STATISTICS

Dalarna University

Master's Thesis 2007

Does a Booking Fee Hold Back Members' Short Trip in Falu Bilpool: A GLMM Approach

Author: **Hao Li**

Registration no.: **850808-T034**

Supervisor: **Kenneth Carling/Ola Nääs**

Abstract

Carsharing is one form of economic association where a number of persons share the use of one or more cars. Carsharing brings a wide variety of positive effects to both the individuals and the society, which is the reason why Carsharing develop with high speed. This paper analyses the data from Falu bilpool[1], by which I try to find how the booking fee affects the booking and use of cars. Because of repeated use of cars for most of members, this paper models this kind of repeated measurements by generalized linear mixed models. As the result, booking fee is found taking negative effects on short-trip use. And the effects are considerable for short-trip during 40 to 80 km and 2 to 5 hours. Comparing the change of short-trip frequency and income of Falu bilpool, this paper finds that increasing booking fee from 0 to 10 SEK was harmful to both Falu bilpool and its members.

KEY WORDS: Carsharing; Generalized linear mixed models; R ;

1 Introduction

1.1 Introduction

"If you live in a city, you do not need to own a car."[2] And if you live in a city… and if you drive less than 10,000 kilometers a year…you should probably consider carsharing.[3] So what is car sharing? One hundred years after the invention of the automobile, which has since had a great impact throughout the industrialized world, especially in the form of the private car, the time grew ripe for the introduction of the public car. At the end of the 1987, an organized form of Carsharing first was found in central Switzerland, and shortly afterwards in Zurich, and around a year later was also introduced in Berlin. The idea behind these developments was very simple: instead of each individual person buying her own car, a large number of people share a small

[1] Falu bilpool is a carsharing association in Falun city, Sweden.
[2] William Clay Ford Jr., CEO, Ford Motor Company Ltd. October 2000.
[3] http://worldcarshare.com/

number of cars that are reserved for them and used individually as required. In this way, the utilization capacity of each car is higher, and considerable benefits can be achieved in terms of both cost savings and reduced impacts on the environment. [5]

1.2 Background

Carsharing is also named carpooling. Swedish National Road Administration gave a formal definition in one report published in 2002 that Car-sharing means that a number of persons share the use of one or more cars.[6] Use of a car is booked beforehand, the user paying a fee based on the distance driven and the length of time the car was made use of.

Eventhough Carsharing is similar in some extent to car rental, there are some distinct differences. Car rental is used mainly outside the urban area with long-distance travel. While most use of Carsharing is around the city and working area. Car rental is not used frequently for some person, while carsharing is used regularly by the club members. When renting a car, you should take the car from the office where cars are collected. Carsharing locate its cars in different places concerning the convenience of members. Carsharing is a kind of club or association not a commercial company as car rental and it just need a contact when joining. So carsharing is easy-entered and easy-used. Member of carsharing association just pay the entrance fee and annual fee if any. That means member need not pay for the petrol, unlike car rental. So carsharing is usually much less costly.

Carsharing has a wide variety of positive effects, which is also why carsharing meet rapid success. Carsharing will reduce the number of private cars, which will result in relax of traffic problem in working time, reduction of energy consumption and carbon dioxide emissions by cars. Individually carsharing will give the member “great freedom of choice thanks to lower fixed costs”.[5] And the member also would save lots of fuel and insurance fees.

In Sweden, carsharing come out very early. The first carpool Majornas Bilkooperativ is formed in 1988. But the rapid growth in national area starts from

1998. [4]

1.3 Falu bilpool

For this paper, I pay attention to carsharing (carpooling) in Sweden, specifically in Falu bilpool where the data used in this paper come from. Falu bilpool is one carsharing economic association formed in Falun at October 11, 2000.[2] It mainly serves the people living in Falun. Today, it has 45 member families and 3 cars available in three different places. The membership is in unit of family. Members of each family can become a user of cars. The second user in a family pay less enter fee than the first one to become user of the car. Annual fee is paid by member. Member also can get lower prices for car rental. So a few members just paid for getting the membership but did not use the carpools' cars.

As a member want to use a car, the member can book it through the on-line booking system. The user uses its family member number to login in the booking system, and chooses the time and car it wants to book. The user must use the car within the duration booked and return it on time. When finishing using the car, users should record the actual duration of use and driven distance on the book located in the car. Officers of carpool regularly check the records in the car and calculate the fees. The fee of each use is calculated according to both time (hours) and distance (Swedish miles[5]) that car was driven. The price per hour booked is called Time fee in this paper and the price per Swedish mile driven is Driving fee in this paper. Start fee (booking fee) started from 2004.10.1. Every booking needs to pay booking fee no matter whether the car is used or not. The member just pay booking fee once if the bookings and uses are connected in time. Every member of the carpool has one account, where fees of each use would be paid.

The problem I focus on is whether and how the change of booking fee affects the members' booking and use of the cars, which I try to find a generalized linear mixed model to analysis. In this way I try to give a pattern of the effects of booking fee

[4] More information can be found in Swedish national road administration whose website is www.vv.se .
[5] 1 Swedish mil = 10 kilometers.

adjustment on activities of car-pool members, and form some useful advices.

2 Data description

Ingemar Elb and Torsten Hylén for offering the original data of Falu bilpool.[6] The original data (see Appendix A) merges facts of information being booking & account records, change of fees and change of car & parking location. Booking & account records consist of the records for every booking (use) of cars or change of account from 2001 to 2006. Booking & account records are saved in Excel file, in which each row is one record of booking or use of cars, or depositing in the account. Every record before 2004.10.1 contains 10 items which are member no. (Nr), name of family member (Namn), date of the record (Datum), explanation (Beskrivnin), duration of use (Tim), driven distance (Mil), bill no. (Vernr), cost or money deposited for this record (Belopp), tax (Moms) and car type (Bil). The booking fee (STARTAVGIF) starts from October 1st, 2004, as a result the booking fee become one more item for each record. Please note that the negative values in Belopp mean what member needs to pay for the use, whereas positive values mean that money deposited in the account. The data of Change of fees show when and how the fees changed, which contain the entrance fee, annual fee, booking fees and the fee for driving per Swedish mile. The data of Change of car & parking not only give the type and use duration of cars, but also clarify when and how the parking location changed.

Looking through the original data, we find there are lots of records, mainly being booking & account records, not relevant for the question of this essay. So according to the question involved, we clean the booking & account records through the following 6 steps. [7]

After looking over all the records from 2001 to 2006, we find some records of 2003 being different from other records. The difference is that a record would calculate the total cost of several booking & use of cars in the last month or several weeks, instead of calculating the cost of each record individually. For the example in

[6] I am grateful for them giving me the opportunity to study the data at the bilpool!
[7] The data was organised by Qiguang Gao, Shanshan Yang, Yiqi Yang and I as a group.

Appendix B, the family No. 12611 booked and used the car for 6 times without paying the cost where the cost (Belopp) was 0. Instead, the total cost for the 6 times of use was paid at 2003.2.25. To make all the records on the same form, for the cases above we delete the record of total cost and calculate and add the cost of each booking and use by the related time price and driving price. This is the first step in making the data in the same form.

This paper focuses on how the fees change affect the use or booking of cars. So the information unrelated to this question can be discarded. By explanations (Beskrivnin) and Ingemar's clarification, we know that all the values of duration and driven distance are actual time and distance used by the member. And we also find there are several kinds of records out of the interest (see Appendix C). One kind of records is members' depositing money in the account and paying annual fees. Another kind of records, where the booking fee (if any), duration and driven distance are set to 0, is for transporting or accounting. Like its members, Falu bilpool itself also books and uses the car for its business. Records with member no. 0 are for Falu bilpool own booking and use. The member SWECO with member no. 46400 is not a family member but a company member, which is of no concern for this paper. All these four kinds of records are not of use for this paper. As the second step, we deleted all the four kinds of records described above, and just keep the records once one of the three items booking fee (if any), duration (Tim) and driven distance (Mil) is a non-zero value.

Next, we find some negative value for duration and distance item (see Appendix D). By the help of Ingemar Elb, we know that the negative values are the corrections for related earlier records. According to the explanation (Beskrivnin) and clarification of Ingemar Elb, we corrected the related earlier records. Through this step, we got the records with duration and driven distance which the members actually paid for. As a result, all the values of duration and driven distance are positive values.

When checking the booking fee in the records, we find some booking and use did not pay the booking fee (see Appendix E). Ingemar explained the reason being that the use without of booking fee was considered an extension of the earlier booking and

use. For example, no. 28638 member Sylvia booked the car from 11:00 to 12:00, but for some reason she did not drive the car. However she still needed to drive the car within the next few hours, so she re-booked the car from 12:00 to 18:00 and used it. In the case above, the booking fee was paid at the first booking. To simplify the data, I decided to merge the two records into one record by adding the duration, driven distance and cost together.

When checking the data, we also found some mistake with 2 records (see Appendix F). Checking the original bill by Ingemar, we found the right record and corrected it.

Next we added the annual fee, household owning car, driving price, booking fee (start fee) and number of members to the booking & use records as two new columns, where I did a simplification of driving price. The driving fee became 2 levels from 2006-9-1 instead of all the cars associated same driving price. I simplified the driving price after 2006-9-1 as 20 for all the cars, instead of 21 for Stor car and 19 for Liten car (see Appendix G). The variable “Household owning car” has two levels where level 0 means the member did not own car when using pool cars and level 1 means the member owned cars when using cars of the carpool. In the data of household owning car, there are just 5 out of 67 members who had or have their own cars.

Then we deleted the column tax (Moms), which is of no concern for this paper and being equal to 20% of the cost (Belopp). By this, the booking & use records contain most of data that we need. To make the unit of driven distance and Mile fee more understandable, we change the unit from Swedish mil to kilometer by multiplying 10 to driven distance and dividing mile fee by 10. We also change the sign of the cost to make it positive, which would make modeling the data easier.

This paper plans to find whether the booking fee reduces the frequency of short-trip use, to which booking fee owns more reasonable effect than long trip. Short-trip variable is defined as 1 if the member used to drive a distance less than 60 km and duration of at most 2 hours, otherwise it was set to 0. Now we got the data we would use, whose example is shown in Appendix H.

Before modeling the data, I give some information of data by following tables.

The information about total cost and its buildup in different years is shown in Table 1. The information on distribution of Cost can be found from quantiles and other statistics of it in different situation shown by Table 2.

I calculated the proportion of number of short-trip taken from total number of booking (trip) related to different 12 months, which is shown in Figure 1. I find short-trip has the highest frequency in November and January, and from February to October members take long trip more likely, especially in April, July and August when Swedish people enjoy the spring and summer. So month also is a factor owning effects on frequency of short-trip.

Table 1: Yearly bookings and number of members associated costs.

Year	Number of members[8]	Number of bookings	Sum of Cost (SEK)	Average cost (SEK)	Cost divided by fees (%)		
					Booking	Driving	Time
2001	21	148	78787.0	532.34	0	61.73%	38.30%
2002	32	395	158453.2	401.15	0	67.55%	31.90%
2003	40	847	203691.4	240.49	0	64.43%	35.57%
2004	49	1041	228887.8	219.87	1.03%	62.01%	36.80%
2005	55	953	233908	245.44	4.07%	63.82%	32.79%
2006	49	845	212618.7	251.62	7.95%	58.67%	33.38%

Table 2: Quantiles of Cost corresponding to different levels of booking fee.

Booking fee	Quantiles of Cost											
	Min	5%	10%	25%	50%	75%	90%	95%	Max	No. of obs	Mean	Sd.
0	5.0	22.6	34.2	65.0	115.9	231.9	529.7	1049.5	9192	2196	281.146	603.483
10	15.0	35.3	45	72.0	123.0	228.3	524.0	789.3	5329	1188	241.019	408.066
20	25.0	46.1	59.4	88.0	141.0	220.0	476.9	743.4	4496	845	251.620	437.961
Overall	5.0	30.0	41.2	72.0	121.2	228.0	515.4	914.6	9192	4229	263.974	523.874

I counted the number of observations for each member (see Appendix I), quantiles of which are given in Table 3. It is obvious that most of members booked the car more than once.

Table 3: Summary of booking frequency for each member

Booking frequency	Number of Members

[8] Number of members in December of each year.

1~3	6
4~9	11
10~30	15
31~60	12
61~100	10
101~200	8
200+	5

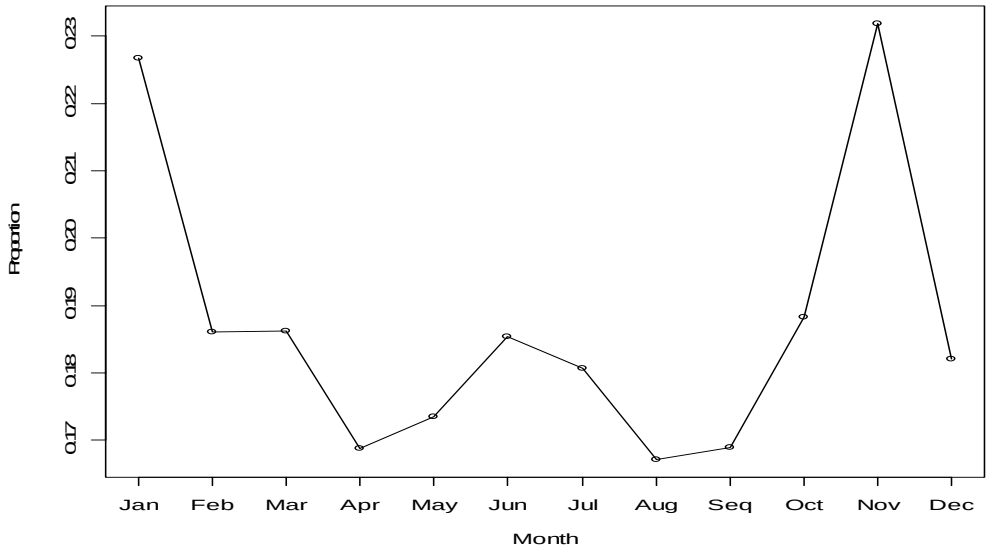

Figure 1. Proportion of number of short-trip taken from total number of booking related to different 12 months.

4 Method and Model specification

4.1 Method

To modeling the data, first thing is to find what kind of data it is and what method is suitable for modeling it. Due to the question of this paper, the response variable should be one indicator variable, named “shtrip”, 1 for short-trip use while 0 for long-trip use. And I use booking fee as the causal variable, while I use annual fee, and driving fee, month, number of member and household owning car as control variables. Because the response variable has a Bernoulli distribution, generalized linear (logistic) model should be used. This data contain 4225 observations of every use in the last 6 years for 67 family members. As the result of Table 3 shows, the observations are grouped by members, and these observations are repeated measurements for each member over time, which make this data be some kind of longitudinal data. So a good

option of modeling method for this data could be a generalized linear mixed model.

Generalized linear mixed models (GLMMs) are generalized linear model with random effects. The basic linear model means $E(y) = \boldsymbol{X}\beta$. Then random effects are incorporated by enlarging the model as $E(y \mid u) = \boldsymbol{X}\beta + \mathrm{Zu}$. If we write a combined model matrix $\mathbf{X}^* = [\mathbf{X} \quad \mathbf{Z}]$ and an enlarged coefficient vector $\boldsymbol{\beta}^* = [\boldsymbol{\beta}' \quad \boldsymbol{u}]$, then it is obvious that $E(y \mid u) = \boldsymbol{X}^* \boldsymbol{\beta}^*$. So GLMMs have conditional distributed response $\mathbf{y}$ given $\mathbf{u}$, each element of which independently have a exponential family distribution

$$y_i \mid u \sim indep. f_{Y_i|u}(y_i \mid u)$$
$$f_{Y_i|u}(y_i \mid u) = \exp\left\{\left[y_i\gamma_i - b(\gamma_i)\right]/\tau^2 - c(y_i, \tau)\right\}$$
$$E(y_i \mid u) = \mu_i \ ; \ \ g(\mu_i) = x_i'\beta + z_i'u \ .$$

And random effects are assigned a distribution $u \sim f_U(u)$.[1]

We will mainly use the R package “lme4” to model this data as following.[2] The factor booking fee has 2 levels i.e. 10 and 20, while the factor annual fee has 3 levels i.e. 200, 400 and 500. Both of factors booking fee and annual fee are fixed effects, which are decided by carpool board and do not have suitable population. Because there are no random factors in the data, a mixed model with only random intercept effects is suitable. The model is in the form:

$$\mathrm{y}_{\mathrm{ijklmn}} = \mu + b_l + \beta_j * \mathrm{Bookingfee} + \beta_i * \mathrm{Annualfee} + \beta_k * \mathrm{Drivingfee} + \beta_m * household car + \beta_n * month + \beta_{\mathrm{o}} * \mathrm{nmember} + \varepsilon_{\mathrm{ijklmn}}$$

, where i=200, 400, 500; j=10, 20; k=1.5, 1.6, 1.8, 2.0; l is group level of Member; b_l are the random-effects intercept; m =0 when member do not own a car, otherwise m =1; n = 1, 2…12, the levels of month; y has a Bernoulli distribution i.e.

$$y = \begin{cases} 1, \text{ for short-trip use} \\ 0, \quad \text{otherwise} \end{cases} .$$

4.2 Result and comments

Firstly I model the data with the above formula. But there is something wrong with

this formula.[9] Error come out with R code above: devLaplace(PQLpars) : Leading minor of order 8 in downdated X'X is not positive definite. Then I do the generalized linear model with same formula, by which I try to find the reason. From the output, the estimates of factor 500 of Annual fee and factor 2 are missing. Checking the model matrix by R code x<-model.matrix(g1); solve(t(x)%*%x), I found the matrix is singular. So checking the data, I found the booking fee turned to 20 as the annual fee became 500, and booking fee started being 10 while driving fee turned to 20. This kind of high correlation makes the model matrix became singular and lead to the error. To avoid this I consider the annual fee and driving fee as continuous variable, and then do the complete model with binomial distribution and logit link, whose output is shown in Table 4. [10]

Table 4: Estimated effects of a booking fee

Fixed effects			
	Estimate	Std.Error	t-value
Causal variable			
factor(Bookingfee)10	-1.747	0.662	-2.636
factor(Bookingfee)20	-1.575	0.380	-4.143
Control variable			
(Intercept)	-15.266	4.585	-3.33
Annual fee	-0.0044	0.0037	-1.189
Drving fee	9.514	3.286	2.895
Household owning car	-0.074	0.333	-0.221
Number of members	-0.039	0.0097	-3.997
Month(1:12)	Included (see detail in Appendix J)		
Random effects			
Groups	Name	Variance	Std.Dev.
Member	(Intercept)	1.00027	1.00014
Residual		0.96702	0.98337
number of obs: 4225, groups: Member, 67			

The AIC of this model is 3761. Because the logit link is the nature link of

[9] Lmer1=lmer(shtrip~factor(Bookingfee)+factor(Annualfee)+factor(owncar)+factor(Milefee)+factor(month)+(1|Member),data,family=quasibinomial(link="logit"))

[10] lmer2<-lmer(shtrip~factor(Bookingfee)+Annualfee+Milefee+factor(owncar)+factor(month)+nmemb+(1|Member), family=quasibinomial(link="logit"), data)

binomial distribution, I take the models with logit link as the most reasonable model. From Table 4, I know the estimate of factor "Household owning car" is not significant with t value -0.221, for which p-value is 0.8251. And the p-vaule for "Annualfee" is 0.2345, which is not significant for significance either. Importantly, both of the estimates of booking fee is significant, and negative which mean the booking fee 10 SEK and 20 SEK have negative effects on short-trip use. The estimate of booking fee 10 SEK is -1.747, whose Odds ratio $e^{-1.747}$ is less than 1, which means the odds of short-trip use when booking fee was 10 SEK is $e^{-1.747}$ (i.e. 0.174) times less than the probability without booking fee. Similarly, the estimate of booking fee 20 SEK is -1.575, whose Odds ratio $e^{-1.575}$ is less than 1, which means the odds of members using cars for short trip when booking fee was 20 SEK $e^{-1.575}$ (i.e.0.207) times less likely than before booking fee started.

4.3 Sensitivity analysis

Because short-trip use can be defined under different driving distance and duration, I want to check the sensitivity of effects of booking fee against different definitions of short-trip use. To show the sensitivity of booking fees' effects more clearly, I simplify the levels of booking fee as two levels rather than three levels, i.e. level 0 means there is no booking fee while level 1 means booking fee exists. Then I model different response variable defined under different driving distance (Km) and duration (hours), and find the estimate of level 1 of booking fee β_1. Then I calculate the effect of paying booking fee by e^{β_1}. The estimates and effects of booking fee related to different short trip definitions are given in Table 5 and shown graphically in figure 2.

From Table 5 and figure 2, we can find all the effects of booking fee on short-trip, defined within different driving distance and duration, are negative. And the curve of effects has a "U" shape, which means the negative effect increases firstly and decreases next, as total cost increases or the trip become longer. I think there are two reasons why negative effects are small for short-trip within 30 km and 1 hour. One is

that number of this kind of short-trip is too small to be well estimated. And the other reason is that total cost for this kind of short-trip is close to minimum cost once car was booked and used. The reason why negative effects are small for short-trip within 150 km and 8 hours or larger distance and duration, is that booking fee takes very small proportion in total cost, in which case booking fee should be expected only marginally.

Table 5: Part of results of models related to different short-trip definitions

Short-trip definitions (<=distance,<= duration)	Estimates of booking fee	Effects of booking fee	t-value of estimates	AIC of model with different short-trip	Number of short-trip
(<=30,<=1)	-0.471	0.624	-1.030	1706	241
(<=30,<=2)	-0.768	0.464	-2.145	3269	636
(<=40,<=2)	-0.861	0.423	-2.414	3393	679
(<=40,<=3)	-0.986	0.373	-3.205	4466	1128
(<=50,<=2)	-1.513	0.220	-4.474	3723	773
(<=50,<=3)	-1.671	0.188	-5.638	4762	1346
(<=60,<=2)	-1.509	0.221	-4.497	3759	791
(<=60,<=3)	-1.677	0.187	-5.783	4822	1387
(<=70,<=2)	-1.517	0.219	-4.526	3768	796
(<=70,<=3)	-1.674	0.188	-5.801	4848	1403
(<=70,<=4)	-1.450	0.235	-0.134	44101	1888
(<=80,<=4)	-1.360	0.257	-0.128	44163	1910
(<=90,<=5)	-0.911	0.402	-0.103	43125	2308
(<=100,<=5)	-0.933	0.394	-0.107	43017	2325
(<=150,<=8)	-0.479	0.620	-2.097	4395	3135
(<=200,<=10)	-0.273	0.761	-1.126	3736	3445

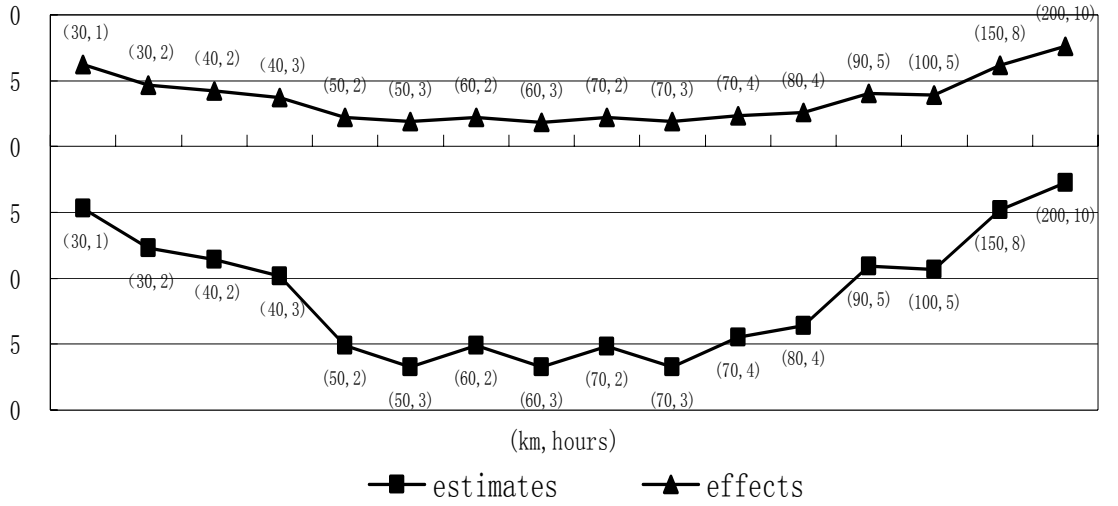

Figure 2. Sensitivity analysis of effects of booking fee against different definition of short trip

4.4 Does booking fee benefit the income of Falu bilpool?

We have found that booking fee have had a negative effects on short-trip use. But maybe booking fee increases the income of car pool. Now I try to find the change of the income and booking frequency as caused by introducing booking fee. Firstly I try to find the change when booking fee became 10 SEK. We know that Odds ratio $= \frac{p_{10}/(1-p_{10})}{p_0/(1-p_0)} = e^{\beta_{10}}$, where p_{10} and β_{10} are the probability of short-trip use and coefficient for booking fee being 10 respectively. $\hat{p}_0$ is the probability of short-trip use as booking fee supposedly being 0 SEK. Since the probability of short-trip for booking fee being 10 p_{10} can be calculated from the data, the value of $\hat{p}_0$ corresponding to p_{10} can be found by

$$\hat{p}_0 = \frac{C}{1+C}, \text{ where } C = \frac{p_{10}}{e^{\beta_{10}}(1-p_{10})}. \qquad \textbf{(1)}$$

Here I assume the booking fee has no effects on long-trip use, which means the number of long-trip use is fixed no matter whether booking fee started or not. Thereafter, I calculate $Nstrip_0$ the number of short-trip supposing booking fee being 0 by $Nstrip_0 = \frac{\hat{p}_0 \cdot Nltrip_{10}}{(1-\hat{p}_0)}$, where $Nltrip_{10}$ is the number of long-trip use. Then the income based on this estimate of number of short-trip use can be found by $Nstrip_0 * avg\cos t_{short} + Nltrip_{10} * avg\cos t_{long}$, where avgcost_i means average cost for i being short and long trip respectively. Then I calculate the 95% confidence interval of $\hat{p}_0$, $Nstrip_0$ and income based on the interval estimates of β_{10} (see Appendix K). The comparison is shown in Table 6.

Next I want to compare the income and booking frequency between booking fee being 10 and 20. Here the odds ratio between booking fee being 10 and 20 is calculated by Odds ratio $= \frac{p_{20}/(1-p_{20})}{p_{10}/(1-p_{10})} = e^{\beta_{20}-\beta_{10}}$, where β_{20} is the coefficient of booking fee being 20 SEK. Similarly as above, the number of short-trip supposing

booking fee 10 can be got through

$$Nstrip_{10} = \frac{\hat{p}_{10} \cdot Nltrip_{20}}{(1-\hat{p}_{10})}, \text{ where } \hat{p}_{10} = \frac{\frac{p_{20}}{e^{\beta_{20}-\beta_{10}}(1-p_{20})}}{1+\frac{p_{20}}{e^{\beta_{20}-\beta_{10}}(1-p_{20})}}.$$

The income can be calculated as above. Then I calculate the 95% confidence interval of $\hat{p}_{10}$, $Nstrip_{10}$ and income based on the interval estimates of β_{20} (see Appendix K). The comparison is shown in Table 7.

Table 6 Comparison between booking fee being 10 and 0 SEK

Comparison between booking fee being 10 and 0SEK Odds ratio = 0.1743761 Average cost of short-trip use = 55.78325 Average cost of long-trip use = 279.8589			
	Booking fee=10	Suppose booking fee=0	
		point estimates	95% confidence interval
Number of long-trip	985	985	-
Number of short-trip	203	1164.151	(1141.161, 1187.643)
Probability of short-trip	0.1708754	0.5416794	(0.5367, 0.5466)
Income of Falu bilpool	286985	319109.6	(318057.0, 320185.2)

Table 7 Comparison between booking fee being 20 and 10 SEK

Comparison between booking fee being 20 and 10 SEK Odds ratio = 1.187402; Average cost of short-trip use=64.8336 Average cost of long-trip use=284.9651			
	Booking fee=20	Suppose booking fee=10	
		point estimates	95% confidence interval
Number of long-trip	717	717	-
Number of short-trip	128	107.7983	(105.325, 110.287)
Probability of short-trip	0.1514793	0.1306966	(0.1281, 0.1333)
Income of Falu bilpool	212618.7	203061	(202925.3, 203197.4)

From Table 6 we find both frequency of short-trip and income of Falu bilpool for booking fee being 10 SEK smaller than that it should be if booking fee still was 0. The decrease of short-trip frequency means the members suffered from booking fee being 10 SEK. At the same time, Falu bilpool did not benefit from booking fee being 10 SEK at all neither, since the income decreased. So in this case, increasing booking fee from 0 to 10 SEK is really a bad decision. Different from Table 6, Table 7 tells us

the booking fee increase the frequency of shirt-trip and income a little. I think the reason for this can be that members became accustomed to the booking fee. The change of booking fee made the member group steadier, maintaining members insensitive to booking fee.

5 Conclusions

This paper has analyzed the data with generalized linear mixed model. The causal variable booking fee, which this paper considers, is found to have a negative effect on the short-trip use. The sensitivity analysis showed the region where short- trip is defined sensitive to the change of booking fee. Booking fee has considerable negative effects on trips between 40 and 90 km and 2 and 5 hours. Comparing the frequency of short-trip use and income with what they would supposedly be at a booking fee of 0, I find the change of booking fee from 0 to 10 SEK reduce both the short-trip frequency and income of the bilpool, which means this change of booking fee is not a good decision for either Falu bilpool or its members.

Appendix A: Original data example (in Swedish)

Table A1: Example of original data.

Nr	Namn	Datum	Beskrivnin	Tim	Mil	Vernr	Belopp	Moms	Bil
15462	Sven	2003-4-14	Rättelse	-2.5	-2.5	03108	70.00	-14.00	CR
19422	Peter Sjö	2003-5-19	Insättning[11]	0.0	0.0	03109	1500.00	0.00	
10517	Sonja Wahlstein	2003-5-20	Insättning	0.0	0.0	03110	700.00	0.00	
12611	Mattias Ahlstedt	2002-8-4	Debitering aug Bojs	10.0	8.0	02224	-244.00	0.00	B
12611	Mattias Ahlstedt	2002-12-13	Debitering dec City	39.0	12.0	03013	-606.00	0.00	CR
26822	Ingemar Elb	2001-3-1	Förskott	0	0	01012	1000.00		
26822	Ingemar Elb	2001-3-3	Hyra	3	8	01052	-150.00		P

Table A2: Example of the records with booking fee (Startavgif).

Nr	Namn	Datum	Bil	Starttid	SlutDatum	Sluttid	Startavgif	Tim	Mil	Beskrivnin	Belopp	Moms	Vernr
23483	Sundh Ewa	2004-5-2			1900-12-31[12]		0.00	0.0	0.0	Årsavgift	-400.00	80.00	04119
24288	Torsten Hylén	2004-10-7	B		1900-12-31		10.00	2.5	1.6	Obokad bil	-67.00	13.40	04311

Table A2: Example of the records with start time (STARTTID) and end time (SLUTTID).

NR	NAMN	DATUM	BIL	START TID	SLUT TID	STARTAV GIF	ANTSTA RT	TIM	MIL	BESKRIVNIN	BELO PP	MO MS	VERN R
286 38	Sylvia Kjellberg	2005-2-9	K			10	1	6	19	Överskriden bokningstid	-446	89	05061
132 28	Anders Ahlén	2005-8-28	B			0	0	0	0	Avgift Överskriden bokningstid	-100	80	05245
132 28	Anders Ahlén	2005-8-28	B			10	1	36	129	Överskriden bokningstid	-2937	587	05244
631 81	Anders Thuresso n	2006-3-10	CFB	11:30	09:00	20	1	13.5	18.1		-517	103. 4	06068
193 36	Fredrik Holmborg	2006-3-11	B	12:00	17:30	20	1	5.5	2.3		-121	24.2	06067

[11] The Swedish word Insättning means depositing money in the account to pay for using the pool car.
[12] This mistake results from booking system error.

Appendix B: Example of special records in 2003

Table B: Example of special records in 2003

Nr	Namn	Datum	Beskrivnin	Tim	Mil	Vernr	Belopp	Moms	Bil
12611	Mattias Ahlstedt	2003-1-8	Debitering januari	4.0	3.4	03031	0.00	0.00	CR
12611	Mattias Ahlstedt	2003-1-11	Debitering januari	4.0	2.4	03031	0.00	0.00	CR
12611	Mattias Ahlstedt	2003-1-13	Debitering januari	8.0	2.8	03031	0.00	0.00	CR
12611	Mattias Ahlstedt	2003-1-24	Debitering januari	4.0	2.3	03031	0.00	0.00	CR
12611	Susanna Ahlstedt	2003-1-27	Debitering januari	0.5	0.6	03031	0.00	0.00	CR
12611	Mattias Ahlstedt	2003-1-29	Debitering januari	4.0	2.2	03031	0.00	0.00	CR
12611	Mattias Ahlstedt	2003-2-25	Debitering januari	0.0	0.0	03031	-502.00	100.40	

Appendix D: Examples of correction related to negative value.

Table D1

NR	NAMN	DATUM	BIL	STARTTID	SLUTDATUM	SLUTTID	STARTAVGIF	ANTSTART	TIM	MIL	BESKRIVNIN	BELOPP	MOMS	VERNR
20737	Ingrid Haglund	2005-7-25	CFR				0	0	0	-4	Rättelse 26 juni	80	-20	05190
32142	Lisbeth Olsson	2005-7-25	CFR				0	0	0	4	Rättelse 26 juni	-80	20	05190

*Ingrid Haglund had been charged for 4 miles too much which Lisbeth Olsson should pay for.

Table D2

26334	Eva Andersson	2006-6-23	B	02:07	2006-06-23	17:06	20	1	10	0	Sen förkortning!	120		06152
26334	Eva Andersson	2006-7-23	B	11:00	2006-07-23	15:00	0	0	-4	0	Åter timavgift	40	-8	06189

*Eva had been charged for 4 hours too much. Her booking were cancelled too late but another person (Fredrik Holmborg) had used the car from 14:00 to the next day. Note that the date in the correction row should be 2006-06-23 and not 2006-07-23!

Appendix C: Example of records discarded.

Table C: Example of records discarded

NR	NAMN	DATUM	BIL	STARTTID	SLUTDATUM	SLUTTID	STARTAVGIF	ANTSTART	TIM	MIL	BESKRIVNIN	BELOPP	MOMS	VERNR
18408	Karin J	2005-1-1					0	0	0	0	IB	36.6	0	05000
6860971	Fredrik Månsson	2005-1-5					0	0	0	0	Inbetalning	3000	0	05004
30194	Kjell Nerhagen	2005-1-10	CFR				0	0	0	0	Hämtning körjournal	0	0	05037
26822	Ingemar Elb	2005-2-5					0	0	0	0	Inköp av lamineringsmaskin	348	0	05031
18408	Karin Jöngren	2005-4-24					0	0	0	0	Årsavgift 2005	-400	80	05104
0	Sven Håkansson	2005-3-11	CFR		2005-03-11		0	0	0	0	Service bilgruppen	0	0	05096
0	Niclas Hedin	2005-12-16	CFB	17:00	2005-12-16	18:00	0	0	0	0	bilvisning	0	0	05338
46400	SWECO VIAK	2005-3-29	CFB		2005-03-29		10	1	10	30	Marie E. 152007810	-704	141	05097
46400	Hanna A	2005-2-4	CFB				10	1	2	4	2412255	-113	23	05062

Appendix E: Example of merging the records belonging to one same booking.

Table E1: Example 1

NR	NAMN	DATUM	BIL	STARTTID	SLUTDATUM	SLUTTID	STARTAVGIF	ANTSTART	TIM	MIL	BESKRIVNIN	BELOPP	MOMS	VERNR
28638	Sylvia Kjellberg	2006-2-18	CFR	11:00	2006-02-18	12:00	20	1	1	0		-30	6	06048
28638	Sylvia Kjellberg	2006-2-18	CFR	12:00	2006-02-18	18:00	0		6	1.1		-82	16.4	06048

Table E2: Example 2

Nr	Namn	Datum	Bil	Starttid	Tim	Mil	Belopp	Moms	Vernr	
26822	Ingemar Elb	2004-11-19	CFR	10	26	0	-270	54	4349	26 hours from 2005-11-19 to 2004-11-20. Max charge on 2004-11-20 is 16 hours. The rest is probably on 2004-11-19. From 13:00 - 23:00?[13]
26822	Ingemar Elb	2004-11-21	CFR	0	3	21	-451	90.2	4349	From 07:00 - 09:30 makes 2.5 hours. Hours during the night 23:00-07:00 is free.
26822	**Ingemar Elb**	**2004-11-19**	**CFR**	**10**	**29**	**21**	**-721**	**144.2**	**4349**	**Total charging for the booking**

*Note the car and bill no. (Vernr) in the record unified is same.

Appendix F: Example of records with mistake.

Table F1

NR	NAMN	DATUM	STARTAVGIF	TIM	MIL	BELOPP	VERNR	BIL
19347	Ulf Söderholm	2006-9-30	0	3.00	0.00	-30	06229	CFR

*He should have paid 20 kr in starting fee![14]

Table F2

Nr	Namn	Datum	Bil	Starttid	SlutDatum	Sluttid	Startavgif	Tim	Mil	Belopp	Moms	Vernr
12611	Mattias	2004-12-30	CR		2004-1-3		0.00	69.0	15.8	-974.40	194.88	04028

*Datum is faulty, should be 2003-12-30 and booking fees start from 2004-10-01.

Table F3

NR	NAMN	DATUM	BIL	SLUTDATUM	STARTAVGIF	TIM	MIL	BESKRIVNIN	BELOPP	MOMS	VERNR
19336	Fredrik Holmborg	2004-10-31		1900-12-31	0.00	0.0	2.3	Statoilbil	-46.00	9.20	04305

[13] All these are explained by Ingemar Elb.
[14] All these below are explained by Ingemar Elb. And kr is the unit of Swedish money SEK.

19336	Fredrik Holmbor	2004-10-31	B	1900-12-31	10.00	5.0	0.3		-66.00	13.20	04311

*The two records were the same booking. There was a problem with the pool car so he could not use it but had driven it 0.3 miles. He hired a car at Statoil instead which he used for 5.0 hours and drove 2.3 miles.

Appendix G

Table G: Changes of annualfee, bookingfee, time price and driving price.

Apply from	Events	Annualfee	Bookingfee	Time price	Driving price	
					Stor car	Liten car[15]
	Free time 00-06		0.00	10.00	15.00	
2001-1-1	Annualfee for 2001	200.00				
2001-6-1					16.00	
2002-3-1	Free time 23-07				18.00	
2002-1-1	Annualfee for 2002	400.00				
2003-1-1	Annualfee for 2003	400.00				
2004-1-1	Annualfee for 2004	400.00				
2004-10-1			10.00		20.00	
2005-1-1	Annualfee for 2005	400.00				
2006-1-1			20.00			
2006-1-1	Annualfee for 2006	500.00				
2006-9-1					21.00	19.00

*I simplified the driving price after 2006-9-1 as 20 instead of 21 and 19.

Appendix H: Example of final data

Table H: Example of final data

Member No.	shtrip	Date	Booking Fee	Time	Km	Cost	BIL	Annual Fee	Mile Fee
23483	0	2001-4-9	0	12	280	540	P	200	1.5
26822	0	2001-10-26	0	17	60	266	P	200	1.6
19422	0	2002-1-30	0	4	90	184	CR	400	1.6
19336	0	2003-6-20	0	22	167	520.6	B	400	1.8
15233	1	2004-3-14	0	2	37	86.6	CFB	400	1.8
19422	0	2005-2-19	10	7	60	203	K	400	2
26334	0	2006-7-13	20	38	197	794	CFR	500	2
20737	0	2006-6-23	20	3	21	92	B	500	2
30194	0	2006-12-19	20	4	42	148.2	CFB	500	2

* The number of observation “shtrip” being 1 is 791.

[15] Stor car and Liten car are two types of car of Falu bilpool.

Appendix I: Number of observations for each Member

Table I: Number of observations for each Member.

	Member	No. of observations		Member	No. of observations		Member	No. of observations		Member	No. of observations
1	10003	5	18	18215	13	35	26225	11	52	63282	85
2	10173	10	19	18408	3	36	26334	21	53	63384	5
3	10424	41	20	19336	125	37	26426	38	54	65358	9
4	10689	11	21	19347	103	38	26822	212	55	69601	31
5	12611	178	22	19422	140	39	27416	6	56	711271	18
6	12849	6	23	19612	4	40	27595	1	57	711286	176
7	13170	36	24	20737	203	41	27702	1	58	711560	85
8	13228	32	25	21503	1	42	27978	13	59	711585	43
9	13281	64	26	22186	116	43	28158	10	60	711641	22
10	13534	6	27	22820	15	44	28485	3	61	770420	48
11	13809	6	28	23433	199	45	28638	84	62	790984	8
12	15233	66	29	23483	47	46	30194	372	63	795204	24
13	15462	92	30	23739	34	47	32142	89	64	2200731	32
14	15940	30	31	24288	430	48	32576	12	65	6127496	18
15	17465	5	32	25353	8	49	39062	3	66	6860971	224
16	17599	75	33	25831	85	50	59132	27	67	8520725	58
17	18054	121	34	26119	40	51	63181	90			

Appendix J: Results of factor “month” in Table 4

Table I: Results of factor “month” in Table 4

	Fixed effects		
	Estimate	Std.Error	t-value
Control variable: month			
February	-0.246464	0.207629	-1.187
March	-0.201296	0.200867	-1.002
April	-0.288303	0.207902	-1.387
May	-0.188443	0.200198	-0.941
June	-0.10987	0.194104	-0.566
July	-0.227804	0.214492	-1.062
August	-0.269445	0.212209	-1.27
September	-0.313553	0.205597	-1.525
October	-0.140915	0.196799	-0.716
November	0.155118	0.194045	0.799
December	-0.111485	0.203763	-0.547

Appendix K: Calculate the 95% confidence interval of probability of short-trip use.

We have found that $p_0(\beta_{10}) = \dfrac{\dfrac{p_{10}}{e^{\beta_{10}}(1-p_{10})}}{1+\dfrac{p_{10}}{e^{\beta_{10}}(1-p_{10})}} = \dfrac{odds_{10}}{e^{\beta_{10}}+odds_{10}}$, $where\ odds_{10} = \dfrac{p_{10}}{(1-p_{10})}$,

and $\sqrt{n}(\hat{\beta}_{10}-\beta_{10}) \to n\left(0,\ \left[Sd(\beta_{10})\right]^2\right)$. So, by the Delta Method (Casella & Berger 2001), we have that $\sqrt{n}(p_0(\hat{\beta}_{10}) - p_0(\beta_{10})) \to n\left(0,\ \left[p_0'(\beta_{10})\,Sd(\beta_{10})\right]^2\right)$

,where $p_0'(\beta_{10}) = -\dfrac{odds_{10} * e^{\beta_{10}}}{\left(e^{\beta_{10}}+odds_{10}\right)^2}$, and $n = 4225$

, which leads to that $\dfrac{\sqrt{n}\left(p_0(\hat{\beta}_{10}) - p_0(\beta_{10})\right)}{\left|p_0'(\beta_{10})\,Sd(\beta_{10})\right|} \to n(0,1)$. p_{10} can be calculated from the data. The estimates of β_{10} and $Sd(\beta_{10})$ can be found in the model results. Now we can find the 95% confidence interval of $p_0(\beta_{10})$ being

$\left(p_0(\hat{\beta}_{10}) - 1.96 * \left|p_0'(\beta_{10})\,Sd(\beta_{10})\right|/\sqrt{n},\ p_0(\hat{\beta}_{10}) + 1.96 * \left|p_0'(\beta_{10})\,Sd(\beta_{10})\right|/\sqrt{n}\right)$.

Similarly, to compare the probability of short trip between booking fee being 20 SEK and being supposedly 10 SEK, we also have that

$p_{10}(\beta_{21}) = \dfrac{\dfrac{p_{20}}{e^{\beta_{21}}(1-p_{20})}}{1+\dfrac{p_{20}}{e^{\beta_{21}}(1-p_{20})}} = \dfrac{odds_{20}}{e^{\beta_{21}}+odds_{20}}$, $where\ odds_{20} = \dfrac{p_{20}}{(1-p_{20})}$; $\beta_{21} = \beta_{20} - \beta_{10}$

, and $\sqrt{n}(\hat{\beta}_{21}-\beta_{21}) \to n\left(0,\ \left[Sd(\beta_{21})\right]^2\right)$ because the sample size 4225 is large. The estimates of β_{10}, $Sd(\beta_{20})$ and $Sd(\beta_{10})$ can be found in the model results. The estimates of $Sd(\beta_{21})$ can be calculated as $Sd(\hat{\beta}_{21}) = \sqrt{Var(\hat{\beta}_{20}) + Var(\hat{\beta}_{10})}$. Similarly, the 95% confidence interval of $p_0(\beta_{10})$ is that

$\left(p_{10}(\hat{\beta}_{21}) - 1.96 * \left|p_{10}'(\beta_{21})\,Sd(\beta_{21})\right|/\sqrt{n},\ p_{10}(\hat{\beta}_{21}) + 1.96 * \left|p_{10}'(\beta_{21})\,Sd(\beta_{21})\right|/\sqrt{n}\right)$.

The 95% confidence interval of income of bilpool can be found based on the confidence interval of $p_0(\beta_{10})$ *and* $p_{10}(\beta_{21})$.

Reference:

[1] Charles E. McCulloch, Shayle R. Searle (2001), *Generalized, linear, and mixed models*, New York: Wiley.

[2] Douglas Bates, Deepayan Sarkar (2007.4). The lme4 Package in R "Linear mixed-effects models using S4 classes", GPL version 2.

[3] www.falubilpool.nu the official website of Falu bilpool.

[4] José C. Pinheiro, Douglas M. Bates (2000). *Mixed-effects models in S and S-PLUS*, Springer, New York.

[5] John Wiley & Sons (2005 April). *Generalized linear mixed model*, *Encyclopedia of Statistics in Behavioral Science*, website: http://www.wiley.com//legacy/wileychi/eosbs/articles.html .

[6] Peter Dalgaard (2002). *Introductory statistics with R*, Springer, New York.

[7] Peter Muheim, Partner (1998). *Carsharing - the key to combined mobility*, Swiss Federal Office of Energy & Energie 2000 Motor fuels section / Mobility, Lucerne.

[8] Per Schillander (2003). *Make space for carsharing*, Vägverket, Publ.no.2003: 88

[9] P. McCullagh, John A. Nelder (1989). *Generalized Linear Models, Second Edition*, Chapman & Hall/CRC.

[10] Shayle R. Searle, George Casella, Charles E. McCulloch (1992). *Variance components*, New York: Wiley.

[11] Ulf Olsson(2002). *Generalized Linear Models: An Applied Approach*, Studentlitteratur AB.

[12] W. N. Venables, B. D. Ripley (2003). *Modern Applied Statistics with S, 4th edition*, Springer.