THE PREDICTION OF OCCUPANCY RATES BASED ON THE KNOWN PARAMETERS FOR DIFFERENT TYPES OF HOUSING

Vladimír Krajčík – Elena Říhová

Abstract

Rapid advances in technology have enabled several service-industry companies to develop innovative business ways. For instance, one of the peer-to-peer market is AirBnB – the largest hotel chain serves in the world. In this paper we try to explore the Airbnb phenomenon and try to predict the occupancy rate with the help of known parameters. The data set is collected with the online serves help, provided by the company Airdna for a period between April 2016 and March 2017. The main aim of current research is predict the future occupancy rate for the most popular type of housing. Based on the main aim is necessary to build the model with the help of LASSO regression, and with the model help be possible to predict the occupation rate. To fulfil the aim, the correlation analysis for the current data set is provided. We discuss these findings as well as implications for practice and policy and offer suggestions for future research.

Key words: sharing economy, AirBnB, regression, LASSO regression

JEL Code: C25, D40

Introduction

Sharing economy is a modern phenomenon, which is rapidly evolving due to the development of information technology and internet. A lot of researches about peer-to-peer markets or sharing economy are published. The most significant studies are made by Rodas & Gosling, Horn, & Merante, Yrigoy, Wegmann and Jiao, Russell. Most of researches stayed, that peer-to-peer economy affects local business negatively, but Horn in his work 'Is home sharing driving up rents? Evidence from Airbnb in Boston' argues, that the impact of sharing economy on the local business is towards to zero. So, this question is under discussion.

This paper aims to provide some empirical research of the structure the Airbnb and try to predict the occupation rate, based on known parameters. And current research is the first step to understand, if sharing economy can impact on local business.

The data set includes the web scraped data on AirBnB in Prague, provided by the company Airdna for a period between April 2016 and March 2017. The data set includes huge number of variables, and for current research the more significant variables (by LASSO regression method) are selected. The size of the data set, including the information about 13918 accommodation units, as like apartments, lofts, bed and breakfast, houses, and etc. To fulfill the goal of research was selected accommodation type – apartments. The main reason is, that this type of accommodation has the most number of units – 12441.

1 Previous research

The sharing economy platforms, as Airbnb or Uber, at large are surveyed in books and papers, including Coyle (Coyle, 2016), Edelman and Gerardin (Edelman and Gerardin, 2015) and Sundararajan (Sundararajan, 2016), Russell (Russell, 2014), and others. The most researches try to understand the causes of rapidly grow of those type of economy, others try to detect, if peer-to-peer economy really affects traditional economy, and the last group of researches try to describe how does peer-to-peer market work, what is the main forces of this market.

Nevertheless, some main finds are discovered. The first one is a fact, which is described by Boswijk in 2016 (Boswijk, 2016) in his work Airbnb: The future of networked hospitality businesses. The successful of Airbnb platform is due to the fact that this concept help to build a positive reputation, like of guests, as of hosts. In this way, it leads to essential trust, and it this turn, it help to build good-working model. The other one find is made by Einav from Stanford University in 2016 (Einav, 2016) in her paper ‚Peer-to-peer market‘. The main statement of this research is there are two kind of players: traditional providers (like hotels), and smaller providers (like Airbnb, in our case), and as Eiven writes: ‘With a peer-to-peer marketplace that charges a break-even transaction fee, the cost of the platform will be split across participants‘. (Eiven, 2016). Or another words, flexible sellers have the opportunity to increased or decreased profits by the entry of additional flexible sellers. (Eiven, 2016) And the last one find is about impact of sharing economy to traditional. And here is two different opinion. Horn and Merante in their work ‘Is home sharing driving up rents? Evidence from Airbnb in Boston’ argue, that the impact of sharing economy on the local business is towards to zero. (Horn and Merant, 2017) So, this question is under discussion.

And Rodas and Gosling prove in their work Sharing Economy In Tourism: A Theoretical Discussion About Collaborative Consumption And Sharing‘‘ from 2017, that there is a destructive influence of peer-to-peer market on traditional economy. (Rodas and Gosling, 2017)

2 Methodology. Least absolute shrinkage and selection operator (LASSO Regression Method)

Regression is a widely accepted statistical method for modelling relations between variables. Linear regression is the most popular and well-studied model in statistics. It can be written in the following form:

$$y = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \varepsilon , \tag{1}$$

where $\left(x_1, x_2, \ldots, x_p\right)$ is an input vector, y is a real-valued output that one wants to predict, $\left(\beta_0, \beta_1, \ldots, \beta_p\right)$ is a coefficient vector and ε is an error. This model is based on the following assumptions:

— the target function is linear w.r.t. the input variables $\left(x_1, x_2, \ldots, x_p\right)$; (a)

— the variables $\left(x_1, x_2, \ldots, x_p\right)$ have the multivariate normal distribution; (b)

— the variables $\left(x_1, x_2, \ldots, x_p\right)$ are not auto-correlated; (c)

— there is no (or little) multicollinearity between $x_j, j = 1, \ldots, p$; (d)

— the variables have the same finite variance (homoscedasticity). (e)

However, even more general types of dependencies can be expressed as a linear regression model. For example, a new input variable x_j can be transformed to approximate polynomial dependencies, e.g., $x_j^{new} = x_j^k$, or others ones as $x_j^{new} = \log\left(x_j\right)$, $x_j^{new} = \sqrt{x_j}$, etc. The interaction between variables can be expressed by an additional new variable as well, e.g., $x^{new} = x_i \cdot x_j$ reflects the dependency between x_i and x_j. Nominal variables can be added to a model by means of ‘‘dummy’’ coding, e.g. a variable g with m qualitative variables can be taken into account in a model as m variables $x_r^{new} = I(x = r)$, where r is a value of g, I is an indicator function.

Building the regression can be described as follows. Having a training set $\left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(n)}, y^{(n)}\right)$, where $x^{(k)}, k = 1, \ldots, n$ satisfies conditions (a)-(e) one aimes to estimate parameters $\left(\beta_0, \beta_1, \ldots, \beta_p\right)$ such that the model (1) describes data as precise as possible. The most popular method is least squares:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y^{(i)} - \sum_{j=1}^{m} \beta_j x_j^{(i)} \right)^2, \qquad (2)$$

In practice, there are a lot of variables that can be used in Formula (1). Building a model based on the whole set of available variables x_i make the coefficients β_i poorly determined and the final model less interpretable. The first issue is usually caused by correlated variables involved in a linear regression model. Widely large coefficients indicates this kind of issues. It is tackled by imposing additional constraints on the coefficients, i.e. by shrinkage methods. The simplest shrinkage methods for least squares optimization of a linear regression in the general form can be expressed as follows:

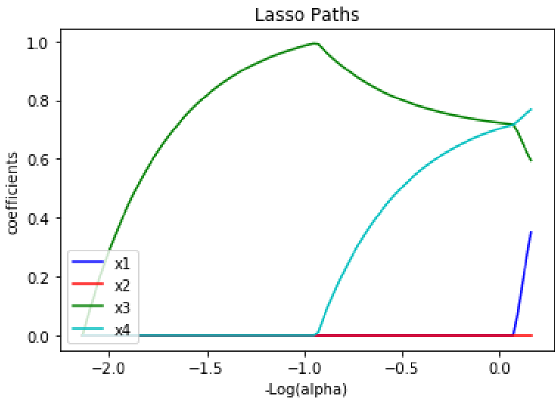
$$\hat{\beta} = \arg\min \left(\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m} \beta_j x_{ij} \right)^2 + \alpha \left|\beta\right|^q \right), \qquad (3)$$

where α is a complexity parameter that controls the amount of shrinkage, q = 2 for ridge regression and q = 1 for lasso regression.

In our experiments we use lasso regression since it allow for both shrinkage and variable selection. Below, we consider a tiny example to illustrate the shrinkage principle of the lasso regression.

Example. Let us consider a linear model $y = 2x_1 + 3x_2$. We also add additional variables that are linearly dependent on x_1 and x_2, i.e. $x_3 = 2x_1 + 5$ and $x_4 = 3x_2$. Figure 1 shows how the values of β_i shrinks with different values of α. It is clear to see that even small values of α allow us to get rid of linearly dependent variables (red line is zeros for small values of α, i.e. for the region where $-\log(\alpha)$ is close to 0), whereas coefficients β_1 and β_3 are shrinked for higher values of α, that indicates a not exact linear dependence between x_1 and x_3.

Fig. 1: The dependence the values of coefficients β_i on values of the regularization constant α



Source:Author

3 Empirical Results and Discussion

For predictive model are included the following variables: Average Daily Rate, Annual Revenue LTM, Number of Bookings LTM, Number of Reviews, Overall Rating, Bedrooms, Bathrooms, Max Guests, Response Rate, Response Time (min), Security Deposit, Cleaning Fee, Extra People Fee, Published Nightly Rate, Published Monthly Rate, Published Weekly Rate, Minimum Stay, Count Reservation Days, Count Available Days, Count Blocked Days, Number of Photos, Latitude, and Longitude.

Based on above-mentioned theoretical aspects the LASSO regression is done. LASSO regression method is shrink toward zero non-informative variables. In Table 1 is shown the results of LASSO regression.

Tab. 1: The results of LASSO regression

Apartments	
Variables	Value
Average Daily Rate	-0.0005
Annual Revenue LTM	0.0000
Number of Bookings LTM	-0.0000
Number of Reviews	0.0002
Overall Rating	0.0006
Bedrooms	0.0029
Bathrooms	0.0006
Max Guests	0.0039
Response Rate	-0.0004
Response Time (min)	0.0000
Security Deposit	-0.0000
Cleaning Fee	0.0000
Extra People Fee	-0.0007

Published Nightly Rate	0.0000
Published Monthly Rate	0.0000
Published Weekly Rate	0.0000
Minimum Stay	0.0019
Count Reservation Days LTM	0.0020
Count Available Days LTM	-0.0019
Count Blocked Days LTM	0.0008
Number of Photos	0.0000
Latitude	-0.3264
Longitude	0.0000
Alfa	16.8735
R squared	0.7811
adjusted_R_squared	0.7772

Source: author

As is described above, LASSO regression tends to assign zero weights to most irrelevant or redundant variables, and hence is a promising technique for variables selection. (Reid, 2016) To find the predictive model is made the following table with deleted zero variables.

Tab. 2: The final results of LASSO regression

Apartments		
X_i	Variables	Value
X_1	Average Daily Rate	-0.0005
X_2	Number of Reviews	0.0002
X_3	Overall Rating	0.0006
X_4	Bedrooms	0.0029
X_5	Bathrooms	0.0006
X_6	Max Guests	0.0039
X_7	Response Rate	-0.0004
X_8	Extra People Fee	-0.0007
X_9	Minimum Stay	0.0019
X_{10}	Count Reservation Days LTM	0.0020
X_{11}	Count Blocked Days LTM	0.0008
X_{12}	Latitude	-0.3264
	Alfa	16.8735
	R squared	0.7811
	Adjusted R_squared	0.7772

Source: author

Thus, our estimated regression line (predicted model for the occupancy rate) can be write as following:

$$\begin{aligned}&16.8735-0.0005x_1+0.0002x_2+0.0006x_3+0.0029x_4+0.0006x_5+0.0039x_6-\\&-0.0004x_7-0.0007x_8+0.0019x_9+0.002x_{10}+0.0008x_{11}-0.3264x_{12}\end{aligned} \tag{10}$$

The value of α = 18.8735 gives the value of $\hat{y}$ for x = 0; that is, it gives the monthly occupancy rate for x=0. Values of b_i give the change in $\hat{y}$ due to a change of one unit in x_i.

Thus, $b_1 = -0.0005$ indicates that, on average, for every extra euro of average daily rate, the monthly occupation rate decreases by 0.0005. Note that when *b* is negative, *y* decreases as *x* increases. With $b_2 = +0.002$, when number of review is growing, increases by one review, the occupancy rate is growing by 0.002. With $b_3 =0.006$ indicates that, on average, for every extra point of rating, the monthly occupation rate increases by 0.006. With $b_4 =0.0029$ indicates that, on average, for every extra bedroom, the monthly occupation rate increases by 0.0029. And other variables can be described the similar way.

The value of $r = 0.7811$ indicates that the (linear) relationship is strong but not very strong. The value of $r^2 = 0.7772$ states that 77.72% of the total variation in occupancy rate is explained by variables and 22.28% is not. The high value of r^2 indicates that there many important variables that contribute to the determination of occupancy rate.

Conclusion

Over the last few years, Airbnb has shaken up the lodging industry. While some might classify these as "disruptive innovations," the authors (Christensen et al., 2015) one is clear: that innovative ideas like Airbnb have the potential to change the very way any industry operates, and the success of Airbnb confirms that once the change is initiated, it is highly unlikely that the industry would revert to the old model. (Varma et al.,2016) Thus, it is necessary to better understand the factors that guide to estimate or predict occupancy rate and understand how much those factors influence the occupancy rate.

The factor, which maximum positive affects the occupancy rate is „Max Guests“ with the *b* value 0.0039, and opposite, minimum positive affects the occupancy rate is „Number of Reviews“ with the *b* value 0.0002. The location of the object is described here by only „Latitude“(„Longitude“ is almost the same). The value of *b* is denied because the zero point is the location of the objects in city center. Our study offers important insight into the Airbnb phenomenon and customer with predicting the occupancy rate of Airbnb apartment. Moreover, the value of $r = 0.7811$ indicates that the (linear) relationship is strong, and along with that the value of $r^2 = 0.7772$, it means, that 77.72% of the total variation in occupancy rate is explained by variables and 22.28% is not. The high value of r^2 indicates that there many important variables that contribute to the determination of occupancy rate. Those variables are the following: Average Daily Rate, Number of Reviews, Overall Rating, Bedrooms, Bathrooms, Max Guests, Response Rate, Extra People Fee, Minimum Stay, Count Reservation Days LTM, Count Blocked Days LTM, and Latitude.

Acknowledgment

The paper was made supported by the Grant Agency of the Academic Alliance, under grant GA /13/2018 „Analysis of the capacity of accommodation facilities in selected European regions“ and performed at University College of Business in Prague.

References

Coyle, Yeung (2016). Understanding AirBnB in Fourteen European cities, *Policy Papers*.

Einav, L., Farronato, C., & Levin, J. (2015). Peer-to-Peer Markets. doi:10.3386/w21496

Edelman, B. G., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *SSRN Electronic Journal*. doi:10.2139/ssrn.2377353

Edelman, B. G., & Geradin, D. (2015). Efficiencies and Regulatory Shortcuts: How Should We Regulate Companies like Airbnb and Uber*? SSRN Electronic Journal*. doi:10.2139/ssrn.2658603

Horn, K., & Merante, M. (2017). Is home sharing driving up rents? Evidence from Airbnb in Boston. *Journal of Housing Economics, 38,* 14-24. doi:10.1016/j.jhe.2017.08.002

Reid, S., Tibshirani, R., & Friedman, J. (2016). A study of error variance estimation in Lasso regression. *Statistica Sinica*. doi:10.5705/ss.2014.042

Rodas V., L. A., & Gosling, M. D. S. (2017). Sharing Economy In Tourism: A Theoretical Discussion About Collaborative Consumption And Sharing. *Revista Eletronica De Estrategia E Negocios-Reen*, *10*(*1*), 226-251.

Seal, H. L. (1967). Studies in the History of Probability and Statistics. XV The historical development of the Gauss linear model. *Biometrika,54*(1-2), 1-24. doi:10.1093/biomet/54.1-2.1

Sundararajan, A. (2016). The Sharing Economy, *MIT Press*.

Varma, A., Jukic, N., Pestek, A., Shultz, C. J., & Nestorov, S. (2016). Airbnb: Exciting innovation or passing fad? *Tourism Management Perspectives,20,* 228-237. doi:10.1016/j.tmp.2016.09.002

Wegmann, J., & Jiao, J. (2017). Taming Airbnb: Toward guiding principles for local regulation of urban vacation rentals based on empirical results from five US cities. *Land Use Policy*, *69*, 494-501. doi:10.1016/j.landusepol.2017.09.025

Yrigoy, I. (2017). Airbnb in Menorca: a new form of touristic gentrification? Distribution of touristic housing dwelling, agents and impacts on the residential rent. *Scripta Nova - Revista Electrónica de Geografía y Ciencias Sociales*, 21(580).

Contact

Vladimir Krajcik
University College of Business in Prague
Spálená 76/14, 110 00 Praha 1 - Nové Město
krajcik@vso-praha.eu

Elena Rihova
University College of Business in Prague
Spálená 76/14, 110 00 Praha 1 - Nové Město
rihova.elena@gmail.com