THESIS

OBSERVING PRICE DISPERSION ON THE BLABLACAR SHARED ECONOMY

Fruzsina Nagy

Supervisor: Prof. Dr. Marco Sahm

In partial fulfillment of the requirements
For the double-degree of

Bachelor of Sciences at
University of Bamberg:
Faculty of Economics,
Program of European Economic Studies

Bachelor of Arts at
Corvinus University of Budapest:
Faculty of Economics,
Program of Applied Economics

Fall, 2016
Acknowledgement

I take this place to express my gratitude. Firstly, I sincerely thank Professor Dr. Marco Sahm, the chair of the Department of Economic Theory at the University of Bamberg. Dear Professor Sahm, I am extremely thankful for agreeing to be my supervisor, and for your feedbacks during the work. Thank you for sharing your expertise, for guiding me since the very beginning and for all the encouragement you extended to me. You supported me through the venture with an enormous contribution. I never could have wished for a greater help.

I also would like to thank my family, especially my mother, Tünde Nagy who unconditionally supported me during the full length of the Bachelor program including this venture. She contributed selfless to my thesis in every way.

Finally, I am thankful for my friends who supported me with love and understanding.
Abstract

According to many experiences, prices of similar goods on a given market do not converge to one price. In this thesis the equilibrium price dispersion of the BlaBlaCar shared economy is observed. The existence and the nature of equilibrium price dispersion is examined on a given route during a given period of time. With the help of the collected data and a transformation of the Varian-model I found evidence for the existence of price dispersion in equilibrium state. It is also proven that the prediction of the Varian-model about the nature of price dispersion does not seem to hold for the BlaBlaCar market on the whole range of prices. Although, I found evidence that the predictions hold between the price quotes 16 and 20 Euros. The results might be implicated to understand the price dispersion on the BlaBlaCar market and on other shared economies.
Table of Content

List of Figures .................................................................................................................. 4

List of Tables ................................................................................................................... 5

1 Introduction .................................................................................................................. 6

2 Theoretical Contribution ............................................................................................... 9
  2.1 Price Dispersion and the evolution of relevant models ................................................. 9
      2.1.1 Statistical Evidence ............................................................................................ 9
      2.1.2 Theoretical Models .......................................................................................... 10
  2.2 The Model of Price Dispersion on the BlaBlaCar Market ........................................ 11
      2.2.1 Choosing the right approach .......................................................................... 11
      2.2.2 A model of the BlaBlaCar market ................................................................... 12

3 Empirical Contribution .................................................................................................. 23
  3.1 BlaBlaCar as a Company and as a Market ................................................................. 23
      3.1.1 The company .................................................................................................... 23
      3.1.2 BlaBlaCar as a Peer-to-Peer Market ............................................................... 24
      3.1.3 The Mechanisms ............................................................................................ 26
  3.2 The Dataset ............................................................................................................... 28
      3.2.1 Clearing the data ............................................................................................ 30
      3.2.2 The Distribution of the Prices ....................................................................... 31

4 Results ......................................................................................................................... 32
  4.1 Hypothesis 1 ............................................................................................................. 32
  4.2 Hypothesis 2 ............................................................................................................. 33

5 Conclusion .................................................................................................................... 36

Bibliography .................................................................................................................... 38

Appendix .......................................................................................................................... 40
List of Figures

**Figure 1.** Screenshot from www.blablacar.de on 30\(^{th}\) November, 2016. ..................11
Source: BlaBlaCar homepage.
Download: 27.11.2016

**Figure 2.** Theoretical Probability Mass Function of Prices. .................................22
Source: Own work with the help of Microsoft Excel.

**Figure 3.** Empirical Probability Mass Function of Prices. .................................31
Source: Own work with the help of Microsoft Excel.

**Figure 4.** Comparison of the Theoretical and Empirical Probability Mass Functions. ......33
Source: Own work with the help of Microsoft Excel.
List of Tables

Table 1. Effects of Different Variables on the Sold Seats. .................................15
    Source: Own work with the help of SPSS.

Table 2. Correlation between the Variables and the Price. .................................17
    Source: Own work with the help of SPSS.

Table 3. Sample of the Raw Data. .................................................................29
    Source: Own work with the help of Microsoft Excel.

Table 4. Descriptive Statistics about the Price Quotes. .....................................32
    Source: Own work with the help of Microsoft Excel.

Table 5. Correlation between the Probability Mass Functions. ............................34
    Source: Own work with the help of SPSS.

Table 6. Correlation between the Probability Mass Functions between 16 and 20 Euros. 35
    Source: Own work with the help of SPSS.

Table 7. Table for the Statistical Analysis. .....................................................40
    Source: Own work with the help of SPSS.

Table 8. Calculations for the Empirical and Theoretical Probability Mass Functions. 41
    Source: Own work with the help of Microsoft Excel.

Table 9. Input Values to calculate the Theoretical Cumulative Density Function. ....42
    Source: Own work with the help of Microsoft Excel.

Table 10. Sample of the Dataset for Observing the Empirical Distribution of Prices. 48
    Source: Own work with the help of Microsoft Excel.

Table 11. T-test for the Significance of the Pearson Correlation Coefficient. ..........49
    Source: Own work with the help of Microsoft Excel.
1 Introduction

During my double-degree program in Bamberg, Germany I often used BlaBlaCars to travel home to Budapest because it was cheaper than any other option and I was able to meet new people during the travel. I often took cars from Munich to Vienna because between these two popular cities there are usually big numbers of offers. The prices I experienced and booked rides with were always different: sometimes very cheap, sometimes average and sometimes more expensive than expected. This phenomenon made me curious. Is it an equilibrium phenomenon that prices are dispersed? What can be the law in the movement of the prices, if there is any? Are there models which describe price dispersion? Assuming similarities in costs for a ride on the same route would predict similarities in the offered prices too like the typical neoclassical model predicting the ‘law of one price’. (Varian, 1993, P. 3, 7-8) However, prices differed in such a range, which could hardly be described by the differences in costs. (Varian, 1980, P. 651) Experiencing this innovative way of shared-car-travelling and dispersion in the prices gave me the idea to do research on the price dispersion of the BlaBlaCar market.

I am devoted to find alternative ways of living in a current wasting economy which can provide a good start for living in harmony with our natural environment. I believe that there are solutions for environmental problems, and BlaBlaCars are good examples for that. Car-sharing platforms facilitate fuller use of cars and fuel. The global effect of the shared-car-traveling as an environmentally friendlier behavior had me committed to the topic of this work.

The analysis of the BlaBlaCar market contains restrictions to make it appropriate for my knowledge and a bachelor’s thesis. Instead of observing the whole BlaBlaCar market, which I would not be capable of, I decided to observe only one route (Munich to Vienna) during a given period of time (24th September to 10th October). The dataset does not contain comprehensively all the information about a given travel either.

To observe price dispersion on the given route on the BlaBlaCar market I decided to compare two probability mass functions. The first function is drawn from existing theoretical models which describe equilibrium price dispersion. I heavily rely on the model of Varian (1980, P. 652) (in following: Varian-model) and the model of Baye et al. (2004, P. 465, 468-470) (in following: Baye-model) to create a model to observe the BlaBlaCar market (in following: BlaBlaCar-model). The theoretical distribution of the prices is derived from the BlaBlaCar-
model. The model predicts a typical U-shaped continuous probability density function of prices: there is a high probability that prices are very high or very low, but there is a small probability for medium prices.

The second function represents the reality. Therefore, data was collected from the internet site www.blablacar.de. A program written in Python language gathered the information twice a day. I use this data to show the empirical distribution of the prices on the route from Munich to Vienna between 24th September and 10th October.

The subjects of the comparison are the former functions. From the probability density function, I estimate the probability of each integer price quote by setting small intervals on the continuous scale. With this method the estimated probability mass of each price can be calculated. The two probability mass functions are comparable. To measure the relationship, I use the Pearson correlation coefficient.

My hypotheses are that (1) the prices are dispersed on the BlaBlaCar market on the given route during the given period of time, and (2) the Pearson correlation coefficient is significantly different from zero at $\alpha = 0.05$ level, when the Pearson correlation coefficient is calculated between the estimated theoretical probability mass function and the empirical probability mass function of prices on the BlaBlaCar market on the given route during the given period of time.

Results show that price dispersion exists on the BlaBlaCar market on the given route during the given period of time in the range of 27 Euros when the mean price is 22.69 Euros. The variance of the observed prices is 17.02 Euros. Nevertheless, price dispersion exists, there is no statistical evidence that the distribution of the observed prices is similar to the predicted theoretical distribution of the prices. The two probability mass functions do not significantly correlate. However, there is significant similarity between the two probability mass functions between 16 and 20 Euros. All in all, the first hypothesis is accepted but the second hypothesis is rejected.

The remainder of the thesis is organized as follows: Section 2 describes the history and the theoretical background behind the Varian- and Baye-model and derives the BlaBlaCar-model to analyze equilibrium price dispersion on the BlaBlaCar market. This section enunciates the theoretical distribution of the prices. Section 3 introduces the BlaBlaCar market and describes the mechanisms how sellers and consumers are able to use the car-sharing platform. It also gives the main idea behind the dataset with its strengths and
limitations. In this section the empirical distribution of the observed prices is given. Section 4 compares the theoretical and the empirical distribution functions and analyses the results. Section 5 concludes and an Appendix can be found at the end of the thesis.
2 Theoretical Contribution

The main purpose of the Theoretical Contribution section is to derive the probability mass function of prices which is assumed to represent the distribution of prices on the BlaBlaCar market.

Firstly, I provide basic information about price dispersion. Then, some seminal works are mentioned which provide the statistical evidence of price dispersion. After, theoretical economic models are brought which describe price dispersion in equilibrium state.

Secondly, the BlaBlaCar-model is derived from the Varian- and the Baye-model. In this part the theoretical cumulative distribution function and the probability mass function are explicated.

2.1 Price Dispersion and the evolution of relevant models

According to many experiences, it seems that prices of similar goods on a given market do not converge to one price, how it is predicted in the neoclassical economic models. As Varian also wrote in his seminal paper The Model of Sales “the law of one price is no law at all”. (1980, P. 651) The phenomenon, when prices do not converge to one price can be described with price dispersion. (Hopkins, 2006, P. 1) According to Hopkins “price dispersion occurs when different sellers offer different prices for the same good in a given market.” (2006, P. 1) Price dispersion measures include the range of prices (Brynjolfsson and Smith, 2000, P. 575), the variance of the prices (Kaplan and Menzio, 2015, P. 1168) or the coefficient of variation of the prices. (Sorensen, 2000, P. 837-838)

To prove the existence of equilibrium price dispersion of similar\(^1\) goods, economists chose two ways. One was to provide statistical evidence. They observed different markets and opposed that the prices converge to one price. The other way was to build theoretical models.

2.1.1 Statistical Evidence

Observations have been provided by for example Lach (2002) on conventional retail stores and by Baye et al. (2004), Ellison and Ellison (2001) and Brynjolfsson and Smith (2000) on online markets. They all observed different markets: Lach (2002) studied the price movements of refrigerators and grocery items, Baye at al. (2004) electronic products on

---

\(^1\) Products are substitutes but can differ for example in label, package, services, etc. (Hirschey et al., 1995, P. 621)
Shopper.com, Ellison and Ellison (2001) computer memories and Brynjolfsson and Smith (2000) books and CD-s on online markets. They all found statistical evidence that the prices of similar goods do not converge to the ‘law of one price’ neither when the goods were sold offline, nor when online. Nevertheless, all the studies came up with different results in the measure of price dispersion. These studies brought the evidence that price dispersion is significant and persistent over time. (Baye et al., 2004, P. 464; Hopkins, 2006, P. 1)

2.1.2 Theoretical Models

The fact that price dispersion exists (over time) attracted a number of economists to provide models which can put theoretical explanations behind the statistical evidence. The biggest challenge was to create a model where price dispersion is an equilibrium phenomenon.

There were two popular approaches. Models of the first approach were usually released beforehand in time than models of the second approach. It is interesting to pay attention to the evolution of the second approach from the first one.

The first approach assumes sequential search and heterogeneities in consumers’ search propensities. Sequential search means that consumers are able to learn one price at once and learning one more price needs extra effort. (Diamond, 1970, P. 164) The differences between consumers stem from different search costs consumers search with. Differences in searching methods or propensities can describe the price dispersion observed in conventional markets, and those internet markets where consumers cover the incremental costs of searching. (Baye et al., 2004, P. 465)

In the second approach the assumption of sequential search disappears. Here, consumers are able to decide if they want to know one price, or all the prices at once for a given cost. (Hopkins, 2006, P. 3) It is not hard to imagine this situation: prices of conventional stores’ products can be listed on price comparison websites, or often in newspapers. Consumers, who prefer to know the whole distribution of prices buy the newspaper or get access to the website. There are also consumers who do not spend on newspapers or do not pay to access those platforms. (Varian, 1980, P. 952)

The model where prices are listed and all consumers have the opportunity to learn all the prices - we call the clearinghouse model. The name ‘clearinghouse model’ comes from Baye and Morgan (2001, P. 454) but the idea of introducing this system appeared first in the theoretical model of Salop and Stiglitz (1977, P. 495). Clearinghouse models, which are for example „price comparison sites such as Shopper.com, mySimon.com, and EvenBetter.com
now make it possible for consumers, to obtain a list of prices for a given product for what is close to a zero marginal cost of obtaining each price quote.” (Baye et al, 2004, P. 465)

The models of the second approach also predict equilibrium price dispersion. The reason for that is that the two types of consumers have different preferences to purchase: who knows the whole list, shops at the lowest available price, and who does not know the list, shops random. (Salop and Stiglitz, 1977. P. 494)

2.2 The Model of Price Dispersion on the BlaBlaCar Market

2.2.1 Choosing the right approach

On the BlaBlaCar market the prices are listed. As Baye et al. states „clearinghouse models [...] more closely match the environment consumers encounter at price comparison sites.” (Baye et al., 2004, P. 466) Let us see a search for the travel from Munich to Vienna on 30th November! The search took place on 25th November and the listed prices are presented in Figure 1.

*Figure 1. Screenshot from www.blablacar.de on 30th November, 2016.*

![Screenshot from www.blablacar.de](image)

Source: BlaBlaCar homepage:

The picture shows that the BlaBlaCar site lists the prices and thus, for the analysis of the BlaBlaCar market it seems to be appropriate to use the clearinghouse model which is the

---

second approach. Note, that the prices are dispersed in the range from 20 Euros to 27 Euros, which makes it considerable to observe price dispersion on the BlaBlaCar market.

2.2.2 A model of the BlaBlaCar market

There are two important models I chose to analyze the BlaBlaCar market. The Varian-model gives the base and the method of my analysis. This model is typical of the second approach, but it is appropriate to apply it only for conventional retail stores or those internet markets where consumers incur the costs of searching. (Varian, 1980, P. 652) The Baye-model is a transformation of the Varian-model, and it is already suitable for online price-comparison sites where getting access to the website is not anymore costly. (Baye et al., 2004, 468-470) Transformations are applied on the Varian-model to make it suitable for the BlaBlaCar market. The method of Baye et al. gave me the idea how to transform the original model.

In the following part the BlaBlaCar-model is derived. Firstly, the consumers’ side is introduced. Here the challenge is to set two new types of consumers with which the model is suitable for the BlaBlaCar market. Secondly, the sellers’ side is described. And finally, the analysis of the equilibrium state of the BlaBlaCar-model can be found in 2.2.2.3 The Analysis part.

2.2.2.1 Consumers’ Side

In this section I heavily rely on Varian’s seminal paper A Model of Sales (1980, P. 652-657). The model assumes that consumers come in large number and exogenously in two types: informed, and uninformed. The maximum amount any consumer would pay to purchase is the reservation value $r$, which is identical, and all consumers have unit demand. (Varian, 1980, P. 652)

Informed consumers have access to the clearinghouse so they know the whole distribution of prices and thus, they also know the lowest available price. We can think that they are the consumers who have subscribed for a newspaper where prices of a given good from different sellers are listed. Since informed consumers are rational agents, they purchase from the seller which offers the lowest price. Uninformed consumers do not have this information. They have not subscribed for the newspaper, yet they did not spend money on subscription but they do not know the list of all the prices. They pick the closest store and purchase there, for

---

3 Later Varian also mentions the case, when informed and uninformed consumers come endogenously. Though, this is depending on their search cost and this is why it is not explicated in this thesis.
example. (Varian, 1980, P. 652) The assumption of the Varian-model is that they shop random when the price of the good they want to purchase is less that $r$ at the seller they happened to appear. (Varian, 1980, P. 652)

Baye et al. prove the validity of the Varian-model for online markets such as the BlaBlaCar website. They transform the Varian-model in such way which matches better the environment of easy-to-access online price comparison sites. When the online price comparison sites are public and costless, the assumption of the existence of uninformed consumers does not seem to hold. (Baye et al., 2002, P. 468-469)

Baye et al. use the Varian model as the base and apply some changes to make it suitable for online price comparison sites. Baye et al. hold the basic assumption that one type of the consumers chooses the lowest available price, and the other type of consumers chooses randomly. Instead of differentiating among information (in the Varian-model there are informed and uninformed consumers) they let consumers choose a seller according to price-sensitiviti. Price-sensitive consumers are the ‘Shoppers’ and price-insensitive consumers are the ‘Loyals’. This means that ‘Shoppers’ choose the lowest prices and ‘Loyals’ choose randomly when the price they choose does not exceed their reservation value. Simplified, informed consumers from the Varian-model become ‘Shoppers’ and uninformed ‘Loyals’ according to the method of choices. (Baye et al., 2004, P. 468-470)

In the Baye-model there are two extra factors compared to the Varian-model, which have to be taken into consideration. Firstly, there is a probability $0 \leq \alpha \leq 1$, with which a consumer has access to the full list of prices (and a probability $1 - \alpha$ with which he does not). (Baye et al., 2004, P. 490) Secondly, they introduce a cost $\phi \geq 0$ for sellers to pay when they choose to list their price. (Baye et al., 2004, P. 468) The Baye-model has the outcome of the Varian-model if $\alpha = 1 \land \phi = 0$. That is in this case, when all consumers have access to the list of prices and sellers pay zero cost to list their prices. (Baye et al., 2004, P. 494) Assuming that $\alpha = 1 \land \phi = 0$ does seem to be appropriate in the observation of the BlaBlaCar market too. Consumers who want to book a ride on BlaBlaCar do not have any other option than visit the website where all the prices are listed. So they see the full list of prices, that is $\alpha =$
1. Sellers do not have any pecuniary costs when they upload a ride. Therefore, $\phi = 0$ can hold\(^4\).

So holding the assumption of lowest-price and random selection makes it possible to create the BlaBlaCar-model with two other types of consumers. An environment has to be ensured where the one type of consumers purchases at the lowest price and the other type of consumers purchases randomly.

‘Loyals’ in the Baye-model, evidently, are loyal to a seller. And this assumption is useful on the market of electronic goods (which they observed) and the writers prove that price dispersion exists on this market. (Baye et al., 2004, P. 466) It is easy to imagine that a consumer continues to shop from the same producer of one’s first USB-stick because the first one was satisfying to one’s needs. But would it make sense to observe loyal consumers on the BlaBlaCar market? According to my experience when traveling from one city to another it does not play a significant role to choose always the same driver. And similarly, based on the dataset the number of drivers offering the ride more than one time in our data appears only 16 times out of 262 offers on the route from Munich to Vienna in the given period of time. This implies that applying the Baye-model which makes difference between ‘Loyals’ and ‘Shoppers’ would not be completely appropriate in the BlaBlaCar-model.

In the BlaBlaCar-model one type of the consumers is the ‘Shoppers’ and other type is the ‘Rest’. I brought a statistical analysis to argue my decision. This analysis is presented in Table 1.

\(^4\) Obviously, there are some costs of searching and offering a ride (for example the time consumed) but they are minimal and do not contain pecuniary factor. Therefore, I assume that the cost of searching and offering a ride on www.blablacar.de is zero.
Table 1. Effects of Different Variables on the Sold Seats

<table>
<thead>
<tr>
<th>Step 1</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PriceEUR</td>
<td>-1.96</td>
<td>.066</td>
<td>8.919</td>
<td>1</td>
<td>.003</td>
<td>.822</td>
</tr>
<tr>
<td>Daynumber</td>
<td>.052</td>
<td>.084</td>
<td>.377</td>
<td>1</td>
<td>.539</td>
<td>1.053</td>
</tr>
<tr>
<td>Age</td>
<td>-0.021</td>
<td>.023</td>
<td>.801</td>
<td>1</td>
<td>.371</td>
<td>.980</td>
</tr>
<tr>
<td>Picture</td>
<td>-7.31</td>
<td>.603</td>
<td>1.469</td>
<td>1</td>
<td>.225</td>
<td>.481</td>
</tr>
<tr>
<td>NumDrives</td>
<td>-0.015</td>
<td>.010</td>
<td>2.190</td>
<td>1</td>
<td>.139</td>
<td>.198</td>
</tr>
<tr>
<td>Rating</td>
<td>.510</td>
<td>.415</td>
<td>1.514</td>
<td>1</td>
<td>.219</td>
<td>1.666</td>
</tr>
<tr>
<td>CarStars</td>
<td>.530</td>
<td>.321</td>
<td>2.737</td>
<td>1</td>
<td>.098</td>
<td>1.699</td>
</tr>
<tr>
<td>Bla</td>
<td>-7.61</td>
<td>.476</td>
<td>2.549</td>
<td>1</td>
<td>.110</td>
<td>.467</td>
</tr>
<tr>
<td>Smoke</td>
<td>-0.045</td>
<td>.860</td>
<td>.003</td>
<td>1</td>
<td>.958</td>
<td>.956</td>
</tr>
<tr>
<td>Pets</td>
<td>-1.121</td>
<td>.716</td>
<td>2.451</td>
<td>1</td>
<td>.117</td>
<td>.326</td>
</tr>
<tr>
<td>Constant</td>
<td>2.691</td>
<td>3.094</td>
<td>.757</td>
<td>1</td>
<td>.384</td>
<td>14.749</td>
</tr>
</tbody>
</table>

a. Variable(s) entered on step 1: PriceEUR, Daynumber, Age, Picture, NumDrives, Rating, CarStars, Bla, Smoke, Pets.

Source: Own work with the statistical program SPSS.

The statistical analysis is a regression analysis to determine which factors have effect and how big on the demanded seats. A sample of the table which Table 1 is created from can be found in Appendix A. In the dataset the supply is given by the listed offers on the internet page but measuring the consumers side is more difficult. To measure the quantity of sold seats I applied a binary logistic regression function which measures the logarithmic relative possibility of whether one seat was sold or not. The regression equation is:

\[
\ln \left( \frac{P(\text{Sold Seat}=1)}{1-P(\text{Sold Seat}=1)} \right) = \beta_0 + \beta_1 \text{PriceEUR} + \beta_2 \text{Daynumber} + \beta_3 \text{Age} + \beta_4 \text{Picture} + \beta_5 \text{NumDrives} + \beta_6 \text{Rating} + \beta_7 \text{CarStars} + \beta_8 \text{Bla} + \beta_9 \text{Smoke} + \beta_{10} \text{Pets}.
\]

Where \( \ln \left( \frac{P(\text{Sold Seat}=1)}{1-P(\text{Sold Seat}=1)} \right) \) is the natural logarithm of the dependent variable. The dependent variable in case of a binary logistic regression is the ratio \( \frac{P(\text{Sold Seat}=1)}{1-P(\text{Sold Seat}=1)} \), which is called the odds. PriceEUR, Daynumber, …, Pets are the observed features of the travels which are the independent variables here. To interpret \( \beta_i \) (the coefficients of the independent variables) the odds ratio has to be defined (OR). \( OR = e^{\beta_i} \). The SPSS output presents the odds ratios in the last column. To understand the effect of the odds ratios, let me define the highlighted row! If the price of a ride enhances by 1 Euro, then the odd (the ratio of the probability that the seat is sold at this price over the probability that it is not) is going to be \( 1 - 0.822 = 0.188 \) times less. (Schwartz, 2016)
An other important data in Table 1 is the P-value (Sig. in Table 1) compared to the significance level. The results of the statistical analysis will be statistically significant if the P-value (Sig.) is less than the pre-set significance level (\(\alpha\)). In our case, \(\alpha = 5\%\). It can be seen in Table 1, that only the effect of PriceEUR is statistically significant. This means that in the dataset only the price has a significant effect on the sold seats. Some consumers’ decisions are influenced by the price, and significantly only by the price. If an offer is more expensive than an other, a part of the consumers has the propensity to buy the cheaper ride, not caring about any other feature of the ride.

Remember, that for changing the types of consumers the assumption that one type of consumers chooses the lowest price and the other type of consumers chooses randomly has to hold.

From the statistical analysis it is straightforward to state that one type of the consumers will be the ones, who are influenced by the price and only by the price. They will be the price-sensitive ones, like in the Varian-model the informed consumers or in the Baye-model the ‘Shoppers’. I will also call them ‘Shoppers’. As Table 1 shows, the effect of the price is significant at a 0,3\% level.

On the other hand, as Table 1 shows the rest of the variables (features of the travel) have effect with a low significance on the probability that one seat was sold relative to that it was not. The low-significance of the factors of the rides implies that there is no common preference which is true for all the rest of the consumers. Every consumer has an individual preference for a travel, but the aggregated preferences do not have a statistically significant effect on the sold seats. Based on this idea I found it appropriate to let the other type of consumers choose randomly from the offers and therefore, meet the assumption of the Varian-model. This can be also consistent with the reality. An offered ride on BlaBlaCar can attract consumers in a lot of different ways. A lot of different factors can play role in the consumers’ decisions.

I decided to call the type of consumers which choose randomly, the ‘Rest’.

For the completeness of random selection condition an endogeneity problem has to be eliminated. Suppose, that a consumer is interested in the driver’s experience (the amount of rides the driver has offered so far). This consumer will have the incentive to purchase from drivers who are very experienced. One is not interested in the price. But what if the experience of the driver is connected to the price? What if for example more experienced
drivers have learned how to price in this market and therefore, they offer lower prices? (Farajallah et al., 2016, P. 19) It would mean that the ones who have the propensity to buy from more experienced drivers would purchase from the lowest prices too subconsciously. To avoid this problem, I created Table 2.

Table 2. Correlation Between the Variables and the Price

<table>
<thead>
<tr>
<th>Price (EUR)</th>
<th>Pearson Correlation</th>
<th>Day (number)</th>
<th>Age</th>
<th>Picture</th>
<th>Num Drives</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price (EUR)</td>
<td>Pearson Correlation</td>
<td>1</td>
<td>.065</td>
<td>.003</td>
<td>.002</td>
<td>.078*</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.099</td>
<td>.930</td>
<td>.950</td>
<td>.049</td>
<td>.004</td>
</tr>
<tr>
<td>N</td>
<td>650</td>
<td>650</td>
<td>650</td>
<td>650</td>
<td>640</td>
<td>650</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rating</th>
<th>Num Rating</th>
<th>Car Stars</th>
<th>Bla</th>
<th>Music</th>
<th>Smoke</th>
<th>Pets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.100*</td>
<td>.139**</td>
<td>.013</td>
<td>-.057</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.030</td>
<td>.003</td>
<td>.769</td>
<td>.145</td>
<td>-.083</td>
<td>.088</td>
</tr>
<tr>
<td>472</td>
<td>472</td>
<td>529</td>
<td>650</td>
<td>306</td>
<td>440</td>
<td>324</td>
</tr>
</tbody>
</table>

Source: Own work with the help of SPSS.

It presents the Pearson correlation coefficients between the price of the travels and the other factors I collected from the BlaBlaCar homepage and could include in the statistical analysis. A sample of the table this analysis is based on can be found in Appendix A. The Driver’s Name was unmeasurable, and the Color Classification and Departure Time were incomplete so they are not included in the analysis. The Pearson correlation coefficient shows the linear relationship between two or more variables in the observed data. (Pearson, 1895, P. 241) SPSS marks the statistically significant correlations between two variables on the $\alpha = 5\%$ significance level with two stars (**). In this case Experience and NumRating variables show significant correlation with the price. Significant correlation between two variables might cause multicollinearity problems in the model and therefore, I do not include them in the analysis. This means that the coincidence, that one of the ‘Rest’ subconsciously but not randomly purchase from the lowest price is eliminated. Now the condition of random decision of the ‘Rest’ is appropriate.

5 It is important to mention that there are no assumptions or proofs for the nature of the random selection neither in the Varian- nor in the Baye-model. It is not clear if every price quote has the same probability or not. In the Varian model it is easy to imagine that consumers’ random selection does not follow the even distribution. There can be hotspots of stores which are more frequently visited than other stores because of nonpecuniary...
To conclude the results of this part, in the BlaBlaCar model consumers come exogenously in two types: ‘Shoppers’ and ‘Rest’. Let $S > 0$ denote the number of ‘Shoppers’ and let $M > 0$ denote the number of the ‘Rest’. Every consumer has access to the full list of prices at zero cost. ‘Shoppers’ are price-sensitive consumers, their only preference is to shop at the lowest available price whilst consumers from the ‘Rest’ are price-insensitive consumers and they shop random. Any consumer shops only if the price of the product is lower than the identical reservation value $r$. This is very similar to the Varian-model. (Varian, 1980, P. 652)

2.2.2.2 Sellers’ Side

On the consumers’ side the Varian-model had to be adjusted to the BlaBlaCar market because the assumptions about the information of the consumers did not fit to the BlaBlaCar market. The sellers’ side is more straightforward and does not need changes of assumptions.

There are $n > 1$ sellers. All sellers have identical, strictly declining average cost curves $c(q)$. (Varian, 1980, P. 952) As observing identical distance at every travel the strictly declining cost curve assumption is restricted only for a fixed cost assumption where $k > 0$. Sellers have a fixed cost of fuel and amortization (because the observations contain the same routes) and the marginal cost of taking an extra person is very close to zero. In my thesis $c(q) = k$ holds.

The sellers are competing on the market. Each seller has a density function $f(p)$. This function represents the probabilities with which the seller charges the price $p$. All sellers choose a price according to that function. The chosen price is presented in the clearinghouse. The prices are chosen randomly from $f(p)$. If a seller manages to have the lowest price being offered on the market (which is not exceeding $r$), it will surely attract all the $(S > 0)$ price sensitive ‘Shoppers’. This seller also attracts a share of the ‘Rest’ $(M > 0)$ consumers who chose randomly from the whole list of prices. Let $R$ denote the number of consumers of the ‘Rest’ per firm, which is given by the formula $R = \frac{M}{n}$. Thus, the seller with the lowest price offered on the market will have $S + R$ consumers. The seller who does not have the reasons. The same holds for the Baye-model: there are companies whose products are more popular and have more brand-loyal consumers than other companies.

---

6 „Retail stores traditionally characterized by fixed costs of rent and sales force, plus constant variable costs of the item being sold” (Varian, 1980, p. 652)

7 Varian (1980) and Baye et al. (2004) calculate with the random-choice assumption where the random choice is drawn from an even distribution.
lowest price will only attract a share of price-insensitive consumers, namely \( R^8 \). (Varian, 1980, P. 652-654)

2.2.2.3 The Analysis

The decision of consumers is clear (‘Shoppers’ shop at the lowest available price in the clearinghouse, consumers from the ‘Rest’ shop randomly) but sellers can decide according to different strategies.

Only the case of the symmetric equilibrium is examined, where each seller chooses the same strategy for pricing\(^9\). (Varian, 1980, P. 652) This means the sellers’ random choice on the price distribution function \( f(p) \). The analysis of the equilibrium state is the analysis of a symmetric monopolistically competitive equilibrium. (Varian, 1980, p. 652) Now let us see the results in the BlaBlaCar-model!

Monopolistic competition is the mixture of a perfect competition and a monopoly. The major difference between the two markets is that the products sellers offer in monopolistically competitive market, are differentiated, while in perfectly competitive market they are homogenous. The major similarity, on which the analysis is based, is that the above-normal profits are only possible in the short-run. In the long-run profits are driven to zero. This implies that entry occurs in the market until monopolistically competitive sellers can take effective countermeasure that is, until the increasing number of sellers reduces each seller’s profit to zero. (Hirschey et al., 1995, P. 621)

The following characteristics have to hold in the market.

1. On the market there are many sellers and many consumers. Each seller offers a small portion of the industry and each consumer buys a small portion too.
2. The products must be heterogeneous.
3. Consumers and sellers have perfect dissemination of information.
4. There is free entry and exit. Sellers are not restricted from entering or leaving the market. (Hirschey et al., 1995, P. 622)

\(^8\) If more than one seller manages to charge the same price, we consider it as a tie, and each seller gets an equal share of consumers. But as Varian shows, the event of a tie has zero probability. The formal proof can be found in Proposition 3. (Varian, 1980, P. 653)

There are many sellers and many consumers on the BlaBlaCar market. According to the dataset there are 650 offered seats and 113 consumers who can choose between them. In the dataset the mean of seats a seller offers is 2.5 and the mean of seats a consumer buys is 1.53.

The product heterogeneity assumption is also applicable for the BlaBlaCars, because the offers not only differ in the prices. As it is presented in for example Picture 1, every travel has a profile in the list. Every travel contains plenty of information which makes small differences between rides. This can be for example the chattiness of a driver, the experience of the driver, the reputation of driver or the quality of the car. All consumers and sellers have the opportunity to be perfectly informed. Every piece of information is presented in the clearinghouse which is public and free for everyone in the market. Entry and exit is also costless, there is no pecuniary factor for any of these movements.

Now, let us see the equilibrium dispersion of prices! All the formal proofs can be found Varian’s *The Model of Sales* (1980, P. 653-656), here only the main steps are mentioned to understand the process of establishing equilibrium.

Varian (1980, P. 654-655) proves that the cumulative distribution function of prices is:

\[
F(p) = 1 - \left(\frac{k-pR}{pS}\right)^{\frac{1}{n-1}}.
\]

The main idea behind the function is that profits are driven to zero on a monopolistically competitive market. There is a probability \((1 - F(p))^{n-1}\) with which a seller charges \(p\) as the lowest price on the market and thus, it attracts all the ‘Shoppers’ plus a share of the ‘Rest’ \((S+R)\). If the seller does not charge the lowest price, it attracts only a share of the ‘Rest’.

The probability of this event is \(1 - (1 - F(p))^{n-1}\). Therefore, the expected profit\(^{10}\) of any seller is:

\[
\int_{p}^{\tilde{p}} \left\{\pi_s(p)(1 - F(p))^{n-1} + \pi_f(p) \left[1 - (1 - F(p))^{n-1}\right]\right\} f(p) \, dp
\]

where the profit of the seller who succeeds to charge the lowest price (and thus, attracts \(S+R\) consumers) is:

\[
\pi_s(p) = p \ast (S + R) - c(S + R) = p \ast (S + R) - k
\]

---

\(^{10}\) See: Varian, 1980, P. 654-655
and the profit of the seller who does not succeed to charge the lowest price (and thus, attracts \( R \) consumers) is:

\[
\pi_f(p) = pR - c(R) = pR - k.
\]

The profits are maximized that is, the derivative of (3) must equal zero:

\[
\left\{ \pi_s(p)(1 - F(p))^{n-1} + \pi_f(p) \left[ 1 - (1 - F(p))^{n-1} \right] \right\} f(p) = 0.
\]

So if \( f(p) > 0 \) then

\[
\pi_s(p)(1 - F(p))^{n-1} + \pi_f(p) \left[ 1 - (1 - F(p))^{n-1} \right] = 0.
\]

Which can be rearranged to the formula, using (4) and (5):

\[
F(p) = 1 - \left( \frac{k - pR}{pS} \right)^{\frac{1}{n-1}}.
\]

It is sure, that \( \pi_f(r) = 0 \) since every profit is driven to zero. Substituting (4) into this equation gives that:

\[
R = \frac{M}{n} = \frac{k}{r}.
\]

For the same reasons Substituting (3) into the equation \( \pi_s(p^*) = 0 \) we have:

\[
p^* = \frac{k}{S + k/r}.
\]

Here, \( p^* \) denotes the lowest available price. (Varian, 1980, P. 656)

In the analysis of Varian (1980, P. 653) the theoretical cumulative density function is a continuous function\(^{12}\). What is needed for the later comparison is a probability mass function which gives the probability that a discrete random variable is equal to a given value. In this case the probability mass function should give the probability that a listed price quote on the BlaBlaCar website is equal to an integer amount of Euros.

To estimate the probability mass function, I set small intervals on the scale of prices that in the middle of every interval there is an integer value. The borders of the interval are going

\(^{11}\) On the interval \([p^*, r]\) each price has to have a positive density. The formal proof can be found in The Model of Sales as Proposition 8 and 9. (Varian, 1980, P. 658)

\(^{12}\) According to Proposition 3. (Varian, 1980, P. 653)
to be \( p_l - 0,4999 \) and \( p_u + 0,5 \). The mathematical fundamental theorem of calculus says that \( \int_a^b f(p) \, dp = F(b) - F(a) \). To every integer I order a probability.

The result is presented in Figure 2 and the table and method according to which it was calculated is presented in the Appendix B.

**Figure 2. Theoretical Probability Mass Function of Prices**

[Graph showing theoretical probability mass function of prices]

Source: Own work with the help of Microsoft Excel.

With this result the *Theoretical Contribution* section accomplished its purpose: the theoretical probability mass function is given.
3 Empirical Contribution

The main purpose of this section is to derive the observed probability mass function of the prices on the BlaBlaCar market. Therefore, firstly, I give basic information about BlaBlaCar as a company and as a market platform. Secondly, I introduce the dataset and in this part the probability mass function is explicated.

3.1 BlaBlaCar as a Company and as a Market

3.1.1 The company

The BlaBlaCar company and the online platform was founded by Frédéric Mozzalla and his two co-founders in France in 2006. They imagined a new way of transporting. The idea was to build a network which connects people who has empty seats in their cars (drivers in the following) with people who does not have a car but are willing to travel on the same route (riders in the following). The shared trips are intercity trips. Riders pay the driver for the empty seats so they share the costs of the travel. The internet based platform matches the riders with the drivers. (Farajallah et al. 2016, P. 6)

On the one hand, sharing a car between more people has economic reasons. The platform makes traveling more affordable for drivers because they earn money, and riders can save money because a typical trip on BlaBlaCar is cheaper than the other transportation options. (Farajallah et al., 2016, p. 6) We talk about a fuller use of economic resources like the car and fuel, which make the transportation more effective.

On the other hand, users can enjoy many non-pecuniary factors of shared-car traveling. The founders imagined a platform which makes traveling more social than ways of traveling. (BlaBlaCar homepage13) The indicator of this feature appears in the name of the company: BlaBlaCar. The ‘Bla’-s are indicating the drivers’ chattiness or talking preference during the ride. If the driver sets ‘Bla’ on his profile means, that he does not like to talk, ‘BlaBla’ means that he likes to talk, and ‘BlaBlaBla’ means that he loves to talk. (Farajallah et al., 2016, P. 6)

Sharing a car on the same route has also environmental benefits. As it is written on the homepage of BlaBlaCar users save approximately 250,000 tons of oil and 1,000,000 tons of

13 https://www.blablacar.co.uk/about-us/the-blablacar-story
CO₂ within a year. This is possible, because of the fuller use of resources: the average car occupancy in a car is 2.8 people instead of 1.6 people average. (BlaBlaCar homepage\textsuperscript{14})

All these factors make the BlaBlaCar car-sharing platform economically more efficient, socially more enjoyable and environmentally friendlier.

BlaBlaCar was told to be the leading long-distance car-sharing platform according to Sundararajan (2016, P. 12). Currently BlaBlaCars are used among 22 countries (mostly in Europe, but also in Russia, India, Mexico and Brazil). The platform has currently approximately 35 million members, and it has 12 million travelers per quarter. (BlaBlaCar homepage\textsuperscript{15}) The current value of the firm reaches approximately 1.6 billion dollars. (Thomson, 2015) BlaBlaCar has not been involved in regulatory issues like Uber, because BlaBlaCar is considered as a non-for-profit service. (Einav et al, 2016, P. 17) This means that drivers and riders only share the costs of the trip and this is the only purpose of the money transfer. On contrast, countries sometimes restrict sharing platforms. (Farajallah et al., 2016, P. 6; Einav et al., 2016, P. 17)

3.1.2 BlaBlaCar as a Peer-to-Peer Market

Einav et al. (2016, P. 4) emphasize the role and the therefore the rise of peer-to peer markets in creating trade between large number of fragmented buyers and sellers. According to the writers BlaBlaCar is considered as a peer-to-peer market as it connects drivers with riders to share empty seats on intercity trips. Einav et al. see BlaBlaCar as a peer-to-peer market, and introduce 3 difficulties markets face to make trade efficient. (Einav et al., 2016, P. 4)

(1) The first of them is to match buyers with special interests to sellers while the searching method is not too complicated. On BlaBlaCar the problem to match buyers with special and unique interests to sellers exists. (Einav et al., 2016, P. 4-5) For example, a rider would like to meet at Munich, Fröttmaning but the driver departs from Hauptbahnhof. The site lists all the offers which are departing from every part of Munich. Obviously the driver and the rider have possibility to communicate, but this is a good example, that BlaBlaCar found the easy search more important than special preferences.

At this point there is an interesting specialty. According to my experience there are such bargains, where the rider or the driver offers an extra amount to pay if the starting or ending

\textsuperscript{14} https://www.blablacar.co.uk/about-us

\textsuperscript{15} https://www.blablacar.co.uk/about-us
point of the travel can be changed. For example, the driver planned to depart from Munich, Hauptbahnhof. The rider offers the driver some extra money to do a bypass at Munich, Fröttmaning because this place is more comfortable for him.

These bargains are not indicated on the BlaBlaCar homepage and are not official transactions. They are not measurable. Therefore, in my analysis I do not take them into consideration.

(2) The second problem is to ensure whether the trade is safe and reliable for buyers and sellers who do not know each other. BlaBlaCar as many internet-based peer-to-peer markets introduced a reputation system. On BlaBlaCar riders can evaluate and rate drivers after a travel and it is published on the drivers’ profile. Although the reputation system exists on BlaBlaCar, the effect of the rating of the driver on his sold seats did not seem to be significant neither in my analysis\textsuperscript{16}, nor in the analysis of Farajallah et al. (2016, P. 17).

Also for safety reasons BlaBlaCar introduced the online payment system in Germany. They created a third party between riders and drivers. Whenever a payment takes place, the money is transferred first to the third party. After the actual travel the driver can call for this money from the third party with a code the rider gave him. Introducing a third party can make travels safer. (BlaBlaCar homepage\textsuperscript{17})

(3) The third difficulty is the pricing mechanism. Different peer-to-peer markets use different mechanisms for pricing to make the trade more efficient. There are price platforms where pricing is centralized, for example Uber. There, a central power decides the price of every offered route. (Gurley, 2014) Pricing on the BlaBlaCar market is decentralized: drivers set their own price which makes the BlaBlaCar market very interesting to observe. Sellers have the opportunity to use different strategies of pricing, and can experiment which price serves their preferences better. (Farajallah et al., 2016, P. 3)

It looks like BlaBlaCar was successful in dealing with the former problems, as it is one of the biggest intercity car-sharing platforms as mentioned before. Personally I think it is interesting to observe such markets, because on a peer-to-peer market sellers do not have such a big competitive advantage over consumers, and because sharing economy gives a

\textsuperscript{16} See: Table 1.

\textsuperscript{17} https://www.blablacar.co.uk/faq/question/how-do-i-get-paid
possible way of fuller use of resources. This can be a solution for global environmental problems which stem from wasting.

### 3.1.3 The Mechanisms

Basically, there are two platforms of BlaBlaCar where users can offer or book rides. One is the homepage of BlaBlaCar, in this case www.blablacar.de. The other platform is the BlaBlaCar application which can be downloaded for IOS and Android operational systems. (BlaBlaCar homepage\(^{18}\)) There is difference only in the configuration, the method of both platforms are the same. Therefore, the operation I am going to explain here is valid for both platforms.

#### 3.1.3.1 The Driver

When a driver would like to share his car on a travel first, he creates a profile on www.blablacar.de. In the profile he provides some basic demographic and contact information. It is possible to upgrade the profile too. According to Farajallah et al. (2016, P. 19) more experienced drivers are more likely to upgrade their profiles than new-coming users. Upgrading the profile includes uploading a profile picture, setting preferences in chattiness, smoking, music and allowing pets in the car and adding a car and details about the car.

After, he creates the profile of the ride. He submits the meeting point, the destination and the stopovers as well as the departure date and time. The driver can choose the price he is willing to take passengers with.

BlaBlaCar's method of pricing is unique among other car-sharing platforms. (Farajallah et al., 2016, P. 6-7) The page recommends a price on each route based on the distance and the estimated costs of the route\(^{19}\). After this the driver is allowed to set his own price but within limits. The minimum price he can set is 50% of the recommended price and the maximum is 150% of the recommended price. When the data collection started the price the driver chose was ordered into 3 categories represented with colors: green orange and red\(^{20}\).

---

\(^{18}\) [https://www.blablacar.co.uk/faq/question/how-do-i-access-blablacar-from-my-mobile-phone](https://www.blablacar.co.uk/faq/question/how-do-i-access-blablacar-from-my-mobile-phone)

\(^{19}\) According to the homepage in 2015 on blablacar.de the recommended price was calculated as the following: 0.05 Euro per kilometer and per seat. The prices are rounded up to integer amount of Euros. ([https://www.blablacar.de/faq/frage/wie-lege-ich-den-preis-meiner-fahrt-fest](https://www.blablacar.de/faq/frage/wie-lege-ich-den-preis-meiner-fahrt-fest))

\(^{20}\) The color classification has been used since February 2012. Green represented the lowest offered prices (below the recommended price), red represented the highest prices (between 125% and 150% of the
(Farajallah et al., 2016, p. 7) The driver can also set the number of the offered seats and write a detailed description about the ride. He has the opportunity to set the size of the luggage, the departure flexibility and his willingness to make a detour or not.

When the driver is complete with his profile and the details of the ride, he can publish his offer on the platform. Most of the information he has given is public except for his personal and contact data.

3.1.3.2 The Rider

When someone would like to travel between two cities with BlaBlaCar he searches on the homepage, or on the application. The search starts with giving the city where the car departs from, the destination, and the date when the rider wants to travel. The platform lists the offers which take place on the given route on the given day. Later riders are also able to specify their search with the advanced search tool.

3.1.3.3 The Match

When a rider finds a travel which satisfies his preferences, he can book the travel. He is also able to communicate with the driver about the details. He also pays immediately. When a travel is booked, the number of free seats in the car in question decreases by the number of the booked seats and the rider receives a code. When the travel actually takes place and the rider is brought to the destination he gives the code the driver. When the driver collects the codes from all passengers, he can request for the money from a third party on BlaBlaCar. The money is going to be transferred to his account within a few days. (BlaBlaCar homepage21)

A special feature of the BlaBlaCar car-sharing market is that communication and the trade start online but finish offline, in the car where the driver and the rider(s) are sitting in the same car for a couple of hours. This feature lets BlaBlaCar users to be motivated by non-pecuniary factors besides the obvious pecuniary gain. (Farajallah et al., 2016, P. 4) After the travel, finally the riders and drivers have the possibility to evaluate each other which is published on their profile on the homepage.

---

21 https://www.blablacar.co.uk/faq/question/how-do-i-get-paid
3.2 The Dataset

The dataset was created to get an overall picture about (1) the riders’ preferences about choosing a travel and (2) about the dispersion of the prices on the market. Therefore, I wrote a program in Python language to collect the needed information to an excel file. The code script can be found in the Appendix C.

The dataset contains 16 features about each ride including the price of the travel, and the quantity of seats offered. The data collection took place from 21st September 2016 to 8th October 2016 which resulted observations from 24th September to 10th October. The method was to download data 2 times for each day: one about the offers that take place three days before the travel and one about the offers that take place two days before the travel. The reason for this was to be able to measure the quantity of sold seats. For example, on the 21st and 22nd September I downloaded data for travels which took place on 24th September. With this method I was able to see the changes in the rides, including how many seats drivers have sold.

Between two polling, there was sometimes change in in the sold seats (which is very positive for us) and sometimes in the price too. Price decrease from one polling day to an other happened two times in my dataset, there was no observation when the price increased between the given days. When the number of offered seats decreased at the same time, the problem was that I did not know for sure on which price the seat was sold. These times I used the price of the later data poll.

There is no special reason for polling the data between 21st September and 8th October. This period of time was comfortable for my thesis work. I chose the route from Munich to Vienna because I had experience on that route. Farajallah et al. found evidence that the mean prices on different travels are not in strict linear relationship with the distance. For example, the mean price on Toulouse – Bordeaux trip (245 kms) is 15,114 Euros, and on Tours – Paris trip (240 kms) is 15,441 Euros. (Farajallah et al., 2016, P. 44) Choosing one trip in the given period of time may reduce the general validity of my results.

I have not found evidence on the effect of the day the ride took place on the prices or on sold seats, but according to my experience in holidays the prices are higher and users are more

---

22 One limitation of the dataset is that it was not automatized. Every day I had to run the program. One weekend I happened to be without internet connection and that weekend the data was not collected. This may reduce the validity of my results.
active on the BlaBlaCar homepage. Obviously, my experience might not be representative, but it is important to state, that choosing the given period of time (24th September to 10th October) in my observation may distort the results.

On the BlaBlaCar homepage there is much information about each travel. I was not able to collect every detail of each travel because of two reasons. Firstly, I had limitations in my programming skills and secondly, there is information which is hidden from the public. I was able to collect 16 features of each travel including the price of the ride and the number of seats sold. The incompleteness of the dataset may also reduce the generality of my results.

A sample of the raw dataset is presented in Table 3.

### Table 3. Sample of the Raw Data

<table>
<thead>
<tr>
<th>Price</th>
<th>Color</th>
<th>Time</th>
<th>Driver Name</th>
<th>Age</th>
<th>Picture</th>
<th>Num Drives</th>
<th>Account Created</th>
<th>Rating</th>
<th>Num Rating</th>
<th>Car Stars</th>
<th>Bla</th>
<th>Music</th>
<th>Smoke</th>
<th>Pets</th>
<th>Seats</th>
<th>Seats 2</th>
<th>Sold Seats</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 €</td>
<td>green</td>
<td>08:00</td>
<td>Peter P</td>
<td>42</td>
<td>0</td>
<td>34</td>
<td>07/03/2016</td>
<td>4.8</td>
<td>22</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34 €</td>
<td>red</td>
<td>16:00</td>
<td>Ruslan E</td>
<td>26</td>
<td>1</td>
<td>3</td>
<td>16/09/2016</td>
<td>-</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 €</td>
<td>orange</td>
<td>15:12</td>
<td>Ivanov Z</td>
<td>32</td>
<td>1</td>
<td>3</td>
<td>15/12/2016</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 €</td>
<td>green</td>
<td>08:30</td>
<td>Daniel S</td>
<td>32</td>
<td>1</td>
<td>6</td>
<td>19/09/2015</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 €</td>
<td>green</td>
<td>10:00</td>
<td>Peter G</td>
<td>66</td>
<td>1</td>
<td>102</td>
<td>03/09/2016</td>
<td>4.8</td>
<td>52</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 €</td>
<td>green</td>
<td>10:30</td>
<td>Elisabeth L</td>
<td>40</td>
<td>1</td>
<td>7</td>
<td>30/07/2016</td>
<td>4.8</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 €</td>
<td>red</td>
<td>11:30</td>
<td>Albert M</td>
<td>32</td>
<td>1</td>
<td>2</td>
<td>21/09/2016</td>
<td>-</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 €</td>
<td>orange</td>
<td>13:00</td>
<td>Marco H</td>
<td>38</td>
<td>1</td>
<td>3</td>
<td>01/09/2016</td>
<td>-</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 €</td>
<td>green</td>
<td>14:00</td>
<td>Lisa S</td>
<td>27</td>
<td>1</td>
<td>5</td>
<td>06/05/2014</td>
<td>4.5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 €</td>
<td>green</td>
<td>15:00</td>
<td>Alpar P</td>
<td>28</td>
<td>1</td>
<td>21</td>
<td>01/11/2015</td>
<td>5</td>
<td>11</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 €</td>
<td>green</td>
<td>15:00</td>
<td>Arnie A</td>
<td>26</td>
<td>1</td>
<td>7</td>
<td>30/01/2014</td>
<td>4.8</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 €</td>
<td>orange</td>
<td>15:00</td>
<td>Frank E</td>
<td>54</td>
<td>1</td>
<td>60</td>
<td>08/05/2007</td>
<td>5</td>
<td>27</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 €</td>
<td>green</td>
<td>16:00</td>
<td>Peter S</td>
<td>29</td>
<td>1</td>
<td>13</td>
<td>06/03/2014</td>
<td>4.7</td>
<td>17</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 €</td>
<td>green</td>
<td>17:00</td>
<td>Gabriel N</td>
<td>26</td>
<td>1</td>
<td>3</td>
<td>11/10/2014</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 €</td>
<td>green</td>
<td>17:00</td>
<td>Alexandra B</td>
<td>26</td>
<td>1</td>
<td>24</td>
<td>26/03/2015</td>
<td>4.8</td>
<td>16</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 €</td>
<td>orange</td>
<td>18:00</td>
<td>Adam N</td>
<td>34</td>
<td>1</td>
<td>17</td>
<td>07/10/2007</td>
<td>4.9</td>
<td>88</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 €</td>
<td>orange</td>
<td>18:00</td>
<td>Marc B</td>
<td>23</td>
<td>1</td>
<td>2</td>
<td>29/06/2016</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 €</td>
<td>red</td>
<td>20:00</td>
<td>Sander G</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>20/09/2016</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 €</td>
<td>orange</td>
<td>22:00</td>
<td>Peter W</td>
<td>40</td>
<td>0</td>
<td>10</td>
<td>26/05/2016</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Own work with the help of Excel.

Table 3 presents a part of the rides which were offered on 25th September. Each row represents an offered travel from the list of the homepage www.blablacar.de on the route from Munich to Vienna. The first column shows the prices of each ride, the second the color classification of the prices, the third the exact departure times, the fourth presents the names of the drivers and the fifth the drivers’ age. The sixth column shows if the driver has a profile picture (1) or not (0). The seventh and the eighth column represent the drivers’ experience: the number or drives and the date when the account was created. (Measuring the experience I inserted a new column, where the amount of the days is shown for which the driver has been registered. This was created by an easy method: subtracting the date of account creation of the second polling day of the travel.) The ninth column shows the quality of the rating measured in decimals from 1 to 5, and the tenth column shows the number of the ratings a driver has received. The eleventh column shows the quality of the cars, measured in stars.
from 1-3. Columns from twelve to the fifteen present the drivers’ preferences about chattiness (1-3, where 3 is the chattiest), music (‘1’ allows, ‘0’ does not allow), smoking (‘1’ allows, ‘0’ bans) and pets (‘1’ allows, ‘0’ bans). The sixteenth column shows the number of offered seats on the first polling day, in this case on 22nd September. The seventeenth column presents the number of offered seats on the second polling day, on 23rd September. The eighteenth column is the subtraction of the offered seats on the first and on the second polling day. This column gives the number of the sold seats in the given car.

Data has been collected on the Munich – Vienna trip for 15 days which resulted 302 offered travels in the observation.

3.2.1 Clearing the data

To be able to analyze the data I cleaned it and created two new tables. The first table was suitable for the statistical analysis and the second table was appropriate to create the probability mass function.

3.2.1.1 Table for the Statistical Analysis

Firstly, to weight the offers, where the number of sold seats was more than zero, I simply added the copy of the offers. If, for example, the driver sold 3 seats in the car, I inserted 2 more rows with the copy of the offer in question. Then I decreased the number of offered seats with the number of the inserted rows. When the number of available seats was zero at the first polling I automatically deleted the given row. It helped to clean the database from the rides, which were already fully booked on the first polling date. I did not include them in the analysis, because I could not be sure on which price did they sold seats, and how many. With this method I could make sure, that the number of seats sold becomes a binary variable. After applying this method, the database contained 650 observations (out of 262 offered travels), in which all the number of the offered seats was one and the number of the sold seats was 1 or 0. Here, I eliminated the rows which were insufficient for the SPSS program. These features were the Departure time and the Color classification. The new database made it possible to create Table 1 to measure the features’ effect on the sold seats with a binary logistic regression and to create Table 2 to measure the correlation between the price and other variables. A sample of the table which was used for the statistical analysis is presented in the Appendix A.
3.2.1.2 Table for the Probability Mass Functions

To create the second table, I used the raw database again. All the offers were eliminated, where the number of the sold seats was zero. It was useful to observe only the offers which were actually booked. After the elimination I was able to analyze the equilibrium state, where supply meets the demand that is, the number of offered seats equals the number of booked seats. This database contained 74 offered travels with 113 booked seats. The analysis of the empirical price dispersion is based on this database. This table also gave the information to substitute the exogenous variables to the theoretical cumulative distribution function. Appendix D presents this table.

3.2.2 The Distribution of the Prices

For this part the relevant information from the second table is the prices and the number of the seats sold on a given price. The aim is to create the probability mass function of the prices. Therefore, with the help of Excel I calculated the frequency of occurrence of a given price and after, the probability with which a price occurs. The probability mass function is presented in Figure 3.

**Figure 3. Empirical Probability Mass Function of Prices**

The chart shows with which possibility is a given price equal to the prices presented on the X axis. On the Y axis these probabilities are presented. For example, according to the graph, the probability that a price at which a travel was bought on the BlaBlaCar homepage is 16 Euros, is 0.0619.
The minimum price in the dataset is 9 Euros and the maximum price is 36 Euros at which seats were sold on the Munich – Vienna trip between 24th September and 10th October. Peaks are recognizable at 16 Euros, 24 Euros, and 28 Euros.

With Figure 3 the main purpose of The BlaBlaCar Market section is achieved. The probability mass function of the prices is present. At the first sight the chart implies that the prices on the BlaBlaCar market do not converge to the ‘law of one price’, since prices occur with different possibilities.

4 Results

In the beginning of my thesis I mentioned two hypotheses:

(1) The prices are dispersed on the BlaBlaCar market on the given route during the given period of time.

(2) The Pearson correlation coefficient is significantly different from zero at \( \alpha = 0,05 \) level, when the Pearson correlation coefficient is calculated between the estimated theoretical probability mass function and the empirical probability mass function of prices on the BlaBlaCar market on the given route during the given period of time.

In this section I find evidence to accept or to reject the hypotheses.

4.1 Hypothesis 1

Price dispersion is a frequently observed phenomenon. (Varian 1980, P. 651) To prove that it exists on the BlaBlaCar market too, Table 4. is brought.

**Table 4. Descriptive Statistics about the Price Quotes**

<table>
<thead>
<tr>
<th>PriceEUR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>22,69026549</td>
</tr>
<tr>
<td>Median</td>
<td>23</td>
</tr>
<tr>
<td>Mode</td>
<td>24</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4,125442933</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>17,01927939</td>
</tr>
<tr>
<td>Range</td>
<td>27</td>
</tr>
<tr>
<td>Minimum</td>
<td>9</td>
</tr>
<tr>
<td>Maximum</td>
<td>36</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>0,18181554</td>
</tr>
<tr>
<td>Count</td>
<td>113</td>
</tr>
</tbody>
</table>

Source: Own work with the help of Microsoft Excel.
Price dispersion measures include the range of prices (Brynjolfsson and Smih, 2000, P. 575), the variance of the prices (Kaplan and Menzio, 2015, P. 1168) or the coefficient of variation of the prices. (Sorensen, 2000, P. 837-838) According to Baye et al. (2004, P. 467) "[w]hen the law of one price holds […] these measures of price dispersion are all zero.” The reversion of the statement is that if the law of one price does not hold (so if there is price dispersion) all these measures differ from zero.

The range of prices on the BlaBlaCar market on the given route on the given period of time is 27 Euros, where the mean price quote is 22.69 and the minimum and the maximum price quotes are 9 and 36 Euros respectively. This represents a variance of 17.02 Euros and a standard deviation of 4.13 Euros. The coefficient of variance is 0.18. As none of the measures are equal to zero, I can conclude that price dispersion exist on the BlaBlaCar market on the given route during the given period of time that is, the first hypothesis is accepted.

4.2 Hypothesis 2

Figure 2 in the 2. Theoretical Contribution section shows the theoretical probability mass function of the prices which are assumed to hold in reality on the BlaBlaCar market. Figure 3 on the other hand presents the empirical observation on the BlaBlaCar market. Now the question is if they have similarities or not. Figure 4 shows the two probability mass functions.

**Figure 4. Comparison of the Theoretical and Empirical Probability Mass Functions**
On the one hand, the blue line represents the empirical data. Very low and very high prices occur with relatively low probability but middle prices occur with relatively higher probabilities. The price quote 24 Euros (which is approximately 2 Euros above the mean) has the highest probability to occur. Other peaks are recognizable at 16, 21 and 28 Euros.

On the other hand, the orange line represents the theoretical distribution of prices in the BlaBlaCar-model. It has a typical U-shape as Varian (1980, P. 651) predicted. Lower and higher prices occur with high probabilities whilst middle prices occur with relatively low probabilities. The highest price occurs with far the highest probability.

Based on Figure 4 the first impression is that the two functions are not similar to each other. To observe the relationship between them, the Pearson correlation coefficient is calculated. It measures the linear relationship between two variables. The advantage of this method that it can be a good measure for the shape of the curves but the disadvantage is that it does not count with the distance between them. But as these functions are probability mass functions, we know that the maximum value they can take is 1. This gives a limit to the shift from each other (and from the X axis). The Pearson correlation-matrix between the two functions is presented in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Empirical f(p)</th>
<th>Theoretical f(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical f(p)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Theoretical f(p)</td>
<td>-0.2260131</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Own work with the help of Microsoft Excel.

The value of the Pearson correlation coefficient is -0.23. The question is if it is significant or not. Testing for the significance I applied a t-test on $\alpha = 5\%$ significance level. The result of the test is that the Pearson correlation coefficient is not significant between the empirical and the theoretical probability mass functions. This means that the second hypothesis is rejected. For the detailed calculation see Appendix E.

Although the two functions do not show similarities on the whole range of prices, it would be interesting to observe the interval between 16 and 20 Euros. The typical U-shape of the BlaBlaCar-model is a very interesting density of the prices. A peak and a slow reduction of the probabilities is also recognizable on the empirical probability mass function. I measured the Pearson coefficient on this interval, which is presented in Table 6.
Table 6. Correlation between the Probability Mass Functions between 16 and 20 Euros.

<table>
<thead>
<tr>
<th></th>
<th>Empirical f(p)</th>
<th>Theoretical f(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical f(p)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Theoretical f(p)</td>
<td>0,937750374</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Own work with the help of Microsoft Excel.

The value of 0,94 is a significant and strong linear relationship between the two functions according to the t-test on a $\alpha = 5\%$ significance level which is presented in Appendix E. The value of the Pearson correlation coefficient shows that the two functions move along, but it is not so straightforward where exactly they are located. This means that in the BlaBlaCar-model and in the observation the probabilities that the price quotes are equal with the prices from 16 to 20 Euros are decreasing in both cases but they are not necessarily equal.

It can be concluded that the first hypothesis is accepted and the second one is rejected. This means that on the BlaBlaCar market on the route from Munich to Vienna between 24\textsuperscript{th} September and 10\textsuperscript{th} October the prices are dispersed and that the shape of the dispersion is not similar to the shape the BlaBlaCar-model predicted. Although, there is an interval where the predictions of the BlaBlaCar model meet the empirical shape of the dispersion of prices.
5 Conclusion

In my thesis I searched the answers for two main questions. Firstly, if there is equilibrium price dispersion on the BlaBlaCar market and secondly, what can be the nature of the dispersion of the prices on the BlaBlaCar market?

To find the answers firstly, I restricted the questions for a given route and for a given period of time because I would not have been able to observe the whole market. Then, I collected data from the BlaBlaCar homepage. The analysis of the observed prices and sold seats made it possible to show the shape of the dispersion of the prices with a probability mass function. A statistical analysis proved the existence of equilibrium price dispersion on the BlaBlaCar market on the given route during the given period of time. The data collection reduced the generality of the results, but the existence of price dispersion on the restricted market might imply that the prices are also dispersed on the BlaBlaCar shared economy. It would be too ambitious to state that my results imply equilibrium price dispersion also on shared economies. Proving this, can be subject of a further analysis.

The second step of my analysis was to compare the empirical data with an already existing model. During my research I found the Varian- and the Baye-model the most applicable. According to these models I created the BlaBlaCar-model. Although the results showed that the shape of the price dispersion is different in the model than in the reality it still contains useful information.

For example, the Varian-model is a 36-year-old model which may be outdated and the assumptions does not seem to hold in the current economy where information is not anymore a scarce resource. (Berman and McClellan, 2002, P. 28) The transformation of Baye et al. also implies that they consider information as an expensive and scarce resource which sellers and consumers are fighting for. I believe that the Varian-model can be a good base for other models to observe price dispersion on online markets where information is not anymore costly. This can also be the question of later research.

Between 16 and 20 Euros the prediction and the observation seemed to have relationship. This may show that the idea to use the transformed Varian-model on the BlaBlaCar market

---

23 Introducing a cost for firms which they pay when they want to list their prices on the price comparison site can confirm this idea. Introducing a variable which measures the probability that a consumer knows the list of the prices also implies that the writer do not count with the trend, that information is becoming abundant and what matter more is the attention. Overloading information makes the attention rather a scarce resource. (Berman and McClellan, 2002, P. 28)
was not a useless idea. Obviously this small match can not surely prove the applicability of the model, but it can be a small sign that making difference between two types of consumers is a good start to analyze shared economy markets.

On the other hand, the fact that the BlaBlaCar-model could not describe perfectly the empirical dispersion of prices can not be ignored. The difference between the two probability mass functions could occur because of many reasons. Firstly, as it was mentioned before, not every factor could have been collected from the homepage. My technical skills are limited in this field. Secondly, the information which was observed sometimes was incomplete and therefore, I could not include those factors in the analysis. These two limitations of the data could cause distortions in my analysis and therefore in the BlaBlaCar-model. I would be curious if my prediction holds with larger and more complete dataset.

The BlaBlaCar-model assumes that sellers’ pricing strategy is to randomize their selection on a probability density function \( f(p) \). To be able to model the dispersion of prices such assumptions have to hold but it is obvious that sellers’ random selection on the same density function is far from reality. It would also be interesting to let sellers come in different types, like consumers do. According to my experience with shared-car traveling I realized for example, that there are drivers who do not care about the amount of received money, they are motivated only by non-pecuniary factors like environmental or social ones. Others are very close to the rational agents and they are only interested in the financial factors. This is only a small example but I am sure that observing the sellers’ behavior and applying the observations in the model can make it more applicable for the BlaBlaCar and other car-sharing markets. This could enhance the generality of the model.

The purpose of this thesis was to show a small insight of price dispersion phenomenon on the BlaBlaCar market by comparing a model with a restricted observation. Knowing that the results have their limitations I hope that my work could prove the importance of observing price dispersion on shared economies in the current wasting economy. I also hope that I could make the analysis interesting and considerable for further research to create a more general picture about peer-to-peer markets.
Bibliography


• EINAV, LIRAN, CHIARA FARRONATO, JONATHAN LEVIN (2016). Peer-to-Peer Markets.
  Source: http://www.nber.org/papers/w21496.pdf
  Download: 23.11.2016


  Source: http://abovethecrowd.com/2014/03/11/a-deeper-look-at-ubers-dynamic-pricing-model/
  Download: 23.11.2016

  Source: http://www.dictionaryofeconomics.com/article?id=pde2008_P000368
  Download: 23.11.2016


• SCHWARTZ, JÜRGEN (2016). Logistische Regressionsanalyse.
  Source: http://www.methodenberatung.uzh.ch/de/datenanalyse/zusammenhaenge/lreg.html
  Download: 24.11.2016


• THOMSON, AMY (2015). BlaBlaCar Value Hits $1.6 Billion After Latest Funding Round.
  Download: 23.11.2016


Appendix

Appendix A

Table 7. Table for the Statistical Analysis.

<table>
<thead>
<tr>
<th>Price (EUR)</th>
<th>Day (number)</th>
<th>Age</th>
<th>Picture</th>
<th>Num Drives</th>
<th>Experience</th>
<th>Rating</th>
<th>Num Rating</th>
<th>Car Stars</th>
<th>Size</th>
<th>Music</th>
<th>Smoke</th>
<th>Pets</th>
<th>If Sold Seats?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>33</td>
<td>1</td>
<td>12</td>
<td>49</td>
<td>4.0</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>24</td>
<td>1</td>
<td>2</td>
<td>313</td>
<td>4.0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>24</td>
<td>1</td>
<td>2</td>
<td>313</td>
<td>4.0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>27</td>
<td>0</td>
<td>2</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>27</td>
<td>0</td>
<td>2</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>35</td>
<td>0</td>
<td>5</td>
<td>84</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>35</td>
<td>0</td>
<td>5</td>
<td>84</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>35</td>
<td>0</td>
<td>5</td>
<td>84</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>23</td>
<td>1</td>
<td>14</td>
<td>911</td>
<td>4.9</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>23</td>
<td>1</td>
<td>14</td>
<td>911</td>
<td>4.9</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>23</td>
<td>1</td>
<td>14</td>
<td>911</td>
<td>4.9</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>42</td>
<td>0</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>42</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>42</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>30</td>
<td>1</td>
<td>5</td>
<td>140</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Source: Own work with the help of Microsoft Excel.

This table represents the base of the statistical analysis. Creating Table 1 I used the statistical program SPSS. The dataset contains 16 features about each ride but I had to eliminate 3 because they were incomplete or unmeasurable (Name of driver, Color classification, Departure time). And to avoid multicollinearity problems in the model I had to eliminate 2 variables (Experience and Number of rating). For that I used Table 2, where I observed the Pearson correlation coefficient. I eliminated the variables which correlated with the Price because in this case the random selection of the second type of consumers would not be appropriate.
Table 8. Calculations for the Empirical and Theoretical Probability Mass Functions

<table>
<thead>
<tr>
<th>Price</th>
<th>Empirical $\mathbf{f(p)}$</th>
<th>Theoretical $\mathbf{f(p)}$</th>
<th>$p(a)$</th>
<th>$p(b)$</th>
<th>$F(a)$</th>
<th>$F(b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0,068925494</td>
<td>4,29</td>
<td>5,5</td>
<td>0</td>
<td>0,068925494</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0,045374427</td>
<td>5,5</td>
<td>6,5</td>
<td>0,068925499</td>
<td>0,114299971</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0,038338251</td>
<td>6,5</td>
<td>7,5</td>
<td>0,114299976</td>
<td>0,152638227</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0,033274933</td>
<td>7,5</td>
<td>8,5</td>
<td>0,152638231</td>
<td>0,185915458</td>
</tr>
<tr>
<td>9</td>
<td>0,008845558</td>
<td>0,029490917</td>
<td>8,5</td>
<td>9,5</td>
<td>0,185915551</td>
<td>0,215406467</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0,026573474</td>
<td>9,5</td>
<td>10,5</td>
<td>0,21540647</td>
<td>0,241979944</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0,024279433</td>
<td>10,5</td>
<td>11,5</td>
<td>0,241979947</td>
<td>0,265248799</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0,022433354</td>
<td>11,5</td>
<td>12,5</td>
<td>0,26524882</td>
<td>0,286688236</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0,020939741</td>
<td>12,5</td>
<td>13,5</td>
<td>0,286688238</td>
<td>0,309627949</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0,019718304</td>
<td>13,5</td>
<td>14,5</td>
<td>0,309627951</td>
<td>0,329346256</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0,018715321</td>
<td>14,5</td>
<td>15,5</td>
<td>0,329346257</td>
<td>0,348061568</td>
</tr>
<tr>
<td>16</td>
<td>0,061946903</td>
<td>0,07891802</td>
<td>15,5</td>
<td>16,5</td>
<td>0,348061567</td>
<td>0,365953372</td>
</tr>
<tr>
<td>17</td>
<td>0,053097345</td>
<td>0,072191347</td>
<td>16,5</td>
<td>17,5</td>
<td>0,365953374</td>
<td>0,383172721</td>
</tr>
<tr>
<td>18</td>
<td>0,042427788</td>
<td>0,06677184</td>
<td>17,5</td>
<td>18,5</td>
<td>0,383172722</td>
<td>0,399849906</td>
</tr>
<tr>
<td>19</td>
<td>0,042427788</td>
<td>0,06250379</td>
<td>18,5</td>
<td>19,5</td>
<td>0,399849908</td>
<td>0,416100287</td>
</tr>
<tr>
<td>20</td>
<td>0,042427788</td>
<td>0,05928642</td>
<td>19,5</td>
<td>20,5</td>
<td>0,416100289</td>
<td>0,432028931</td>
</tr>
<tr>
<td>21</td>
<td>0,097345133</td>
<td>0,05705605</td>
<td>20,5</td>
<td>21,5</td>
<td>0,432028932</td>
<td>0,447734538</td>
</tr>
<tr>
<td>22</td>
<td>0,079646018</td>
<td>0,05578476</td>
<td>21,5</td>
<td>22,5</td>
<td>0,447734539</td>
<td>0,463313015</td>
</tr>
<tr>
<td>23</td>
<td>0,141592922</td>
<td>0,05438006</td>
<td>22,5</td>
<td>23,5</td>
<td>0,463313016</td>
<td>0,478861022</td>
</tr>
<tr>
<td>24</td>
<td>0,159292035</td>
<td>0,05618222</td>
<td>23,5</td>
<td>24,5</td>
<td>0,478861024</td>
<td>0,494479846</td>
</tr>
<tr>
<td>25</td>
<td>0,079646018</td>
<td>0,05800818</td>
<td>24,5</td>
<td>25,5</td>
<td>0,494479848</td>
<td>0,510280034</td>
</tr>
<tr>
<td>26</td>
<td>0,042427788</td>
<td>0,06107399</td>
<td>25,5</td>
<td>26,5</td>
<td>0,510280036</td>
<td>0,526378435</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>0,06564221</td>
<td>26,5</td>
<td>27,5</td>
<td>0,526378437</td>
<td>0,542951658</td>
</tr>
<tr>
<td>28</td>
<td>0,088495575</td>
<td>0,07207081</td>
<td>27,5</td>
<td>28,5</td>
<td>0,542951659</td>
<td>0,560158741</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>0,08092618</td>
<td>28,5</td>
<td>29,5</td>
<td>0,560158742</td>
<td>0,578251361</td>
</tr>
<tr>
<td>30</td>
<td>0,017695115</td>
<td>0,019311977</td>
<td>29,5</td>
<td>30,5</td>
<td>0,578251362</td>
<td>0,597563339</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0,02102009</td>
<td>30,5</td>
<td>31,5</td>
<td>0,597563341</td>
<td>0,618583438</td>
</tr>
<tr>
<td>32</td>
<td>0,026548673</td>
<td>0,02302431</td>
<td>31,5</td>
<td>32,5</td>
<td>0,618583441</td>
<td>0,642085872</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
<td>0,02735146</td>
<td>32,5</td>
<td>33,5</td>
<td>0,642085874</td>
<td>0,669437399</td>
</tr>
<tr>
<td>34</td>
<td>0</td>
<td>0,03405398</td>
<td>33,5</td>
<td>34,5</td>
<td>0,669437402</td>
<td>0,703941321</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>0,04895031</td>
<td>34,5</td>
<td>35,5</td>
<td>0,703941325</td>
<td>0,752441641</td>
</tr>
<tr>
<td>36</td>
<td>0,008845558</td>
<td>0,247553353</td>
<td>35,5</td>
<td>36,5</td>
<td>0,752441647</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Own work with the help of Microsoft Excel.

The first column of the table shows a list of price quotes, the second row shows the empirical probability masses generated as follows: first, with the CountIf function the number of each observed price quotes (on which consumers bought seats) are calculated and then this number is divided by the number of all sold seats. The third column shows the theoretical
probabilities of each prices quote which are estimated by the cumulative density function of prices. The fourth and the fifth columns how the lower and the upper values of the intervals which are used to estimate the probability of every integer price quote. The sixth and seventh columns show the lower and the upper value of the cumulative distribution function, where \( k, R, S, n \) are estimated by the following table.

Table 9. Input Values to calculate the Theoretical Cumulative Density Function

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>5</td>
</tr>
<tr>
<td>( M )</td>
<td>2</td>
</tr>
<tr>
<td>( R-M/n )</td>
<td>0,4</td>
</tr>
<tr>
<td>( S )</td>
<td>3</td>
</tr>
<tr>
<td>( r )</td>
<td>36,5</td>
</tr>
<tr>
<td>( k )</td>
<td>14,6</td>
</tr>
<tr>
<td>( p(\text{min}) )</td>
<td>4,29411765</td>
</tr>
<tr>
<td>( F(\text{max}) )</td>
<td>1</td>
</tr>
<tr>
<td>( F(\text{min}) )</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Own work with the help of Microsoft Excel.

Like in the Varian-model (Varian, 1980, P. 657) the exogenously given values (which come from my observation) are: \( M = 2, S = 3, r = 37 \). Substituting them into (8) and (9) we have:

\[
k = 14,6, n = 5, R = 0,4 \text{ and } p^* = 4,29.
\]

It is surprising that the number of the firms is only 5 and the number of ‘Shoppers’ and the ‘Rest’ is 3 and 2 respectively. The Varian-model is well useable for small number of firms (and therefore for small number of the consumers) because in \( n \to \infty \) then \( \lim_{n \to \infty} 1 - \left(\frac{k-pR}{pS}\right)^{\frac{1}{n-1}} = 0 \) and only if \( p = r \), \( F(r) = 1 - \left(\frac{k-rR}{rS}\right)^{\frac{1}{n-1}} = 1 \). This would mean that every price quote has a very low probability to occur (near 0), but the probability that the reservation price occurs is close to 1. This would not be suitable for the BlaBlaCar market so I decided to cluster sellers and buyers. Therefore, I used the color classification. In the observation every price under 25 Euros were ‘green’, after this until 29 Euros ‘orange’ and higher prices were ‘red’. I estimated the color classification of the offer which did not have the classification. Out of the 113 offered seats 82 were sold on a ‘green’ price and 31 on ‘orange’ or on ‘red’. According to these results I assumed that the ratio of the ‘Shoppers’ over ‘Rest’ is \( \frac{3}{2} \) and the number of the clusters of sellers is 5.
The reservation value comes from the dataset. The highest price consumers bought was 36 Euros. To be able to estimate the theoretical probability of 36 Euros I set the reservation price at 36,5 Euros.

The amount of cost and the minimum price is calculated endogenously with the formula (8) and (9) respectively. The lowest price is 4,29 in the model so estimating the probability of 5 Euros I set the lower border of the interval at 4,29 not at 4,4999. This may distort the theoretical probability mass function.

To check that \( F(p) \) is a cumulative density function I calculated the value if \( p \to 0 \) and if \( p \to r \). The results were satisfying and are presented in Table 9.
Appendix C

Script of the program written in Python language:

```python
from lxml import html
import requests
import xlwt
import os
import locale
import calendar
import datetime
locale.setlocale(locale.LC_ALL, 'deu_deu')

date_format = xlwt.XFStyle()
date_format.num_format_str = 'dd/mm/yyyy'
time_format = xlwt.XFStyle()
time_format.num_format_str = 'h:mm'

def parseInt(sin):
    import re
    m = re.search(r'([\d.]+)\[.,\]?', str(sin))
    return int(m.groups()[0]) if m and not callable(sin) else None

def get(l, idx, default):
    try:
        return l[idx]
    except IndexError:
        return default

# main function
def parseSearchResults(filename, excel):
    sheet = book.add_sheet(filename)
    f = open('files/' + filename + '.html', 'r', encoding='utf-8')
    search = f.read()
    search_tree = html.fromstring(search)
    
    # get page of each result
    pages = search_tree.xpath('//a[@class="trip-search-oneresult"]/@href')
    
    sheet.write(0, 0, "Price")
    sheet.write(0, 1, "Color")
    sheet.write(0, 2, "Time")
    sheet.write(0, 3, "Driver Name")
    sheet.write(0, 4, "Age")
    sheet.write(0, 5, "Picture")
    sheet.write(0, 6, "Num Drives")
    sheet.write(0, 7, "Account Created")
    sheet.write(0, 8, "Rating")
    sheet.write(0, 9, "Num Rating")
    sheet.write(0, 10, "Car Stars")
    sheet.write(0, 11, "Bla")
    sheet.write(0, 12, "Music")
    sheet.write(0, 13, "Smoke")
    sheet.write(0, 14, "Pets")
    sheet.write(0, 15, "Seats")
    sheet.write(0, 16, "URL")

    # open and parse individual pages for search results
    num = 1
    for url in pages:
```

44
print('Processing ' + filename + ':' + str(num) + '/' + str(len(pages)) + '\r', end='')
page = requests.get(url)
page_tree = html.fromstring(page.content)

price = get(page_tree.xpath('//span[@class="Booking-price u-block"]//text()'), 0, '').strip()

price_color = ''
if len(page_tree.xpath('//div[@class="Booking Booking--green Block"]')) > 0:
    price_color = 'green'
else:
    if len(page_tree.xpath('//div[@class="Booking Booking--orange Block"]')) > 0:
        price_color = 'orange'
    else:
        if len(page_tree.xpath('//div[@class="Booking Booking--red Block"]')) > 0:
            price_color = 'red'

price_color = price_color.upper()
time = get(page_tree.xpath('//strong[@class="RideDetails-infoValue"]/span/text()'), 0, '').strip().partition(' - ')[2].partition(' ')[0].strip()

name = get(page_tree.xpath('//h4[@class="ProfileCard-info ProfileCard--info u-truncate"]//a/text()'), 0, '').strip()

age = parseInt(get(page_tree.xpath('//div[@class="ProfileCard-info"]//text()'), 0, ''))

pic = 1
if len(page_tree.xpath('//img[@src="https://d1ovtcjitiy70m.cloudfront.net/vi-1/images/avatar/driver-male-72.png"]')) > 0:
    pic = 0

numdrives = parseInt(get(page_tree.xpath('//ul[@class="main-column-list unstyled"]//li/text()'), 0, ''))

membersince_str = get(page_tree.xpath('//li[contains(text(),"Angemeldet seit:")]//li/text()'), 0, '').partition(': ')[2].split(' ')
membersince = ''
try:
    month = list(calendar.month_abbr).index(membersince_str[0][:3])
except ValueError:
    month = 3
if len(membersince_str) > 1:
    membersince = datetime.date(int(membersince_str[1]), month, 1)

rating = float(get(page_tree.xpath('//p[@class="ratings-container tip"]//span/text()'), 0, '-1').partition('/')[0].replace(',', '.')).strip()

if rating == -1: rating = '-'
numrating = parseInt(get(page_tree.xpath('//p[@class="ratings-container tip"]//span/text()'), 1, ''))
if numrating == None: numrating = '-'
carstars = '-'
if len(page_tree.xpath('//span[@class="tip rating-car star_1"]')) > 0:
carstars = 1
else:
    if len(page_tree.xpath('//span[@class="tip rating-car star_2"]')) > 0:
carstars = 2
else:
if len(page_tree.xpath('//span[@class="tip rating-car star_3"]')) > 0:
    carstars = 3
else:
    if len(page_tree.xpath('//span[@class="tip rating-car star_4"]')) > 0:
        carstars = 4
    else:
        blaicon = 
        if len(page_tree.xpath('//span[@class="bla prefs tip"]')) > 0:
            blaicon = 1
        else:
            if len(page_tree.xpath('//span[@class="blabla prefs tip"]')) > 0:
                blaicon = 2
            else:
                if len(page_tree.xpath('//span[@class="blablabla prefs tip"]')) > 0:
                    blaicon = 3
                else:
                    musicicon = 
                    if len(page_tree.xpath('//span[@class="music prefs tip"]')) > 0:
                        musicicon = 1
                    else:
                        smokeicon = 
                        if len(page_tree.xpath('//span[@class="smoking prefs tip"]')) > 0:
                            smokeicon = 1
                        else:
                            if len(page_tree.xpath('//span[@class="no-smoking prefs tip"]')) > 0:
                                smokeicon = 0
                            else:
                                peticon = 
                                if len(page_tree.xpath('//span[@class="pet prefs tip"]')) > 0:
                                    peticon = 1
                                else:
                                    if len(page_tree.xpath('//span[@class="no-pet prefs tip"]')) > 0:
                                        peticon = 0
                                    else:
                                        seats_str = get(page_tree.xpath('//span[@class="Booking-seats u-block"]//b/text()'), 0, 
                                        if seats_str == "Ausgebucht":
                                            seats = 0
                                        else:
                                            seats = parseInt(seats_str)

sheet.write(num, 0, price)
sheet.write(num, 1, price_color)
sheet.write(num, 2, time, time_format)
sheet.write(num, 3, name)
sheet.write(num, 4, age)
sheet.write(num, 5, pic)
sheet.write(num, 6, numdrives)
sheet.write(num, 7, membersince, date_format)
sheet.write(num, 8, rating)
sheet.write(num, 9, numrating)
sheet.write(num, 10, carstars)
sheet.write(num, 11, blaicon)
sheet.write(num, 12, musicicon)
sheet.write(num, 13, smokeicon)
sheet.write(num, 14, peticon)
sheet.write(num, 15, seats)
sheet.write(num, 16, url)
num = num + 1

print()

# program entry point
book = xlwt.Workbook(encoding="utf-8")

# call with filename of downloaded html file of search results,
# individual pages will be opened from there
parseSearchResults('10.07_10.09', book)
parseSearchResults('10.07_10.10', book)

# save results
try:
    book.save('output.xls')
except PermissionError:
    book.save('output2.xls')

print('Done!')
## Appendix D

Table 10. Sample of the Dataset for Observing the Empirical Distribution of Prices

<table>
<thead>
<tr>
<th>Price $/lit.</th>
<th>Estimated Color</th>
<th>Driver Nat. Age</th>
<th>Picture</th>
<th>Num Drive Account</th>
<th>Credit Rating</th>
<th>Num Rating Car</th>
<th>Stars</th>
<th>Bio</th>
<th>Music</th>
<th>Smoke</th>
<th>Pets</th>
<th>Seats</th>
<th>Seats 2</th>
<th>Sold Seats</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 green</td>
<td>Delor G</td>
<td>42</td>
<td>0</td>
<td>2</td>
<td>54</td>
<td>-</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16 green</td>
<td>Zujko P</td>
<td>33</td>
<td>1</td>
<td>3</td>
<td>175</td>
<td>-</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16 green</td>
<td>Peter O</td>
<td>37</td>
<td>1</td>
<td>2</td>
<td>442</td>
<td>4.1</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15 green</td>
<td>Delor A</td>
<td>24</td>
<td>1</td>
<td>4</td>
<td>125</td>
<td>4.7</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16 green</td>
<td>Petra A</td>
<td>33</td>
<td>1</td>
<td>0</td>
<td>933</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17 green</td>
<td>Pavel M</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1211</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>17 green</td>
<td>Delor A</td>
<td>24</td>
<td>1</td>
<td>4</td>
<td>125</td>
<td>4.6</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17 green</td>
<td>Jana M</td>
<td>27</td>
<td>1</td>
<td>2</td>
<td>247</td>
<td>4.9</td>
<td>112</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17 green</td>
<td>Ivan T</td>
<td>23</td>
<td>1</td>
<td>2</td>
<td>242</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18 green</td>
<td>Roman G</td>
<td>34</td>
<td>1</td>
<td>1</td>
<td>522</td>
<td>-</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18 green</td>
<td>Robert H</td>
<td>40</td>
<td>0</td>
<td>6</td>
<td>359</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18 green</td>
<td>Malja</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>345</td>
<td>5</td>
<td>21</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19 green</td>
<td>Cuber R</td>
<td>44</td>
<td>1</td>
<td>1</td>
<td>935</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18 green</td>
<td>Toly A</td>
<td>31</td>
<td>0</td>
<td>5</td>
<td>381</td>
<td>4.7</td>
<td>69</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>19 green</td>
<td>Tomas P</td>
<td>38</td>
<td>0</td>
<td>5</td>
<td>182</td>
<td>4.5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>20 green</td>
<td>Alex M</td>
<td>43</td>
<td>0</td>
<td>6</td>
<td>822</td>
<td>4.9</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>20 green</td>
<td>Johannes I</td>
<td>41</td>
<td>1</td>
<td>1</td>
<td>230</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21 green</td>
<td>Oskar C</td>
<td>25</td>
<td>1</td>
<td>5</td>
<td>481</td>
<td>-</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21 green</td>
<td>Greta H</td>
<td>29</td>
<td>1</td>
<td>5</td>
<td>1227</td>
<td>3.1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21 green</td>
<td>Hans-Peter</td>
<td>49</td>
<td>0</td>
<td>6</td>
<td>192</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21 green</td>
<td>David S</td>
<td>34</td>
<td>1</td>
<td>4</td>
<td>710</td>
<td>4</td>
<td>16</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Own work with the help of Microsoft Excel.
Appendix E

T-test for the Significance of the Pearson Correlation Coefficient

The null hypothesis of the t-test is that there is no significant correlation between the two variables. Here the two variables are the empirical and the theoretical probability mass functions. The formula for the t-test for the correlation coefficient is \( t^* = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} \) where \( r \) is the Pearson correlation coefficient, \( n \) is the number of the variables (which is in this case 31) and \( n-2 \) is the degrees of freedom. In this case \( t^* = -1.249 \). Since \( \alpha = 5\% \) the critical values of the t-test are approximately \( \pm 2.045 \). Below or above these values the null hypothesis is rejected. Now \( t^* \) is between the critical values so the null hypothesis is accepted which means that there is no significant linear relationship between the two observed variables.

The same method is applied to show the significance of the Pearson correlation coefficient between the theoretical and the empirical probability mass functions on the price range from 16 to 20 Euros. The values are presented in Table 11.

| \( t(*) \) | \(-1,249447946\) |
| \( t \) | \( 2,0452 \) |
| \( t(*) \) | \( 14,54015233 \) |
| \( t \) | \( 3,1824 \) |

Source: Own work with the help of Microsoft Excel
Declaration of Own Work Statement

I, Fruzsina Nagy confirm that all this work is my own except where indicated and that I have:

- Clearly referenced/listed all sources as appropriate
- Referenced and put in inverted commas all quoted text of more than three words (from books, journals, the internet, etc.)
- Given the sources of all figures, tables data, etc. that are not my own
- Not made any use of the essay(s) or other work of any other student(s) either past or present at this or any other educational institution
- Not submitted for assessment work previously submitted for any other course, degree or qualification at this or any other educational institution
- Acknowledged in appropriate places any help that I have received from others
- Complied with any other plagiarism criteria specified in the Course or Program handbook
- Included an accurate word count, if requested.

Fruzsina Nagy

28.11.2016