LUNDS UNIVERSITET
Ekonomihögskolan

Uncertainty in online markets

An empirical study of asymmetric information on the Airbnb market

The Department of Economics
Bachelor Thesis
January 2016

Author:
Jakob Wahlberg
Viktor Jakobsson

Supervisor:
Alessandro Martinello

Abstract

This thesis examines how asymmetric information causes issues in online markets and how it leads to mispricing between buyers and sellers. Asymmetric information and its negative effect on traditional markets have been examined by earlier studies but we strive to test the the theory on an online market. This is tested by looking at how review systems affects the market and whether this information can mitigate the problem of asymmetric information. This is done by looking at data from Airbnb and from this data we build a multivariate regression to see how reviews affect prices. Our results show that a higher review score corresponds to a higher price and this indicates that buyers value the information provided by the review system. We can also see that an increase in the number of reviews has a significant impact on prices and this allows us to determine that asymmetric information is a problem in the market. By observing this we can conclude that online markets are affected by asymmetric information and that review systems is a method for mitigating the problem. The study does not include time series data or data from other online markets.

Keywords: *asymmetric information, online markets, pricing behavior, uncertainty.*

Table of Content

1. Introduction

Today we do a large part of our consumption through online marketplaces. Websites like eBay, Amazon and Airbnb connect buyers with sellers worldwide and provide a chance for individuals to enter new markets. However, an issue with online markets is the lack of information between buyers and sellers. In a regular market quality is observed by inspecting the good visually and physically. In an online market the buyer can only decide on the quality of the good based on the information provided by the seller.

Economic theory defines this as information asymmetries (Varian, 2010). More precisely, it means that one part possesses more information about the quality of the good than the other does. Online markets recognize the problem with asymmetric information and try to reduce the asymmetry with the use of review systems. The role of a review system is to decrease the information asymmetries by providing buyers with valid information about the quality of the good.

Given this background, the goal of this thesis is to empirically test if asymmetric information is present in an online market and if a review system can alleviate the problem. The purpose is to test whether microeconomic theory can help us understand issues of contemporary online markets.

We chose Airbnb as a representation of an online market. The website is an example of a sharing-economy, which allocates resources more efficiently among individuals through peer-to-peer matching. Travelers in search of a room match with a host, without the involvement of a major third party such as hotels and hostels (Byers, Prosperpio and Zervas, 2013). The Airbnb market is of relevance to the thesis since it is an online market where trust and review systems are vital. This is due to the uncertainty of renting an unknown person's residence without any guarantees that the quality is as promised.

Our data set is cross-sectional and contains over 25 500 listings posted in London. The data includes complete information on listed prices, reviews, location, type of property, the host's profile etc. We use a multivariate regression model to study how review systems affect prices and see if information, in the shape of reviews, can alleviate the problem of asymmetric information. The study does not include data from other online markets, time-series data or final prices.

1.2 Disposition

The thesis is structured as follows: In section 2 we present theory behind asymmetric information and discuss how it applies to Airbnb. In section 3 we present the data and method used. In section 4 we present our results. In section 5 we discuss results and limitations. In section 6 we conclude the study.

2. Theory

The theory section consists of two main parts. First, we present how asymmetric information causes market inefficiencies. We discuss ways of mitigating the issue of asymmetric information and then examine how these theories apply to Airbnb. Second, we state our hypotheses and discuss how to test them with the help of our theoretical framework. Subsequently we formulate predictions and test them empirically.

2.1 Theoretical framework

To understand the role of asymmetric information in online markets it is useful to obtain a theoretical background of the issue. In microeconomics a condition for a market to be in equilibrium is the presence of symmetric information (Varian, 2010). More specifically, this means that both parties have access to an equal amount of information about the good and that there are no additional costs for obtaining this information. If one party possesses more information about the good than the other does, there is a risk that they will not disclose the information in order to obtain a better price.

The thesis will look at an example based on George A. Akerlof's paper "The Market for Lemons: Quality Uncertainty and the Market Mechanism" to see how asymmetric information can affect a market explicitly. The example is; assume a second hand car market with one hundred cars available. The sellers are fully aware of the cars' quality while the buyers are not. In our example, half of the cars in the market are of good quality, defined as plums, and half of the cars are of bad quality, defined as lemons. Sellers are willing to sell a plum for $2000 and a lemon for $1000. Buyers are on the other hand willing to buy a plum for $2400 and a lemon for $1200. If symmetric information was present, the price on the market for any given plum would be between $2000 and $2400 and the price for any given lemon would be between $1000 and $1200. However, because of the information asymmetries regarding

quality the buyer cannot tell the difference between a plum and a lemon. Therefore, buyers are only willing to pay the expected value of any given car on the market. The expected value of a car on the market is calculated:

$$EV(car) = 0.5 * 2400 + 0.5 * 1200 = \$1800$$

At the price of $1800, the buyers will accept either a plum or a lemon. But in order for trade between buyers and sellers to be possible, sellers who own plums have to lower their prices from $2000 to $1800. This creates an incentive for sellers to choose the worse quality plums and the definition of the process is adverse selection, which is a consequence of asymmetric information (Varian, 2010).

When the next trade takes place the quality has fallen, due to adverse selection, and this causes the expected value to go down even more. This example allows us to understand that when sellers and buyers value a good differently, due to the imbalance of information available, the market will be in disequilibrium. This insight will be important for later when we apply our theoretical framework.

Having exemplified how asymmetric information affects markets, we now move on to discuss how to alleviate it. The natural solution to asymmetric information is to return the information state on the market to something that resembles information symmetry. Varian (2010) presents signaling as a way of alleviating asymmetric information by informing the part with an information disadvantage. The signal might be very costly (in terms of time and/or money) to acquire, but provides an assurance of quality for the other part. Akerlof (1970) provides an example of how signaling is used against asymmetric information and defines it as counteracting institutions. He mentions guarantees as a counteracting institution since they signal a certain quality for the buyer. The guarantee makes the seller obliged to compensate the buyer if the quality is not sufficient. This creates an incentive for the seller to sell high-quality cars since selling low-quality cars would be costly in the long run.

Having presented the theoretical background of the issue we will now turn to how it relates to Airbnb. As mentioned in the introduction, the buyers in an online market

only possess information about quality provided by the seller. The uncertainty increases because buyers and sellers rarely meet in person and play non-repeated games. This applies to Airbnb since the seller possesses more information about his or her accommodation than the buyer, and their real-life interactions are limited. The risk is therefore, in similarity with Akerlof's example, that the seller uses this imbalance to rent out low-quality accommodations, depicted as a high-quality accommodation in the ad. However, Airbnb mitigate this issue by encouraging renters to leave reviews about hosts and their accommodation (Airbnb, 2016). The review system works as a signal that provides information about quality to the buyers. For the sellers, signaling becomes a cost when not delivering the quality promised in the ad. This behavior will most likely result in bad reviews, which incur future losses as fewer people make reservations.

2.2 Applying theoretical framework

This thesis aims to test if asymmetric information is present in the Airbnb market and if a review system can alleviate the issue of asymmetric information. By applying our theoretical framework on our thesis statement we conduct a thought experiment. The theory tells us that a condition for a market to be in perfect equilibrium is that buyers and sellers have the same amount of information about the quality of the good. If this is true then the price on the market should reflect this information. However, when a market is in disequilibrium due to asymmetric information buyers and sellers value the good differently which leads to mispricing. This means that the market price does not reflect all information about the good.

From this reasoning, we test if asymmetric information is present in the market by providing more information about the good, in form of reviews, and then observe if this information has a significant impact on prices. If there is no asymmetric information in the market prices will not change because the market already incorporates all information about the good. But if we do have asymmetric information in the market the reviews will give the buyers better information about the quality of the good and the price will move closer to equilibrium. However, in order for this to work the information provided must be useful to the buyers.

This reasoning allows us to state the following hypotheses:

Hypothesis 1. The review system provides useful information for the buyers.

If this hypothesis holds true then we can predict that buyers will reward a higher rated listing with a higher price.

Hypothesis 2. Asymmetric information is present in the Airbnb market

If this hypothesis holds true we predict that more information, in form of reviews, changes the price towards equilibrium.

3. Data and Methodology

In order to test our hypotheses we use a large data set with listings from the Airbnb market. From this data set we obtain information about each listings price, review score and number of reviews. By using a multivariate regression model including these independent variables, with price as our dependent variable, we observe how the amount of information in the market, in form of reviews, affects prices. To get the "all else being equal effect" of reviews we include a number of control variables.

This section consists of two main parts, data and methodology. The data section examines our source material and evaluates our independent variables and control variables. The methodology section introduces our choice of econometrical model, considers potential econometrical problems with the model and finally introduces the specifications that allow us to test our hypotheses.

3.1 Data

Airbnb is an adequate representation of a market where uncertainty is present since it is highly based on trust between buyers and sellers. To examine Airbnb's online marketplace we use insideAirbnb, a website that provides information on Airbnb listings in larger cities around the world. The website provides cross-sectional data by collecting all listings on Airbnb at different dates in 30 different cities around the world. We use a data set with over 25 500 listings available in London on the Airbnb website the 2 September 2015 (Cox, 2016). In order to validate the data we looked at a number of listings from the insideAirbnb data and found that they were still available on Airbnb's website and that the information in the listings matched.

3.2 Variables

Table 1 provides a brief presentation of the review variables. We provide a more detailed description of each variable in order to get a better understanding of the review system.

Table 1: Review variables

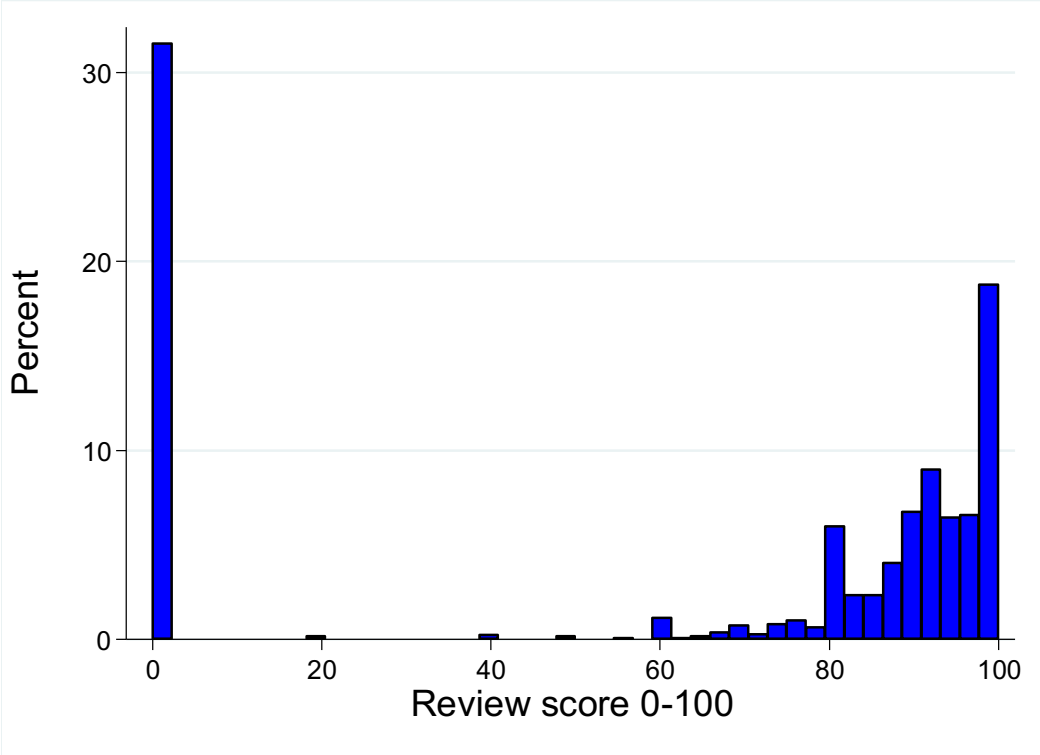
Variable	Description
Review scores rating	Review score between 0-100
Number of reviews	Number of reviews on the listing
Zero reviews	Dummy variable that accounts for users with zero number of reviews.

The review score rating is an overall score based on six criteria (Airbnb, 2016). The first variable is “accuracy” and we interpret this as how well the listing matches the actual experience. The second variable is “communication”, how fast and accurate the host has answered questions. The third variable is “cleanliness”. The fourth variable is “location”, proximity to attractive locations. The fifth variable is “check in”, rating how smooth the check in process was. The sixth and last variable is “value”, presenting if the price was representable of what the buyer received. The website then provides an overall review score in the form of stars, 0-5, on these criteria. InsideAirbnb transforms the stars data into a scale from 0-100. Table 2 shows a description of the variable Review score rating and we notice that the standard deviation is high. To investigate this we created a histogram examining the distribution of the variable (Figure 1)

Table 2: Review scores rating

Variable	Observations	Mean	Std. Dev.	Min. Value	Max. Value
Review scores rating	25361	61.91041	42.88494	0	100

Figure 1: Distribution of Review score



The distribution in Figure 1 explains the high standard deviation. Distributions of review scores are unevenly spread; there is a large group with a zero review score and a large group with a high review score 80-100. Resnick and Zeckhauser (2002) find similar patterns when looking into eBay's review score system. They observe that only 50% of the buyers leave a review after purchase and if they leave a review, it is very likely to be a positive one. Their data set shows that the feedback provided by buyers was 0.6 % negative, 0.3% neutral and 99.1 % positive.

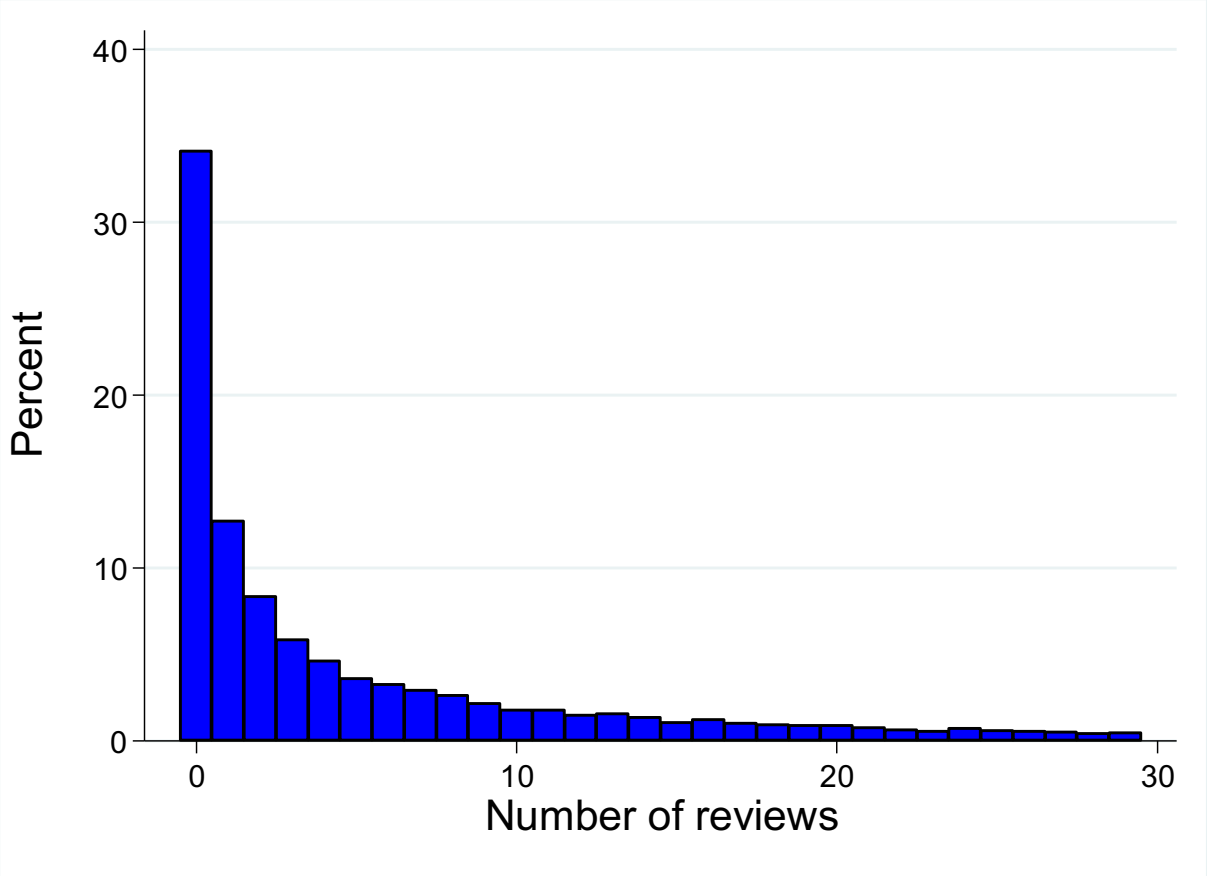
The number of reviews variable tells us the number of reviews of each individual listing. From Figure 2 we observe that the distribution is similar to the one of the review score variable with a large group with a zero number of reviews. The mean of eleven review scores in table 3 shows that the average seller on the market is rather experienced.

Table 3: Number of reviews

Variable	Observations	Mean	Std. Dev.	Min. Value	Max. Value

Number of reviews	25361	11.44549	22.50431	0	336

Figure 2: Distribution of Number of reviews



To account for the skewness in the review score and number of reviews variable we define a dummy variable, zero reviews, that takes a value of 0 if a user has zero number of reviews and 1 if the user has one review or more.

3.3 Control variables

A control variable is a variable held constant in order to evaluate the relationship between dependent and independent variables. By including control variables that affect prices, we receive a more accurate estimate of our review variables and at the same time alleviate the problem of omitted variable bias. Control variables' traits are in no correlation with the independent variables but possess considerable explanatory power for the dependent variable (Angrist and Pischke, 2008).

The question is which variables have a significant impact on prices and should be included in the regression model. To answer this question we use a hedonistic price model that differs from traditional consumer theory. The hedonistic price model accounts for the value by looking at different characteristics of the good. Different characteristics create different utility for each individual. This provides a more versatile valuation of the good, and is not limited to only physical properties (Lancaster, 1966). This model can be used when evaluating real estate. The characteristic then includes both internal and external factors. Internal factors include physical features such as size, quality and type of accommodation. External factors on the other hand are more abstract features, features which individuals perceive as positive. Examples of these features are location and closeness to transportation systems (Monson 2009).From this theoretical model, we conclude that the control variables in Table 4 are of interest. Table 4 presents a detailed description of each control variable.

Table 4: Control variables

Variable	Description
Distance from center	Distance from the center in London measured in km
Borough	A dummy variable that accounts for each of the 33 boroughs in London.
Room type	Type of room; entire home/apartment, private room or shared room
Property type	Type of property ranging from apartment to yurt
Accommodates	Number of people the listing accommodates

An important factor when it comes to choosing accommodation in a city is location. We assume that the average traveler prefers housing that is close to the city center. The listings do not provide this information so we had to create a new variable, Distance from center, by inserting each listing's coordinates and the coordinates of

the city center[1] into a distance formula in Excel. We also considered the preferences of consumers when it comes to which areas they wish to stay in London. The large variation in prices between neighborhoods in London shows the difference in preferences among consumers. In our data, this information is available so we chose to create a dummy-variable, Borough, to control for the borough[2] of each listing.

We know that different consumers have different preferences when it comes to housing options. The preferences we chose to include consist of type of room, property type and customers accommodated. Type of room itself consists of entire home, private room or shared housing. Property type consists of apartment, house, bed & breakfast, etc. "Accommodates" denotes number of people accommodated.

3.4 Choice of model

In order to look at the relationship between prices and reviews we use a multivariate regression model. By adding other factors than reviews into the model, we control for the affect these factors have on prices. By getting this "all else equal effect", we receive a better estimate on how review systems affect prices in the market (Wooldridge, 2002).

When using a linear regression model with OLS estimate of the parameters, a number of criteria need to hold. By meeting these five OLS criteria, the model meets the criteria for BLUE (best linear unbiased estimation). Acquiring BLUE helps our regression model to better estimate the relationship between reviews and prices in the market. If one of the OLS criteria does not hold, we can conclude that the estimators might be biased or inefficient (Wooldridge, 2002). We state the criteria below and evaluate how they connect to our model.

1. Linearity in parameters

   This implies that the regression model represents the studied population. The model has an intercept and needs to be linear in parameters.

$$y = a + \beta_1 x + \varepsilon$$

[1] Center is defined as location of Sharing Cross, London.
[2] London consists of 33 boroughs, see appendix.

Where a is the intercept and β_1 is the slope parameters of the population respectably. This criterion holds true in our model.

2. The expected value of the error term is zero, regardless of the value of the independent variable.

$$\varepsilon(\varepsilon X) = 0$$

This implies that the independent variables are uncorrelated with the error term. A correlation with the error term would be problematic since it suggests that there are variables that are not included in the model. In our regression model, this could be an issue since there are many factors that affect prices in a market and there is a risk that we have not accounted for these.

3. The variance of the error term is the same, regardless of the value of the independent variable.

$$Var(\varepsilon X) = \sigma^2$$

This implies that the variance of the residuals has to be constant. Otherwise, the model suffers from heteroscedasticity. If the variance changes in each observation, it will result in unreliable standard errors. We do not know whether our data suffers from heteroscedasticity. However, by adjusting for heteroscedastic consistent standard errors we can alleviate the problem.

4. The error terms in separate periods are uncorrelated.

$$Corr(\varepsilon_i \varepsilon_j) = 0 \text{ for all } \quad i \neq j.$$

Correlation between error terms ε_i and ε_j implies serial correlation. However, this will not be an issue in our model since we use cross section data and not time series data.

5. No perfect collinearity

This implies that there cannot be a perfect linear relationship between each independent variable. If perfect collinearity existed, it would reduce the prediction reliability of the individual parameters. This is not

an issue regarding our regression model since none of our independent variables are perfectly correlated.

6. The error terms are normally distributed.

   This regression requirement is only necessary for BLUE, normally distributed error terms are not required for an adequate OLS estimation of the parameters.

3.5 Econometric issues

We need to consider the second and third criterion in our regression. The second criterion considers heteroscedasticity and the third criterion considers the issue of omitted variable bias.

One assumption that needs to hold in order to get a good OLS estimation of the parameters is homoscedasticity. Homoscedasticity is a constant variance of the unobservable values in the error term, related to the independent variables (Wooldridge, 2002). However, homoscedasticity condition does not show whether our parameters of the dependent variables are biased. The homoscedasticity condition only states the efficiency of OLS estimation of the parameters. The opposite of homoscedasticity is heteroscedasticity, an inconsistent variance of the unobserved values in the error term.

Given heteroscedasticity, the OLS estimators of the parameters are inefficient due to an undervaluation of covariance and variance. A wrong estimation of covariance and variance result in wrong estimation of standard errors. The wrong estimation of standard errors will possibly lead to the wrong interpretation of the effect of given independent variables. For example, independent variables might seem significant when they in fact they are not (White, 1980). In our regression model, we consider this issue by adjusting for heteroscedasticity consistent standard errors, also known as Huber-White standard errors.

The second econometric issue we consider is misspecification of our regression model. In our model, we assume the presence of omitted variable bias since there are many factors that affect prices in the market that we do not have data on. When we state our multivariable regression model there is a risk that we have variables that explain our dependent variable, but that are not included in the model. We can

illustrate this with an example (Studendmund, 2013). If our true model is the following:

$$Y_i = a + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

ε_i is our error term. We have not included $\beta_2 X_2$ in our model because it is not in our data or we somehow missed it then our model we be the following:

$$Y_i = a + \beta_1 X_{1i} + \varepsilon_i^*$$

The $\beta_2 X_2$ will now be included in our error term ε_i instead:

$$\varepsilon_i * = \varepsilon_i + \beta_2 X_{2i}$$

This will result in our error term being dependent on our explanatory variables, unless they are fully uncorrelated with each other, which is very unusual. This violates our second assumption that the error term should be independent of our explanatory variable, $\varepsilon \perp X$. To mitigate this problem, we chose to include as many control variables in our model as possible.

3.6 Model specification

We are now ready to specify our regression models in order to test our predictions. By including different amounts of controls in our specifications, we can observe whether we have mitigated the problem of omitted variable bias or not. In the last model, we run a White-test in Stata in order to evaluate the potential problem for heteroscedasticity.

Model 1: We regress the price of listing *i*

$$P_i = \alpha + \beta_1 number\ of\ reviews_i + \beta_2 reviewscore_i + B_3 zeroreviews_i + \varepsilon_i$$

Model 2: We regress the price of listing *i* in borough *b* as

$$P_{i,b} = \alpha + \beta_1 number\ of\ reviews_i + B_2 reviewscore_i + B_3 zeroreviews_i + \beta_4 distance\ to\ center_i + \sum_b \beta_b borough_{ib} + \varepsilon_i$$

Where *b* indicates borough in question.

Model 3: We regress the price of listing *i* in borough *b,* type of property type *p*, room type t

$$P_{i,b} = \alpha + \beta_1 number\ of\ reviews_i + \beta_2 reviewscore_i + \beta_3 zeroreviews_i + \beta_4 distance\ to\ center_i + \beta_5 accommodates_i + \sum_b \beta_b borough_{ib} + \sum_p \beta_p propertytype_p + \sum_t \beta troomtype_t + \varepsilon_i$$

Where *p* indicates type of property type and *t* type of room

Model 4: We regress model 3 with the white test for heteroscedasticity.

4. Results

The results are presented in Table 5, it contains our four model specifications with the three independent review variables. Model 1 is without control variables, Model 2 includes the control variables distance from center and boroughs and Model 3 and 4 contains all the control variables. The results show that all beta-coefficients are significant with a p-value less than 0.01.

Table 5: Regression results

Variable	Model (1)	(2)	(3)	(4)
Number of reviews	-0.177 (6.24)**	-0.279 (10.24)**	-0.302 (14.64)**	-0.302 (15.42)**
Review score rating	0.124 (2.75)**	0.214 (5.00)**	0.256 (8.09)**	0.256 (8.37)**
Zero reviews	19.215 (4.61)**	28.707 (7.22)**	42.002 (14.30)**	42.002 (13.30)**
Constant	89.336 (22.20)**	148.929 (5.79)**	84.451 (4.46)**	84.451 (7.58)**
R^2	0.00	0.10	0.52	0.52
N	25,357	25,357	25,351	25,351

Notes: t-values in parentheses. *p<0.05; ** p<0.01,

Model 1 contains no control variables.

Model 2 includes the control variables distance to center and borough.

Model 3 includes the control variables distance to center, borough, room type, property type and accommodates.

Model 4 includes the same control variables as Model 3.

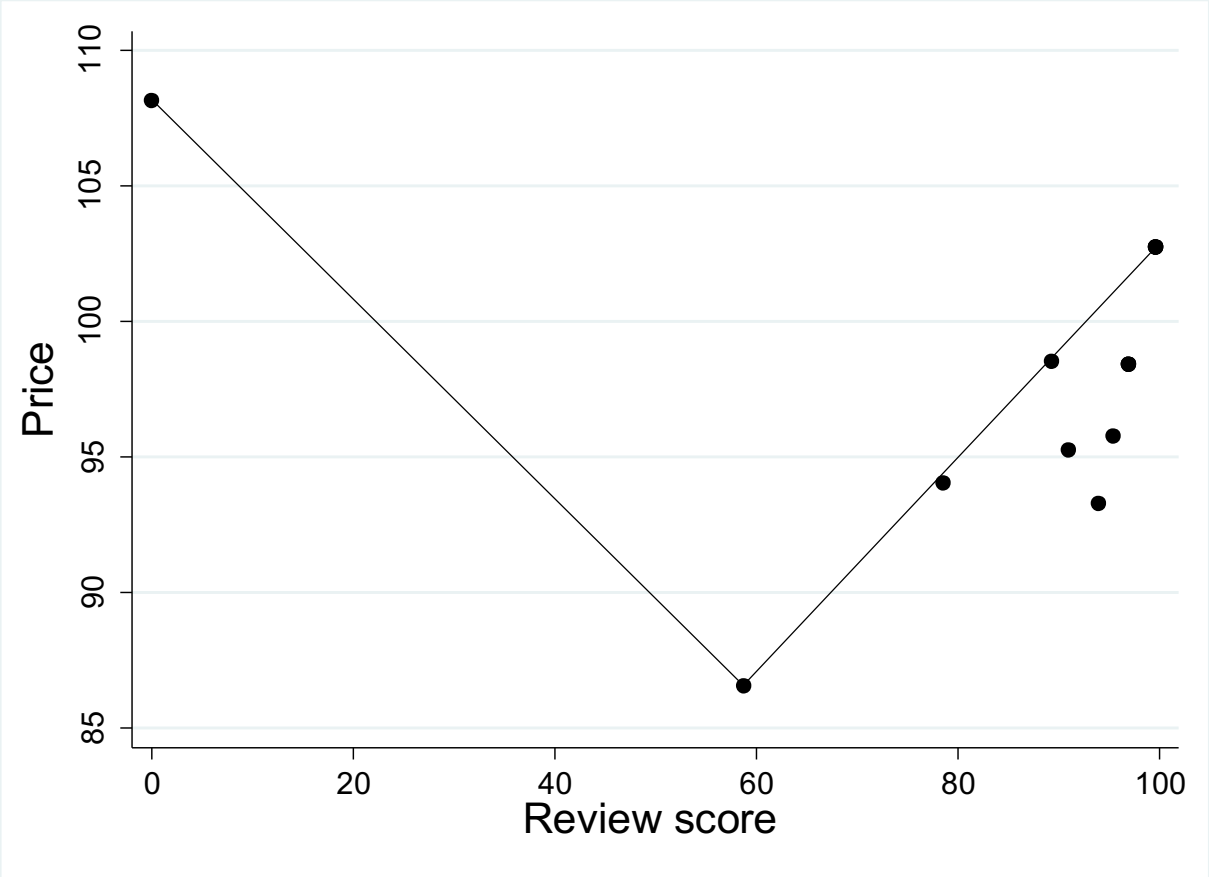
Hypothesis 1. The review system provides useful information for the buyers.

If the first hypothesis would hold true, we predicted that buyers would reward a higher rated listing with a higher price. The results show that the variable review score rating in model 3 has a positive marginal effect of +$0.256 for every increase in review score point. This result tells us that buyers are willing to pay a higher price for a higher review score, which shows that they value the information that reviews provide as useful. Consequently, we cannot reject our first hypothesis.

Figure 3 presents the relationship between price and review score, dividing seller into quantiles. Figure 3 displays that most sellers receive a review score between 80 and

100. A second observation displays that sellers with a zero review score set a price above $105 whereas sellers with a review score between 80 and 100 set a price around $97.

Figure 3: Relationship of Price ($) and Review score

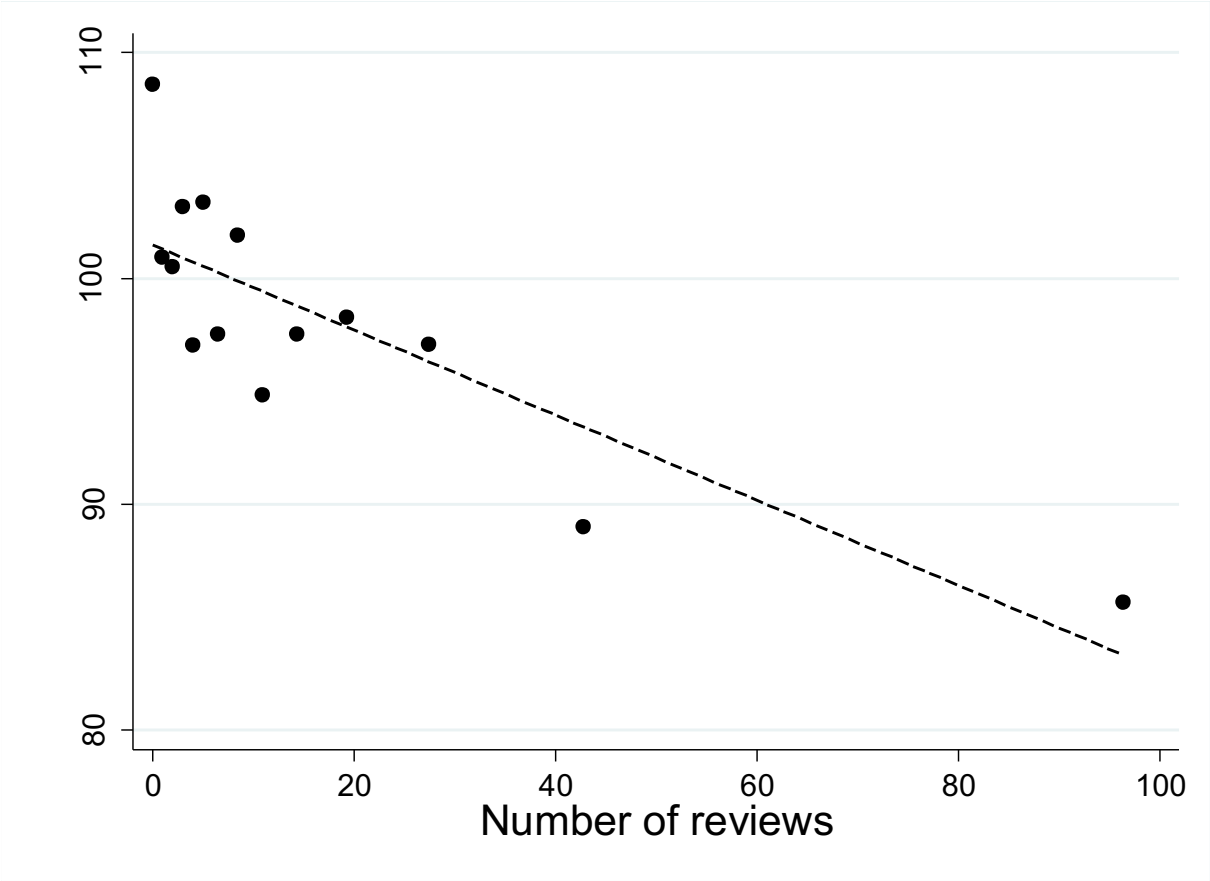


Hypothesis 2. Asymmetric information is present in the Airbnb market

The theoretical framework tells us that if we have asymmetric information it will cause the market to be in disequilibrium. Because of this, the price does not reflect the true quality of the good due to mispricing. By increasing the amount of useful information in the market, in form of reviews, we predicted that prices will better reflect the quality of the good and would therefore move closer to equilibrium. From model 3 in table 5 we observe a significant negative marginal effect of -$0.302 on price with each additional review. This shows that the information provided has a significant impact on prices. Consequently, we cannot reject our second hypothesis; asymmetric information is present in the market.

Figure 4 presents the relationship between prices and number of reviews, dividing sellers into quantiles. In Figure 4 a fitted line displays a negative relationship between prices and number of reviews. It shows that more reviews results in a lower price. A second observation displays that sellers without reviews set a price above $105.

Figure 4: Relationship between Price ($) and Number of reviews



4.1 Limitations

A fourth model complements the third model specification in table 5 in order to test for potential heteroscedasticity. By adding the robust command after the regression, Stata considers heteroscedasticity consistent standard errors (Angrist and Pischke, 2008). This eliminates incorrect interpretations while testing hypotheses (White, 1980).Table 5 visualizes the result of the new specification.

The difference between model 3 and model 4 is different t-values, due to the consideration of potential heteroscedasticity in model 4. For example, the t-value of number of reviews is now 15.42 instead of 14.64. Note that the table shows no difference in significance level between the model 3 and model 4.

To address the problem of omitted variable bias we complement our regression model with control variables to achieve a more accurate estimation of our review variables. Studendmund (2013) states four types of specification criteria used to evaluate if control variables have generated a better regression. The first criterion considers the theoretical perspective and questions whether it is logical to include the control variable. The second criterion considers the significance of the explanatory variables effect, examining whether it remains significant after the control variables are included. The third criterion considers R^2 and whether R^2 adjusted for degrees of freedom improves after adding the control variables. The fourth and final criterion is bias. It considers whether the explanatory variables' coefficients change significantly after adding the control variables.

As stated in the method section, the risk of spurious results due to omitted variable bias is present in the model since we assume that there are many other factors affecting prices of listings. This implies that the first criterion holds, it is logical to include control variables since some explanatory variables are omitted. By observing our regression results in Table 5 we can also conclude that our second criterion holds, our t-test shows that all coefficients are still significant when we add more control variables. The third criterion holds as well when looking at Table 5, our adjusted R^2 has increased with each specification. The final criterion used to evaluate if the control variables have generated a better result hold as well. We observe changes in the coefficients when including more control variables. Investigating each specification criteria for a better regression indicates that adding control variables has provided a less spurious result.

5. Discussion

The results stated that buyers value the information in form of reviews as useful. The marginal effect of +$0.256 per increase in review score confirms our first hypothesis. At first, it appears that this effect is rather low. However, it is important to consider that the variable is based on a scale from 0 to 100 review points. These review points are then transformed which to a 5-star score system on Airbnb's online market. A one-star increase in review rating is equivalent to a 20 points increase, which would generate a price increase of $5.12 in total. The results confirm our previous theoretical discussion that a review system is a form of signaling that assures quality

for the buyers. Our results are in line with those of Lewis (1987) who studied the second hand car market on eBay and found that by including more pictures and disclosing personal information in the listing, the buyers were willing to pay a higher price. Lewis (1987) argues that this is a method for mitigating the problem with asymmetric information in online markets.

Our second hypothesis stated that asymmetric information is present in the Airbnb market and the reasoning behind was that if a market possesses information asymmetries it is in disequilibrium. By adding useful information to the buyers, there will be a significant effect on prices as the market moves closer to equilibrium. This appears to hold true when examining our variable number of reviews, which has a negative marginal effect of -$0.302 per additional review.

By observing Figure 3 and Figure 4 it becomes clear that sellers with a zero in review score and a zero in number of reviews tend to set a price that is significantly higher than other sellers[3]. Our independent variable Zero reviews in shows the overpricing effect in table 5, which in model 3 has a marginal effect of +$42. From this, we conclude that new entry sellers tend to overprice in comparison with experienced sellers. The reason for this behavior is ambiguous but experiments conducted show that when individuals are given a new asset to sell, they tend to overprice that asset (Kahneman, D., Knetsch, J. L., & Thaler, R H., 1980). Economic theory defines this behavior as the endowment effect. If new sellers overprice, it seems natural that the marginal effect of number of reviews is negative since the equilibrium price would be lower than the overprice set by the new entry sellers. However, this is just our interpretation, something our result does not confirm.

Given the results, it is important to be critical to our model as well as our methods. We will therefore discuss potential issues and limitations in both our theoretical framework and with the data. It could be questioned how well a review system is in fact a solution for decreasing asymmetric information. Observing the distribution of review scores in Figure 1 we see that there is a skewness in review scores. If the review scores were an adequate representation of quality, it would state that almost all listings are of excellent quality, which is between 80 and 100 in review score. This is however very unlikely. Resnick and Zeckhauser (2002) observes that up to 50 % of

[3] By using an if-function in Excel we can conclude that sellers with zero review score also has a zero in number of reviews.

eBay users do not leave reviews at all. We suspect that many users do not leave reviews on Airbnb either. These circumstances indicate that review systems are limited when it comes to an accurate representation of quality. If reviews were a better representation of quality, we assume that our results would show a smaller change in prices as the number of reviews increases.

In the section where we apply our theory, we stated that price is a representation of quality and we further stated that in equilibrium, price carries all the information about the good. This is problematic because prices on differentiated goods, such as real estate, are determined by the utility different factors create for each individual. It is therefore unlikely that all buyers will value the information given in a similar way. However, we assume most individuals prefer similar characteristics such as cleanliness, location and closeness to transportation systems.

It could also be questioned if the data is optimal for the purpose of the thesis. Cross sectional data collects all information from a specific time and from that data we build our case. However, if we had access to Panel data we could obtain a better measurement of how each single review affects prices in a certain time period. From this information, we would see if the price change is greater for the first number of reviews in comparison with the hundredth number of review. To be able to observe this over a time period would have strengthened our results.

It would also be useful to have data on booked prices instead of prices provided by the sellers. However, since our sample is relatively large, over 25 500 listings, we assume that the prices are an adequate representation of the booked prices on the market. If we had data available on the number of completed bookings for each listing it would have helped us to further validate our results. By using this data, we could have controlled for the risk of renters not leaving a review and thus get a more robust result. It would also be useful to have data from other online markets to further validate our results.

6. Conclusion

This thesis sought to study empirically the presence of asymmetric online markets and if review systems can alleviate this issue. We establish a theoretical framework

in order to understand the concept of asymmetric information and its implications. Applying this framework on the Airbnb market lead us to two hypotheses: asymmetric information is present in the Airbnb market and the review system can alleviate this issue by providing useful information for buyers.

By using a multivariate regression model on cross sectional data from the Airbnb market, our results show that our two hypotheses cannot be rejected. The results indicate that asymmetric information exists in the market and that it can be alleviated with a review system. In our discussion, we mention limitations regarding our method and our results. For instance, our data shows that a vast majority of the reviews tend to be overly positive, which indicates that it is not an optimal reflection of quality and thereby not optimal for alleviating asymmetric information. For future research it would be interesting to conduct a similar study with the help of time series data and panel data. This would allow for a more robust examination of how a review system affects the market.

References

Airbnb. (2016) *How do review works?* [online]. Airbnb, Inc. Available at: https://www.airbnb.se/help/article/13/how-do-reviews-work

Akerlof, G A. (1970) The market for “lemons”: Quality uncertainty and the market mechanism. *The quarterly journal of economics*, pp.488-500.

Angrist, J. D., & Pischke, J. S. (2008) *Mostly Harmless Economics: An Empiricist’s Companions*, (1st edition). Princeton University Press, New Jersey, pp. 18-36.

Cowen, T and Tabarrock, A. (2011) *Modern Principles of Economics*, (2nd edition). Worth Publishers, London, pp.456-460.

Cox, B. (2016) *Get the data* [online] . Insideairbnb. Availble at: http://www.insideairbnb.com/get-the-data.html

Kahneman, D., Knetsch, J. L., & Thaler, R H. (1980) Experimental tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy,* pp. 1325-1348.

Lancaster, K. J. (1966) A new approach to consumer theory. *The journal of political economy*, pp.132-157.

Lewis, G. (1987) Asymmetric Information, Adverse Selection, and Online Disclosure: The Case of eBay Motors. *The American Economic Review,* pp.1535-1546.

Monson, M (2009) Valuation Using Hedonic Pricing Models. *Cornell Real Estate Review* 7.1: 10.

Resnick, P., & Zeckhauser, R. (2002) Trust among strangers in internet transactions: Empirical analysis of eBay‘s reputation system. *The economics of the internet and E-commerce, 11*(2), pp.23-25.

Studenmund A.H (2013) *Using Econometrics: Pearson New international Edition*, (6th edition). Pearson, pp. 156-197.

Varian, H. R. (2010) *Intermediate Microeconomics: A modern approach*, (8th edition). W.W. Norton & Company Inc, New York, pp.719-738.

White, H (1980) A heteroscedasticity- consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica,* pp. 817-838.

Wooldridge, J. M. (2002) *Introductory Econometrics: A Modern Approach*, (2nd edition). South-Western, New Hampshire, pp. 45-353.

Zervas, G., Proserpio, D. & and Byers, J. (2014) The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Boston U. School of Management Research Paper,* (2013-16).

Appendix

List of London's 33 boroughs organized in alphabetical order

Boroughs	Included in sample	Location in London
Barking and Dagenham	X	Outer
Barnet	X	Outer
Bexley	X	Outer
Brent	X	Outer
Bromley	X	Outer
Camden	X	Inner
City of London	X	Inner
Croydon	X	Outer
Ealing	X	Outer
Enfield	X	Outer
Greenwich	X	Inner
Hackney	X	Inner
Hammersmith and Fulham	X	Inner
Haringey	X	Outer
Harrow	X	Outer
Havering	X	Outer
Hillingdon	X	Outer
Hounslow	X	Outer
Islington	X	Inner
Kensington and Chelsea	x	Inner
Kingston upon Thames	X	Outer
Lambeth	X	Inner
Lewisham	X	Inner
Merton	X	Outer
Newham	X	Outer
Redbridge	X	Outer
Richmond upon Thames	X	Outer
Southwark	X	Inner
Sutton	X	Outer
Tower Hamlets	X	Inner
Waltham Forest	X	Outer
Wandsworth	X	Inner
Westminister	X	Inner

Type of rooms available at Airbnb

Room type	Included in sample
Entire home	X
Private room	X
Shared room	X

Type of property available at Airbnb

Property type	Included in sample
Apartment	X
Bed & Breakfast	X
Boat	X
Bungalow	X
Cabin	X
Camper/RV	X
Chalet	X
Condominium	X
Dorm	X
House	X
Island	X
Loft	X
Other	X
Parking Space	X
Tent	X
Treehouse	X
Villa	X
Yurt	X

Linnear regression results with control variables

```
Linear regression                               Number of obs     =     25,351
                                                F(56, 25290)      =          .
                                                Prob > F          =          .
                                                R-squared         =     0.5193
                                                Root MSE          =     65.698
```

Price	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
number_of_reviews	-.3017962	.0195723	-15.42	0.000	-.340159	-.2634333
review_scores_rating	.2558082	.0305601	8.37	0.000	.1959085	.3157078
zeroreviews	42.00172	3.159114	13.30	0.000	35.80967	48.19376
1.superhost_dummy	12.71179	1.326631	9.58	0.000	10.11151	15.31206
Distancefromcenterinkm	-6.284794	.2725953	-23.06	0.000	-6.819096	-5.750491
daysfromtoday	.0103595	.0012629	8.20	0.000	.0078841	.012835
accommodates	21.51523	1.146474	18.77	0.000	19.26808	23.76239
morelistingcounts	1.591859	.8084298	1.97	0.049	.0072903	3.176429
neighbourhood						
Barnet	-16.32607	10.5832	-1.54	0.123	-37.06975	4.417606
Bexley	35.24283	13.37942	2.63	0.008	9.018402	61.46726
Brent	-39.10161	9.961241	-3.93	0.000	-58.62622	-19.577
Bromley	-16.58339	10.56287	-1.57	0.116	-37.28723	4.120438
Camden	-34.13971	10.10144	-3.38	0.001	-53.93912	-14.3403
City of London	-21.54916	10.7299	-2.01	0.045	-42.58039	-.5179347
Croydon	-30.21514	10.92063	-2.77	0.006	-51.6202	-8.810074
Ealing	-19.39771	9.819353	-1.98	0.048	-38.64421	-.151214
Enfield	-1.972776	10.50029	-0.19	0.851	-22.55396	18.60841
Greenwich	-22.98308	9.956221	-2.31	0.021	-42.49785	-3.468315
Hackney	-48.52567	9.853212	-4.92	0.000	-67.83853	-29.2128
Hammersmith and Fulham	-30.82751	9.845092	-3.13	0.002	-50.12446	-11.53056
Haringey	-38.11011	9.976779	-3.82	0.000	-57.66518	-18.55505
Harrow	6.520464	11.25812	0.58	0.562	-15.5461	28.58703
Havering	62.77251	10.57845	5.93	0.000	42.03813	83.50689
Hillingdon	43.95462	11.73933	3.74	0.000	20.94486	66.96437
Hounslow	6.519761	11.10806	0.59	0.557	-15.25268	28.29221
Islington	-43.7958	9.920996	-4.41	0.000	-63.24153	-24.35008
Kensington and Chelsea	-.997771	10.0376	-0.10	0.921	-20.67205	18.67651
Kingston upon Thames	10.57995	10.0873	1.05	0.294	-9.191747	30.35164
Lambeth	-50.6483	9.919082	-5.11	0.000	-70.09028	-31.20633
Lewisham	-45.82899	10.11401	-4.53	0.000	-65.65304	-26.00495
Merton	-.8082348	9.946084	-0.08	0.935	-20.30314	18.68667
Newham	-30.21025	9.875026	-3.06	0.002	-49.56587	-10.85463
Redbridge	-10.21484	11.32596	-0.90	0.367	-32.41437	11.98468
Richmond upon Thames	27.45087	10.41756	2.64	0.008	7.031847	47.86989
Southwark	-48.63403	10.07961	-4.82	0.000	-68.39065	-28.87742
Sutton	-1.397535	12.69014	-0.11	0.912	-26.27093	23.47587
Tower Hamlets	-45.49386	9.835114	-4.63	0.000	-64.77125	-26.21646
Waltham Forest	-35.33549	10.12555	-3.49	0.000	-55.18215	-15.48884
Wandsworth	-34.57864	9.828957	-3.52	0.000	-53.84397	-15.31332
Westminster	-19.29806	9.982715	-1.93	0.053	-38.86475	.2686422

propertytype						
Bed & Breakfast	2.985938	2.045902	1.46	0.144	-1.024148	6.996023
Boat	22.22969	10.9212	2.04	0.042	.8234985	43.63587
Bungalow	5.38898	14.27803	0.38	0.706	-22.59679	33.37475
Cabin	10.64316	8.770451	1.21	0.225	-6.547432	27.83375
Camper/RV	1.445284	11.23608	0.13	0.898	-20.57809	23.46866
Chalet	-1.715857	3.80523	-0.45	0.652	-9.174327	5.742613
Condominium	-5.577532	7.800658	-0.72	0.475	-20.86727	9.712207
Dorm	-29.288	7.648892	-3.83	0.000	-44.28027	-14.29573
House	15.57978	1.310407	11.89	0.000	13.01131	18.14826
Island	-33.19502	51.86996	-0.64	0.522	-134.8631	68.4731
Loft	12.0041	5.090879	2.36	0.018	2.025684	21.98252
Other	60.91343	62.9196	0.97	0.333	-62.41262	184.2395
Parking Space	36.78527	1.979761	18.58	0.000	32.90482	40.66571
Tent	-45.31069	5.027266	-9.01	0.000	-55.16442	-35.45696
Townhouse	13.07504	11.12816	1.17	0.240	-8.736804	34.88688
Treehouse	4953.138	1.121336	4417.18	0.000	4950.94	4955.336
Villa	-3.165915	19.68356	-0.16	0.872	-41.74683	35.415
Yurt	176.4065	2.661955	66.27	0.000	171.1889	181.6241
roomtype						
Private room	-36.96014	2.071533	-17.84	0.000	-41.02046	-32.89981
Shared room	-62.60334	3.222914	-19.42	0.000	-68.92044	-56.28624
_cons	84.45136	11.1403	7.58	0.000	62.61573	106.287