ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis: Data Science & Marketing Analytics

Persuading customers by appealing to credibility, emotion, or logic: Evidence from the Airbnb marketplace

Author: Daan Straatman (424817)

Supervisor: Dr. N.M. (Nuno) Almeida Camacho

Co-reader: Dr. P. (Phyllis) Wan

December 10, 2020

Abstract

Despite the growth in popularity of the upcoming peer-to-peer (P2P) marketplaces, only little research has focussed on the role of persuasion on customer decision making in such environments. This research analyses data from Airbnb, a prominent P2P platform and example of the sharing economy, to understand the effects of persuasion by hosts on bookings. Through exploring the dynamics around persuasion and the booking performance, this paper offers valuable insights for Airbnb hosts and more generally consumer behaviour in P2P marketplaces. The occupancy rate is chosen as performance measure and, due to its absence in the dataset, is calculated using the number of reviews throughout the year. The effects of persuasion are quantified using Aristotle's rhetoric theory, in which three modes of persuasion are identified: ethos (appealing to credibility), pathos (appealing to emotion) and logos (appealing to facts and logic). Based on previous theory, the variables in the dataset are either carefully matched to one of the appeals or used as potential control variables. Since pathos (emotion) is measured using the textual descriptions of the host, I first apply sentiment analysis to measure the emotive tone. Also, elastic net regression is employed for variable selection of the control variables and the final model is estimated using tobit regression. Overall, each mode of persuasion affects the booking performance. Appealing to emotion (pathos), compared to not appealing to emotion, has a negative effect on the number of bookings, whereas appealing to logic and facts (logos) can positively influence guests' booking intentions. Credibility appeals (ethos) are deemed most persuasive, meaning that the positive effects on bookings are the highest. Based on these findings, I make various recommendations for hosts and suggestions for future research.

Key words: Sharing economy, Aristotle's rhetoric theory, Persuasion, Peer-to-peer accommodation, Airbnb

Table of Contents

1. Introduction 1

2. Theoretical Framework 8

2.1 Aristotle's Rhetoric Theory in Diverse Contexts 8

2.2 Aristotle's Appeals in the Airbnb Setting 10

2.2.1 Persuasion through Ethos (credibility) 11

2.2.2 Persuasion through Pathos (emotion) 12

2.2.3 Persuasion through Logos (facts and logic) 14

3. Data 17

3.1 Pre-Processing and Data Cleaning 19

4. Methodology 22

4.1 Sentiment Analysis 22

4.2 Elastic Net 26

4.3 Tobit Regression 29

5. Results 31

5.1 Descriptive Statistics 31

5.2 Results of Sentiment Analysis 32

5.3 Results of Elastic Net Regression 36

5.4 Results of Tobit Regression 37

6. Robustness Check 42

7. Conclusion 43

7.1 Implications 43

7.2 Limitations and Future Research 47

References 49

Appendix 54

1. Introduction

Over the past forty years our advances in technology have led to a new production and employment sector, the so-called sharing economy (Sundararajan, 2016). This concept is best defined as a new set of business activities in which firms, such as Uber or Airbnb, allow consumers to create value for one other by using their own possessions on online platforms (Dellaert, 2019). More specifically, the peer-to-peer (P2P) accommodation market is growing at a significant pace (MarketWatch, 2020) and has a considerable impact on the conventional tourism and hospitality industry (Zervas et al., 2017). The success of, for example, the popular P2P accommodation platform Airbnb is so big that some see it as a disruptive innovation as it alters the way that consumers and business in the hotel industry operate (So et al., 2018). In the United States alone, the growth of Airbnb in terms of home-sharing is responsible for roughly twenty percent of the average annual increase in rents and about fourteen percent of the average annual increase in housing prices (Barron et al., 2019). On the other hand, one can argue that Airbnb should not be seen as a disruptive technology as it is too small in comparison to its rivals in the hotel industry (e.g. in New York City the share of Airbnb of total hotel bookings was just 14%). Moreover, Airbnb listings in major cities grew in 2016, which slowed in 2017 and several markets were considered saturated or declining in 2018 and 2019 (Muller, 2019).

Geared towards better understanding the theory behind the sharing economy, it is crucial to further investigate how consumers make decisions in these markets. Hence, through the P2P accommodation platform Airbnb I will research the dynamics of the decision-making process. On the Airbnb platform, guests and hosts are able to find and interact with each other online. The guests, however, start the decision-making process not only by considering the various features of the place, but also by considering features about the hosts themselves. For the latter, attributes that evoke trust in a host seem to be influential (Ert et al., 2016). Also, it is suggested that trust contributes towards a more efficient marketplace since buyers associate a trustworthy source with lower risk of poor performance (Grewal et al., 1994). Therefore, Airbnb enables guests to acquire information, such as the various amenities, star rating, or reviews, along with many other attributes.

Furthermore, this also means that hosts are provided with the opportunity to be perceived as trustworthy by guests and hence to positively influence customers' booking intentions. It could be interesting to know what features drive the differences between a well performing Airbnb listing and a bad performing one. Mostly, research in this field focusses on specific features, of which the price (Wang & Nicolau, 2017) or the host profile (Ert et al., 2016) is often investigated. However, although some literature exists in the field of host self-representation, a potentially important factor in the seller's ability to persuade buyers in the sharing economy is scarcely considered: The influence of a host's argumentative capability on persuasiveness towards guests.

Persuasion is defined as a process through which a persuader moves people to a desired position that they currently do not hold (Conger, 1998). This process involves good preparation, adequate framing of arguments, delivering proper evidence, and accurately matching with the audience's emotions. Very generally, well implemented persuasion strategies can have a significant impact on a business' performance, which applies to a wide range of business. One can think about using evocative text on a menu to pique interest in a dish or to emphasize characteristics in commercials to increase sales (Guitart et al., 2018). It is important for our information gathering behaviour to know what other people think about a product or service. As more online reviews and blogs becomes available, particularly the rise of text analytics provides new opportunities (Ludwig et al., 2013). To further delve into the art of persuasion, I am leaded by Aristotle's rhetorical theory, which is a broad concept from a multidisciplinary field.

In short, the structure of one's argument is central in this theory as text or speech can be framed in different ways to 'convince' someone. These argumentative strategies are based on Aristotle's three modes of persuasion: ethos, pathos and logos. *Ethos* appeals to credibility and authority, *pathos* appeals to emotion and sympathy, and *logos* appeals to facts and logic (Kotler & Levy, 1973). In my context, of

Airbnb hosts promoting their listing, (1) *ethos* implies the credibility of the Airbnb host (e.g., hosts rewarded with a super host badge have more credibility compared to hosts without such badge), (2) *pathos* implies the ability of the host to move the audience to the desirable emotional actions, and (3) *logos* means that the host persuades by using rational arguments that appeal to logic (e.g., "the property includes many provisions"). This research is set to investigate which argumentation modes, when adopted by an Airbnb host, significantly influence guests' decision making (i.e., their decision to rent or not rent the host's property), which leads to the following research question:

> *Which persuasion strategies based on Aristotle's appeals, implemented by the Airbnb host, are most effective for the listing's performance?*

To elaborate on the performance measure, I believe that occupancy rate is the best metric. By using this, a luxury villa with a higher price can be compared to a small apartment with a lower price as occupancy rate should be based on all important factors in the customer-decision making process. In other words, the price quality of listings is more easily assessed. Moreover, in the hotel industry it is seen as an objective, key performance measure (Agarwal et al., 2013).

Furthermore, I use a rich dataset that contains information on Airbnb listings in Amsterdam. Here, a big proportion of the listings is lowly available with less than 60 days a year. This proportion reaches 76% of all listings. In addition, estimated occupancy is still only 20% (Inside Airbnb, 2019). On the other hand, the occupancy rate of hotels in Amsterdam is estimated to be 82% in 2019 (Statista, 2020). Amsterdam is a popular destination for tourists and tourism numbers keep on growing yearly. With more expected tourists, a relatively low occupancy rate for Airbnb listings and a higher occupancy rate for hotels, there is quite some ground to be gained by Airbnb hosts.

To investigate the research question, I will attribute specific variables to either ethos (credibility), such as an indication of being a super host, or logos (logic and facts), such as enumerating property amenities. This will be further specified in the theoretical framework. Pathos represents the appeal to emotion and therefore a simple variable will not be used. Instead, I apply text analytics to determine the emotional orientation in a text provided by the host. More specifically, I use sentiment analysis through a polarity[1] algorithm, which measures the emotive tone within a text and allows me to determine to what extend the host appeals to pathos (emotion). The idea is very similar to a more well-known analysis software, called linguistic inquire word count (LIWC), which also measures the emotional orientation. This approach is adopted by different researchers in diverse contexts, such as Ludwig et al. (2013), who used LIWC on book reviews to study the influence of affective content on retail websites' conversion rates, and Humphreys (2010), who analysed how newspapers framed the casino gambling industry throughout decades and the consequences on the acceptance of this industry.

In addition, to properly compare different Airbnb listings I need to control for variables in the dataset that might affect the occupancy rate, other than the so-called core variables that are assigned to the three appeals. However, there are many variables and thus first a regularization technique is applied, namely elastic net. This enables variable selection as coefficients can shrink to zero. Moreover, it reduces the complexity of the model and decreases the risk of overfitting. Note that this is only done on the control variables as the core variables should be included in the final model.

Next, I use tobit's regression with the core variables and the control variables that are significant after regularization to construct the final model. The reason for a tobit's regression model is that it accounts for situations in which the dependent variable is censored in either the lower- or upper bound and for situations where selection bias occurs due to the inability to establish the actual value of the dependent variable. In this study, the occupancy rate is proxied by yearly reviews and bound to a maximum value, making tobit regression rather suitable. This is further elaborated in the Data section.

As for this paper's relevance, it certainly contributes to the existing literature. Much of the previous research is relatively recent, showing the growing popularity and need to understand the upcoming sharing economy market. Table 1 shows an overview of various studies that have contributed

[1] The polarity algorithm measures sentiment by counting positive and negative words in a text, whilst also considering (de-)amplifications and negations.

in further understanding this concept of the sharing economy. Note that this table does not comprise all existing literature but is an attempt to cover various prominent ones. In previous research, often the effects of numerous attributes on price or trust and online review analyses on P2P markets are investigated through a survey or quantitative design.

However, prior research lacks the role of persuasion or more specifically the role of persuasion techniques and their influence on an objective performance measure. This includes both persuasion through text as persuasion through various signalling attributes. Still, some researches have embraced the role of persuasion in a P2P context. Otterbacher (2011) uses text analytics in combination with argumentation theory to study the prominence of page reviews on popular forums. Yet, findings from this study might not hold because reviews are written after a purchase or transaction has taken place, whereas this research investigates the persuasive power of the three appeals prior to the actual transaction. In other words, there might be a difference between the effectiveness of ethos (credibility), pathos (emotion) and logos (facts and logic) in different phases of the customer journey. Furthermore, the study by Yang et al. (2018) takes Aristotle's rhetoric theory in the context of the sharing economy into account. They explore role of the three appeals in the formation of guests' trust in Airbnb (Yang et al., 2018). Still, besides a different output variable various differences are evident. Their analysis is based on a survey design with relatively few people that answer questions using a Likert scale. With thousands of observations, this study adopts a quantitative approach with different core variables that are assigned to each of the appeals. Moreover, this research's model includes control variables to strengthen the results. Also, the role of persuasion through text (pathos; appeal to emotion) is empathized in this study. Hence, different results are likely to appear.

Table 1. Previous literature on the sharing economy

Source	Cites	Context	Research Design	Key Findings
Cheng & Jin (2019), *International Journal of Hospitality Management*	156	Text mining and sentiment analysis on reviews is employed to find out which attributes affect Airbnb guests' experience.	Quantitative	Three major attributes are location, amenities and reviews about the host. Negative emotions are primarily due to noise, whereas a general positive bias is present. Guests also tend to compare their experience to recent hotel stays.
Dellaert (2019), *Journal of the Academy of Marketing Science*	50	About the need for firms to account for consumer co-production in their marketing process.	Literature Review	A two-layered conceptual framework is provided that considers customer co-production activities. The framework is based on two known theories and research in the field of consumer behaviour and leads to several implications.
Eckhardt et al. (2019), *Journal of Marketing*	86	A literature review on what the sharing economy is and its impact on three vital foundations of marketing.	Literature Review	As the sharing economy touches on all bases of marketing, new questions should be asked and new frameworks should be developed. With the help of new data and methods the full potential of the sharing economy should be further investigated.
Ert et al. (2016), *Tourism Management*	919	The effect of photos on the guests' decision making in peer-to-peer accommodation.	Experiment/Survey	Guests infer trustworthiness from photos of the host. The more trustworthy the photo, the higher the price of the listing and the likelihood that the listing is booked. Reputation based on review scores has no effect on trustworthiness or likelihood to book.
Guitiérrez et al. (2017), *Tourism Management*	373	Spatial patterns of Airbnb are compared with hotels and sightseeing spots.	Quantitative	Results suggest a close relationship between Airbnb and hotels. Moreover, Airbnb exploits the locations of main tourist attractions more than hotels.
Herzenstein et al. (2011), *Journal of Marketing Research*	355	The effect of identity claims on peer-to-peer lending decisions.	Quantitative	A higher number of identity claims, which decreases the loan performance, increases the loan funds. Furthermore, lending decisions are affected by unverifiable information beyond what is rational.
Lamberton & Rose (2012), *Journal of Marketing*	840	A framework on understanding commercial sharing systems.	Experiment/Survey	Cost-related benefits and perceived risk of scarcity related to sharing are determinants of its attractiveness.
Ludwig et al. (2013), *Journal of Marketing*	526	The influence of semantic content and style properties in customer reviews on conversion rates.	Quantitative	Positive affective content on conversion rates is asymmetrical. Managers should promote influential reviews, stimulate reviewers to write reviews, and adapt their style in own editorial reviews.
Wang & Nicolau (2017), *International Journal of Hospitality Management*	288	The difference in price determinants of the sharing economy based accommodation offers and hotels.	Quantitative	For sharing economy based accommodation, host attributes are important price determinants. In hotels, stars and chain affiliation are important for price.

Yang et al. (2018), *Journal of Travel and Tourism Marketing*	31	Aristotle's rhetorical theory on guests' trust in the sharing economy.	Survey	Appeals are positively associated with trust in Airbnb hosts, leading to trust in the brand Airbnb.
Yang et al. (2017), *Journal of Services Marketing*	111	Determinants of customer loyalty in sharing economy services and the mediating effect of commitment.	Qualitative/Survey	Confidence, together with social and safety benefits positively, affect commitment. Also, commitment mediates between these determinants and customer loyalty.
Zervas et al. (2017), *Journal of Marketing Research*	1833	The economic effects of Airbnb on the hotel industry in Texas.	Quantitative	The entry of Airbnb has a negative economic impact on the hotel industry. Low-end hotels are most affected by the entry of Airbnb.

2. Theoretical Framework

The goal of this paper is to identify how persuasion strategies, based on Aristotle's three appeals, affect the performance of an Airbnb listing. This performance is measured by the occupancy rate, which is an objective, key performance measure in the hotel industry (Agarwal et al., 2013). It is imperative to examine the use of Aristotle's rhetoric theory in previous literature and to delve into previous research that contributed towards our better understanding of the sharing economy. First, Aristotle's rhetoric in diverse contexts is reviewed. Then, the scope is narrowed down to the appeals in the context of the sharing economy and potential indicators for each of the persuasive modes are matched.

2.1 Aristotle's Rhetoric Theory in Diverse Contexts

Being able to persuade someone is possibly one's greatest asset in today's economy (Gallo, 2019). When doing so, holding on to a strategy could be valuable. A persuasion strategy is defined as 'one that attempts to produce a favourable response in the seller through identifying his natural interests with the transaction' (Kotler & Levy, 1973). In Aristotle's rhetorical theory, three modes of persuasion are represented, based on Aristotle's conceptualization of what the key elements are of credible communications. The three classical dimensions include persuasion through one's credibility (*ethos*), by appealing to emotion (*pathos*) and providing logical arguments (*logos*) (Kotler & Levy, 1973). Table 2 conceptualizes this by considering the context of an employee who introduces a green initiative and tries to persuade the board to take on this project.

Table 2. Example of different arguments for each mode of persuasion

Persuasion strategy	**Argument**
Ethos (credibility)	"As a climate expert, I am certain that this initiative benefits all."
Pathos (emotion)	"Can you imagine the environmental and economic consequences if we do not step up now? This is all about contributing to a better world."
Logos (facts and logic)	"Actually, research shows that companies engaging in green initiatives increase profits by 7% due to the effects of goodwill and tax benefits."

Furthermore, the importance of rhetoric has been stressed throughout multiple decades, in various contexts, of which the most relevant for this research are rhetoric in advertising (Ertimur & Gilly, 2012) and in marketing (Pollay, 1985). There are, however, differences in which appeals are deemed most effective depending on the time that we live in (Brown et al., 2018; Pollay, 1985). Pollay (1985) investigated the rhetorical focus of ads in American magazines over multiple decades. Ads in the first two decades of the twentieth century primarily utilize logos (facts and logic), different from the 1930's and 1940's in which The Great Depression and The Second World War are used to evoke emotions, appealing to pathos. After that, again logical and factual appeals (logos) remain most popular (Pollay, 1985). Nowadays, pathos (emotion) seems most prominent in both the marketing and political setting, as people forget communication that does not evoke emotions (Brown et al., 2018).

Contrary to written ads, Ertimur & Gilly (2012) investigate how consumers respond to three types of digital ads. These ads include both contest and voluntary consumer-generated ads (CGA) and ads produced by a company itself. The message, described as the rhetorical focus, is measured by the appeals. They found that all three types of ads intensely use an affective rhetoric style, whereas voluntary CGA also make use of source appeals, appealing to both pathos (emotion) and ethos (credibility). However, the public does not perceive voluntary CGA as credible, yet only authentic (Ertimur & Gilly, 2012). These insights in consumer-to-consumer communications suggest that persuasion through credibility (ethos) is ineffective when it involves a P2P situation, such as Airbnb.

In addition to that, an established brand enjoys the benefits of its reputation when persuading consumers, which increases market performance (Dawar & Parker, 1994). Since small companies and P2P markets cannot rely solely on reputation, a difference in the effectiveness of the persuasive cues is expected. Ruokolainen & Aarikka-Stenroos (2016) provide a framework that highlights which aspects a customer's reference should focus on in the case of a start-up and which corresponding rhetoric principles, measured by Aristotle's three appeals, fortify the argumentation.

Moreover, they mention that the quantity of 'evidence' in one's argument does not per se strengthen its persuasive power, suggesting the importance of quality and structure. Herzenstein at al. (2011) confirm this in their study on how borrowers' identity claims in stories affect lenders' decisions to grant personal loans in the P2P market. Results indicate that lenders are affected by these stories beyond rational and verifiable information, unjustly granting higher personal loans to borrowers who present themselves as trustworthy, appealing to ethos (credibility). This indeed indicates the value of quality and structure that one's argumentation should have on P2P platforms such as Airbnb.

2.2 Aristotle's Appeals in the Airbnb Setting

On the Airbnb platform, individuals who like to rent out their place, the hosts, need to make an account on the online platform by registering first. In this process, hosts are provided with the opportunity to share information about both the place and themselves. This is vital as self-representation of the host contributes towards the general performance of a listing (Liang et al., 2020). On the other hand, potential guests evaluate the place by various indicators such as the characteristics of the listing, host attributes, reputation and competition. Because this is the only information available, it is paramount in the customer-decision making process. These indicators are reviewed and matched to Aristotle's three appeals. An overview of this is depicted by Figure 1.

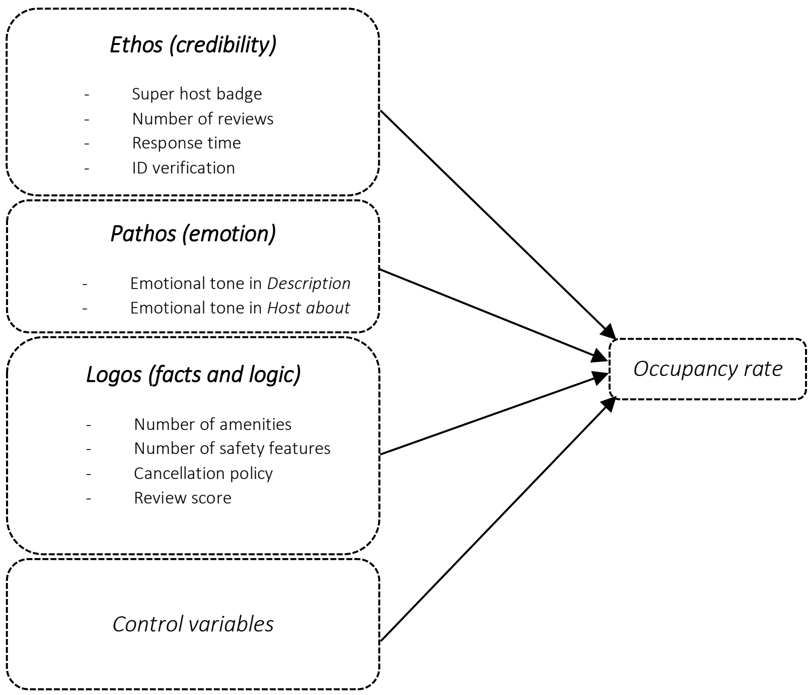

Figure 1. Conceptual model

2.2.1 Persuasion through Ethos (credibility)

In Airbnb, persuasion through ethos can be achieved by strengthening the credibility of the host. This evokes trust, which in turn decreases the guests' perceived risk of bad performance whilst booking (Grewal et al., 1994). In online and offline communications, there is also a positive association between reputation and credibility (Marshal & WoonBong, 2003). Mauri et al. (2018) highlight the importance of reputation in Airbnb, stating that it is the most important factor in explaining popularity variations. In this research, four indicators through which credibility can be established are identified; a super host badge, the number of reviews, response time and host ID verification.

Firstly, a super host badge is rewarded to hosts with excellent hospitality skills. To earn a super host badge, Airbnb requires the host to satisfy four specific criteria (Airbnb, 2020a):

- Complete at least ten reservations or three reservations with a total of at least 100 nights;
- Uphold a response rate of at least 90 per cent;
- Keep up a cancellation rate of 1 per cent, excluding extenuating circumstances;
- Out of five, maintain an overall rating of at least 4.8 in the last year.

Moreover, these requirements are checked every three months and thus the super host status is updated several times a year (Airbnb, 2020a). The super host badge can be seen as a form of advertising. Also, it is demonstrated that online advertisements can increase sales offline (Dinner et al, 2014) and online (Kumar et al., 2016). Therefore, the advertising effects of the super host badge can improve the occupancy rate (Liang et al., 2020). This literature suggests that guests could perceive the super host badge as a credible indication for the quality of both the listing and the host.

Secondly, guests can only write a review when they have stayed at a place, making it a good proxy for the number of online bookings and hence reliable indicator of the host's experience (Ye et al., 2009). Additionally, a large number of reviews can serve as signal of popularity (Zhu & Zhang, 2010) and there exists a positive relationship between the number of reviews and trustworthiness, which is more evident in combination with high ratings than with low ratings (Gavilan et al., 2018). Furthermore, Chevalier & Mayzlin (2006) find a positive association between the relative difference in number of reviews and relative sales when comparing two online book providers. This could suggest that guests, when comparing listings on Airbnb, regard the number of reviews as a reliable indicator for the listing's quality and the host's trustworthiness.

Then, the response time is calculated based on the average time it takes a host to respond to all new messages over the past thirty days. Various studies have shown the importance of a quick response time in the hospitality industry (Sparks et al., 2016). Sparks et al. (2016) investigate the effects of a hotel's responding time to negative reviews on guests' trust inferences. They find that a fast response time (within one day) has a positive effect on guests' trust as compared to a slow response time (with a thirty day delay). However, there is no difference between 'shorter' responses, such as an one day delay and a seven day delay (Sparks et al., 2016). This could suggest that the credibility of an Airbnb host is only negatively affected in case the response time is unacceptably long. Also, response time can be attributed to the personal reputation of a host (Mauri et al., 2018), which is in turn positively associated with credibility (Marshal & WoonBong, 2003). Therefore, it can serve as a good signal for a host's trustworthiness.

Lastly, on Airbnb there are two ways to verify the host's identity. The first is providing a legal name and address, which should match with banking documents or utility bills. The other is providing a picture of a government ID, occasionally with a new photo to confirm it is a match (Airbnb, 2020c). Ert et al. (2016) mention that the formation of trust between parties depends on the level of identification and communication. Since ID verification is a quality attribute that guarantees that a host is real, it stimulates trust building and consequently improves the booking performance.

2.2.2 Persuasion through Pathos (emotion)

Pathos implies the approach of appealing to one's emotion and sympathy. The idea is that Airbnb guests are encouraged to book by reading affective content. Thus, this would enable a host to actively persuade guests by writing in a rather emotive style. Previous studies have shown the importance of how hosts articulate themselves through Airbnb (Liang et al., 2020). Lench et al. (2011) suggest that the use of emotive words in text drives the customer decision making, which is more prevalent in the case of limited information (Ludwig, 2013). In Airbnb, guests are also provided with limited information regarding the host or listing, which implies the potential value of emotion-based appeals in the host written description. Furthermore, this value is particularly important in Airbnb as presenting oneself online is regarded to be more difficult when there are multiple audiences (Labrecque et al., 2011). Advancing, the importance of emotional content is applied in diverse contexts. Chandy et al. (2001) look at toll-free referral services and show that emotion-based appeals can stimulate the behavioural response through personal involvement. When considering reviews, it appears that reviews that include emotive words, compared to strictly informational reviews, have a stronger effect on customers' attitude towards a product (Zablocki et al., 2019).

Yet, previous research does not always advocate this type of persuasive marketing communication and highlights the role of both informative and persuasive information. According to Narayana et al. (2005), informative communication allows customers to adjust their earlier beliefs and increase certainty about a product's actual quality, e.g. by providing practical information. On the other hand, persuasive communication alters consumers' preferences through goodwill accretion, such as appealing to one's emotion. Narayana et al. (2005) found both types of marketing communication to be evident, however, in different stages of the product. To elaborate, salesforce effort (detailing) on the number of sales is primarily affected by informative communication in the introductory phase, explained by the customer's limited experience and need for certainty regarding the 'true' quality of a new product. Moreover, persuasive communication dominates consecutive stages as uncertainty naturally decreases (Narayana et al., 2005). Hence, based on the aforementioned literature, engaging emotionally with guests by using emotive words could be a valuable asset as it potentially affects the number of bookings.

2.2.3 Persuasion through Logos (facts and logic)

Persuading through logos is achieved by presenting facts and logic that show whether an option or choice is logically appealing, e.g. showcasing the features or benefits of the property. It has proven to be a valuable mode of persuasion when customers make decisions (Ruokolainen & Aarikka-Stenroos, 2016). In Airbnb, objective information that is provided mostly stems from accommodation characteristics. In this research, four indicators that comprise objective information regarding the listing are identified; number of amenities, number of safety features, cancellation policy and review score.

First off, when listing a house or apartment, the host is asked to select which amenities are present in the place. There are many amenities of which a few examples are; basic amenities such as soap and toilet paper (Airbnb Essentials); a kitchen; Wi-Fi; a washer; safety features such as a first aid kit. It is a small sample of the total options. Previous research finds the significant effect of amenities on the listing's price (Wang & Nicolau, 2017). Even more, amenities are one of the most popular topics that guests talk about in reviews (Cheng & Jin, 2019). Nevertheless, this research assesses the effects on the booking performance, which is also investigated by Hamilton et al. (2017), who conclude that amenities can increase the number of stays and enhance customer retention in the hotel industry. This strengthens the believe that amenities can also improve the booking performance on Airbnb.

In addition, a potentially important factor in determining the benefits of a listing for a guest, and is sometimes overlooked, is facts and data about the property's safety and security. Hosts can mention both benefits in their description (e.g., "safe building" or "safe neighbourhood") or they can indicate so-called safety features, such as a smoke detector or fire extinguisher. These features are categorized within the list of amenities on Airbnb, that distinguishes between five different safety features. Several studies have emphasized the lack of security and safety on Airbnb properties, which is still one of the biggest barriers for guests (Mao & Lyu, 2017). This also implies that hosts who provide more safety features could have the competitive edge over hosts who do not. Thus, both the number of amenities and the number of safety features can be persuasive factors in the customer decision making process.

Another fact or data the host can utilize to increase the logical appeal of the listing is the cancellation policy. Specifically, the host determines the cancellation policy and these choices can help increase (or decrease) the appeal of the listing by transferring risk from/to the host and to/from the guest. The host faces three standardized options for the cancellation policy that differ in time and amount of refund. The flexible and moderate cancellation policy allow free cancellation at least 24 hours and five days respectively before the check-in time. The flexible policy refunds the full amount minus the first night and service fee if the guest cancels after this period, but before the check-in. In the same situation, the moderate policy refunds half the amount minus the first night and service fee. The strict policy only allows guests to cancel 48 hours after the booking as long as they cancel at least 14 days before the check-in[2]. Also, guests cancelling after this period, but up to seven days before the check-in, receive a fifty percent refund (Airbnb, 2020e). Chen et al (2011) stress that, unlike the amount of cancellation fee, a more lenient policy can positively influence guests to book hotels, suggesting that cancellation policy can be used as tool for revenue management. From the guest's point of view it would be logical to prefer a flexible cancellation policy in the case of unforeseen events. Thus, a more lenient policy could more effectively persuade guests to book.

A last logical appeal that guests can use when evaluating a property are reviews from other guests. Even though this is not under the direct control of the host, it does influence the appeal of a property. Specifically, guests can give a rating for the overall experience and for categories, such as cleanliness and location, after their stay. The most prominent rating is displayed at the top and comprises the aggregate of the overall experience, which ranges from zero to five and is based on at least three different guest ratings (Airbnb, 2020b). Generally, online ratings are considered to be a quickly available informational source that can influence customer decisions (Chen & Xie, 2008). Gavilan

[2] In special circumstances and by invitation only, a host can apply the super strict 30/60 days policy. However, since this occurs very rarely this research considers these policies as a regular strict policy.

et al. (2018) find a nuance between high and low ratings, where people tend to trust low ratings more than high ratings. Moreover, merely high ratings in combination with a sufficient number of reviews are deemed as reliable indicator for the quality (Gavilan et al., 2018). This idea is supported by the social proof heuristic, which implies that the more people agree on something the more likely an individual will think likewise. This heuristic is especially evident in situations where easily processable information is provided by peers (Rao et al., 2001), such as ratings provided by other guests on Airbnb. This enhances the belief that on the Airbnb platform guests find the rating an important information source regarding the quality of listing, which could positively influence their booking behaviour.

3. Data

To identify the persuasive power of the three aforementioned appeals in the sharing economy, I exploit a rich dataset with Airbnb listings in Amsterdam, extracted from Inside Airbnb[3]. Inside Airbnb provides publicly accessible data that is verified, analysed and combined for major Airbnb cities. Moreover, this data is periodically compiled, allowing to investigate specific time periods (Inside Airbnb, 2020). In this research, I use data that is extracted on December 4th 2017 and look at the occupancy rate over the following year, until December 4th 2018. Before clarifying this specific time period I should first elaborate on the use of occupancy rate as performance metric.

Unfortunately, the occupancy rate is not publicly available and hence not present in the data provided by Inside Airbnb. It is, however, possible to deal with this by proxying the occupancy with the number of reviews over a period. Previous literature shows the association of the number of reviews with performance in the hospitality industry, identifying this as a good proxy for hotel sales (Ye et al., 2009). Therefore, it is a realistic estimate of the occupancy rate, given that only guests that have successfully booked a listing can write reviews. In addition to that, estimating the occupancy rate for Airbnb is quite challenging and numerous methods have been discussed and applied (Inside Airbnb, 2020; Budget and Legislative Analyst's Office, 2015; Marqusee, 2015). Marqusee (2015) uses three methods for measuring the occupancy rate and demonstrates an increasing difference in estimates between these methods as an Airbnb listing is more frequently used. To specify, for the top ten percent most frequently used listings, the least conservative method produces an occupancy rate that is roughly sixty percent higher than the most conservative method. An important assumption is the review rate, which is the rate at which customers actually leave reviews, since it is not mandatory to do. Eventually, Marqusee (2015) uses a rate of seventy-two percent, based on information provided by Airbnb's CEO Brian Chesky. However, it should be seen as an unreliable source as the rate might be overestimated

[3] Data used in this research can be found through the following link: http://insideairbnb.com/get-the-data.html

due to legal concerns[4]. Furthermore, the Budget and Legislative Analyst's Office of San Francisco (2015) comments on this by showing that their estimate of the occupancy rate in New York is thirty and a half percent, based on data from Inside Airbnb. They attribute the large difference with Brian Chesky to a rather dubious statement that Airbnb might remove and alter the number of reviews. Nevertheless, the analyst's office neglected lost reviews due to the deletion of listings, which would likely increase the review rate. Therefore, this research adopts fifty percent as review rate, which is in between the former two estimates and is in line with Inside Airbnb (2020).

Consequently, this implies that the estimated number of bookings is twice the number of reviews a listing receives. Yet, to get a more truthful estimate of the occupancy rate, the length of each stay is also taken into consideration. A research by consultancy agency Ecorys (Briene et al., 2018), published and funded by Airbnb, estimates that guests in Amsterdam stay for an average of 3.4 nights. The latter is adopted in this research, unless the minimum required number of nights exceeds this number. If so, that minimum nights value is used. In the end, the occupancy rate is approximated by multiplying estimated bookings by the minimum nights, weighted over a year and capped at one, resulting in 404 observations that are censored.

With this in mind, a substantial difference between hotels and Airbnb listings still exists, which is the availability to book. Where it is fairly safe to assume that hotels offer rooms on a continuous basis, Airbnb listings are often available for a specific period of time when it suits the owner. Even though the data provides insights into the availability to book the place on a monthly and yearly basis, some concerns arise. Admittedly, it is unlikely that hosts indicate on which days a guest can book the place one year in advance. However, taking only shorter periods into account is more troublesome because the Airbnb calendar lacks differentiation between booked days and unavailable days (Inside

[4] In 2012 Brian Chesky posted this statistic on https://www.quora.com/profile/Brian-Chesky. In many countries there are strict Airbnb regulations that impose a maximum number of nights a host can rent the property per year. Since the occupancy rate is not publicly available, the best proxy is the number of reviews. Therefore, it could be in the interest of Airbnb to hold on to a overestimated review rate, which decreases the total number of proxied stays and avoids legal conflicts.

Airbnb, 2020). Thus, when taking shorter time periods into account it is likely that more popular listings are already booked instead of unavailable, which would produce misleading results. Also, Airbnb enforces regulations in 2018 indicating that, for entire homes and apartments, there is a limit of renting out one's property of sixty nights per calendar year[5] (Airbnb, 2020d). Therefore, the most logical starting date for a yearly analysis is at the beginning of the calendar year, of which December 4th 2017 is the closest possibility. Note that the year 2019 is deliberately not being used as the regulations became even stricter with a maximum of thirty nights a year (Airbnb, 2020d) and 2020 is expected to give less generalizable (yet interesting) results due to the global Corona crisis.

Lastly, because difficulties arise with lowly available listings, due to the indistinguishable difference between already booked and actually unavailable listings, it makes sense to only include listings that are highly available and active for a proper comparison. Here, I define highly available as a listing that can be booked on more than 60 days in the year. Active is defined as a listing that has received at least one review in the year prior or after December 4th 2017, which is necessary as some hosts quit renting out their place whilst not deleting the listings and not updating the availability accordingly.

3.1 Pre-Processing and Data Cleaning

To ensure that the data is useful and interpretable, some pre-processing and data cleaning procedures are applied. The original dataset has roughly twenty thousand observations and nearly one hundred variables. Firstly, the number of reviews from December 4th 2017 to December 4th 2018 comes from counting each review that occurs in this period from a dataset that is compiled on December 6th 2018. Subsequently, these are matched with the original dataset by the listing ID, assigning a value of zero to those listings with no reviews. Then, the occupancy rates are calculated by the aforementioned method.

[5] This maximum of sixty night, however, is not properly enforced, according to analyses conducted by Inside Airbnb.

Furthermore, separate text descriptions with the title, summary and space description are merged into one variable (*Description*). There is a deliberate choice to separately analyse the emotional tone of *Description* and the textual description about the host (*Host about*). This stems from the fact that both texts are far apart from each other, which potentially affects how guests perceive these, and secondly, to investigate whether there is a difference in affective content when selling the apartment compared to the host selling him- or herself. One point of notice is that there seems to be a maximum amount of text that Inside Airbnb scrapes, resulting in some abrupt endings in the text in case of very long descriptions.

In addition, when examining the textual data it also appeared that some texts included no more than three words, such as ‘Hi I am’ or that a text was not (completely) written in English. This is problematic when applying sentiment analysis and hence these are encoded to missing values. Besides, because the textual description about the host is not always present, but can be particularly important in the guest decision making process (Liang et al., 2020), I add a binary variable (*Host has description*) that indicates whether this text is present. To continue, missing values for *Security deposit* and *Cleaning fee* are assumed to be zero as the host did not mention it and guests evaluate the listing as if there are no additional costs. At last, the original data contains a list with amenities, from which the number of amenities and number of safety features are extracted and turned into separate numerical variables.

In terms of text pre-processing, I remove NA’s resulting from merging, any unknown characters and most punctuation marks, with some exceptions that are elaborated in the Methods section. Moreover, purely for the elastic net the data is randomly split into a training set of eighty percent and a test set of twenty percent and then separately standardized, needed to assess the out-of-sample performance of the model.

Then, a substantial number of variables are removed from the dataset because they; contain no, the exact same or no useable information (e.g. URLs or dates), are already represented by another variable (e.g. there are many variables that contain similar information about the neighbourhood of the property). At last, when checking for missing values the variable containing information about the square feet is removed due to too few observations (over ninety percent is missing). Also, every categorical variable with a category that encompasses less than one percent of the total observations is encoded to 'Other', if possible. Table 3 provides a description of all control variables used in this research.

Table 3. Description of all potential control variables used in this research

Variable	**Description**	**Levels**
Host response rate	Indication how often a host responds to inquiries and reservations	Between zero and one
Neighbourhood	The area of a listing	Twenty-two areas are identified
Exact location	Whether the location is precisely displayed	True/False
Property type	The type of accommodation that is offered	Seven property types are identified
Room type	What is available to the guests	Entire property, private room, shared room
Accommodates	Number of guests the property can accommodate	Numeric
Bathrooms	Number of bathrooms	Numeric
Bedrooms	Number of bedrooms	Numeric
Beds	Number of beds	Numeric
Bed type	The kind of bed that is provided	Pull-out sofa, real bed, other
Guest included	Number of guests that can initially book	Numeric
Extra people	Cost of extra people beyond guests included	Numeric
Minimum nights	Minimum number of nights a guest can book	Numeric
Maximum nights	Maximum number of nights a guest can book	Numeric
Price	Price per night in euro	Numeric
Security deposit	Down payment required by the host as measure of security	Numeric
Cleaning fee	Additional cleaning cost on top of the total rent	Numeric
Instant bookable	Whether guests can directly book without a host's approval	True/False
Guest picture	Whether a guest's profile picture is required	True/False
Guest phone verification	Whether guest phone verification is required	True/False
Host has description	Whether the host provides a textual description about him/her self	True/False

4. Methodology

In this section, the methods that are applied in this research are outlined. For a quick overview to which methods are used, their corresponding goal and which variables are included in the analyses, I refer to Table 4.

Table 4. Outline of methods used in this research

Method	**Goal**	**Variables included**
Sentiment analysis	Extract sentiment scores from each text in *Description* and *Host about.*	*Description, Host about*
Elastic net regression	Select variables that significantly affect the occupancy rate.	All control variables
Tobit regression	Estimate a linear relationship between variables whilst accounting for a censored dependent variable.	All core variables, all significant control variables

4.1 Sentiment Analysis

The first analysis that I apply is a form of text mining, namely sentiment analysis. The fundamental goal of sentiment analysis is to classify which sentiments are present in texts and whether expressions have a positive or negative attitude towards a subject matter (Nasukawa & Yi, 2003). Hereafter, this information can be employed to track product and brand attitudes in the online market place. In this research, sentiment analysis is purely applied to measure the emotional tone in the two earlier specified textual descriptions. Many researchers have come up with resourceful dictionaries and algorithms that are able to track sentiment, among which the Linguistic Inquiry and Word Count and the National Research Council (NRC) Canada are prominent examples. I use a dictionary developed by Minqing Hu & Bing Liu (2004b), also known as the Bing dictionary.

The Bing dictionary contains 6786 words, from now on referred to as affective words, of which roughly two thousand are classified as positive words and the rest as negative words. The word list is generated by researchers and is still being expanded. The basic idea is that affective words are identified in a text and are given a score of 1 (for positive words) or -1 (for negative words). Then, the sum of these scores shows the overall emotional tone of the text (Hu & Liu, 2004a). Yet, since this method only detects predefined affective words, it neglects potentially important factors such as negations and (de-)amplifiers. These could have a tremendous effect as the Bing dictionary identifies 'the apartment is not very spacious' as a positive text whereas the tone is clearly the other way around. Therefore, I incorporate these factors by calculating the polarity score (δ), which is represented in equation (1). Note that the list of words containing negations and (de-)amplifications is derived from the QDAP dictionary, which is composed of multiple text analysis dictionaries and word lists.

$$\delta = \frac{x_i^T}{\sqrt{n}} \quad (1)$$

Where:

$$x_i^T = \sum((1 + c\left(x_i^A - x_i^D\right))\,\omega(-1)^{\sum x_i^N}) \quad (2)$$

$$x_i^A = \sum(\omega_{neg} x_i^a) \quad (3)$$

$$x_i^D = \max\,(x_i^{D'}, -1) \quad (4)$$

$$x_i^{D\prime} = \sum(-\omega_{neg} x_i^a + x_i^d) \quad (5)$$

$$\omega_{neg} = (-1)^{\sum x_i^N} \quad (6)$$

Here, x_i^a, x_i^d and x_i^N take on the value of one whenever an amplifier, de-amplifier and negator is encountered respectively. x_i^T is the context cluster in which a specific length of text, ranging from four words prior and two words after an affective word is encountered, is scanned for amplifiers (x_i^a), de-amplifiers (x_i^d) and negators (x_i^N) and theoretically ranges from -5.8 to 5.8 (an example is provided later). Furthermore, x_i^A represents the sum of all encounters with an amplifier (x_i^a) in a context cluster (x_i^T), which can theoretically occur a maximum of six times (e.g. four amplifiers before and two amplifiers after an affective word is encountered) and is weighted by ω_{neg}. The latter represents the overall positive or negative effect of the negators in a context cluster (x_i^T), e.g. 'not good' gives an overall effect of $(-1)^1$, which is minus one and hence a negative effect. x_i^D is in essence the same as x_i^A, but then for de-amplifiers. Important, however, is that x_i^D is assumed to be zero whenever no de-amplifier (x_i^d) is encountered. Besides, the maximum negative effect x_i^D can have is constrained to minus one. Moreover, c is the amplifier weight between zero and one, but in this research is set to 0.8 and ω is the weight of the detected affective word, which is either one (for a positive word) or minus one (for a negative word), based on its classification in the Bing dictionary. At last, n is the number of words in the full text.

To elaborate, consider the slightly negative sentence 'the house is not very ideal for families'. The algorithm encounters the affective word 'ideal', which is positive in the Bing dictionary and consequently gets a weight of one ($\omega = 1$). The context cluster (x_i^T) is 'house is not very ideal for families', in which there is one negator ($x_i^N = 1$) and one amplifier ($x_i^a = 1$). Hence, the overall effect of negators in this cluster is minus one ($\omega_{neg} = -1$), which leads to $x_i^A = -1$ and as no de-amplifiers are encountered $x_i^d = 0$. This means that the raw polarity score in the context cluster equals -0.2 ($x_i^T = 1 + \left(0.8 * (-1 - 0)\right) * 1 * (-1)^1$). Now, we control for the number of words in the total text ($n = 7$), which gives the polarity score of -0.076 ($\delta = \frac{-0.2}{\sqrt{7}}$). Theoretically, this also means that the raw polarity score (δ) in a context cluster (x_i^T) can range between -5.8 and 5.8, e.g. by using a dubious sentence like 'very seriously enormously immense spacious severely truly', which gives 5.8. In practice, however, these scores are more centred.

Also, it seems counterintuitive to measure and interpret the effect of using an additional affective word, e.g. one would infer that the occupancy rate increases/decreases by an estimated value when one simply adds affective words to a text. This might hold in theory, but in practice it is about the overall valence of a text. Therefore, it is more relevant to compare all textual descriptions and categorize them based on their polarity score (δ). In the Theoretical Framework I refer to the potentially important distinction between persuasive and informative marketing communication by Narayanan et al. (2005). This research makes a similar distinction and categorizes the texts as persuasive or informative. Methodically, this means that first all polarity scores are estimated for each text in *Description* and *Host about*. Then, the average is set to zero by standardizing these scores, meaning that negative scores have a low polarity score and positive scores have a high polarity score, on average. Consequently, all scores of zero or lower are categorized as informative since these text contain no affective words, fewer positive affective words or more negative effective words. In addition, all scores above zero are labelled persuasive as these text include more positive affective words or fewer negative affective words. This is summarized in Table 5. Note that I cannot infer the reason why a score is higher or lower before the actual analysis, which is therefore discussed in the Results.

Table 5. Text categories

Category	**Standardized polarity score (δ)**	**Potential reasons**
Informative	$\delta \leq 0$	▪ No affective words ▪ Fewer positive affective words ▪ More negative affective words
Persuasive	$\delta > 0$	▪ More positive affective words ▪ Fewer negative affective words

In terms of text pre-processing, it is important not to stem words, as it is not recognized by the Bing dictionary, and to keep all original words in the text. Even though a neutral word holds no value in the context cluster (x_i^T), it does affect the total word count (n). Hence, pre-processing procedures are limited to removing various punctuation marks, NA's that appear as a result of merging missing values into *Description* and excess spaces. It is, however, imperative to keep all comma's and transform all dots, colons and semicolons to comma's in the text as the polarity algorithm can differentiate between separate sentences solely based on comma's. To further specify, if a comma is encountered before an affective word in context cluster (x_i^T), then only words after the comma are taken into account. This is necessary because these marks can signal a shift in one's thoughts, meaning that those should not be taken into account as they can untruthfully influence the polarity score (δ). Finally, the single acute accent (') remains in the text as negation words often use those (don’t, isn’t, hasn’t, etc.) and the algorithm cannot recognize these negators otherwise.

4.2 Elastic Net

The elastic net regression is a fairly new regularization technique that was first introduced by Zou & Hastie (2005). In essence, it combines two well-known regularization techniques; the ridge regression and the lasso regression. Firstly, the ridge regression is introduced to counter accuracy and interpretability problems that arise using the simple linear regression, such as the ordinary least squares (OLS) that minimizes the residual sum of squares. This is done through adding a penalty parameter, lambda (λ), that penalizes the size of the coefficients. Mathematically, the loss function (L) of the ridge regression is represented by equation (7).

$$L(b_1, \dots b_m) = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{m} x_{ij} b_j)^2 + \lambda \sum_{j=1}^{m} b_j^2 \quad (7)$$

Where b_j represents the unknown regression weights for variable $j = 1, \dots m$. x_{ij} are elements from the $n\ x\ m$ matrix of explanatory variables X and y_i implies the value of the response variable for object $i = 1, \dots, n$. When λ approaches zero the estimation of b_j acts more like an OLS regression, whereas when it reaches infinity the coefficient weights are being penalized more strongly towards zero. This is also called L_2-shrinkage.

Yet, even though the ridge regression reduces the complexity of a model and increases the predictive accuracy compared to OLS, the interpretation and sparsity remains an issue as no coefficients are reduced to zero (Tibshirani, 1996). Therefore, Tibishirani (1996) introduces the least absolute shrinkage and selection operator (LASSO) to account for this problem. This is mathematically displayed by equation (8).

$$L(b_1, \dots b_m) = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{m} x_{ij} b_j)^2 + \lambda \sum_{j=1}^{m} |b_j| \quad (8)$$

To elaborate, everything is the same as for the ridge regression, with the only exception that the penalty is the sum of absolute values of b_j instead of the square of the coefficients. Hence, LASSO puts more emphasis on the magnitudes and now when λ approaches infinity the size of the coefficients can be penalized to be exactly zero, effectively performing variable selection. This is referred to as a L_1-shrinkage.

However, despite LASSO being more attractive compared to ridge in modern data research due to sparsity concerns (Zou & Hastie, 2005), previous literature points out there is no method that is consistently superior to the other (Tibshirani, 1996). With the objective of using regularization purely for variable selection in this research, LASSO seems to be the best choice. Still, Zou & Hastie (2005) point out shortcomings of LASSO when using it for variable selection that potentially apply for my data. Most importantly, when pairwise correlations among the same group of variables are high, LASSO tends to randomly choose one of those variables as predictor. One can expect high correlations in the Airbnb dataset between, for example, the number of beds, bedrooms and accommodates. This is an undesirable consequence since all control variables that significantly influence the occupancy rate should be taken into account in the final model. Therefore, the elastic net is the preferred regularization technique as it provides a solution to the aforementioned problems.

Elastic net regression performs variable selection through shrinkage whilst accounting for highly correlated variable groups, combining the positives of ridge and LASSO. This is achieved by adding the penalty terms of (7) and (8) together and by introducing a mixing parameter, alpha (α). In equation (9), the mathematical representation of the elastic net regression is shown.

$$L(b_1, \dots b_m) = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{m} x_{ij} b_j)^2 + \lambda(\alpha \sum_{j=1}^{m}|b_j| + (1-\alpha)\sum_{j=1}^{m} b_j^2) \quad (9)$$

Again, the interpretation is identical to the ridge and LASSO equations, except for parameter α, which is bound between zero and one. Intuitively, when α is zero it behaves just like a ridge regression and when it equals one is functions as a LASSO. Every value between zero and one indicates that both L_1 and L_2 penalties are applied.

Furthermore, now that the idea behind the elastic net is explained, some challenges arise regarding the tuning of λ and α. Generally, regularization techniques reduce the variance of an OLS model by introducing bias, called the bias-variance trade off. The high variance that can be observed in OLS models stems from the (often) unjust assumption that real world relationships are linear. Thus, when there are many predictor variables, or in the case of multicollinearity, the variance is overly high. Regularization techniques such as ridge and LASSO handle these problems by finding optimal levels of bias and variance through the tuning of λ. The elastic net, however, has an extra tuning parameter, α, that should also be taken into account. There are multiple ways of determining the optimal levels, such as looking at the smallest information criterion of which AIC and BIC are frequently used. This research sets out to tune both parameters through minimizing the mean squared error (MSE).

Moreover, training the model is done through K-fold cross-validation, which randomly splits the data in K number of parts, called folds. Consecutively, all but one folds are used to estimate values of λ, given a set of values of α, that minimize the MSE in each particular iteration. The remaining fold is used as an out of sample testing set to validate the results. This process is iterated K times and eventually the λ and α with the overall lowest MSE is the preferred model. Yet, this research adopts the λ that is one standard error away from the best cross-validated lambda as it tends to decrease model complexity by excluding considerably more variables. Also, a K of 10 is chosen, which is regarded common practice. Note that one hundred different values of α, from zero to one with steps of 0.01, are tested.

In addition to that, elastic net regression cannot manage variables with more than two categories. To account for this, each individual category is encoded to a binary variable. Note that these remain encoded as binary variable, in order to select only those control variables that significantly affect the dependent variable and hence decrease complexity. Lastly, considering that the elastic net puts constraints on the magnitude of the coefficients related to each variable, it is imperative to standardize the data. This is necessary for both numeric and categorical variables (encoded to dummy variables) to overcome harmful effects from variables differing in size and to have a fair penalization scheme (Tibshirani, 1997). One reduces interpretability of the model as the relative scaling between different types of variables is arbitrary to some extent. However, because the sole goal is variable selection this does not outweigh the alternative of not scaling categorical variables.

4.3 Tobit Regression

The final step of this research is to estimate the effects of the core variables and remaining control variables on the occupancy rate. To arrive at these effects, I estimate a tobit model, also called a censored regression model. This type of regression is used when there is either upper- or lower-censoring, meaning that a value is cut off at a certain threshold, in a non-negative response variable. Because I proxy the occupancy rate by the number of reviews it is not exactly equal to the actual number of stays. Yet, due to the inclusion of an assumed review rate and minimum nights in calculating the occupancy rate, some cases where the occupancy rate exceeds one appear. These values, however, are censored to a maximum value of one. Moreover, previous research also suggests the apparent problem of selection bias, which occurs in situations where it is not possible to establish the actual value of the dependent variable (Qazi et al., 2016). The tobit model estimates a linear model that accounts for these situations.

The tobit regression behaves similarly to a multiple regression model. However, due to the aggregation of observations that is caused by censoring the dependent variable at the upper bound, the assumption of multiple regression that the maximum value of a dependent variable equals its expected value, given any value of the independent variables, does not hold (Tobin, 1958). As a result, there is a linear estimation of the nonlinear aggregation of observations. Mathematically, tobit regression starts by considering the stochastic model in equation (10).

$$y_t = X_t\beta + u_t \quad (10)$$

Here, y_t is the dependent variable, X_t and β are vectors of predictor variables and coefficient estimates respectively, where $t = 1, 2, \ldots, N$ and N equals the number of observations. Last, u_t represents the independently distributed error term, inferring that the variance (σ^2) of u_t is constant with a mean of zero. Note that equation (10) only holds when $X_t\beta + u_t > 0$ and in any other instances ($X_t\beta + u_t \leq 0$) equals 0 due to the presumption of a non-negative response variable (McDonald & Moffitt, 1980). Then, the expected value of all observations (Ey) is given by (Tobin, 1958):

$$Ey = X\beta F(z) + \sigma f(z) \quad (11)$$

Where, $z = X\beta/\sigma$ and $f(z)$ implies its normal distribution. Moreover, $F(z)$ represents the cumulative normal distribution function (excluding singular subscripts), indicating the probability that a predictor variable, weighted at x, assumes a value less than or equal to x . Now, equation (12) estimates the expected value of y for those observations that exceed the boundary, denoted as Ey^*.

$$Ey^* = X\beta + \sigma f(z)/F(z) \quad (12)$$

To elaborate, the expected value of y for observations that exceed the limit (Ey^*) is estimated by $X\beta$ plus the expected value of the error term that follows: the so-called truncated normal distribution. This truncated normal error term ($\sigma f(z)/F(z)$) is obtained in a similar fashion as a normally distributed error term, but by bounding the variable from above and/or below. Consequently, when considering a censored, non-negative dependent variable, its expected value depends on:

$$Ey = F(z)Ey^* \ (13)$$

Thus, the expected value of all observations (Ey) is affected by the expected value of those observations outside the boundary (Ey^*) and the probability of exceeding this boundary ($F(z)$).

5. Results

In this section the results of the three aforementioned analyses are presented. First, I start with descriptive statistics of the core variables, followed by the sentiment analysis, elastic net regression and thereafter the tobit regression containing the final results.

5.1 Descriptive Statistics

To gain a better understanding of the data, I provide several descriptive statistics concerning the core variables of this research (see Table 6).

Table 6. Descriptive statistics of the core variables (N = 3636)

		Min	Max.	Freq. of 1 / Mean	Percentage / SD
Ethos (credibility)					
Super host badge	(0: No badge, 1: Badge)	0	1	925	25.4%
Number of reviews		0	578	39.11	59.14
Response time	A few days or more	0	1	45	1.2%
	Within a day	0	1	841	23.1%
	Within a few hours	0	1	955	26.3%
ID verification	Within an hour	0	1	1495	41.1%
Pathos (emotion)					
Description	Informative	0	1	1883	51.8%
	Persuasive	0	1	1555	42.8%
	No text	0	1	198	5.4%
Host about	Informative	0	1	1162	32.0%
	Persuasive	0	1	957	26.3%
	No text	0	1	1517	41.7%
Logos (facts and logic)					
Amenities		0	68	18.55	8.49
Safety features		0	5	2.29	1.43
Cancellation policy	Flexible	0	1	646	17.8%
	Moderate	0	1	1292	35.5%
	Strict	0	1	1698	46.7%
Review score		20	100	95	5.47
Occupancy Rate		0	1	0.33	0.33

Note that missing observations are included in calculating the percentages in the last column, meaning that totals might not add up to one hundred percent. In addition, all missing values occurring in *Description* and *Host about* are encoded to *No text* as excluding these observations in the final model results in a considerable loss of information. For any other core variables the number of missing observations remain low. To delve into the statistics, the mean occupancy rate is 0.33, meaning that the average host receives guests overnight for 120 days per year, which is quite substantially. Also, almost every host responds within a day, which is not surprising as previous literature suggests the positive effect of a response time within one day on guests' trust (Sparks et al., 2016). In addition, the review score is really high with 95 on average, which is roughly the same in other big cities, such as London, Paris or Barcelona (Ert et al., 2016).

5.2 Results of Sentiment Analysis

After the text was properly pre-processed, the polarity scores for each individual listing were calculated. The distribution of these scores are given in Figure 2.

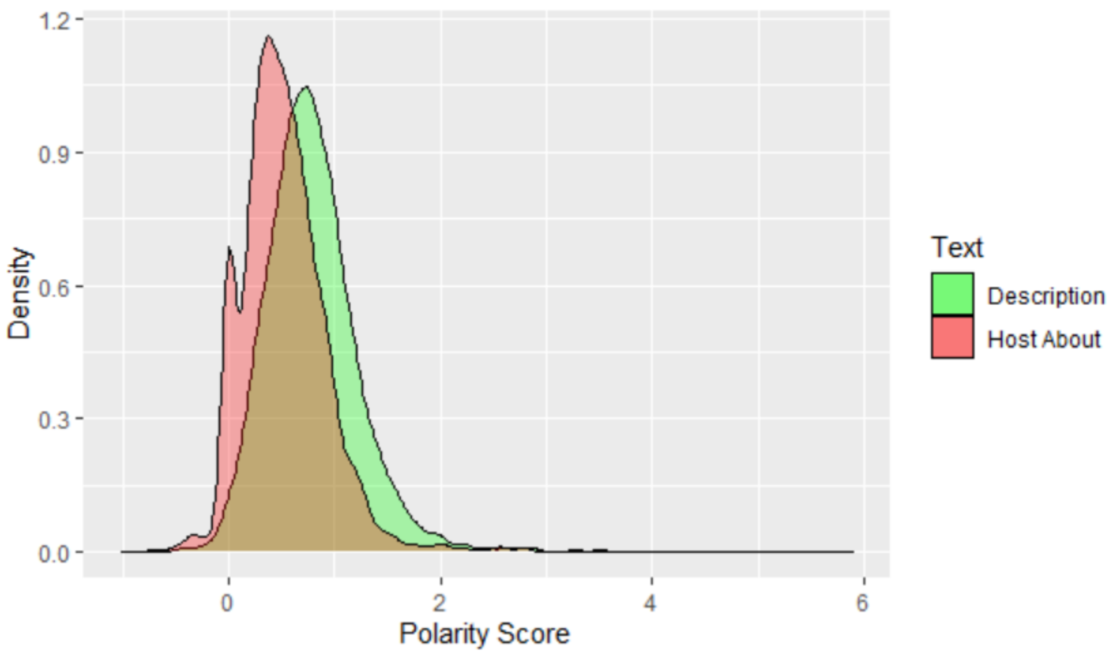

Figure 2. Distribution of sentiment scores using the polarity algorithm

From this figure it becomes clear that the overall positive sentiment is higher when hosts write a description as compared to writing about themselves, with a mean polarity score of 0.804 for *Description* and 0.525 for *Host about*. However, it is unclear whether this is attributed to a relative use of more positive words, or by including fewer negative words, which decreases the polarity score. To infer this, the number of positive and negative emotional words is counted for each text, which is shown in Table 7 along with basic descriptive statistics.

Table 7. Descriptive statistics for positive and negative affective words

Variable	*Mean*	*SD*	*Min*	*Max*	*Informative mean*	*Persuasive mean*
Positive words description	8.98	5.31	0	31	5.83	12.80
Negative words description	0.37	0.75	0	6	0.36	0.38
Positive words host about	4.21	4.06	0	41	2.14	6.71
Negative words host about	0.21	0.65	0	12	0.18	0.24

On average, hosts include more positive and negative emotional words in *Description* compared to *Host about*, which is not surprising as the texts are usually longer. Therefore, I consider the ratio of positive words to negative words, which is higher in *Description* (24.27) than in *Host about* (20.05), suggesting that the higher polarity scores in the description of the place, compared to writing texts about the hosts, is due to the usage of relatively more positive affective words.

Moreover, in the Methodology I mention the distinction that is made between informative and persuasive classified texts, based on the standardized polarity scores. Their frequencies can be found in Table 6. Note that a text in *Description*/*Host about* is classified as informative if the polarity score is 0.804/0.525 (mean polarity score) or lower and as persuasive if higher. Yet, I could not provide the reason for the difference between an informative text (low polarity score) and persuasive text (high polarity score). Based on the mean polarity score for informative and persuasive texts, in Table 7, one can infer that the primary reason that a text is labelled informative or persuasive is due to the relative usage of positive affective words[6]. To shortly elaborate, the difference in mean between negative affective words in informative and persuasive texts is very low, whereas this difference between positive words is fairly high, which rules out the possibility that the categories differ because of the usage of negative words. Furthermore, the mean value for negative words is very low, compared to a high value for positive words, which rules out that a text is informative (low polarity score) because of

[6] (De-)amplifications and negations are not taken into consideration here.

ambivalence[7]. Specifically for *Host about*, besides the relative usage of positive affective words, texts classified as informative can also be partially explained by the absence of affective words in these texts, because the mean value (2.14) is close enough to zero for this to occasionally occur.

Finally, to get an idea of what specific words are used in informative and persuasive texts, I consider the words that are frequently mentioned in each category. However, common words such as Amsterdam or apartment are likely to occur in both kind of texts, e.g. 'the apartment was very spacious' (persuasive) or 'the apartment was big enough' (informative). Thus, looking at the difference in word frequencies between persuasive and informative texts is much more insightful. This is visualized for *Description* and *Host about* in Figure 3 and 4 respectively. Note that the words are ordered from most specific to least specific, meaning that 'house' in the left panel of Figure 3 is the most specific word that is used in informative texts about the description.

Firstly, for *Description* you can see the word 'apartment' in the right panel of Figure 3, meaning that this word is typically used in persuasive texts, such as the previous example 'the apartment was very spacious'. Moreover, in the left panel of Figure 3 the word 'Amsterdam' appears, implying that this is more often used in informative texts, such as 'the house is located in the city centre of Amsterdam'. Naturally, words that appear relatively often in persuasive texts are positive adjectives to describe the property or place, such as 'spacious' or 'comfortable', which you can see in the right panel in Figure 3. In contrast, informative texts are more practically focused, e.g. by providing geographical information or clarifying the time it takes to go somewhere. One can infer this by looking at the words in the left panel of Figure 3, such as 'station', 'minutes', '3' (minutes) or 'centre'.

Furthermore, typical persuasive and informative words in *Host about* differ with those for

[7] In this context, ambivalence means that there is a polarity score around zero, caused by using roughly the same amount of positive affective words as negative affective words.

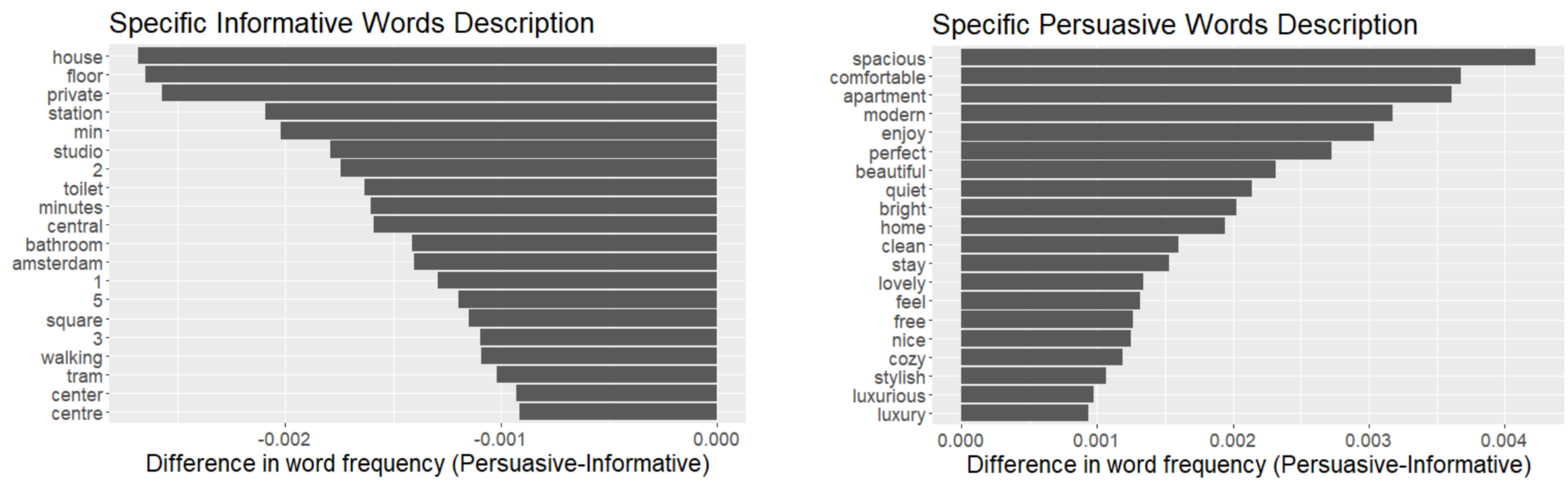

Figure 3. Specific informative and persuasive words in *Description*

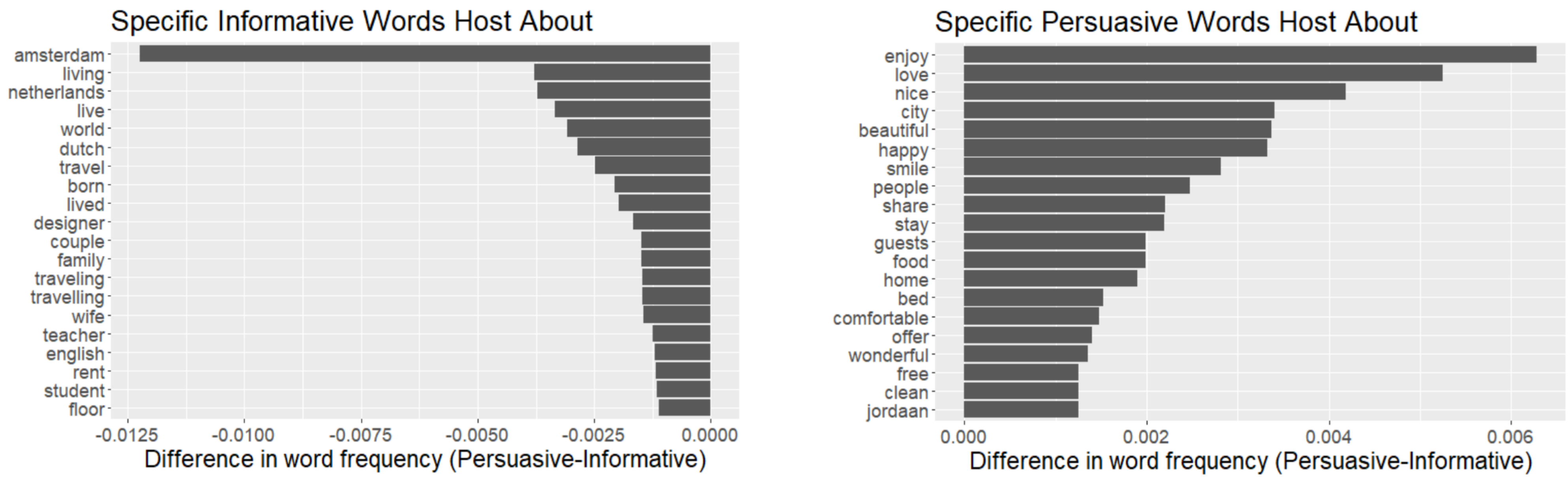

Figure 4. Specific informative and persuasive words in *Host about*

Description. In the right panel of Figure 4, one can see a mix of positive adjectives, such as 'nice' and 'beautiful', adverbs like 'enjoy' and 'share, but also nouns, e.g. 'home' and 'guests'. One can think of a text like 'we love to share our beautiful home with guests'. On the other hand, informative texts deal with the personal life of the host, such as hobbies, origin or profession. In the left panel of Figure 4 you can see that most words are related to these topics, like 'travelling', 'born' or 'teacher'. Thus, this indicates that texts in which these topics are mentioned leave out affective words and are rather formal, such as 'I am born in the Netherlands, work as a teacher and travel a lot'.

5.3 Results of Elastic Net Regression

In this section, merely results are presented and shortly elaborated. First, the lowest mean squared error (MSE) is obtained for an alpha of 0.70 (see Figure 5). Right beneath, Figure 6 displays the corresponding cross-validated lambdas.

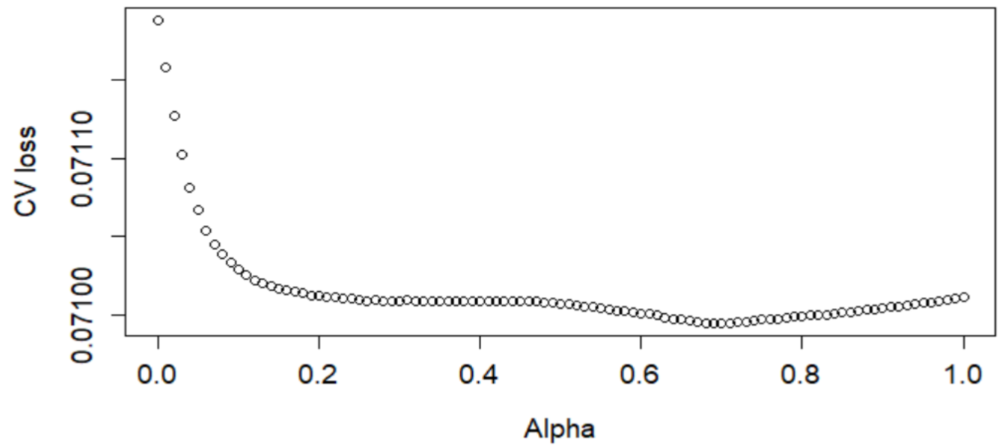

Figure 5. Lowest cross validated MSE for each value of alpha

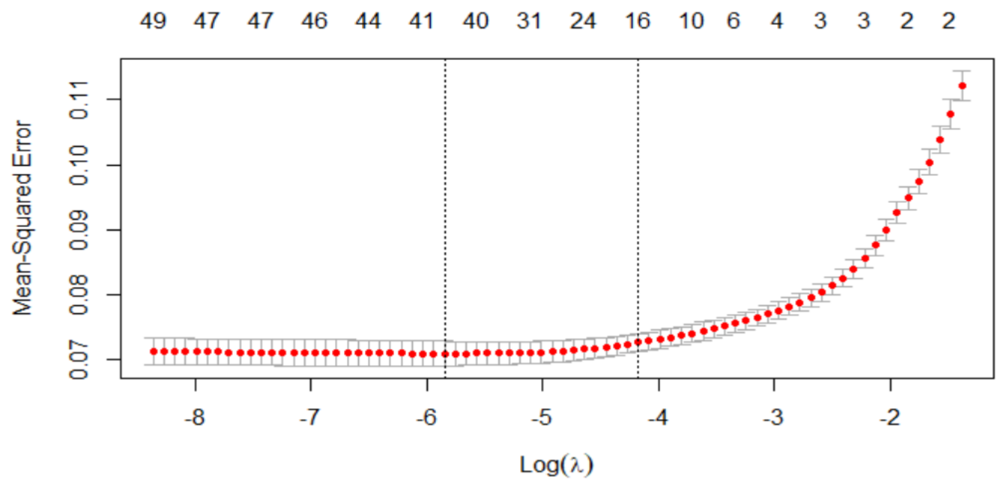

Figure 6. MSE for various cross-validated lambda's, given the optimal value of alpha

Here, the left vertical line is the best cross-validated lambda with a value of 0.0029. Yet, the lambda that is one standard error away from this lambda is favored, corresponding to the right vertical line with a value of 0.0153. The implementation of this lambda results in sixteen significant control variables, of which the coefficients are shown in Table 8 (eight variables in the left part and eight variables in the right part). For a table with all control variables, including the insignificant ones, consult Appendix A. Note that (Nb), (Pt) and (Rt) represent neighbourhood, property type and room type.

Table 8. Estimated coefficients of elastic net regression on occupancy rate

Variable	*Coefficient*	*Variable*	*Coefficient*
Host response rate	0.0183	(Pt) Other	0.0007
(Nb) Bos en Lommer	-0.0035	(Rt) Entire property	-0.0494
(Nb) Centrum-Oost	0.0126	(Rt) Private room	0.0846
(Nb) Centrum-West	0.0181	Beds	0.0038
(Nb) Gaasperdam - Driemond	-0.0086	Price	-0.0311
(Nb) Ijburg - Zeeburgereiland	-0.0014	Security deposit	-0.0018
(Nb) Oostelijk Havengebied – Indische Buurt	-0.0003	Cleaning fee	-0.0049
Exact location	0.0065	Instant bookable	0.0708

Regarding the interpretation, if *Host response rate* increases by one unit of standard deviation (due to standardized data), the occupancy rate increases by 0.0183. A binary variable is interpreted differently, e.g. if a property in Amsterdam is located in Centrum-West, the occupancy rate is 0.0181 higher as compared to a listing that is not located here. Lastly, the in-sample RMSE of the model is 0.2664, whereas the RMSE on test data is 0.2687, implying that the model did not overfit and performed reasonably well given that solely control variables are used.

5.4 Results of Tobit Regression

Finally, the Tobit regression combines all core variables and significant control variables into one model to assess the overall effect of *ethos* (credibility), *pathos* (emotion) and *logos* (facts and logic) on the occupancy rate, whilst accounting for a censored occupancy rate at one. The results can be found in Table 9. At the bottom, the Pseudo R2 represents the model's goodness of fit. Note that the calculation is based on McFadden's R squared, which considers the likelihood of the full model compared to the likelihood of a model with solely an intercept[8]. A value of 0.785 implies that the model fit is pretty well relative to the null model. However, the value itself is relative and does not tell whether the model it sufficiently precise. Therefore, I look at *LogSigma*, which represents the logarithmic standard deviation of the residuals and can also be regarded as a goodness of fit measure. The advantage is that this is an absolute measure as compared to the relative measure provided by the Pseudo R2. After exponentiation the value becomes 0.245, which indicates the average difference between the true occupancy rates and the model's estimates. This value suggests that the model performs reasonably well.

Regarding *ethos* (credibility), a super host tends to have an occupancy rate that is 0.053 (p<.001) higher compared to hosts without the super host badge. Moreover, the number of reviews have a positive effect on the occupancy rate, namely 0.003 (p<.001) per review, which is considerable as the average number of reviews approximates forty. An unexpected result, however, is the significant negative effect of a host ID being verified, compared to hosts without ID verification. In addition, a misleading coefficient estimate due to multicollinearity is ruled out. When contemplating, a potential explanation is that hosts willingly exploit their higher reputation, obtained by having a verified ID, and thus increase their price (Wang & Nicolau, 2017). Also, since this model shows a significant negative coefficient for price of -0.001 (p<.001) for each additional euro, this clarification is very plausible. Finally, the average response time does not affect the occupancy rate.

Secondly, persuasion through *pathos* (emotion) produces interesting results. Where most literature expects persuasion through affective content to positively influence the guest's decision making, it is actually the other way around for *Description*. Here, writing a less emotional orientated

[8] $R^2_{McFadden} = 1 - \frac{\log\,(Likelihood\ current\ model)}{\log\,(Likelihood\ null\ model)}$

Table 9. Results of tobit regression on occupancy rate

		Coefficient	Std. Err.	Pr (> t)
Constant		-0.100	0.109	0.360
Ethos (credibility)				
Super host badge		0.053***	0.011	0.000
Number of reviews		0.003***	0.000	0.000
Response time	Within a day	-0.058	0.061	0.340
	Within a few hours	-0.003	0.064	0.964
	Within an hour	0.065	0.064	0.307
ID verification		-0.022*	0.009	0.019
Pathos (emotion)				
Description	Persuasive	-0.033***	0.009	0.000
	No text	-0.041*	0.020	0.044
Host about	Persuasive	0.011	0.012	0.358
	No text	0.052***	0.011	0.000
Logos (facts and logic)				
Amenities		0.002**	0.001	0.004
Safety features		0.004	0.004	0.255
Cancellation policy	Moderate	0.016	0.014	0.235
	Strict	-0.000	0.013	0.999
Review score		0.001	0.001	0.142
Control variables				
Host response rate		0.112	0.060	0.064
Exact location		-0.007	0.010	0.475
Beds		0.017***	0.003	0.000
Price		-0.001***	0.000	0.000
Security deposit		-0.000	0.000	0.545
Cleaning fee		-0.000*	0.000	0.023
Instant bookable		0.130***	0.013	0.000
(Nb) Bos en Lommer		-0.060**	0.022	0.007
(Nb) Centrum Oost		0.053***	0.015	0.000
(Nb) Centrum West		0.071***	0.013	0.000
(Nb) Gaasperdam Driemond		-0.099*	0.041	0.016
(Nb) Ijburg Zeeburgereiland		-0.081**	0.028	0.004
(Nb) Havengebied Ind. Buurt		-0.068**	0.023	0.003
(Pt) Other		0.038	0.022	0.083
(Rt) Entire property		0.008	0.061	0.900
(Rt) Private room		0.171**	0.061	0.005
LogSigma		-1.406***	0.014	0.000

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; Log likelihood = -408.24; Pseudo R2 = 0.785

text benefits the host because, compared to an informative text, the occupancy rate is -0.033 (p<.001) lower for persuasive texts and -0.041 (p=.044) lower when there is no text. Hence, this supports the belief that guests prefer to have more practical information and are not persuaded by more positive affective words.

Furthermore, no significant effects were found between persuasive and informative texts in *Host about*. Yet, having no host description has a positive effect of 0.052 (p<.001) on the occupancy rate as compared to having an informative host description, which is rather unexpected. As earlier mentioned in the Data section, the texts of *Description* and *Host about* are far apart from each other. In fact, the text about the host is at the bottom of the page, for which you usually have to scroll down quite a bit. Therefore, based on all the available information before the text about the host, a guest could already have formulated a decision on whether to book or not before reading the *Host about*. Besides, there might be selection bias stemming from the supposition that foremost poor performing hosts write a host about as a final attempt to persuade potential guests.

As for *logos* (facts and logic), the only significant variable is the number of amenities that are present in one's place, increasing the occupancy rate by 0.002 (p=.004) per extra amenity. The number of safety features is not significant. To explain this, an important factor could be that the Netherlands is considered to be a safe country, decreasing the importance of safety enhancing features. Moreover, the insignificant effect of the review score is in line with previous theory, such as Gavilan et al. (2018), who stress that people tend to solely trust low ratings. Since the average rating is 95, it is no surprise that adverse effects of low ratings on occupancy rate are not found. At last, the cancellation policy also has no significant effect on the occupancy rate.

Continuing, most of the estimated effects of the control variables are quite straightforward, with significant negative effects for *Price* and *Cleaning fee* and a mixture of significant effects depending on the neighbourhood of the listing. The large positive effect of *Instant bookable* of 0.130 (p<.001) is because hosts forgo the opportunity to turn down any guest, making it rather attractive for 'less-desirable' visitors, such as partying students or population groups that are more prone to being discriminated against. In addition, the number of beds increase the occupancy rate by 0.017 (p<.001) per bed. A possible explanation is that Airbnb offers relatively fewer beds compared to normal hotels (Gutiérrez et al., 2017), suggesting that guests more frequently have to share a bed or sleep on an alternative such as a bunk-bed, which is naturally less optimal. Hence, offering more beds positively influences the guests' decision making. A final point of notice is the large effect of renting out a private room as compared to renting out something else (often an entire home or apartment), namely 0.171 (p=.005). In the Data section, I mention that Airbnb enforces regulations in 2018, which indicate a limit of renting out an entire home or apartment up to sixty nights a year, which likely explains the size of the coefficient.

Overall, when comparing *ethos* (credibility), *pathos* (emotion) and *logos* (facts and logic) with each other there are clear differences. First of all, *pathos* (emotion) has a relative negative effect on the occupancy rate as writing an informative description persuades guests rather than appealing to emotion by using positive affective words, which makes it the only ineffective persuasion strategy. After that, *logos* (facts and logic) is effective in persuading guests though the number of amenities. Even more, the relative positive effects of writing an informative description and appealing to facts and data (*logos*) are on average equally high, assuming an average number of nearly twenty amenities. Yet, if one offers more than the average number of amenities the persuasive power of *logos* (facts and logic) surpasses that of writing an informative description. Note that I exclude the positive effects of having no text compared to an informative text in *Host about* as I believe this is due to the previously considered reasons. Furthermore, in terms of positive effects on the occupancy rate, *ethos*, appealing to credibility, works best. To specify, being a super host and assuming an average number of reviews of roughly forty results in a positive effect on the occupancy rate that is approximately twice as high as the combined positive effects of an informative text and *logos* (facts and logic). Moreover, even when including the potential negative effects of *ID verification*, the positive effects of *ethos* (credibility) prevail.

6. Robustness Check

In this part, two variations in the estimation of the occupancy rate are applied and their final models are compared to the main one. This is necessary to establish whether the main model is robust enough, given its various assumptions; (1) a review rate fifty percent and (2) a multiplication of the number of reviews with an average minimum nights value of 3.4, unless the minimum required number of nights exceeds this number.

The first variation comprehends no review rate and no use of minimum nights, meaning that solely the number of reviews that are actually written are taken into account, which previous literature suggests to be a reliable indicator for the number of bookings (Ye et al., 2009). Moreover, for this particular model the occupancy rate is not weighted by 365 days, because the maximum occupancy rate would be 0.5, which makes the interpretation of the coefficient estimates more baffling for comparison. Secondly, I apply a variation where the occupancy rate is estimated without using a review rate, whilst multiplying the number of reviews according to the same minimum nights value that is used in the main model, weighted by 365 days. Note that by doing so, the percentage of censored observations decreases from roughly eleven percent to around a half percent. The results for both variations are shown and shortly elaborated in Appendix B, Table 2 and Table 3 respectively. Finally, due to very similar results for various estimations of the occupancy rate one can be confident in the main model's robustness.

7. Conclusion

In this paper, I investigate how hosts can effectively persuade potential guests to book their property by answering the following research question: "Which persuasion strategies based on Aristotle's appeals, implemented by the Airbnb host, are most effective for the listing's performance?". To derive at this, various variables were carefully matched to *ethos* (credibility), *pathos* (emotion) and *logos* (facts and logic) to assess their effects on the occupancy rate, whilst also account for a large range of possibly important control variables. Overall, each of the appeals affect the occupancy rate to some extent. However, appeals to credibility are deemed most dominant compared to appealing to emotion or facts and logic.

7.1 Implications

This research objectively analyzes the role of persuasion on booking performance in Amsterdam. Most literature that is related to this topic focusses on the effects on price or trust, rather than a performance measure such as the occupancy rate. Moreover, even fewer papers discuss the role of persuasion in the sharing economy. Hence, this study offers meaningful insights in our understanding of the customer decision making, measured by the occupancy rate, in the sharing economy. A comprehensive summary of all implications in this research is shown on the next page (Table 10).

To start, several suggestions arise when fully answering "Which persuasion strategies based on Aristotle's appeals, implemented by the Airbnb host, are most effective for the listing's performance?". I have mentioned that persuasion through credibility appeals are most potent in terms of increasing the number of bookings. Yet, I have to make a nuance here as it is not directly the most effective persuasion strategy for Airbnb hosts. To elaborate, a distinction should be made between short term persuasive strategies that hosts can implement immediately and long term persuasive strategies. Firstly, I advise all hosts on Airbnb to benefit from short term strategies, such as writing a practical text about the place and its surroundings (not solely appealing to emotion) or equipping the property with

Table 10. Summary implications

Implications in Terms of	Findings	Strategies for Hosts	General Implications
Ethos (credibility)	• Being a super host has a substantial positive effect on the number of bookings. • More reviews improve the booking performance. • Hosts with a verified ID perform worse than hosts without a verified ID.	• New hosts should make an effort to quickly receive a sufficient number of reservations and reviews. • Seasoned hosts should improve their response rate, cancellation rate or review rate in case they are no super host yet.	• Introduce symbols or badges that indicate superiority for products and services. • Stimulate customers to write reviews for any product or service to enhance persuasion through credibility.
Pathos (emotion)	• Writing an informative text in the description enhances the number of bookings compared to writing an persuasive text or no text at all. • Hosts without a text about themselves perform better than hosts with an informative text about themselves.	• Write a textual description that is focussed on providing practical information rather than solely appealing to guests' emotion. • Do not put much time and effort in writing the host about.	• On peer-to-peer marketplaces suppliers should ensure that texts about their product or service focus on presenting practical and functional information.
Logos (facts and logic)	• The number of amenities that are available positively affect the booking performance.	• Invest in equipping the property with more amenities.	• Other hotel-like facilities also likely benefit from equipping more amenities.

additional amenities (appealing to facts and data). I believe the latter strategy can similarly be effective for many hotel-like facilities, since the preference for amenities should not be too different depending on where you spend the night, e.g. having air conditioning in an Airbnb property is likely equally important when considering hostels.

Regarding texts, a best practice that is extracted from the original dataset would be ‘This modern room is 12 m2 and 15 minutes from the center of Amsterdam by tram, which runs every 6 minutes. It is part of a recently built house, which includes a condo with private gardens’. A less good example from the data is ‘Lovely house in a beautiful historical building that used to be a water and fire house. It is situated in the Jordaan, known for its cozy atmosphere’. Note that the best practice example makes use of an affective word (modern), which is surely all right to include. However, the focus is on providing practical information. In contrast, the lesser example merely focusses on persuading readers through appealing to emotion and actually conveys no practical information. I believe that the same holds in other P2P marketplaces, where texts should be practically focussed as there is an apparent need for more certainty regarding the ‘true’ quality of the product or service in P2P situations.

Advancing, it is not feasible that new hosts can directly benefit from persuasion through credibility appeals because this is a long term strategy, e.g. obtaining a super host badge requires at least ten completed reservations or alternatively three reservations with a total of at least one hundred nights, along with other requirements. Moreover, the number of reviews itself also have a relatively large positive effect on the number of bookings, highlighting the essentiality of guests leaving reviews. Therefore, I recommend that new hosts should make an effort to achieve the required number of reservations in combination with receiving reviews, simply by asking guests to write them. In addition, since the review score does not directly impact the booking intention of guests there seem to be no adverse consequences, albeit the average minimum review score should be 96 for a super host badge. Furthermore, to speed up the number of reservations, and hence reviews, it could be valuable to temporarily charge low prices, no cleaning fee and enable guests to instantly book the property as these measures increase the number of bookings. In the context of seasoned hosts the aforementioned reasoning can similarly be applied. Yet, the dire need for a number of reservations and reviews is absent and in case these hosts are still to receive a super host badge their focus should shift to improving the response rate, cancellation rate or review rate, which are the other requirements for the super host badge. These credibility appeals might be elevated to more general implications for other P2P marketplaces or even regular markets, albeit these should be investigated in future research as it remains situational. For instance, the observed effects of the super host badge sparks the believe that consumers are more willing to purchase products and services that are distinguished by a symbol or badge, indicating superior quality and hence appeals to credibility (ethos). Also, suppliers of any product or service should stimulate their customers to leave reviews to strengthen credibility and positively influence the customer decision making, given that the ratings are good.

Advancing with reviews, I urge Airbnb to introduce mandatory guest reviews, e.g. each guest either leaves a review with text and ratings or leaves a 'blank' review. This not only benefits hosts as the number of reviews increases the booking performance through credibility appeals, but also researchers, regulators and policymakers are benefitted by much more accurate data concerning the number of actual bookings. For example, the municipality of Amsterdam can easier assess whether regulations regarding the maximum nights for entire homes or apartments are properly enforced and make better judgement calls in future regulatory affairs.

Lastly, a fairly important contribution to the existing literature is how this study only compares the performance between highly available and active listings. In contrast, other studies often do not account for this, which could lead to misleading results. Moreover, the number of reviews turns out to be a reliable proxy for the number of bookings, which is used to calculate the occupancy rate. I strongly recommend to adopt the same approach in future research. To conclude, since this is the first paper to assess the effects of persuasion on an objective performance measure in the sharing economy, using a comparable sample, it advances our understanding of consumer behaviour in the P2P marketplace.

7.2 Limitations and Future Research

There are some limitations with regard to this study that future research can implement. First, even though the final model contains quite some explanatory variables, there remain unconsidered factors that are beyond the scope of this research and are deemed important in previous research. One can think of including the quality of room pictures or the facial expression of a host profile picture using face and object detection algorithms. Going on further, instead of number of amenities one might assess which specific amenities persuade guests to book. I specifically encourage future research to further explore possible heterogeneity effects to improve our general understanding of booking performance in the P2P marketplace.

Secondly, text processing algorithms are quite ingenious, but still far from perfect. Context is very important in identifying whether a word is meant to be positive, negative or even an affective word at all. One can think of irony or sarcasm, which could change the sentiment of a text, but is very difficult to detect. Moreover, in this research the two most encountered negative affective words that were found during the initial sentiment analysis were 'sink' and 'bunk', which are indeed negative English adjectives. However, in the context of Airbnb these are very likely to be used as nouns and hence these were encoded to synonymous words. Still, it is unrealistic to verify this for every word that is encountered, which means that the sentiment analysis is prone to error to some extent. In addition, there is a maximum amount of text that Inside Airbnb scrapes, resulting in some abrupt endings in the text in case of very long descriptions, which in turn limits the sentiment analysis. My advice is to scrape the textual data yourself and match it to existing data with help of the listing ID for completeness purposes. Also, applying more context specific text processing algorithms can improve the sentiment analysis.

Third, this research uses a dataset that is extracted on December 4th 2017 and derives the number of reviews from a dataset compiled on December 6th 2018. Here, the assumption is that hosts do not change any information throughout the year. Still, there is a considerable chance that hosts adjust various attributes, e.g. to increase the number of bookings. If the occupancy rate would improve because of these changes, the model would erroneously associate this success to the prior attributes. Given that Inside Airbnb provides monthly updated data, future research might account for this, for example by applying a difference-in-differences estimation.

Finally, the findings of this research are realized by using an extensive dataset of Amsterdam. A comparison across cities or countries could yield interesting results, given that a similar approach is maintained. To ensure the latter, the complete code that is used in this research is uploaded on Github[9] and researches are welcomed to work with this and make further improvements.

[9] On Github, search for Daanstr-MSc-Thesis

References

Agarwal, Sanjeev, M. Krishna Erramilli, and Chekitan S. Dev. (2003), “Market Orientation and Performance in Service Firms: Role of Innovation,” *Journal of Services Marketing,* 17(1), 68-82.

Airbnb (2020a), “How do I Become a Superhost?” (accessed August 2, 2020), https://www.airbnb.com/help/article/829/how-do-i-become-a-superhost.

Airbnb (2020b), “How do Star Ratings Work for Stays?” (accessed August 20, 2020), https://www.airbnb.com/help/article/1257/how-do-star-ratings-work-for-stays.

Airbnb (2020c), “How does it Work when Airbnb Verifies your Identity?” (accessed August 2, 2020), https://www.airbnb.com/help/article/1237/how-does-it-work-when-airbnb-verifies-your-identity.

Airbnb (2020d), “I Rent out my Home in Amsterdam. What Short-Term Rental Laws Apply?” (accessed September 20, 2020), https://www.airbnb.com/help/article/1624/i-rent-out-my-home-in-amsterdam-what-shortterm-rental-laws-apply.

Airbnb (2020e), “What Is the Airbnb Cancellation Policy for Stays?” (accessed August 20, 2020), https://www.airbnb.com/home/cancellation_policies.

Barron, Kyle, Edward Kung, and Davide Proserpio (2019), “Research: When Airbnb Listings in a City Increase, So Do Rent Prices,” *Harvard Business Review*.

Briene, Michel, Elvira Meurs, Daan Krins and Nick Rundberg (2018), “Tourism in Amsterdam Today and Tomorrow,” (accessed September 22, 2020), https://news.airbnb.com/wp-content/uploads/sites/4/2020/02/03122018_Tourism-in-Amsterdam_.pdf.

Brown, Stephen, Chris Hackley, Shelby D. Hunt, Charles Marsh, Nicholas O’Shaughnessy, Barbara J. Phillips, David Tonks, Chris Miles, and Tomas Nilsson (2018), “Marketing (as) Rhetoric: Paradigms, Provocations, and Perspectives,” *Journal of Marketing Management*, 34 (15-16), 1336–78.

Budget and Legislative Analyst’s Office (2015), “Analysis of the Impact of Short-Term Rentals on Housing,*” City and County of San Francisco Board of Supervisors*.

Chandy, Rajesh K., Gerard J. Tellis, Deborah J. Macinnis, and Pattana Thaivanich (2001), “What to Say When: Advertising Appeals in Evolving Markets,” *Journal of Marketing Research*, 38 (4), 399–414.

Chen, Chih-Chien, Zvi Schwartz, and Patrick Vargas (2011), “The Search for the Best Deal: How Hotel Cancellation Policies Affect the Search and Booking Decisions of Deal-Seeking Customers,” *International Journal of Hospitality Management*, 30 (1), 129–35.

Chen, Yubo and Jinhong Xie (2008), “Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix,” *Management Science*, 54 (3), 477–91.

Cheng, Mingming and Xin Jin (2019), “What do Airbnb Users Care About? An Analysis of Online Review Comments,” *International Journal of Hospitality Management*, 76, 58–70.

Chevalier, Judith A. and Dina Mayzlin (2006), “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research*, 43 (3), 345–54.

Conger, Jay A. (1998), “The Necessary Art of Persuasion,” *Harvard Business Review*, 76, 84-97.

Crowley, Ayn E., and Wayne D. Hoyer (1994), "An Integrative Framework for Understanding Two-Sided Persuasion," *Journal of Consumer research* 20 (4), 561-574.

Dawar, Niraj and Philip Parker (1994), “Marketing Universals: Consumers’ Use of Brand Name, Price, Physical Appearance, and Retailer Reputation as Signals of Product Quality,” *Journal of Marketing*, 58 (2), 81–95.

Dellaert, Benedict G. C. (2018), “The Consumer Production Journey: Marketing to Consumers as Co-Producers in the Sharing Economy,” *Journal of the Academy of Marketing Science*, 47 (2), 238–54.

Dinner, Isaac M., Harald J. Heerde Van, and Scott A. Neslin (2013), “Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising,” *Journal of Marketing Research*, 51 (5), 527–45.

Eckhardt, Giana M., Mark B. Houston, Baojun Jiang, Cait Lamberton, Aric Rindfleisch, and Georgios Zervas (2019), “Marketing in the Sharing Economy,” *Journal of Marketing*, 83 (5), 5–27.

Ert, Eyal, Aliza Fleischer, and Nathan Magen (2016), “Trust and Reputation in the Sharing Economy: The Role of Personal Photos in Airbnb,” *Tourism Management*, 55, 62–73.

Ertimur, Burçak and Mary C. Gilly (2012), “So Whaddya Think? Consumers Create Ads and Other Consumers Critique Them,” *Journal of Interactive Marketing*, 26 (3), 115–30.

Gallo, Carmine (2019), “The Art of Persuasion Hasn’t Changed in 2,000 Years,” *Harvard Business Review*.

Gavilan, Diana, Maria Avello, and Gema Martinez-Navarro (2018), “The Influence of Online Ratings and Reviews on Hotel Booking Consideration,” *Tourism Management*, 66, 53–61.

Grewal, Dhruv, Jerry Gotlieb, and Howard Marmorstein (1994), “The Moderating Effects of Message Framing and Source Credibility on the Price-Perceived Risk Relationship,” *Journal of Consumer Research*, 21 (1), 145.

Guitart, Ivan A., Jorge Gonzalez, and Stefan Stremersch (2018), “Advertising Non-Premium Products as if They Were Premium: The Impact of Advertising up on Advertising Elasticity and Brand Equity,” *International Journal of Research in Marketing*, 35 (3), 471–89.

Gutiérrez, Javier, Juan Carlos García-Palomares, Gustavo Romanillos, and María Henar Salas-Olmedo (2017), “The Eruption of Airbnb in Tourist Cities: Comparing Spatial Patterns of Hotels and Peer-to-Peer Accommodation in Barcelona,” *Tourism Management*, 62, 278–91.

Hamilton, Rebecca W., Roland T. Rust, Michel Wedel, and Chekitan S. Dev (2017), “Return on Service Amenities,” *Journal of Marketing Research*, 54 (1), 96–110.

Herzenstein, Michal, Scott Sonenshein, and Utpal M. Dholakia (2011), “Tell Me a Good Story and I May Lend you Money: The Role of Narratives in Peer-to-Peer Lending Decisions,” *Journal of Marketing Research*, 48 (SPL), S138-S149.

Hu, Minqing, and Bing Liu (2004a), "Mining and Summarizing Customer Reviews," *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177.

Hu, Minqing, and Bing Liu (2004b), "Mining Opinion Features in Customer Reviews," *AAAI*, 4 (4), 755-760.

Humphreys, Ashlee (2010), “Megamarketing: The Creation of Markets as a Social Process,” *Journal of Marketing*, 74 (2), 1–19.

Inside Airbnb (2019), “How is Airbnb Really Being Used in and Affecting your Neighbourhoods?” (accessed June 25, 2020), http://insideairbnb.com/amsterdam/?neighbourhood=&filterEntireHomes=false&filterHighlyAvailable=false&filterRecentReviews=false&filterMultiListings=false.

Inside Airbnb (2020), “About Inside Airbnb” (accessed September 20, 2020), http://insideairbnb.com/about.html.

Kotler, Philip and Sidney J. Levy (1973), “Buying is Marketing Too!,” *Journal of Marketing*, 37 (1), 54–59.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.k. Kannan (2016), “From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior,” *Journal of Marketing*, 80 (1), 7–25.

Labrecque, Lauren I., Ereni Markos, and George R. Milne (2011), “Online Personal Branding: Processes, Challenges, and Implications,” *Journal of Interactive Marketing*, 25 (1), 37–50.

Lamberton, Cait Poynor and Randall L. Rose (2012), “When is Ours Better than Mine? A Framework for Understanding and Altering Participation in Commercial Sharing Systems,” *Journal of Marketing*, 76 (4), 109–25.

Lench, Heather C., Sarah A. Flores, and Shane W. Bench (2011), “Discrete Emotions Predict Changes in Cognition, Judgment, Experience, Behavior, and Physiology: A Meta-Analysis of Experimental Emotion Elicitations,” *Psychological Bulletin*, 137 (5), 834–55.

Liang, Sai, Markus Schuckert, Rob Law, and Chih-Chien Chen (2020), “The Importance of Marketer-Generated Content to Peer-To-Peer Property Rental Platforms: Evidence from Airbnb,” *International Journal of Hospitality Management*, 84, 102329.

Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels, and Gerard Pfann (2013), “More than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates,” *Journal of Marketing*, 77 (1), 87–103.

Mao, Zhenxing and Jiaying Lyu (2017), “Why Travelers Use Airbnb Again?,” *International Journal of Contemporary Hospitality Management*, 29 (9), 2464–82.

Marketwatch (2020), “Peer-To-Peer Accommodation Market Growth Analysis by Revenue, Size, Share, Scenario on Latest Trends, Types and Applications Forecast” (accessed on June 28, 2020), https://www.marketwatch.com/press-release/peer-to-peer-accommodation-market-growth-analysis-by-revenue-size-share-scenario-on-latest-trends-types-and-applications-forecast-2020-04-07.

Marqusee, Alex (2015), “Airbnb and San Francisco: Descriptive Statistics and Academic Research,” *San Francisco Planning Department*.

Marshall, Roger and Na Woonbong (2003), “An Experimental Study of the Role of Brand Strength in the Relationship between the Medium of Communication and Perceived Credibility of the Message,” *Journal of Interactive Marketing*, 17 (3), 75–79.

Mauri, Aurelio G., Roberta Minazzi, Marta Nieto-García, and Giampaolo Viglia (2018), “Humanize your Business. The Role of Personal Reputation in the Sharing Economy,” *International Journal of Hospitality Management*, 73, 36–43.

Mcdonald, John F. and Robert A. Moffitt (1980), “The Uses of Tobit Analysis,” *The Review of Economics and Statistics*, 62 (2), 318.

Muller, Eitan (2020), “Delimiting Disruption: Why Uber is Disruptive, but Airbnb is not,” *International Journal of Research in Marketing*, 37 (1), 43–55.

Narayanan, Sridhar, Puneet Manchanda, and Pradeep K. Chintagunta (2005), “Temporal Differences in the Role of Marketing Communication in New Product Categories,” *Journal of Marketing Research*, 42 (3), 278–90.

Nasukawa, Tetsuya, and Jeonghee Yi (2003), "Sentiment Analysis: Capturing Favorability Using Natural Language Processing," *Proceedings of the 2nd International Conference on Knowledge Capture*, 70-77.

Otterbacher, Jahna (2011), "Being Heard in Review Communities: Communication Tactics and Review Prominence," *Journal of Computer-Mediated Communication*, 16 (3), 424–44.

Pollay, Richard W. (1985), "The Subsiding Sizzle: A Descriptive History of Print Advertising, 1900-1980," *Journal of Marketing*, 49 (3), 24.

Qazi, Aika, Karim Bux Shah Syed, Ram Gopal Raj, Erik Cambria, Muhammad Tahir, and Daniyal Alghazzawi (2016), "A Concept-Level Approach to the Analysis of Online Review Helpfulness," *Computers in Human Behavior*, 58, 75–81.

Rao, Hayagreeva, Henrich R. Greve, and Gerald F. Davis (2001), "Fools Gold: Social Proof in the Initiation and Abandonment of Coverage by Wall Street Analysts," *Administrative Science Quarterly*, 46 (3), 502.

Rucker, Derek D., Richard E. Petty, and Pablo Briñol (2008), "Whats in a Frame Anyway?: A Meta-Cognitive Analysis of the Impact of One Versus Two Sided Message Framing on Attitude Certainty," *Journal of Consumer Psychology*, 18 (2), 137–49.

Ruokolainen, Jari and Leena Aarikka-Stenroos (2016), "Rhetoric in Customer Referencing: Fortifying Sales Arguments in Two Start-Up Companies," *Industrial Marketing Management*, 54, 188–202.

Sparks, Beverley A., Kevin Kam Fung So, and Graham L. Bradley (2016), "Responding to Negative Online Reviews: The Effects of Hotel Responses on Customer Inferences of Trust and Concern," *Tourism Management*, 53, 74–85.

Statista (2020), "Occupancy Rate of Hotels in Amsterdam from 2011 to 2019" (accessed September 15, 2020), https://www.statista.com/statistics/545191/hotel-occupancy-rate-amsterdam/.

Sundararajan, Arun (2016), *The Sharing Economy: the End of Employment and the Rise of Crowd-Based Capitalism*, Cambridge, MA: The MIT Press.

Tibshirani, Robert (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58.1, 267-288.

Tibshirani, Robert (1997), "The Lasso Method For Variable Selection In The Cox Model," *Statistics in Medicine*, 16 (4), 385–95.

Tobin, James (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26 (1), 24-36.

Wang, Dan and Juan L. Nicolau (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, 120–31.

Yang, Shuai, Yiping Song, Sixing Chen, and Xin Xia (2017), "Why are Customers Loyal in Sharing-Economy Services? A Relational Benefits Perspective," *Journal of Services Marketing*, 31 (1), 48–62.

Yang, Sung-Byung, Hanna Lee, Kyungmin Lee, and Chulmo Koo (2018), "The Application of Aristotle's Rhetorical Theory to the Sharing Economy: an Empirical Study of Airbnb," *Journal of Travel & Tourism Marketing*, 35 (7), 938–57.

Ye, Qiang, Rob Law, and Bin Gu (2009), "The Impact of Online User Reviews on Hotel Room Sales," *International Journal of Hospitality Management*, 28 (1), 180–82.

Zablocki, Agnieszka, Katerina Makri, and Michael J. Houston (2019), “Emotions Within Online Reviews and their Influence on Product Attitudes in Austria, USA and Thailand,” *Journal of Interactive Marketing*, 46, 20–39.

Zervas, Georgios, Davide Proserpio, and John W. Byers (2017), “The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry,” *Journal of Marketing Research*, 54 (5), 687–705.

Zhu, Feng and Xiaoquan (Michael) Zhang (2010), “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics,” *Journal of Marketing*, 74 (2), 133–48.

Zou, Hui, and Trevor Hastie (2005), "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67.2, 301-320.

Appendix

Appendix A

Table 1. Estimated coefficients of all variables from elastic net regression on occupancy rate

Variable	*Coefficient*	*Variable*	*Coefficient*	*Variable*	*Coefficient*
Host response rate	0.0183	(Nb) Oud-Oost	.	Bedrooms	.
(Nb) Bijlmer-Centrum	.	(Nb) Slotervaart	.	Beds	0.0038
(Nb) Bijlmer-Oost	.	(Nb) Watergraafsmeer	.	(Bt) Other	.
(Nb) Bos en Lommer	-0.0035	(Nb) Westerpark	.	(Bt) Pull-out Sofa	.
(Nb) Buitenveldert – Zuidas	.	(Nb)Zuid	.	(Bt) Real Bed	.
(Nb) Centrum-Oost	0.0126	Exact Location	0.0065	Price	-0.0311
(Nb) Centrum-West	0.0181	(Pt) Apartment	.	Security Deposit	-0.0018
(Nb) Aker – Nieuw Sloten	.	(Pt) Bed & Breakfast	.	Cleaning Fee	-0.0049
(Nb) Baarsjes – Oud-West	.	(Pt) Boat	.	Guests Included	.
(Nb) Pijp - Rivierenbuurt	.	(Pt) Other	0.0007	Extra People	.
(Nb) Gaasperdam - Driemond	-0.0086	(Pt) House	.	Minimum Nights	.
(Nb) Geuzenveld - Slotermeer	.	(Pt) Loft	.	Maximum Nights	.
(Nb) Ijburg - Zeeburgereiland	-0.0014	(Pt) Townhouse	.	Instant Bookable	0.0708
(Nb) Noord-Oost	.	(Rt) Entire Property	-0.0494	Guest Picture	.
(Nb) Noord-West	.	(Rt) Private Room	0.0846	Guest Phone Verification	.
(Nb) Oud-Noord	.	(Rt) Shared Room	.	Host Has Description	.
(Nb) Osdorp	.	Accommodates	.		
(Nb) Oostelijk Havengebied – Indische Buurt	-0.0003	Bathrooms	.		

Appendix B

The results of the main model are almost identical to these two models in terms of significant variables, sign and relative size of the coefficients. The only difference is that, for *Description*, the main model finds a significant negative effect of having no text, as compared to an informative text, whereas the two other models do not.

Table 2. Tobit model of the first robustness check

		Coefficient	Std. Err.	Pr (> t)
Constant		-5.064	6.720	0.451
Ethos (credibility)				
Super host badge		3.284***	0.669	0.000
Number of reviews		0.180***	0.005	0.000
Response time	Within a day	-2.884	3.783	0.446
	Within a few hours	-0.050	3.952	0.990
	Within an hour	4.062	3.987	0.308
ID verification		-1.181*	0.572	0.039
Pathos (emotion)				
Description	Persuasive	-2.214***	0.572	0.000
	No text	-1.941	1.253	0.121
Host about	Persuasive	0.168	0.710	0.813
	No text	3.349***	0.663	0.000
Logos (facts and logic)				
Amenities		0.095*	0.037	0.011
Safety features		0.116	0.219	0.595
Cancellation policy	Moderate	0.682	0.827	0.410
	Strict	-1.109	0.809	0.170
Review score		0.082	0.053	0.123
Control variables				
Host response rate		5.579	3.751	0.137
Exact location		-0.416	0.595	0.485
Beds		0.662***	0.199	0.001
Price		-0.024***	0.004	0.000
Security deposit		-0.001	0.001	0.325
Cleaning fee		-0.041**	0.013	0.001
Instant bookable		8.040***	0.771	0.000
(Nb) Bos en Lommer		-4.448**	1.354	0.001
(Nb) Centrum Oost		1.922*	0.910	0.035
(Nb) Centrum West		2.359**	0.781	0.003
(Nb) Gaasperdam Driemond		-7.092**	2.544	0.005
(Nb) Ijburg Zeeburgereiland		-4.311*	1.702	0.011
(Nb) Havengebied Ind. Buurt		-4.239**	1.398	0.002
(Pt) Other		2.104	1.308	0.108
(Rt) Entire property		-0.417	3.751	0.911
(Rt) Private room		9.833**	3.752	0.009
LogSigma		2.725***	0.013	0.000

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; Log likelihood = -13214.09; Pseudo R2 = 0.201

Table 3. Tobit model of the second robustness check

		Coefficient	Std. Err.	Pr (> t)
Constant		-0.042	0.064	0.514
Ethos (credibility)				
Super host badge		0.030***	0.006	0.000
Number of reviews		0.002***	0.000	0.000
Response time	Within a day	-0.032	0.036	0.369
	Within a few hours	-0.004	0.038	0.912
	Within an hour	0.038	0.038	0.315
ID verification		-0.013*	0.005	0.014
Pathos (emotion)				
Description	Persuasive	-0.019***	0.005	0.001
	No text	-0.023	0.012	0.058
Host about	Persuasive	-0.000	0.007	0.948
	No text	0.029***	0.006	0.000
Logos (facts and logic)				
Amenities		0.001***	0.000	0.000
Safety features		0.001	0.002	0.775
Cancellation policy	Moderate	0.005	0.008	0.532
	Strict	-0.005	0.008	0.489
Review score		0.001	0.001	0.272
Control variables				
Host response rate		0.063	0.036	0.079
Exact location		-0.004	0.006	0.453
Beds		0.008***	0.002	0.000
Price		-0.000***	0.000	0.000
Security deposit		-0.000	0.000	0.547
Cleaning fee		-0.000**	0.000	0.002
Instant bookable		0.075***	0.007	0.000
(Nb) Bos en Lommer		-0.040**	0.013	0.002
(Nb) Centrum Oost		0.025**	0.009	0.004
(Nb) Centrum West		0.034***	0.007	0.000
(Nb) Gaasperdam Driemond		-0.067**	0.024	0.006
(Nb) Ijburg Zeeburgereiland		-0.046**	0.016	0.005
(Nb) Havengebied Ind. Buurt		-0.042**	0.013	0.002
(Pt) Other		0.014	0.013	0.275
(Rt) Entire property		0.010	0.036	0.785
(Rt) Private room		0.094**	0.036	0.009
LogSigma		-1.926***	0.013	0.000

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; Log likelihood = 1574.491; Pseudo R2 = -2.932[10]

[10] Because the dependent variable is continuous, the log likelihood can also be positive, which in turn means that the Pseudo R2 can take on a negative value or a value higher than one.