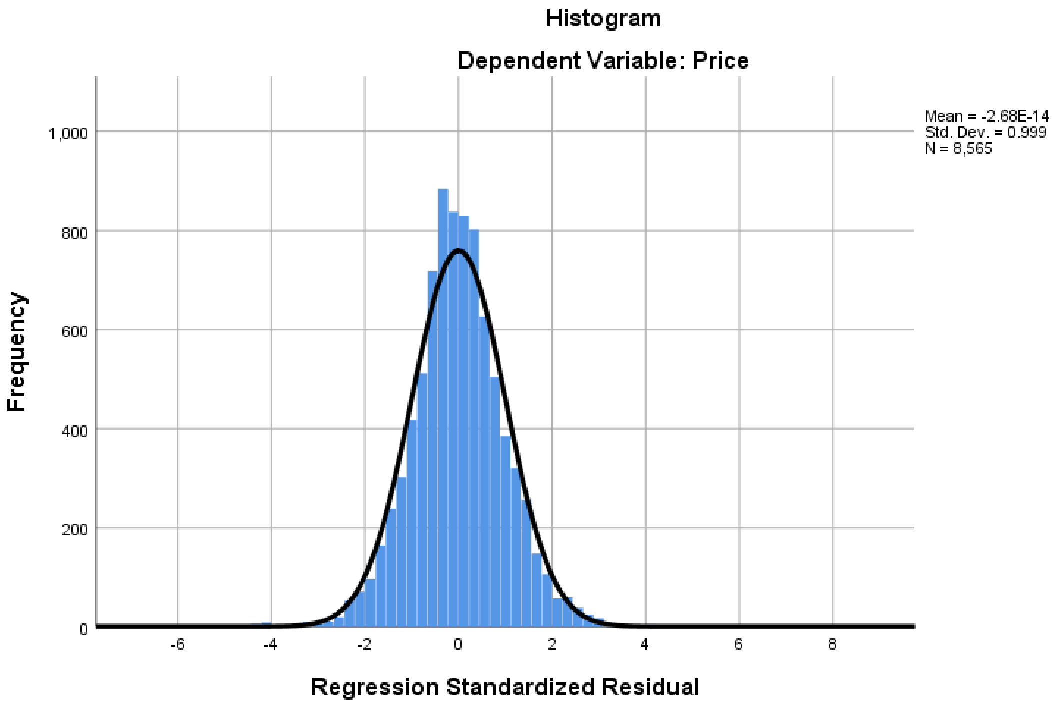

FIGURE B-3. HISTOGRAM OF STANDARDIZED RESIDUALS (MADRID).

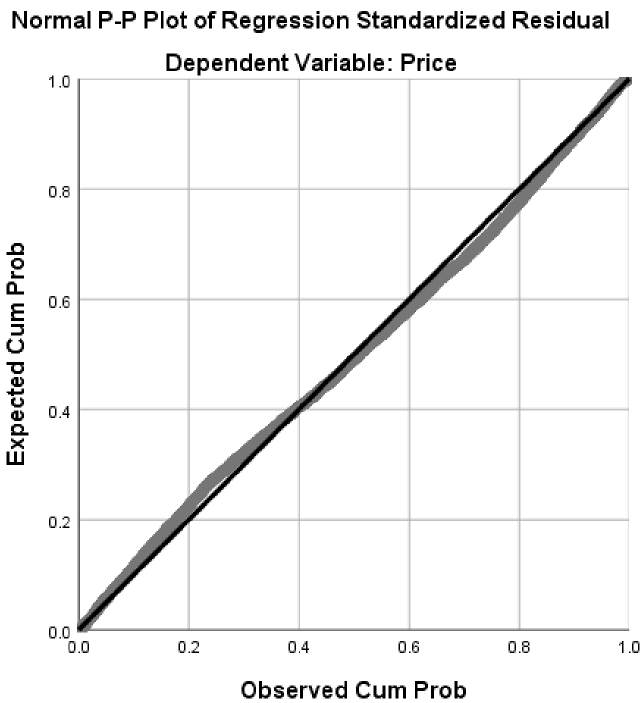

FIGURE B-4. NORMAL P-P PLOT OF STANDARDIZED RESIDUALS (MADRID).

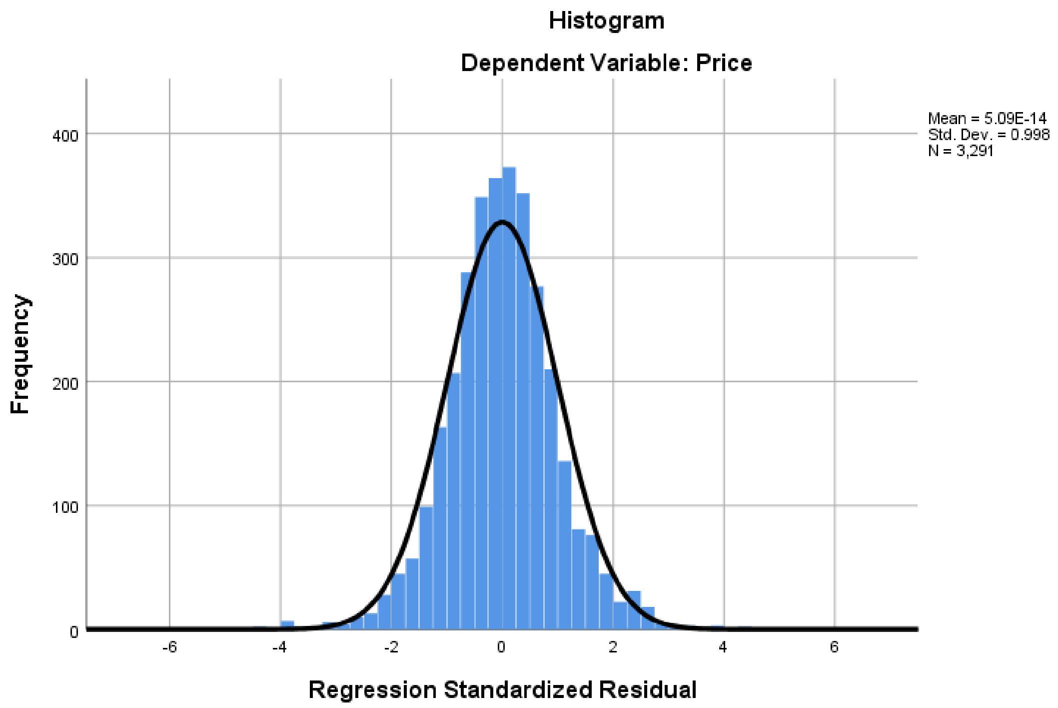

FIGURE B-5. HISTOGRAM OF STANDARDIZED RESIDUALS (SEVILLE).

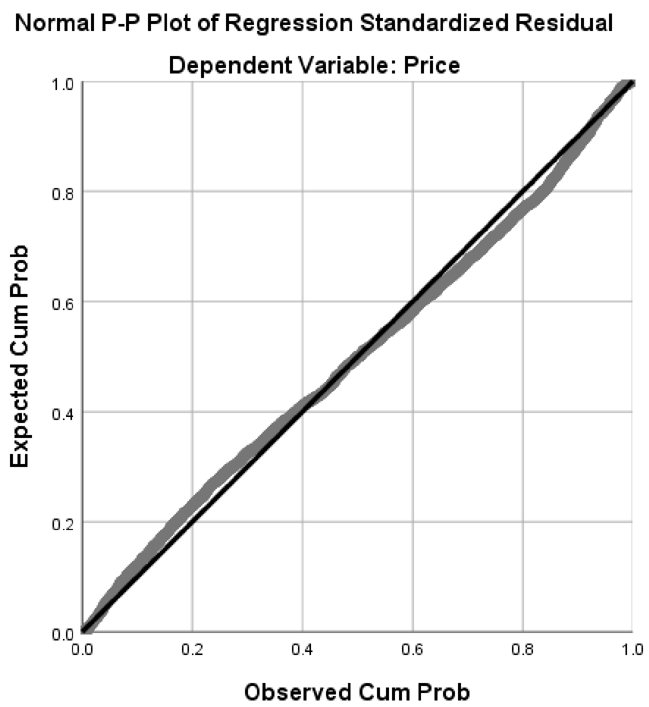

FIGURE B-6. NORMAL P-P PLOT OF STANDARDIZED RESIDUALS (SEVILLE).

APPENDIX C

TABLE C-1. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING PUBLISHED NIGHTLY RATE (BARCELONA).

Coefficients[a]

	Coefficients	Std. Error	t	Sig.
(Constant)	2.946	0.74	39.879	.000
Cancellation Policy	.026	.007	3.545	.000
Listing Type	.674	.009	74.906	.000
Number of Reviews	-.083	.002	-34.391	.000
Overall Rating	.340	.020	16.876	.000
Bedrooms	.138	0.11	12.820	.000
Bathrooms	.190	.010	18.779	.000
Max Guests	.320	.009	36.179	.000
Response Rate	.028	.014	1.947	.052
Response Time	-.002	.002	-.969	.333
Superhost Status	.121	.024	5.144	.000
Number of Photos	.063	.005	11.851	.000
Instantbook Enabled	.019	.008	2.338	.019
Professional Status	-.084	.007	-12.479	.000
Distance from City Center	-.144	.004	-32.212	.000

a. Dependent Variable: Published Nightly Rate

TABLE C-2. MODEL SUMMARY USING PUBLISHED NIGHTLY RATE (BARCELONA).

Number of observations	18738
R	.820
R^2	.673
Adjusted R^2	.672
Std. Error of the Estimate	.4049
R^2 Change	.673
F Statistic	2748.211
Degrees of Freedom	14
Sig. *F* Change	.000

TABLE C-3. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING PUBLISHED NIGHTLY RATE (MADRID).

Coefficients[a]

	Coefficients	Std. Error	t	Sig.
(Constant)	2.905	.123	23.629	.000
Cancellation Policy	-.019	.009	-2.056	.040
Listing Type	.686	.011	59.723	.000
Number of Reviews	-.054	.003	-16.446	.000
Overall Rating	.320	.027	11.702	.000
Bedrooms	.356	.014	25.797	.000
Bathrooms	.160	.014	11.691	.000
Max Guests	.175	.011	16.265	.000
Response Rate	-.008	.025	-.323	.746
Response Time	.006	.002	3.163	.002
Superhost Status	.119	.014	8.556	.000
Number of Photos	.060	.007	9.258	.000
Instantbook Enabled	-.032	.011	-3.041	.002
Professional Status	-.014	.009	-1.683	.092
Distance from City Center	-.100	.004	-22.382	.000

a. Dependent Variable: Published Nightly Rate

TABLE C-4. MODEL SUMMARY USING PUBLISHED NIGHTLY RATE (MADRID).

Number of observations	9103
R	.829
R^2	.687
Adjusted R^2	.687
Std. Error of the Estimate	.3737
R^2 Change	.687
F Statistic	1425.433
Degrees of Freedom	14
Sig. *F* Change	.000

TABLE C-5. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING PUBLISHED NIGHTLY RATE (SEVILLE).

Coefficients[a]

	Coefficients	Std. Error	t	Sig.
(Constant)	2.863	.190	15.093	.000
Cancellation Policy	.056	.015	3.789	.000
Listing Type	.569	.020	28.533	.000
Number of Reviews	-.073	.005	-13.923	.000
Overall Rating	.288	.050	5.752	.000
Bedrooms	.261	.022	11.718	.000
Bathrooms	.181	.021	8.466	.000
Max Guests	.280	.019	14.644	.000
Response Rate	-.002	.039	-.040	.968
Response Time	.004	.003	1.388	.165
Superhost Status	.068	.025	2.715	.007
Number of Photos	.042	.012	3.558	.000
Instantbook Enabled	-.013	.018	-.746	.456
Professional Status	-.029	.014	-2.061	.039
Distance from City Center	-.148	.010	-15.415	.000

a. Dependent Variable: Published Nightly Rate

TABLE C-6. MODEL SUMMARY USING PUBLISHED NIGHTLY RATE (SEVILLE).

Number of observations	3448
R	.821
R^2	.674
Adjusted R^2	.673
Std. Error of the Estimate	.3822
R^2 Change	.674
F Statistic	506.430
Degrees of Freedom	14
Sig. *F* Change	.000

TABLE C-7. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING BEDROOMS ONLY (BARCELONA).

Coefficients[a]

	Coefficients	Std. Error	t	Sig.
(Constant)	2.910	.091	31.903	.000
Cancellation Policy	.154	.009	17.124	.000
Number of Reviews	-.092	.003	-30.485	.000
Overall Rating	.118	.025	4.766	.000
Bedrooms	.971	.008	115.481	.000
Response Rate	.133	.018	7.425	.000
Response Time	.011	.002	5.505	.000
Superhost Status	.059	.028	2.069	.039
Number of Photos	.189	.006	29.526	.000
Instantbook Enabled	.086	.010	8.645	.000
Professional Status	-.061	.008	-7.393	.000
Distance from City Center	-.200	.005	-36.725	.000

a. Dependent Variable: Price

TABLE C-8. MODEL SUMMARY USING BEDROOMS ONLY (BARCELONA).

Number of observations	17984
R	.742
R^2	.550
Adjusted R^2	.550
Std. Error of the Estimate	.4860
R^2 Change	.550
F Statistic	1996.338
Degrees of Freedom	11
Sig. *F* Change	.000

TABLE C-9. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING BEDROOMS ONLY (MADRID).

Coefficients[a]

	Coefficients	Std. Error		Sig.
(Constant)	3.088	.162	19.058	.000
Cancellation Policy	.101	.012	8.447	.000
Number of Reviews	-.084	.004	-19.348	.000
Overall Rating	.117	.036	3.250	.001
Bedrooms	.891	.013	67.060	.000
Response Rate	.052	.033	1.562	.118
Response Time	.014	.003	5.567	.000
Superhost Status	.070	.018	3.907	.000
Number of Photos	.066	.014	4.860	.000
Instantbook Enabled	.227	.008	27.566	.000
Professional Status	-.081	.011	-7.386	.000
Distance from City Center	-.199	.006	-35.423	.000

TABLE C-10. MODEL SUMMARY USING BEDROOMS ONLY (MADRID).

Number of observations	8667
R	.722
R^2	.521
Adjusted R^2	.520
Std. Error of the Estimate	.4737
R^2 Change	.521
F Statistic	854.930
Degrees of Freedom	11
Sig. *F* Change	.000

TABLE C-11. ESTIMATED RESULTS FROM THE HEDONIC PRICING MODEL USING BEDROOMS ONLY (SEVILLE).

	Coefficients	Std. Error	t	Sig.
(Constant)	2.701	.241	11.215	.000
Cancellation Policy	.151	.019	8.065	.000
Number of Reviews	-.091	.007	-13.613	.000
Overall Rating	.242	.066	3.649	.000
Bedrooms	.821	.020	40.100	.000
Response Rate	.120	.049	2.442	.015
Response Time	.002	.004	.471	.638
Superhost Status	.067	.031	2.127	.033
Number of Photos	.166	.015	11.343	.000
Instantbook Enabled	.073	.023	3.247	.001
Professional Status	-.096	.018	-5.396	.000
Distance from City Center	-.288	.012	-24.983	.000

TABLE C-12. MODEL SUMMARY USING BEDROOMS ONLY (SEVILLE).

Number of observations	3220
R	.718
R^2	.516
Adjusted R^2	.514
Std. Error of the Estimate	.4747
R^2 Change	.516
F Statistic	320.168
Degrees of Freedom	11
Sig. *F* Change	.000