"Reputation For What? The Impact of Online Reviews in Airbnb"*

Michelangelo Rossi†

January 2017

Abstract

This article studies the review system of Airbnb. The main goal is to identify the impact of guests' reviews over the performance of dwellings' owners in Airbnb. The novelty of my approach is twofold: firstly, I use a sentimental analysis of the reviews' content and the reviews expressing concerns about the fixed dwellings' characteristics are separated from the reviews about hosts' misbehaviour. Secondly, I identify a credible proxy for the value of the owners' exit option so as to control for the hosts' selection process using the Heckman two-step procedure. Thus, the impacts of different types of negative reviews are empirically analysed over three dimensions of the hosts' performances: the exit decision, the pricing decision, and the probability of transactions.

*I am grateful to my advisors, Matilde Machado and Natalia Fabra for their valuable comments and numerous consultations. I want to specially thank for useful feedback conferences participants at the Carlos III Phd Workshop, Madrid, Spain and I also thank my friends, Ruben Veiga Duarte and Michele Rosemberg for their helpful advice. Special gratitude to my family and to Elizaveta Pronkina for their enormous support and encouragement. All errors are mine.

†Universidad Carlos III, Department of Economics, Madrid, Spain. E-mail: mrossi@eco.uc3m.com

1 Introduction

This work studies the review system of Airbnb, a digital platform where dwellings' owners can rent their apartments to visitors for small periods of time; and, after the visit, both parties have the possibility to write a review about their experience.

The main objective of this paper is to identify the impact of guests' reviews over the performance of dwellings' owners. My research focuses on the identification of the reviews' impact over the hosts' performance and how the impact changes due to the reviews' content.

Most of the articles regarding online reputation and its impact over agents' performance models reputation combining characteristics that are proper of two theoretical settings: hidden information and hidden action (Cabral and Hortacsu, 2010).

Models of hidden information emphasize the signalling nature of reputation: users learn about the quality of items from posted online reviews. Conversely, models of hidden action consider reputation as a sanctioning device used to implement over time trigger strategies that prevent agents' misbehaviours (Dellarocas, 2006; Bar-Isaac and Tadelis, 2008). These two approaches lead to significant differences in the expected impact of reviews over agents' performances and how the review systems can discipline the competition among agents through the revelation of new pieces of information. Defining models where these two approaches are present (Cabral and Hortacsu, 2010) may lead to theoretical predictions that are broad and not able to define and select specific behaviours by agents.

Thanks to the specific features of the Airbnb platform, the approach of my analysis is reversed as I analyse separately reviews that regards unknown hosts' characteristics (hidden information) and hosts' misbehaviour (hidden action).

The reviews' content is studied using a lexicon-based sentimental analysis that focuses on the identification of negative reviews. Reviews are defined to be negative whether a set of most frequent negative adjectives is present or not in the associated comment. Based on the same method the negative reviews are divided in two type expressing negative claims about the house's hardly-modifiable characteristics (hidden information) and about the hosts' effort and misbehaviour (hidden action).

These two different types of negative reviews and their impacts over the performance of Airbnb owners are studied separately. The obtained results are then compared with some theoretical predictions regarding pure hidden information and pure hidden action, respectively.

This approach appears to be systematically applicable for all those emerging online platforms (broadly defined *Sharing Economy* platforms [1]) where service providers repeatedly use the same object over time for many transactions; and, at the same time, the service can be improved through effort (Uber, Lyft, BlaBlaCar).

The analysis of this work focuses on the impact of the negative reviews (and in particular of the first negative reviews observed in the Airbnb hosts' webpage as in the work by Cabral and Hortacsu (2010)) over three dimensions of the hosts' performances:

1. Firstly, I consider the impact of negative reviews over the exit decision of hosts from the digital platform;

2. Secondly, I analyse the relationship between the listings' pricing policies with the occurrence of negative reviews;

3. Finally, I study the impact of negative reviews over the probability of hosts' transactions with guests.

Studying separately the impacts of negative reviews over these three dimensions could lead to inconsistent results. In particular, the hosts' selection process may affect the impact of negative reviews over the probability of transactions and over the pricing decisions of the hosts. Accordingly, I study the impact over these two measures of the hosts' performance controlling for the

[1] For an analysis on definitions regarding the Sharing Economy, see Sundararajan (2013).

selection process using an Heckman two-step approach (see Heckman, 1977). The feasibility of this procedure is strongly related to the possibility to find a credible proxy for the value of the exit option for dwellings' owners (that defines the necessary exclusion restriction in the Heckman correction): in my work this proxy is the average monthly price rent that the owners could potentially charge in the rental market.

The same analysis is conducted for the two types of negative reviews; and finally the results are paired with some theoretical predictions pointed out in the literature (Bar-Isaac and Tadelis, 2008, Mailath and Samuelson, 2006).

The structure of the work is the following: in Section 2, I provide a brief literature review about the empirical papers that studied the impact of online reviews over the agents' performance. In Section 3, I describe the Airbnb dataset used in my analysis and the procedures followed for the sentimental analysis of the reviews. In Section 4, I present the empirical exercise about the impact of the two different types of negative reviews over hosts' performance and I report the main results. Finally in Section 5, I conclude with a comment about the obtained results in line with established theoretical arguments about the contexts of pure hidden information and pure hidden action; and with the further steps that are necessary to refine and improve my analysis.

2 Literature Review

The impact of online feedbacks over the users' performance has been studied on different digital platforms: consumer-to-consumer retail platforms; e-commerce platforms; and Sharing Economy platforms.

The vast majority of articles regarding the impact of online reputation in digital platforms investigates C2C retail platforms where mostly non-professional sellers and buyers exchange goods: Ebay is the most studied platform. Many articles analyse how the buyers' reviews affect the auctions for the sellers' objects. Dellarocas (2003) provides a complete summary of the first attempts to measure how the sellers' feedback affect prices and the probability of sale using cross-sectional regressions of sale prices on feedback characteristics. The cross-sectional approach to estimate the impact of reputation was discarded after the field experiment by Resnick, Zeckhauser, Swanson, and Lockwood (2006). They point out the presence of an omitted variable bias associated with these methods. In order to eliminate this bias, more recent works adopted different strategies: Cabral and Hortacsu (2010) construct a panel using feedback histories as proxies for seller transaction histories and their work provides systematic empirical evidence regarding the change of sellers' incentives in response of the change in reputations. Some articles such as Luca (2011) and Anderson and Magruder (2012) use a regression discontinuity design to exploit some special institutional settings; others considered field experiments following Resnick et al. (2006).

The number of analyses on the role of feedbacks in e-commerce platforms such as Amazon or Flipkart is limited as feedbacks are not the unique sources of information about the quality of the items sold in this setting: Chevalier and Mayzlin (2006) examine the effect of consumer reviews on relative sales of books at Amazon.com and Barnesandnoble.com; Dellarocas and Wood (2008) study whether a significant reviews' bias is present in this framework and how it affects the behaviours of buyers.

Finally, Sharing Economy platforms have been only recently studied by scholars. Few works focus on the matching process among users (Dinerstein, Einav, Levin, and Sundaresan (2014), Fradkin (2014)); and the structural estimation of supply and demand on these platforms (Cullen and Farronato (2014)). Fradkin, Grewal, Holtz, and Pearson (2015) study online reviews on Airbnb; still their main interest regards the bias of reviews and how a different feedback mechanism may affect the bias. In this sense, the study of the impact of online reputation in these types of platforms have not been yet researched extensively.

3 The Airbnb Dataset

3.1 The Airbnb Framework

According to its online description, Airbnb is a "trusted community marketplace for people to list, discover, and book unique accomodations around the world". The digital platform has been active since 2008 and in these eight years of activity more than 30 millions transactions have been successfully operated with more than 1 million of listings inscribed around the world.

In every transaction two parties are involved: the host, the owner or the manager of the listing; and the guest, who has booked the listed dwelling (from now on called listing). After the visit, the host and guest have 14 days to review each other and the feedbacks will be exchanged only after both parties have completed the reviews so as to reduce potential retaliation between agents. Guests feedback consists of four elements:

1. A written comment;
2. Private comments to the host;
3. A one-to-five star rating regarding the overall experience;
4. Six specific ratings regarding the host and the listing:
   - The accuracy of the listing description compared with the guest's expectations;
   - The communicativeness of the host;
   - The cleanliness of the listing;
   - The listing location;
   - The value of the listing and the quality of the amenities provided by the listing.

Similarly, the host can review the guest answering whether or not she would recommend the guest; writing a comment; and rating the guest considering the communicativeness, the cleanliness and how well the guest respected the rules of the house.

Not all these elements are published on the web-site and for what concerns the guest feedback, only the written comment is directly published on the host web-page. The other ratings are not displayed singularly with the comments as only the rounded average of the score and sub-scores are published on the listing and the host web-pages. In the same way, only the comment written by the host is published in the guest web-page.

3.2 The Structure of the Dataset

In this work I exploit public information compiled from the Airbnb web-site regarding all the listings present in a specific location over time. In my analysis, the dataset is formed by seventeen snapshots of all the listings in the New York area at seventeen particular dates (2nd January 2015, 2nd March 2015, 2nd April 2015, 2nd May 2015, 2nd June 2015, 2nd August 2015, 2nd September 2015, 2nd October 2015, 2nd November 2015, 20th November 2015, 2nd December 2015, 2nd January 2016, 2nd February 2016, 2nd April 2016, 2nd May 2015, 2nd June 2016, 2nd July 2016)[2].

For each listing I have the following information about all the past feedbacks left by guests:

- The date of each feedback posted on the listings' webpages;
- The comment written by the guest and published on the listing web-page.

Accordingly, I can track the complete history of past feedbacks for those listings present at the snapshot dates and detect what listings have exited, entered or remained in the platform at these seventeen dates. Still, the dataset results to be truncated since the exact date of listings' exit cannot be observed:

[2] All the data has been dowloaded from the web-site of InsideAirbnb. InsideAirbnb is "an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world" and all the data on the web-site is available under a Creative Commons CC0 1.0 Universal (CC0 1.0) license (`www .insideairbnb.com`).

I can only check whether or not the listings are active on Airbnb at the snapshot dates (while I have the date of the first feedback). Moreover, I cannot observe those listings that entered and exited in-between the snapshot dates. Finally, no information about non-reviewed transactions between hosts and guests can be obtained from the web-site as it is not publicly available.

Finally, for all the listings present in the platform from the 2nd January 2015 to the 2nd July 2016 I have the following information updated at different snapshot dates; i.e. almost each month:

- The current posted price for a night at the listing;
- Geographical information about the position of the listing (zipcode of the residential area where the listing is located);
- Information about the type of dwelling corresponding to the listing: hosts can rent the entire flat or house; or only some rooms.

In my analysis I restrict the focus on hosts that rent the entire apartment (the 98% of hosts present on Airbnb).

The unique data that is not compiled from the Airbnb web-site regards the average monthly rental price for each zipcode area of New York from January 2015 to July 2016. This data is publicly available and published by the Zillow Group Research Center. [3]

3.3 Identifying Negative Reviews: A Lexicon-Based Sentimental Analysis

The empirical exercises presented in this work are all studying the impact of guest negative reviews on the performance of the hosts. The focus on negative reviews is motivated by a common feature of feedbacks in almost all the digital platforms: feedbacks are generally positive. Cabral and Hortacsu (2010)

[3] All the data has been downloaded from the Zillow Group Research Center and its use is in line with the terms of use of Zillow Group (`http://www.zillow.com/research/data/`).

explain this disproportion between positive and negative reviews with the capacity of the feedback to eliminate (or at least control) the most fraudolent behaviours by users: in this sense, feedbacks are positive because users behave well and complaints are rare. Differently, Fradkin et al. (2015) associate the small number of negative reviews in Airbnb with socially induced reciprocity.[4] Still, independently of the different explanations regarding the disproportion between the number of positive and negative reviews, studying the impact of negative reviews seems to be crucial.

As mentioned at the beginning of this section, my dataset does not allow to identify whether or not a single review is negative since no unique criteria can be exploited: only the written comments are published on the web-site and ratings of the guests are only displayed as rounded averages. In this sense the identification of negative reviews has to pass by the content analysis of the comments that is briefly described here.

The approach used in the content analysis is based on the identification of negative adjectives that are expressing negative characteristics of the service provided by the host: all the single words present in the comments are ordered from the most to the least frequent; then, all the negative adjectives written at least 1000 times are considered; that is, 30 negative adjectives. In my work, for each negative adjective I consider all the comments that contain the adjective controlling for expressions that could change or compromise the semantic of the word in the sentences. [5]

In this way I can remove many comments that used the selected adjectives, but not to express negative feedbacks. Moreover, for each adjective I select

[4] Retaliation is also studied by Fradkin et al. (2015) and they observe that the feedback protocol implemented on August 2014 diminuished the fear of retaliation by users: with this protocol, users cannot observe the comment of the other part before having completed and submitted their feedback.

[5] For example, for the adjective "dirty" I did not consider comments that reported the word "dirty" in expressions as "not dirty", "not so dirty", "not very dirty", "not that dirty", "not really dirty", "n't dirty", "n't so dirty", "n't very dirty", "n't that dirty", "n't really dirty" and "Not dirty", "Not so dirty", "Not very dirty", "Not that dirty", "Not really dirty".

randomly 100 comments and I manually check whether the comments are expressing negative remarks or not: if I find that more than 5% of the comments are not negative I discard the adjective. In this way, the number of adjectives reduced to 15: "small", "far", "old", "noisy", "hard", "basic", "loud", "dirty", "loud", "difficult", "smaller", "uncomfortable", "negative", "disappointed", "older", "Small". Based on this content analysis selects I define a set of comments, called "negative comments", that I use in this research.

In this article I do not discuss extensively the procedure of my sentiment analysis of the reivews. This content analysis allows to differentiate negative reviews focusing on some specific words expressing the reasons of the guests' complaints; still it is restrictive and not all the potentially negative comments are selected.

As focus of my future research I am going to analyse comments with a more comprehensive lexicon-based sentiment analysis following the approach used by Taboada, Brooke, Tofiloski, Voll, and Stede (2011).

Asymmetries of information in the forms of hidden information and hidden action may affect the transactions in Airbnb; in this line, the negative comments written by the guests may regard some specific characteristics of the house (distance from the center) or the behaviour of the hosts (cleaningness of the house). Different types of negative reviews can be selected considering some words that are clearly referring to house characterstics or the host's service: many adjectives selected for the content analysis are not sufficiently specific and they cannot be used to identify hidden information or hidden action; still, some adjectives are clearly referring to intrinsic house characteristics, while others deal with host's misbehaviour.

In the next session, the presented empirical exercises will be repeated considering negative reviews that explicitly account for hidden information and hidden action. In particular, I consider negative reviews associated with the adjectives "noisy", "old" and "small" to identify those reviews regarding listing's features impossible to change by the hosts. I define these reviews as hidden information negative reviews (from now on HI negative reviews). In the same way, I consider negative comments including the adjectives "dirty", "not helpful" and "not friendly" in order to identify negative reviews expressing concerns about the host's services[6]: I name the reviews associated with these comments as hidden action negative reviews (from now on HA negative reviews).

3.4 Descriptive Statistics

Given the characteristics of the available data regarding Airbnb, I report here the panel structure of the dataset that is used for all the empirical exercises presented in this work:

- The 40,237 listings appearing on the Airbnb from January 2015 to July 2016 represent the cross-sectional units of the panel;
- 79 weekly time observations (from the first week of January 2015 to the first week of July 2016) per listing.

This panel structure allows to observe whether or not the listing was present on Airbnb in a given week; whether one or more reviews appeared on its Airbnb webpage in a given week; and whether one or more of these reviews were negative, HI negative, and HA negative. Moreover the weekly price posted for one night per listing and the rental price in the same zipcode area are observed. Finally, the total number of listings and the average posted price by listings in the same zipcode area can be calculated.

Before passing to the analysis of the empirical exercises, I conclude this session presenting some important descriptive statistics of the presence of listings on Airbnb; the Airbnb reviewing process; the prices posted by listings and the rental prices.

The Airbnb data consider 40,237 single listings and a total of 409,349 written comments. The average number of comments per listing is 10.17 while

[6]The adjectives "not helpful" and "not friendly" have been selected as they are the negations of two of the most common positive adjectives expressing the quality of the host's services.

the median is 4. Such difference reflects the asymmetry of the distribution of comments per listing as it can be observed in Figure 1.

Negative reviews are much less frequent than reviews: the negative reviews selected are 28,426 (less than the 7% of the total reviews) while the HI and HA negative reviews are 17,568 and 4,818 ,respectively.

The negative reviews found through the content analysis can be compared with the descriptive statistics by Fradkin et al. (2015) defined using Airbnb proprietary data. Fradkin et al. (2015) observed that guests "submit a five star overall rating 74% of the time and a four star rating 20% of the time"; accordingly, if we believe in the methodology used for the content analysis of negative reviews we may associate negative reviews with (unobserved) rating lower than four stars.

The high number of listings with only few reviews is mainly due to high attrition, with the vast majority of listings entering and exiting after few months;[7] and partially to the truncation of our dataset: listings only present at the last snapshot date have few feedbacks because of their recent entry.

Among the 40,237 single listings, 15,748 listings are present at snapshot January 2015 and 14,336 are present at the snapshot of July 2016, while only 1.099 are always present at all the snapshots dates.

The average number of weeks of listings' persistence on Airbnb is 27.94. This low value, together with the distribution of weeks of persistence on Airbnb per listings in Figure 2 shows the great importance of attrition and the entry and exit decisions by listings. On average, each week 345 listings enter and 302 listings exit the platform.

The average price charged for one night by listings is 141.6 US dollars and the distribution of prices is quite dispersed (as observed in Figure 3). Still the prices posted by each listing are quite sticky: the within standard deviation of the price posted by listing is 15.02, while the between standard deviation is 96.50.

[7] Listings are present on the Airbnb web-site if they are, actively or not, inscribed and their web-page can be visited through the platform.

The rental price per zipcode area represents the proxy I use in my empirical exercises for the value of the exit option by dwellings' owners: exiting the platform, the owners can rent the apartment in the usual rental market. The average monthly rental price in the NY area between January 2015 and July 2016 is 1797.6 dollars and the unimodal distribution of the monthly rental price for this period is described in Figure 4. In this sense, the average potential monthly revenues for a listing in Airbnb greatly overcome the average monthly rental price charged on the rental market.

4 The Impact of Negative Reviews

In this section I present the impact of negative reviews over three dimensions of the hosts' performance: the permanence of the hosts on the Airbnb platform; the probability of a transaction; and the pricing decision. In all these exercises presented I will focus on the impact of a particular set of negative reviews: following the approach used by Cabral and Hortacsu (2010) I concentrate my attention on the impact of the first five negative reviews that are published on the Airbnb webpage of the listing. This focus can be justified with two different arguments:

- Almost the 90% of listings have no more than 5 negative reviews published on their webpages;
- Following the theoretical predictions about the HI and HA frameworks (Bar-Isaac and Tadelis, 2008) the most straightforward differences should be observed considering the first negative reviews.

The structure of this section proceeds as follows: in the next subsections the three empirical exercises corresponding to the three different measure of the hosts' performance will be reported, together with the associated results. All the exercises are firstly executed for all the negative reviews; then replicated considering only the HI and the HA negative reviews.

4.1 The Impact over the Listings' Persistence on Airbnb

The first dimension of the hosts' performance that will be studied regards the persistence of the listings on Airbnb: in the exercise presented below I will identify the impact of several variables over the probability for the listings to be present on the Airbnb platform. Some variables may affect this probability in two different ways: they affect the listing's probability to entry into the platform or the listing's probability to exit from the platform. At this step of my analysis I do not explicitly differentiate between entry and exit as this represents one of the most important lines for further investigation.

The results of this exercise provide the first guidance about the study of the Airbnb selection process of listings; in this sense, this first exercise also represents the first stage of the Heckman two-step procedure I will use in order to define the impact of negative reviews over the probability of having transactions and over the pricing decisions controlling for the selection process.

The probability of being present on the platform for listing i in week t is defined using a listing-specific binary variable indicating whether listing i is present in week t or not, equal to 1 and 0, respectively. It is indicated with $\pi_{i,t}$.

The explanatory variables that will be used are the following for each listing i in week t:

- $\mathbb{1}_{i,t}$: a binary variable indicating the occurrence of at least one review in the previous three weeks;
- $\mathbb{1}^{neg1}_{i,t}$,$\mathbb{1}^{neg2}_{i,t}$,$\mathbb{1}^{neg3}_{i,t}$,$\mathbb{1}^{neg4}_{i,t}$,$\mathbb{1}^{neg5}_{i,t}$: binary variables indicating the occurrence of the first, the second, the third, the fourth and the fifth negative review in the previous three weeks[8];
- $p^{rent}_{i,t}$: the rental price in week t for the zipcode area where listing i is located;

[8]The same exercise has been repeated considering different periods of lag. Results are robust to variations of lags of two weeks.

- $\bar{p}_{i,t}^{zipcode}$: the average price posted in week t by all the listings located in the zipcode area where listing i is located;

- $n_{i,t}^{zipcode}$: the number of listings located in week $t-1$ in the zipcode area where listing i is located;

- $g_{i,t}^{n}$: the monthly growth rate of the number of listings located in week t in the zipcode area where listing i is located [9].

In this sense, the following reduced-form equation can be defined:

$$\begin{aligned}\pi_{i,t} = \alpha &+ \beta_1 \mathbb{1}_{i,t}^{neg1} + \beta_2 \mathbb{1}_{i,t}^{neg2} + \beta_3 \mathbb{1}_{i,t}^{neg3} + \beta_4 \mathbb{1}_{i,t}^{neg4} + \beta_5 \mathbb{1}_{i,t}^{neg5} \\ &+ \beta_6 \mathbb{1}_{i,t} + \beta_7 p_{i,t}^{rent} + \beta_8 \bar{p}_{i,t}^{zipcode} + \beta_9 n_{i,t}^{zipcode} + \beta_{10} g_{i,t}^{n} + \epsilon_{i,t}. \end{aligned} \tag{1}$$

In order to estimate the impact of these variables on the persistence of listings on Airbnb, I use the following econometric models: a linear probability model with robust standard errors, probit model with robust standard errors; random effect probit and logit models; fixed effect logit model (all reported in Tables 1 and 2).

The effects of interest are robust to different model specifications and here I report the results with short comments (extensive comments about all the results with reference to the associated theoretical predictions are present in Section 5):

- The occurrence of at least one review in the previous three weeks affects positively and significantly the probability of listings to stay on the platform; that is, it decreases the probability to exit the platform;

- The occurrence of all the first five negative reviews affects negatively and significantly the probability of listings to stay on the platform; that is, it increases the probability to exit the platform;

- An increase in the rental price in the zipcode area where the listing is located affects negatively and significantly the probability of listings to

[9] The monthly growth rate $g_{i,t}^{n}$ is defined as follows: $g_{i,t}^{n} = \ln(n_{i,t-1}^{zipcode}) - \ln(n_{i,t-4}^{zipcode})$.

stay on the platform: the rental price defines the exit option of listings and at the same time it decreases the probability to enter the platform and it increases the probability to exit the platform. Thanks to the statistical significance of this parameter, the rental price $p_{i,t}^{rent}$ can be used as the exclusion restriction in the third exercise where the impact of negative reviews over the probability of transactions is studied controlling for the listings' selection process;

- An increase in the average price posted by listings located in the same zipcode area or an increase in the number of listings located in the same zipcode area both affect positively the probability of listings to stay on the platform. This result can be interpreted considering these two parameters $\bar{p}_{i,t}^{zipcode}$ and $n_{i,t}^{zipcode}$ as a proxy for the attractiveness of the area in which the listing is located. A more attractive area induces more listings to be on the platform through an increase in the probability to entry the platform;

- A higher growth rate of listings in the zipcode area where the listing is located has a not significant impact on the probability of staying (the unique significant result is the one associated with the RE logit model). The interpretation of this result can be the following: the competition among close hosts does not have a real impact on the listings' entry/exit decisions.

The same exercise can be repeated considering the first five HI and HA negative reviews. The occurrences of the first five HI and HA negative reviews are defined with $\mathbb{1}_{i,t}^{neg1HI}$, $\mathbb{1}_{i,t}^{neg2HI}$, $\mathbb{1}_{i,t}^{neg3HI}$, $\mathbb{1}_{i,t}^{neg4HI}$, $\mathbb{1}_{i,t}^{neg5HI}$ and $\mathbb{1}_{i,t}^{neg1HA}$, $\mathbb{1}_{i,t}^{neg2HA}$, $\mathbb{1}_{i,t}^{neg3HA}$, $\mathbb{1}_{i,t}^{neg4HA}$, $\mathbb{1}_{i,t}^{neg5HA}$, respectively. Therefore two different reduced-form equations can be presented:

$$\pi_{i,t} = \alpha + \beta_1 \mathbb{1}_{i,t}^{neg1HI} + \beta_2 \mathbb{1}_{i,t}^{neg2HI} + \beta_3 \mathbb{1}_{i,t}^{neg3HI} + \beta_4 \mathbb{1}_{i,t}^{neg4HI} + \beta_5 \mathbb{1}_{i,t}^{neg5HI} \\ + \beta_6 \mathbb{1}_{i,t} + \beta_7 p_{i,t}^{rent} + \beta_8 \bar{p}_{i,t}^{zipcode} + \beta_9 n_{i,t}^{zipcode} + \beta_{10} g_{i,t}^{n} + \epsilon_{i,t}. \tag{2}$$

$$\pi_{i,t} = \alpha + \beta_1 \mathbb{1}_{i,t}^{neg1HA} + \beta_2 \mathbb{1}_{i,t}^{neg2HA} + \beta_3 \mathbb{1}_{i,t}^{neg3HA} + \beta_4 \mathbb{1}_{i,t}^{neg4HA} + \beta_5 \mathbb{1}_{i,t}^{neg5HA}$$
$$+ \beta_6 \mathbb{1}_{i,t} + \beta_7 p_{i,t}^{rent} + \beta_8 \bar{p}_{i,t}^{zipcode} + \beta_9 n_{i,t}^{zipcode} + \beta_{10} g_{i,t}^{n} + \epsilon_{i,t}. \quad (3)$$

In Tables 3 and 4 I only report the results related to the occurrence of the first five HI and HA negative reviews since the parameters associated with the other variables have the same qualitative behaviour. Here I report and briefly comment the main results:

- The occurrence of the first five HI negative reviews has a negative and significant effect on the probability of listings to stay on the platform; that is, it increases the probability to exit the platform. Moreover, the greater is the order of the HI negative review, the more negative is the impact over the probability to stay (for instance, the impact of the first HI negative review is lower than the impact of the second HI negative review);

- The occurrence of the first five HA negative reviews affects more ambiguously the probability of listings to stay on the platform: the occurrence of the first two HA negative reviews has a significant and negative impact on the permanence of the listings (the parameters are almost always more negative with respect to the one associated with the HI negative reviews). Differently, the occurrence of the following HA negative review has an insignificant effect. This is also due to the extremely low number of HA negative reviews: only 138 listings have more than two HA negative reviews published in their webpages.

4.2 The Impact over the Hosts' Posted Prices

In the remaining part of this section I study the impact of negative reviews over the hosts' posted prices and the probability of transactions faced by hosts.

In both cases, the listings' selection process may severely affect the estimates about the impact of negative reviews: as observed in the previous exercises, negative reviews do have an effect over the persistence of listings on

Airbnb, inducing hosts who receive negative reviews to quit their online account. In this sense, when we study the impact of negative reviews over the hosts' pricing decisions, we have to recall that the impact that is measured regards only those hosts who decided not to exit the platform. As the selection process may be endogenous, with only listings with particular features exiting the market, estimates of the negative reviews' impact may be inconsistent without controlling for the selection.

Before presenting the variables I briefly discuss the inconsistency of the results when the selection process is endogenous and how the Heckman two-step procedure solves this issue.

The variables considered in order to measure the hosts' performance (defined here generically with $y^*_{i,t}$) are observed only when the listings are present in Airbnb; yet, in the previous exercise the probability to be on the platform by listings has been studied in relation with certain observables (called here $z_{i,t}$) so that:

$$y_{i,t} = \begin{cases} y^*_{i,t} & \text{if } z'_{i,t}\beta + \epsilon_{i,t} > 0 \\ - & \text{otherwise} . \end{cases}$$

Accordingly, estimating the parameters in the following equation (where $x_{i,t}$ represents the set of other observables):

$$y^*_{i,t} = x'_{i,t}\gamma + \nu_{i,t} \tag{4}$$

with Least Squares estimation no longer provides consistent estimates of $\hat{\gamma}$, even if the linear model is correctly specified. The bias comes from the fact that $E(y_{i,t}|x_{i,t})$ is not equivalent to $x'_{i,t}\gamma$ if $\epsilon_{i,t}$ and $\nu_{i,t}$ are correlated:

$$E(y_{i,t}|x_{i,t}) = E(y^*_{i,t}|x_{i,t}, z'_{i,t}\beta + \epsilon_{i,t} > 0) = x'_{i,t}\gamma + E(\nu_{i,t}|z'_{i,t}\beta + \epsilon_{i,t} > 0). \tag{5}$$

In order to control for endogenous selection (that is $E(\nu_{i,t}|z'_{i,t}\beta + \epsilon_{i,t} > 0) \neq 0$), it is possible to write equation 5 in the following way, assuming a joint normality distribution between $\epsilon_{i,t}$ and $\nu_{i,t}$:

$$\begin{aligned}
E(y_{i,t}|x_{i,t}) &= E(y^*_{i,t}|x_{i,t}, z'_{i,t}\beta + \epsilon_{i,t} > 0) \\
&= x'_{i,t}\gamma + E(\nu_{i,t}|z'_{i,t}\beta + \epsilon_{i,t} > 0) \\
&= x'_{i,t}\gamma + E(\rho\sigma\epsilon_{i,t} + \xi|z'_{i,t}\beta + \epsilon_{i,t} > 0) \\
&= x'_{i,t}\gamma + \rho\sigma\lambda(z'_{i,t}\beta) \qquad (6)
\end{aligned}$$

where $\lambda(z'_{i,t}\beta) = \frac{\phi((z'_{i,t}\beta)}{\Phi((z'_{i,t}\beta)}$.

The Heckman procedure (Heckman, 1977) consists of two stages:

1. A first stage in which $\lambda(z'_{i,t}\hat{\beta})$ is constructed through predictions of the probit estimations obtained in the first exercise regarding the persistence of listings in Airbnb (see Equations 1, 2 and 3);

2. A second stage which consists of the linear estimation of $y^*_{i,t} = x'_{i,t}\gamma + \nu_{i,t}$ augmented with the constructed regressor $\lambda(z'_{i,t}\hat{\beta})$ estimated in the first stage. The significance of the coefficients corresponding with $\lambda(z'_{i,t}\hat{\beta})$ provides a test for endogenous selection.

It is crucial for the identification that $z_{i,t}$ includes some regressors that are not in $x_{i,t}$: it is necessary to have variation in the selection equation that only affects the outcome through the selection, but not directly through the outcome equation. These are called *exclusion restrictions*.

In the next two exercises the monthly rental price calculated at the zipcode level, $p^{rent}_{i,t}$, plays the role of the exclusion restriction: in this sense, I impose that the hosts' pricing decision as well as the probability of transactions are affected by variations in the rental prices only through the selection channel as it represents the value of the hosts' exit option.

In this subsection I study the impact of negative reviews over the pricing decions by hosts: the price posted by listing i in week t is indicated by $p_{i,t}$. In the following regressions I use the logarithm transformation of this value in order to interpret the resulting parameters as the percentage changes on the posted price; the explanatory variables considered for the following regressions are the same of the ones used for the previous exercise except for the rental price $p_{i,t}^{rent}$ (the exclusion restriction).

Accordingly, I show how the selection process affects the estimates considering first POLS, random effect and fixed effect regression models without controlling for the selection:

$$\begin{aligned}\ln(p_{i,t}) = \alpha &+ \gamma_1 \mathbb{1}_{i,t}^{neg1} + \gamma_2 \mathbb{1}_{i,t}^{neg2} + \gamma_3 \mathbb{1}_{i,t}^{neg3} + \gamma_4 \mathbb{1}_{i,t}^{neg4} + \gamma_5 \mathbb{1}_{i,t}^{neg5} \\ &+ \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \nu_{i,t}. \end{aligned} \tag{7}$$

Finally, I consider the selection process calculating the parameter $\lambda(z_{i,t}'\hat{\beta})$ from the probit random effect model presented in Equation 1; and then proceeding with a random effect model of the following equation:[10]

$$\begin{aligned}\ln(p_{i,t}) = \alpha &+ \gamma_1 \mathbb{1}_{i,t}^{neg1} + \gamma_2 \mathbb{1}_{i,t}^{neg2} + \gamma_3 \mathbb{1}_{i,t}^{neg3} + \gamma_4 \mathbb{1}_{i,t}^{neg4} + \gamma_5 \mathbb{1}_{i,t}^{neg5} \\ &+ \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \gamma_{11} \lambda(z_{i,t}'\hat{\beta}) + \nu_{i,t}. \end{aligned} \tag{8}$$

Tables 5 and 6 present the results associated with the reduced-form Equations 7 and 8 where the impact of the generic first five negative reviews is studied.

Here I report and briefly comment the main results:

- Controlling for the selection process has a great effect on the parameters regarding the occurrence of reviews in the previous three weeks. In particular, the estimate of the coefficient regarding $\mathbb{1}_{i,t}^{neg1}$ becomes significant (and negative); and the estimate of the coefficient regarding $\mathbb{1}_{i,t}$ reverts the sign becoming positive and significant;

- In all the models, the coefficients regarding $\mathbb{1}_{i,t}^{neg2}$, $\mathbb{1}_{i,t}^{neg3}$, $\mathbb{1}_{i,t}^{neg4}$, $\mathbb{1}_{i,t}^{neg5}$ are always positive and often significant: the occurrence of negative reviews (that are not the first one) have a positive effect on prices. The sign of this effect may be explained recalling that negative reviews may not be perceived as really negative when they report an information already

[10] Regarding the feasibility of this approach with panel data, see `http://www.stata.com/statalist/archive/2005-06/msg00456.html`.

present in previous reviews. No further information is disclosed and no changes in price are necessary;

- A higher average price $\bar{p}_{i,t}^{zipcode}$ or a higher number of listing in the zipcode area of the listing $n_{i,t}^{zipcode}$ lead to higher posted prices as they are a measure of the attractiveness of the area;

- An increase in the growth rate of listings in the zipcode area $g_{i,t}^{n}$ leads to a greater degree of competition among listings and a lower price;

- The coefficient regarding the Heckman lambda $\lambda(z'_{i,t}\hat{\beta})$ is significant and the selection appears to be endogenous.

The same exercise is repeated considering the first five HI and HA negative reviews, defining the occurrence of the HI and HA negative reviews as in the previous subsection. The reduced-form equations used in the first three coloumns of Tables 7 and 8 are defined as follows (as before I report only the estimates of the coefficients related to the HI and HA negative reviews as the other coefficients behave similarly with respect to the previous regressions):

$$\begin{aligned}\ln(p_{i,t}) &= \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1HI} + \gamma_2 \mathbb{1}_{i,t}^{neg2HI} + \gamma_3 \mathbb{1}_{i,t}^{neg3HI} + \gamma_4 \mathbb{1}_{i,t}^{neg4HI} + \gamma_5 \mathbb{1}_{i,t}^{neg5HI} \\ &\quad + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \nu_{i,t} \qquad (9)\end{aligned}$$

$$\begin{aligned}\ln(p_{i,t}) &= \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1HA} + \gamma_2 \mathbb{1}_{i,t}^{neg2HA} + \gamma_3 \mathbb{1}_{i,t}^{neg3HA} + \gamma_4 \mathbb{1}_{i,t}^{neg4HA} + \gamma_5 \mathbb{1}_{i,t}^{neg5HA} \\ &\quad + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \nu_{i,t}. \qquad (10)\end{aligned}$$

The last coloumns of Tables 7 and 8 represent the estimates of the second step of the Heckman procedure using random effect:

$$\begin{aligned}\ln(p_{i,t}) &= \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1HI} + \gamma_2 \mathbb{1}_{i,t}^{neg2HI} + \gamma_3 \mathbb{1}_{i,t}^{neg3HI} + \gamma_4 \mathbb{1}_{i,t}^{neg4HI} + \gamma_5 \mathbb{1}_{i,t}^{neg5HI} \\ &\quad + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \gamma_{11} \lambda^{HI}(z'_{i,t}\hat{\beta}) + \nu_{i,t} \qquad (11)\end{aligned}$$

$$\begin{aligned}\ln(p_{i,t}) &= \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1HA} + \gamma_2 \mathbb{1}_{i,t}^{neg2HA} + \gamma_3 \mathbb{1}_{i,t}^{neg3HA} + \gamma_4 \mathbb{1}_{i,t}^{neg4HA} + \gamma_5 \mathbb{1}_{i,t}^{neg5HA} \\ &\quad + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \gamma_{11} \lambda^{HA}(z'_{i,t}\hat{\beta}) + \nu_{i,t}. \qquad (12)\end{aligned}$$

where the parameters $\lambda^{HI}(z'_{i,t}\hat{\beta})$ and $\lambda^{HA}(z'_{i,t}\hat{\beta})$ are calculated from the models defined in Equations 2 and 3 with probit random effect.

The results concerning HI and HA negative reviews are the following:

- The impacts of HI and HA negative reviews result to be significantly different;
- The HI negative reviews show a similar behaviour with respect to the generic negative reviews: after controlling for the selection process, the occurrence of the first HI negative review has a significant negative impact on the posted price; while the occurrence of the other following HI negative reviews has a significant positive impact: as explained above, the HI negative reviews after the first HI negative review may not disclose additional information and they may be perceived as positive;
- The HA negative reviews have a non-significant impact over the price posted by the hosts;
- For both types of negative reviews, the coefficients regarding the Heckman lambdas ($\lambda^{HI}(z'_{i,t}\hat{\beta})$ and $\lambda^{HA}(z'_{i,t}\hat{\beta})$) are significant and the selection appears to be endogenous.

4.3 The Impact over the Probability of Transactions

In this last subsection I analyse the impact of negative reviews over the listing' probability of having a transaction. As reported in Section 3, the public information available on the Airbnb listings' webpages does not include the number of transactions occurred and associated dates, but it only considers the number of reviewed transactions that appears online. In this sense, my study will take the reviewed transactions present on the listings' webpages as a proxy for the total transactions between hosts and guests occurred in a period. This approach is based on the following assumption: the frequency of guests reviews is orthogonal with respect to the characteristics of the host (including reputation).

Thanks to the work by Fradkin et al. (2015) I can motivate the credibility of this assumption. Fradkin et al. (2015) studied the bias and the reciprocity of online reviews using an Airbnb proprietary dataset. They show that reviews' bias exists in the dataset because of fear of retaliation and prosocial reciprocity; still, at the aggregate level, Airbnb reviews result to be informative.

Accordingly, the probability of having a transaction for listing i in week t is defined using a listing-specific binary variable indicating whether a review appear on the listing i's webpage in week t or not; and it is denoted with $\phi_{i,t}$.

The explanatory variables that are used in this set-up coincide with the ones used in the previous exercise, with the rental price $p_{i,t}^{rent}$ playing the role of the exclusion restriction. The unique difference with respect to the previous exercise is represented by the period of time in which the first five negative reviews occur before the week studied. In the previous two analyses $\mathbb{1}_{i,t}^{neg1}$, $\mathbb{1}_{i,t}^{neg2}$, $\mathbb{1}_{i,t}^{neg3}$, $\mathbb{1}_{i,t}^{neg4}$, $\mathbb{1}_{i,t}^{neg5}$ defined the occurrence of the first five negative reviews during a period of three weeks before week t. In this exercise I consider also a shorter period of time, with $\mathbb{1}_{i,t}^{neg1_1}$, $\mathbb{1}_{i,t}^{neg2_1}$, $\mathbb{1}_{i,t}^{neg3_1}$, $\mathbb{1}_{i,t}^{neg4_1}$, $\mathbb{1}_{i,t}^{neg5_1}$ representing the occurrence of the first five negative reviews in week $t-1$.

Tables 9 and 11 report the results regarding the three-weeks parameters; while Tables 10 and 12 report the results regarding the one-week parameters. In all the tables, the first three coloumns include the estimates of the parameters shown in the following equations considering random effect probit and logit models and fixed effect logit model:

$$\begin{aligned}\phi_{i,t} &= \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1} + \gamma_2 \mathbb{1}_{i,t}^{neg2} + \gamma_3 \mathbb{1}_{i,t}^{neg3} + \gamma_4 \mathbb{1}_{i,t}^{neg4} + \gamma_5 \mathbb{1}_{i,t}^{neg5} \\ &\quad + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \nu_{i,t}\end{aligned} \tag{13}$$

$$\begin{aligned}\phi_{i,t} &= \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1_1} + \gamma_2 \mathbb{1}_{i,t}^{neg2_1} + \gamma_3 \mathbb{1}_{i,t}^{neg3_1} + \gamma_4 \mathbb{1}_{i,t}^{neg4_1} + \gamma_5 \mathbb{1}_{i,t}^{neg5_1} \\ &\quad + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \nu_{i,t}.\end{aligned} \tag{14}$$

Finally, I consider the selection process calculating the parameters $\lambda(z_{i,t}'\hat{\beta})$ and $\lambda_1(z_{i,t}'\hat{\beta})$ from the model presented in Equation 1 with probit random effect; and then proceeding with a linear random effect model of the following equations:

$$\begin{aligned}\phi_{i,t} = \alpha &+ \gamma_1 \mathbb{1}_{i,t}^{neg1} + \gamma_2 \mathbb{1}_{i,t}^{neg2} + \gamma_3 \mathbb{1}_{i,t}^{neg3} + \gamma_4 \mathbb{1}_{i,t}^{neg4} + \gamma_5 \mathbb{1}_{i,t}^{neg5} \\ &+ \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \gamma_{11} \lambda(z_{i,t}' \hat{\beta}) + \nu_{i,t} \end{aligned} \tag{15}$$

$$\begin{aligned}\phi_{i,t} = \alpha &+ \gamma_1 \mathbb{1}_{i,t}^{neg1_1} + \gamma_2 \mathbb{1}_{i,t}^{neg2_1} + \gamma_3 \mathbb{1}_{i,t}^{neg3_1} + \gamma_4 \mathbb{1}_{i,t}^{neg4_1} + \gamma_5 \mathbb{1}_{i,t}^{neg5_1} \\ &+ \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \gamma_{11} \lambda_1(z_{i,t}' \hat{\beta}) + \nu_{i,t}. \end{aligned} \tag{16}$$

Here I report and comment the main results:

- All coefficients associated with the occurrence of negative reviews in the previous three weeks are always positive and significant; as well as positive and significant is the coefficient associated with the occurrence of a generic review in the previous three weeks;

- Controlling for selection the coefficient of $\mathbb{1}_{i,t}^{neg1_1}$ becomes negative and significant; while the rest remains significant and positive; not controlling for selection, all the coefficients associated with the occurrence of negative reviews in the previous week are always positive and significant;

- The coefficient regarding $\lambda_1(z_{i,t}'\hat{\beta})$ is always significant expressing the presence of endogenous selection.

Finally, I repeat the same exercise for the HI and HA negative reviews studying only the occurrence of reviews one week before the one studied. In Tables 13 and 14 the coefficients about the HI and the HA are reported considering the random effect probit without controlling for the selection:

$$\begin{aligned}\phi_{i,t} = \alpha &+ \gamma_1 \mathbb{1}_{i,t}^{neg1HI_1} + \gamma_2 \mathbb{1}_{i,t}^{neg2HI_1} + \gamma_3 \mathbb{1}_{i,t}^{neg3HI_1} + \gamma_4 \mathbb{1}_{i,t}^{neg4HI_1} + \gamma_5 \mathbb{1}_{i,t}^{neg5HI_1} \\ &+ \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \nu_{i,t} \end{aligned} \tag{17}$$

$$\begin{aligned}\phi_{i,t} = \alpha &+ \gamma_1 \mathbb{1}_{i,t}^{neg1HA_1} + \gamma_2 \mathbb{1}_{i,t}^{neg2HA_1} + \gamma_3 \mathbb{1}_{i,t}^{neg3HA_1} + \gamma_4 \mathbb{1}_{i,t}^{neg4HA_1} + \gamma_5 \mathbb{1}_{i,t}^{neg5HA_1} \\ &+ \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \nu_{i,t}. \end{aligned} \tag{18}$$

and with the control for the selection using, as a first stage the results obtained from the Regressions 2 and 3:

$$\phi_{i,t} = \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1HI_1} + \gamma_2 \mathbb{1}_{i,t}^{neg2HI_1} + \gamma_3 \mathbb{1}_{i,t}^{neg3HI_1} + \gamma_4 \mathbb{1}_{i,t}^{neg4HI_1} + \gamma_5 \mathbb{1}_{i,t}^{neg5HI_1} + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \gamma_{11} \lambda_1^{HI}(z'_{i,t}\hat{\beta}) + \nu_{i,t} \tag{19}$$

$$\phi_{i,t} = \alpha + \gamma_1 \mathbb{1}_{i,t}^{neg1HA_1} + \gamma_2 \mathbb{1}_{i,t}^{neg2HA_1} + \gamma_3 \mathbb{1}_{i,t}^{neg3HA_1} + \gamma_4 \mathbb{1}_{i,t}^{neg4HA_1} + \gamma_5 \mathbb{1}_{i,t}^{neg5HA_1} + \gamma_6 \mathbb{1}_{i,t} + \gamma_8 \bar{p}_{i,t}^{zipcode} + \gamma_9 n_{i,t}^{zipcode} + \gamma_{10} g_{i,t}^{n} + \gamma_{11} \lambda_1^{HA}(z'_{i,t}\hat{\beta}) + \nu_{i,t}. \tag{20}$$

The main results are briefly commented as follows:

- Controlling for the selection process (always endogenous since the Heckman coefficients are always significant) makes the impact of HI negative reviews significant: the occurrence of the first negative review affects negatively the probability to have a transaction; while the following others have positive effects.
- The occurrence of HA negative reviews is rarely significant: they do not really have an impact over the probability to have a transaction.

5 Conclusion

I conclude this paper with the obtained results associated with the impact of HI and HA negative reviews and the connected theoretical predictions. Finally I propose the further steps of my research.

5.1 In Search of a Theory

From the previous three analyses some results can be evidenced in relation to the impact of the two different types of negative reviews:

1. Both types of negative reviews show a negative significant impact on one dimension of the hosts' performance: the persistence on the Airbnb platform;

2. Negative reviews have an increasing negative impact over the persistence: it is identifiable in relation to all the first five HI negative reviews; while only in relation to the first two HA negative reviews;

3. HA negative reviews have a very limited impact over the hosts' pricing decision and the probability to have a transaction;

4. The impact of HI negative reviews differs with respect to the order of the negative reviews: the first HI negative reviews has a significant negative effect over prices and (with the lag of just one week) over the probability to have a transaction. Conversely, the following HI negative reviews have a positive effect on both dimensions of the hosts' performance.

In order to reconcile the great negative impact over the persistence of listings with the limited impact over the other measures of the hosts' performance, I develop the following theory: dwelling's owners enter the platform as they are attracted by the possibility to earn more money with respect to the monthly rent they get in the rental market. Still, at the beginning, and for a significant part of their life on Airbnb, the price they charge does not cover the additional costs they have to face staying on Airbnb with respect to the ones they face renting the dwelling. In this sense, they need to increase the price over time, but this profitability improvement may not be enough to repay the additional costs causing the hosts to exit the market. I can support this theory observing that the positive coefficients related to the occurrence of negative reviews are always lower than the ones associated with the occurrence of generic reviews. Similarly, the probability of having a transaction increases over time as hosts start with no reputation and the guests' beliefs for the listings' quality associated with no reputation could be very low. Accordingly, reviews expressing some concerns about the quality of the goods are improving the probability of stay, but not as much as completely positive reviews.

Yet, as pointed out before, these improvements may not be sufficient to overcome the additional costs and hosts may decide to exit even when their performance is increasing.

This reasoning is applicable to all the negative reviews and it is not considering the specific nature of the content of negative reviews. HI negative reviews seem to have a greater impact as they provide information about the dwelling's characteristics that are persistent over time and that could directly affect the future guests. Differently, HA negative reviews seem to have a lower impact in all the dimensions of the hosts' performance as they report concerns that could be temporal and they may not affect future guests.

5.2 Three Directions for the Future

There are three main directions that I will follow in order to strengthener my results and refine my analysis:

1. *The Sentimental Analysis*: as reported before, my content analysis is restrictive and a greater accuracy may be necessary to further improve the capacity to detect all negative reviews and their topics: using Taboada et al. (2011) I am going to select negative reviews using the established vocabularies defined in the literature of lexicon-based sentimental analysis. With similar technique I may proceed also in the extraction of the negative review's topic;

2. *The Theoretical Model*: I devoted my attention to sound empirical findings. The next point of my research relates to the development of a dynamic theoretical model able to explain the results with a unified theory;

3. *The Empirical Strategy*: The Heckman two-step procedure is a powerful econometric technique that allows to get more reliable results. In the future I am going to strengthen it with further computational and technical investigations.

References

Michael Anderson and Jeremy Magruder. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The*

Economic Journal, 122(563):957–989, 2012.

Heski Bar-Isaac and Steven Tadelis. *Seller reputation*. Now Publishers Inc, 2008.

Luis Cabral and Ali Hortacsu. The dynamics of seller reputation: Evidence from ebay*. *The Journal of Industrial Economics*, 58(1):54–78, 2010.

Judith A Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354, 2006.

Zoë Cullen and Chiara Farronato. Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Job Market Paper*, 2014.

Chrysanthos Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science*, 49(10):1407–1424, 2003.

Chrysanthos Dellarocas. Reputation mechanisms. *Handbook on Economics and Information Systems*, pages 629–660, 2006.

Chrysanthos Dellarocas and Charles A Wood. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 54(3):460–476, 2008.

Michael Dinerstein, Liran Einav, Jonathan Levin, and Neel Sundaresan. Consumer price search and platform design in internet commerce. Technical report, National Bureau of Economic Research, 2014.

Andrey Fradkin. Search frictions and the design of online marketplaces. *NBER Working Paper*, 2014.

Andrey Fradkin, Elena Grewal, Dave Holtz, and Matthew Pearson. Bias and reciprocity in online reviews: Evidence from field experiments on airbnb. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 641–641. ACM, 2015.

James J Heckman. Sample selection bias as a specification error (with an application to the estimation of labor supply functions). 1977.

Michael Luca. Reviews, reputation, and revenue: The case of yelp. com. *Com (September 16, 2011). Harvard Business School NOM Unit Working Paper*, (12-016), 2011.

George J Mailath and Larry Samuelson. *Repeated games and reputations: long-run relationships*. Oxford university press, 2006.

Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. The value of reputation on ebay: A controlled experiment. *Experimental economics*, 9(2):79–101, 2006.

Arun Sundararajan. From zipcar to the sharing economy. *Harvard Business Review*, 1, 2013.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

Appendix A: Tables

Table 1: Estimates associated with the Regression 1 considering the impact of the occurrence of the first five negative reviews and a generic review in the previous three weeks.

	(POLS robust se)	(probit robust se)	(RE probit)	(RE logit)	(FE logit)
	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$
$\mathbb{1}_{i,t}^{neg1}$	-0.0566***	-0.307***	-0.307***	-0.617***	-0.694***
	(0.00373)	(0.0298)	(0.0154)	(0.0299)	(0.0298)
$\mathbb{1}_{i,t}^{neg2}$	-0.0472***	-0.298***	-0.298***	-0.578***	-0.656***
	(0.00440)	(0.0386)	(0.0211)	(0.0412)	(0.0412)
$\mathbb{1}_{i,t}^{neg3}$	-0.0392***	-0.299***	-0.299***	-0.594***	-0.679***
	(0.00520)	(0.0485)	(0.0283)	(0.0551)	(0.0554)
$\mathbb{1}_{i,t}^{neg4}$	-0.0378***	-0.317***	-0.317***	-0.638***	-0.711***
	(0.00628)	(0.0595)	(0.0367)	(0.0715)	(0.0723)
$\mathbb{1}_{i,t}^{neg5}$	-0.0291***	-0.315***	-0.315***	-0.644***	-0.771***
	(0.00772)	(0.0767)	(0.0470)	(0.0923)	(0.0928)
$\mathbb{1}_{i,t}$	0.388***	1.550***	1.550***	2.916***	2.876***
	(0.00200)	(0.00602)	(0.00275)	(0.00540)	(0.00539)
N	3,162,225	3,162,225	3,162,225	3,162,225	3,139,620

Note: POLS linear probability model with robust standard error, probit with robust standard error, random effect probit and logit models, and fixed effect logit model are reported from the first to the fifth coloumn. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 2: Estimates associated with the Regression 1 considering the impact of the rental price at the zipcode area, the average price posted by listing located in the zipcode area, the number of listings located in the zipcode area, the growth rate of listings.

	(POLS robust se)	(probit robust se)	(RE probit)	(RE logit)	(FE logit)
	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$
$p_{i,t}^{rent}$	-0.0000970***	-0.000648***	-0.000648***	-0.00120***	-0.00133***
	(0.00000443)	(0.0000282)	(0.00000933)	(0.0000175)	(0.0000259)
$\bar{p}_{i,t}^{zipcode}$	0.000419***	0.00220***	0.00220***	0.00361***	0.00110***
	(0.0000626)	(0.000385)	(0.000119)	(0.000223)	(0.000250)
$n_{i,t}^{zipcode}$	0.000669***	0.00487***	0.00487***	0.00923***	0.0109***
	(0.00000732)	(0.0000681)	(0.0000169)	(0.0000334)	(0.0000368)
$g_{i,t}^{n}$	-0.000582*	0.00386**	0.00386*	0.0129***	0.00282
	(0.000276)	(0.00139)	(0.00150)	(0.00273)	(0.00273)
N	3,162,225	3,162,225	3,162,225	3,162,225	3,139,620

Note: POLS linear probability model with robust standard error, probit with robust standard error, random effect probit and logit models, and fixed effect logit model are reported from the first to the fifth coloumn. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 3: Estimates associated with the Regression 2 considering the impact of the occurrence of the first five HI negative reviews in the previous three weeks.

	(POLS robust se)	(probit robust se)	(RE probit)	(RE logit)	(FE logit)
	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$
$\mathbb{1}_{i,t}^{neg1HI}$	-0.0576***	-0.320***	-0.320***	-0.651***	-0.731***
	(0.00402)	(0.0324)	(0.0171)	(0.0331)	(0.0330)
$\mathbb{1}_{i,t}^{neg2HI}$	-0.0597***	-0.422***	-0.422***	-0.829***	-0.912***
	(0.00541)	(0.0448)	(0.0253)	(0.0487)	(0.0488)
$\mathbb{1}_{i,t}^{neg3HI}$	-0.0469***	-0.386***	-0.386***	-0.784***	-0.887***
	(0.00687)	(0.0616)	(0.0377)	(0.0725)	(0.0729)
$\mathbb{1}_{i,t}^{neg4HI}$	-0.0441***	-0.426***	-0.426***	-0.880***	-0.997***
	(0.00886)	(0.0791)	(0.0538)	(0.104)	(0.105)
$\mathbb{1}_{i,t}^{neg5HI}$	-0.0496***	-0.552***	-0.552***	-1.079***	-1.211***
	(0.0133)	(0.115)	(0.0710)	(0.139)	(0.141)
N	3,162,225	3,162,225	3,162,225	3,162,225	3,139,620

Note: POLS linear probability model with robust standard error, probit with robust standard error, random effect probit and logit models, and fixed effect logit model are reported from the first to the fifth coloumn. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 4: Estimates associated with the Regression 3 considering the impact of the occurrence of the first five HA negative reviews in the previous three weeks.

	(POLS robust se)	(probit robust se)	(RE probit)	(RE logit)	(FE logit)
	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$	$\pi_{i,t}$
$\mathbb{1}_{i,t}^{neg1HA}$	-0.0438**	-0.269**	-0.269***	-0.525***	-0.556***
	(0.0169)	(0.101)	(0.0545)	(0.105)	(0.104)
$\mathbb{1}_{i,t}^{neg2HA}$	-0.168*	-0.999***	-0.999***	-1.835***	-1.888***
	(0.0696)	(0.282)	(0.137)	(0.269)	(0.269)
$\mathbb{1}_{i,t}^{neg3HA}$	-0.00507	5.111***	5.111	16.92	16.41
	(0.0408)	(0.319)	(4929.4)	(8014.6)	(8014.3)
$\mathbb{1}_{i,t}^{neg4HA}$	0.00442	4.878***	4.878	17.47	17.33
	(0.00375)	(0.194)	(19465.0)	(17364.8)	(17365.2)
$\mathbb{1}_{i,t}^{neg5HA}$	0.000190	4.831***	4.831	17.33	17.23
	(0.00301)	(0.171)	(35430.0)	(26724.4)	(26724.9)
N	3,162,225	3,162,225	3,162,225	3,162,225	3,139,620

Note: POLS linear probability model with robust standard error, probit with robust standard error, random effect probit and logit models, and fixed effect logit model are reported from the first to the fifth coloumn. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 5: Estimates associated with the Regressions 7 and 8 considering the impact of the occurrence of the first five negative reviews and a generic review in the previous three weeks.

	(POLS)	(RE)	(FE)	(Heckman - RE)
	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$
$\mathbb{1}_{i,t}^{neg1}$	-0.00110	-0.00149*	-0.000958	-0.00148***
	(0.00113)	(0.000709)	(0.000705)	(0.000214)
$\mathbb{1}_{i,t}^{neg2}$	0.00326*	0.00221*	0.00363***	0.00289**
	(0.00135)	(0.000919)	(0.000915)	(0.000923)
$\mathbb{1}_{i,t}^{neg3}$	0.00602***	0.00489***	0.00630***	0.00563***
	(0.00171)	(0.00119)	(0.00119)	(0.00119)
$\mathbb{1}_{i,t}^{neg4}$	0.00321	0.00259	0.00354*	0.00281
	(0.00203)	(0.00152)	(0.00152)	(0.00153)
$\mathbb{1}_{i,t}^{neg5}$	0.00627*	0.00567**	0.00640***	0.00587**
	(0.00252)	(0.00190)	(0.00189)	(0.00190)
$\mathbb{1}_{i,t}$	-0.00158**	-0.00204***	-0.00114***	0.000755**
	(0.000525)	(0.000213)	(0.000212)	(0.000190)
N	1105212	1105212	1105212	1105212

Note: POLS panel model with robust standard error and no selection; random effect panel model and no selection; fixed effect panel model and no selection; and random effect panel model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 6: Estimates associated with the Regressions 7 and 8 considering the impact of the average price posted by listing located in the zipcode area, the number of listings located in the zipcode area, the growth rate of listings and the Heckman coefficient.

	(POLS)	(RE)	(FE)	(Heckman - RE)
	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$
$\bar{p}_{i,t}^{zipcode}$	0.00174***	0.00165***	0.00123***	0.00174***
	(0.0000776)	(0.0000168)	(0.0000171)	(0.0000165)
$n_{i,t}^{zipcode}$	0.0000791***	0.0000682***	0.0000821***	0.0000842***
	(0.00000936)	(0.00000216)	(0.00000213)	(0.00000243)
$g_{i,t}^{n}$	-0.0000426***	-0.0000255***	-0.0000473***	-0.0000428***
	(0.00000585)	(0.00000515)	(0.00000511)	(0.00000513)
$\lambda(z_{i,t}'\hat{\beta})$				0.00250***
				(0.000589)
N	1105212	1105212	1105212	1105212

Note: POLS panel model with robust standard error and no selection; random effect panel model and no selection; fixed effect panel model and no selection; and random effect panel model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 7: Estimates associated with the Regressions 9 and 11 considering the impact of the occurrence of the first five HI negative reviews in the previous three weeks.

	(POLS)	(RE)	(FE)	(Heckman - RE)
	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$
$\mathbb{1}_{i,t}^{neg1HI}$	0.00148	0.00148	0.00163*	-0.00125**
	(0.00120)	(0.000778)	(0.000774)	(0.000384)
$\mathbb{1}_{i,t}^{neg2HI}$	0.00423*	0.00423***	0.00448***	0.00392***
	(0.00166)	(0.00113)	(0.00112)	(0.00113)
$\mathbb{1}_{i,t}^{neg3HI}$	0.00719**	0.00719***	0.00740***	0.00693***
	(0.00237)	(0.00160)	(0.00159)	(0.00160)
$\mathbb{1}_{i,t}^{neg4HI}$	0.00739*	0.00739**	0.00747***	0.00708**
	(0.00305)	(0.00227)	(0.00226)	(0.00227)
$\mathbb{1}_{i,t}^{neg5HI}$	0.0104*	0.0104***	0.0101***	0.0101**
	(0.00464)	(0.00308)	(0.00307)	(0.00309)
N	1105212	1105212	1105212	1105212

Note: POLS panel model with robust standard error and no selection; random effect panel model and no selection; fixed effect panel model and no selection; and random effect panel model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 8: Estimates associated with the Regressions 10 and 12 considering the impact of the occurrence of the first five HA negative reviews in the previous three weeks.

	(POLS)	(RE)	(FE)	(Heckman - RE)
	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$	$\log p_{i,t}$
$\mathbb{1}_{i,t}^{neg1HA}$	0.000219	0.000219	0.000538	-0.000274
	(0.00453)	(0.00294)	(0.00293)	(0.00295)
$\mathbb{1}_{i,t}^{neg2HA}$	0.0226	0.0226*	0.0229*	0.0215*
	(0.0264)	(0.00959)	(0.00955)	(0.00961)
$\mathbb{1}_{i,t}^{neg3HA}$	-0.00422	-0.00422	-0.00213	-0.00284
	(0.00872)	(0.0265)	(0.0263)	(0.0265)
$\mathbb{1}_{i,t}^{neg4HA}$	0.00356**	0.00356	0.00380	0.00531
	(0.00130)	(0.0374)	(0.0372)	(0.0374)
$\mathbb{1}_{i,t}^{neg5HA}$	0.0189***	0.0189	0.0153	0.0216
	(0.00131)	(0.0528)	(0.0526)	(0.0529)
N	1105212	1105212	1105212	1105212

Note: POLS panel model with robust standard error and no selection; random effect panel model and no selection; fixed effect panel model and no selection; and random effect panel model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 9: Estimates associated with the Regressions 13 and 15 considering the impact of the occurrence of the first five negative reviews and a generic review in the previous three weeks.

	(RE probit)	(RE logit)	(FE logit)	(Heckman - RE)
	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$
$\mathbb{1}_{i,t}^{neg1}$	0.140***	0.227***	0.283***	0.0152***
	(0.0104)	(0.0172)	(0.0172)	(0.00339)
$\mathbb{1}_{i,t}^{neg2}$	0.180***	0.289***	0.247***	0.105***
	(0.0133)	(0.0219)	(0.0218)	(0.00441)
$\mathbb{1}_{i,t}^{neg3}$	0.126***	0.201***	0.129***	0.124***
	(0.0173)	(0.0284)	(0.0283)	(0.00572)
$\mathbb{1}_{i,t}^{neg4}$	0.0737***	0.113**	0.0228	0.129***
	(0.0221)	(0.0364)	(0.0362)	(0.00733)
$\mathbb{1}_{i,t}^{neg5}$	0.0729**	0.107*	0.0106	0.165***
	(0.0276)	(0.0456)	(0.0454)	(0.00910)
$\mathbb{1}_{i,t}$	0.294***	0.501***	0.317***	0.262***
	(0.00333)	(0.00569)	(0.00565)	(0.000844)
N	1120999	1106800	1094253	1120999

Note: random effect probit model and no selection; random effect logit model and no selection; fixed effect logit model and no selection; and random effect probit model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 10: Estimates associated with the Regressions 14 and 16 considering the impact of the occurrence of the first five negative reviews and a generic review in the previous three weeks.

	(RE probit)	(RE logit)	(FE logit)	(Heckman - RE)
	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$
$\mathbb{1}_{i,t}^{neg1_1}$	0.0955***	0.161***	0.246***	-0.0506***
	(0.0175)	(0.0291)	(0.0289)	(0.00587)
$\mathbb{1}_{i,t}^{neg2_1}$	0.119***	0.191***	0.191***	0.0398***
	(0.0223)	(0.0369)	(0.0367)	(0.00760)
$\mathbb{1}_{i,t}^{neg3_1}$	0.0278	0.0421	0.00930	0.0558***
	(0.0288)	(0.0474)	(0.0471)	(0.00980)
$\mathbb{1}_{i,t}^{neg4_1}$	0.0505	0.0743	0.0190	0.0970***
	(0.0369)	(0.0609)	(0.0605)	(0.0125)
$\mathbb{1}_{i,t}^{neg5_1}$	-0.0286	-0.0567	-0.114	0.0989***
	(0.0462)	(0.0763)	(0.0759)	(0.0157)
$\mathbb{1}_{\mathbb{1}\,i,t}$	0.228***	0.370***	0.224***	0.292***
	(0.00328)	(0.00546)	(0.00540)	(0.00208)
N	1120999	1106800	1094253	1120999

Note: random effect probit model and no selection; random effect logit model and no selection; fixed effect logit model and no selection; and random effect probit model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 11: Estimates associated with the Regressions 13 and 15 considering the impact of the average price posted by listing located in the zipcode area, the number of listings located in the zipcode area, the growth rate of listings and the Heckman coefficient.

	(RE probit)	(RE logit)	(FE logit)	(Heckman - RE)
	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$
$\bar{p}_{i,t}^{zipcode}$	0.00152***	0.00254***	0.0274***	0.0000380***
	(0.0000860)	(0.000149)	(0.000499)	(0.0000100)
$n_{i,t}^{zipcode}$	-0.000269***	-0.000405***	0.000656***	-0.0000228***
	(0.0000162)	(0.0000286)	(0.0000689)	(0.00000196)
$g_{i,t}^{n}$	-0.203***	-0.420***	-0.306***	-0.0647***
	(0.00231)	(0.00662)	(0.00666)	(0.000626)
$\lambda(z_{i,t}'\hat{\beta})$				.0339813***
				(.0017349)
N	1120999	1106800	1094253	1120999

Note: random effect probit model and no selection; random effect logit model and no selection; fixed effect logit model and no selection; and random effect probit model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 12: Estimates associated with the Regressions 14 and 16 considering the impact of the average price posted by listing located in the zipcode area, the number of listings located in the zipcode area, the growth rate of listings and the Heckman coefficient.

	(RE probit)	(RE logit)	(FE logit)	(Heckman - RE)
	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$	$\phi_{i,t}$
$\bar{p}_{i,t}^{zipcode}$	0.00176***	0.00297***	0.0285***	0.0000560***
	(0.0000901)	(0.000157)	(0.000499)	(0.0000122)
$n_{i,t}^{zipcode}$	-0.000281***	-0.000415***	0.000706***	-0.0000593***
	(0.0000168)	(0.0000299)	(0.0000689)	(0.00000521)
$g_{i,t}^{n}$	-0.190***	-0.408***	-0.303***	-0.0550***
	(0.00231)	(0.00668)	(0.00670)	(0.000633)
$\lambda(z_{i,t}'\hat{\beta})$				0.0375***
				(0.00204)
N	1120999	1106800	1094253	1120999

Note: random effect probit model and no selection; random effect logit model and no selection; fixed effect logit model and no selection; and random effect probit model with selection are reported from the first to the fourth coloumn. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

Table 13: Estimates associated with the Regressions 17 and 19 considering the impact of the occurrence of the first five HI negative reviews.

	(RE probit)	(Heckman - RE)
	$\phi_{i,t}$	$\phi_{i,t}$
$\mathbb{1}_{i,t}^{neg1HI_1}$	0.0847***	-0.0423***
	(0.0192)	(0.00644)
$\mathbb{1}_{i,t}^{neg2HI_1}$	0.0414	0.0253**
	(0.0274)	(0.00931)
$\mathbb{1}_{i,t}^{neg3HI_1}$	0.0132	0.0650***
	(0.0389)	(0.0132)
$\mathbb{1}_{i,t}^{neg4HI_1}$	0.00931	0.0998***
	(0.0553)	(0.0186)
$\mathbb{1}_{i,t}^{neg5HI_1}$	-0.145	0.0857***
	(0.0747)	(0.0254)
N	1120999	1120999

Note: random effect probit model and no selection; and random effect probit model with selection are reported. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

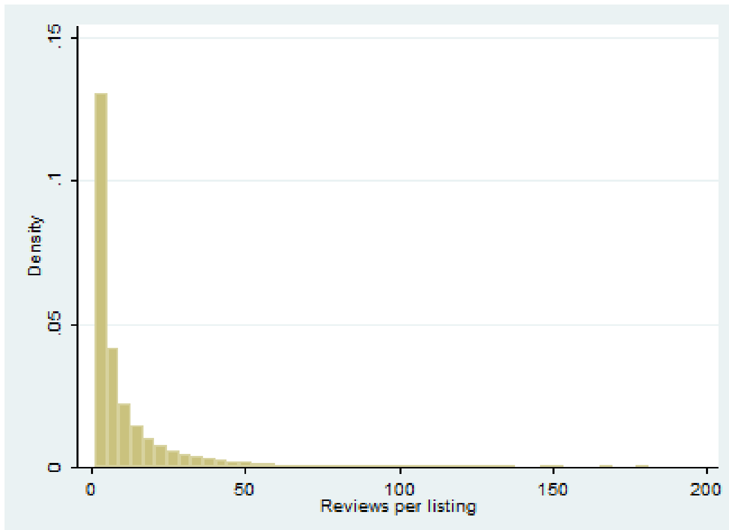
Table 14: Estimates associated with the Regressions 18 and 20 considering the impact of the occurrence of the first five HA negative reviews.

	(RE probit)	(Heckman - RE)
	$\phi_{i,t}$	$\phi_{i,t}$
$\mathbb{1}_{i,t}^{neg1HA_1}$	-0.0279	-0.0332
	(0.0721)	(0.0241)
$\mathbb{1}_{i,t}^{neg2HA_1}$	0.371	0.205**
	(0.235)	(0.0760)
$\mathbb{1}_{i,t}^{neg3HA_1}$	0.461	0.254
	(0.689)	(0.222)
$\mathbb{1}_{i,t}^{neg4HA_1}$	-0.573	-0.0471
	(0.868)	(0.313)
$\mathbb{1}_{i,t}^{neg5HA_1}$	-5.807	-0.608
	(285.4)	(0.443)
N	1120999	1120999

Note: random effect probit model and no selection; and random effect probit model with selection are reported. The parameters in the last column are the ones obtained in the second-step regression of the Heckman procedure considering random effect. The results associated with the first-step are the ones reported in the third coloumn of Table 1 and 2. Standard errors in parenthesis. 1,2,3 stars are used if the parameters are significant at the 10%, 5%, 1% level.

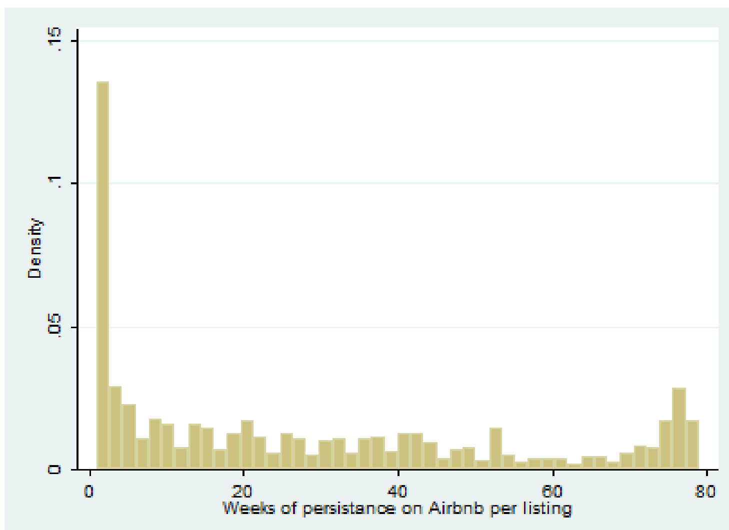
Appendix B: Figures

Figure 1: Distribution of reviews per listing



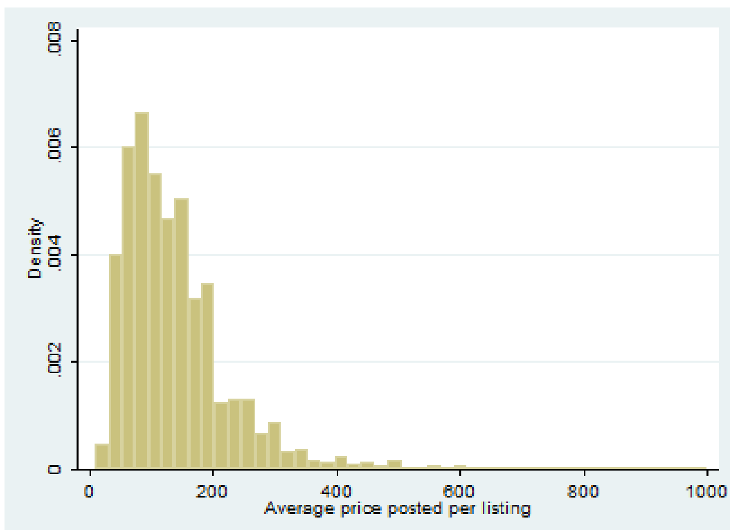
Note: the figure reports the percentage distribution of feedbacks per listing: almost the 15% of listings has only one feedback. This results may be due to the truncation of dataset of the last snapshot date; or to the attrition of listings over time.

Figure 2: Distribution of weeks of persistence on Airbnb per listing



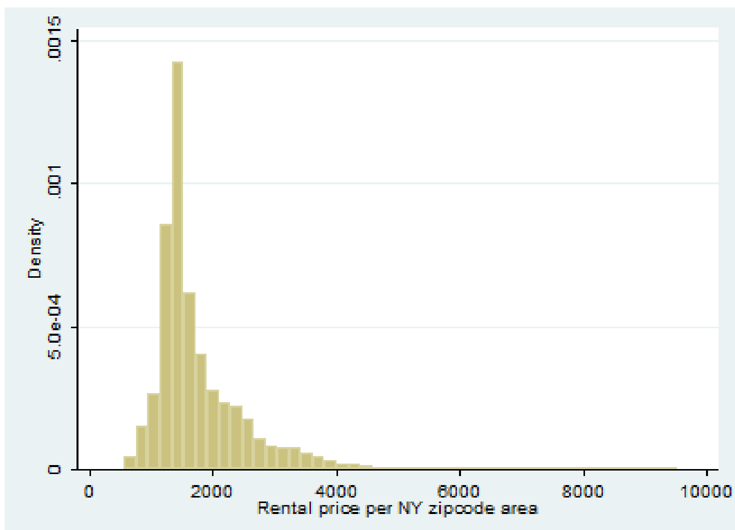
Note: the figure reports the percentage distribution of weeks of persistence on Airbnb per listing: almost 30% of listings survives less than 5 weeks on the Airbnb and less than 20% stays more than a year on the plaftform.

Figure 3: Distribution of average posted price per listing



Note: the figure reports the percentage distribution of the average posted prices on Airbnb per listing: the distribution is unimodal with half of the listings charging an average price that is lower than 120 dollars per night.

Figure 4: Distribution of rental price per zipcode areas



Note: the figure reports the percentage distribution of the rental prices per NY zipcode area: the distribution is unimodal with half of the rental prices that is lower than 1,512 dollars per month.