What factors affect drivers' choice between petrol and ethanol?
--------An empirical analysis with refueling data from Falu Bilpool

Author:Qiguang Gao

Supervisor: Kenneth Carling/ Ola Nääs

D-level essay in Statistics, Spring Semester 2007

Department of Economics and Society, Dalarna University

Abstract

A lot of scholars want to find the underlying reason for drives' choice between petrol and ethanol. Falu Bilpool, which is a typical non-profit car-sharing company which has a few fuel-flexible vehicles give a good opportunity to analyze this problem. This essay is an empirical analysis with refueling data from Falu Bilpool and there are 508 observations from 2004 to 2006 in the dataset. I set up a generalized linear mixed model(GLMM) to analyze this problem. Then I find that two variables which are quantity of refueling and number of ethanol pumps have significant influence on the choice, while the price variable have not. Therefore I give advises that governments and fuel companies should increase the number of ethanol pumps and bilpools may encourage members to refuel fuel of larger quantity. At the same time, it would be better if governments and bilpools pay more attention to promote people's environmental perspective.

Key words: Choice between fuels, GLMM, R, Falu Bilpool

1. Introduction

"The Swedish Parliament has decided to follow UN recommendations concerning the reduction of carbon dioxide emissions. Further reductions must now be made in emissions from traffic." As pointed out by the website of SEKAB which is a company aim to create the conditions for actively promoting sustainable transport for the future with the help of long-term and sustainable biofuels, Swedish people have paid more and more attention to the issue of reducing the burning of fossil fuels. [1]

Not only in Sweden, as the problem of global greenhouse effect becoming more and more serious, people of all over the world have attached increasingly more importance to how to decrease the consumption of fossil fuels. A lot of alternative fuels come into people's vision, and they are improved rapidly and used more widely. [2]

Biomass fuels show their advantages in causing less negative effects to the environment. The most widespread used biological fuel is E85, which is a mixture fuel constituted of 85% ethanol and 15% petrol in volume. E85 is environmental friendly fuel which burns cleaner than petrol; decreases emissions of harmful gas such as hydrocarbon and benzene, and release less greenhouse gas especially carbon dioxide. Unlike fossil fuels, ethanol is totally renewable. Ethanol, which is a biomass fuel, is produced among other thing from cereal and cellulose through fermentation and distillation. Therefore, to promote the use of E85 contributes to ease the worldwide energy crisis. It is well known that E85 is widely used in Sweden for Sweden walks in forefront of the trend of biomass fuels.

In order to reduce pollution and improve the status of public transportation, more and more "bilpools" are set up by environmentalists. A bilpool, is usually a non-profit car-sharing company. The members of a bilpool usually pay member's fee to attend it, share vehicles, and pay time and distance fee for using the car each time. The board of bilpool is in charge of purchase and maintenance of vehicles; however, members are responsible to refuel vehicles. Different bilpools have different rules of fee and refueling. By sharing vehicles and arrange usages of vehicles effectively, bilpools can remarkably reduce members' expenses on transportation, and decrease consumption of fuel, especially fossil fuel.

Falu Bilpool is a typical bilpool in Sweden. Falu bilpool was founded in 2000, and open for residents, companies and organizations in Falun. It is a non-profit organization, operated by a board and constituted by three parts: economic department, marketing department and car department. It usually has three or four cars which are located in different parts of Falun. Most of the cars are fuel-flexible vehicles and can drive on the E85 or petrol or mixture of such two fuels of any ratio.

At the same time, the price per liter of E85 is remarkably lower than petrol too. The relative prices of ethanol and petrol are thought to be a factor affecting people's choices towards fuel. The price of per liter petrol is higher than the price of per liter E85. At the same time, the energy of 1 liter petrol is also higher than that of 1 liter E85. The energy value of E85 is approximately, 25.2MJ/L, while the energy value of petrol is 34.8. The energy value of E85 is about 0.72 time of that of petrol. And it is generally believed that the efficiency of burning E85 is about 4 to 5 percent higher than that of petrol. When consumers make their choice

[1] http://www.sekab.com/default.asp?id=1351&refid =1346, (2007-07-17, 2:30 GMT+8)
[2] http://en.kikipedia.org/wiki/E85, (2007-7-17, 2:39 GMT+8)

between fuels, they consider cost performance of fuels rather than prices. Falu bilpool collected all of receipts of fuel purchase since its foundation. So types and real prices of fuels can be determined.

Analyzing the refueling data of Falu Bilpool provides us a very good opportunity to find what factors affect people's choices between ethanol and petrol under the same condition. It helps to find an effective way to promote E85's consumption towards petrol, so as to reduce negative effects to the environment.

2. Data description

This thesis makes use of refueling data of Falu Bilpool.[3] The data is collected from original refueling receipts from Jan 1, 2004 to Dec 31, 2006, together with corresponding records of those 3 years' booking found in data files provided by Falu Bilpool.[4]

Several records was chosen from the refueling data and showed in Table 1 to show the form of the data. The columns named YEAR and DATE demonstrate the exact date of refueling. The column NO denotes the identifying number of the original receipt in the documented data of the board of Falu Bilpool, in which the specific fueling is recorded. Column CAR denotes the number of the car refueled. STATION denotes the fuel station where the refueling took place, in which the first letter O, V and S denotes OKQ8, Volvo and Statoil, respectively; and the word behind the first dot identify the town of station, and sometimes the road of station is also recorded, for example VASAG denotes Vasagatan.

Column FUEL specifies what kind of fuel was chosen at that time, in which E denotes E85 and 95 denotes petrol, usually 95#, sometimes 96# or 98#. Sometimes from a receipts the type of fuel is recorded as DRIVMEDEL (Swedish word which means fuel) or OTHER, so the type of fuel can not be determined and I delete them from the dataset. There are only 2 records with diesel in the 3 years, so I excluded them too. With column QUANTITY and AMOUNT with unit liter and SEK respectively, how much volume of fuel and how much cost can be known so as to calculate the real price of each fuel at that time. Column MEMBER denotes the member number of who refueled the vehicle. The value of MEMBER is assessed by relating refueling data to booking data by the same date and the same vehicle of refueling and booking. A sample of the booking data is shown in Table 2.

Table 1 Example of refueling data of Falu Bilpool

YEAR	DATE	NO	CAR	STATION	FUEL	QUANTITY	AMOUNT	PRICE	MEMBER
2004	0616	196	665	O.FALUN	E	40.69	304.77	7.49	26822
2004	0812	252	147	V.FALUN	95	31.06	309.98	9.98	18054
2004	*1204*	*9*	*665*	*O.FALUN*	*E*	*38.69*	*302.94*	*7.83*	*10424*
2005	0209	66	975	O.HEDEMORA	95	31.85	316.59	9.94	28638
2005	0829	235	147	O.ORNSKOLDSV	95	32.77	389.96	11.90	13228
2006	0213	300	665	S.FALUN.VASAG	E	47.24	370.36	7.84	12849
2006	0618	155	136	O.FALUN	E	43.03	343.81	7.99	13228
2006	0626	155	778	V.FALUN	DRIVMEDEL	20.55	248.66	12.10	19336
2006	1003	255	136	V.FALUN	OTHER	35.08	294.67	8.40	24288

[3] The original data is provided by Ingemar Elb. I am very grateful for his help!
[4] The booking data is organized by Shanshan Yang, Yiqi Yang, Hao Li and Qiguang Gao.

Table 2 Example of booking data of Falu Bilpool

MEMBER	CAR	Date	CarSymbol	StartingFee	Time	Km	Amount	Tax	DocNO
24288	147	2004-12-3	B	10.00	7.5	16	-117.00	23.40	04365
30194	136	2004-12-3	CFB	10.00	4.0	43	-136.00	27.20	04366
46400	136	2004-12-3	CFB	10.00	3.0	3	-46.00	9.20	04366
711286	147	2004-12-4	B	10.00	6.0	23	-116.00	23.20	04365
59132	136	2004-12-4	CFB	10.00	4.0	44	-138.00	27.60	04366
10424	*665*	*2004-12-4*	*CFR*	*10.00*	*9.0*	*120*	*-340.00*	*68.00*	*04367*
19347	975	2004-12-4	K	10.00	9.0	26	-152.00	30.40	04368
24288	147	2004-12-5	B	10.00	2.5	8	-51.00	10.20	04365
59132	665	2004-12-5	CFR	10.00	5.5	31	-127.00	25.40	04367

For example, the two Italian records in Table 1 and Table 2 can be related to each other, for they have the same car number and date. Therefore I can find the member with number 10424 refueled the No.665 vehicle on Dec 4, 2004.

Most of members of refueling records can be specified with the method above. Unfortunately, sometimes it can not be totally sure who indeed refueled the car among two or more members. On those occasions I chose the possible member with largest driving distance, which can be find in column Km in Table 2. This method is obtained under the intuitive hypothesis that the member who drove for the longest distance has the largest probability to refuel the car. If we can not specify who refueled the car by the two method above, which is usually caused by the missing data when transforming data or lack of corresponding records, the member of refueling is set as "?". Only very few observations of refueling data can not be identified with former two methods. When that happens, I delete the corresponding record form the dataset.

The fuels petrol and E85 accounts for more than 90% of total volume of fuel, therefore, I chose the refueling records with petrol or E85, and let 1 denote E85 and 0 denote petrol. Then the binary series can denote which type of fuel is chosen at each time of refueling. I name the binary series as FUEL.

Then I integrate those records with the average temperature and average relative price of the month in which the refueling took place. Since I did not find the historical temperature records of Falun from 2004 to 2006, I make use of the temperature data of the town Borlange, 18 km away on the southwest of Falun. For this specific problem, the 18 km's distance can be ignored in meteorology. Since the winter of Falun is cold and E85 is more difficult to ignite compared with petrol, the temperature might have effects on the usage of E85. I selected monthly average temperature of Borlange named TEMP to serve as a measure of the monthly temperatures of Falun.

In terms of relative price of E85 to petrol, first there is a fact that should be noticed that these two fuels of the same volume have different energy, as pointed out in introduction part. According to previous demonstration of their differences in energy value and burning efficiency, it can be calculated that 1 liter E85 equals to 1*0.72*1.05=0.756 liter petrol. I divide the original price per liter of E85 by 0.756, and then get a price that can be compared with the original price of petrol. I compute a These two prices have the same unit SEK/(34.9MJ), which can be express as cost-energy ratio. The two series of ratio of E85 and

petrol in the 36 months are shown in Figure 1. The thick line denotes the cost-energy ratio of E85, while the thin line denotes petrol. The cost-energy ratios measures cost performance of two fuels, in which the lower ratio indicates better cost performance. However, we can not intuitively find which fuel has higher cost performance if we look the 3 years as a whole. In order to compare the prices of two fuels, I standardize the price of E85 by making price of E85 divided by the product of 0.72 and 1.05, and I name this rice as STPE. STPE is comparable to the price of petrol, for they both measure the cost-energy ratio.

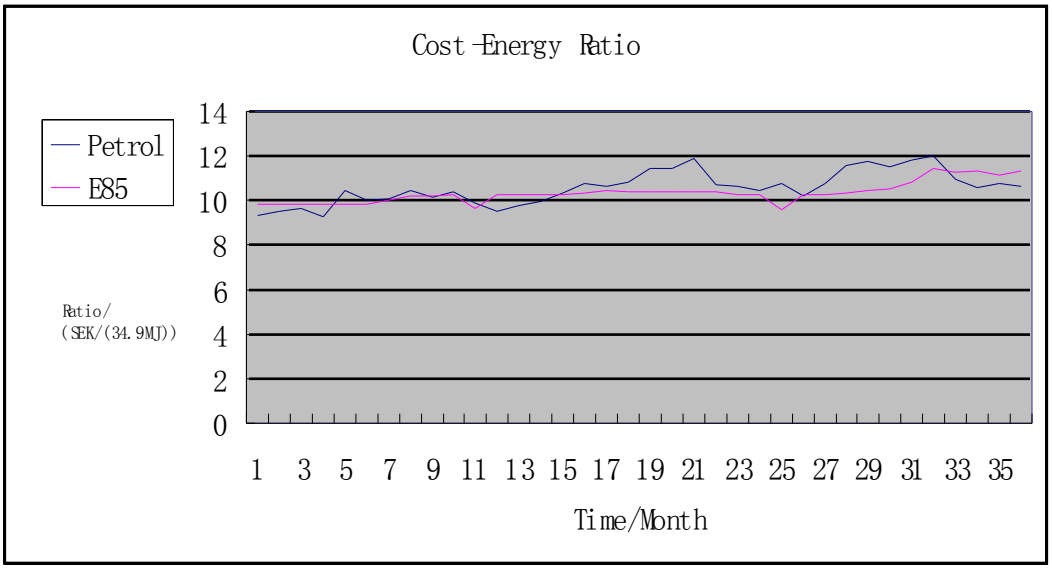

Figure 1 Cost-Energy Ratio of E85 and petrol during 2004 and 2006

By using two-tail T test, I can find that the corresponding P-value is 0.00566, which means the null hypothesis should be rejected and two series of ratio does not come from the population with the same mean. The mean of cost-energy ratio of petrol is 10.625, and the mean of cost-energy ratio of E85 is 11.034. It means that petrol is slightly more economic than E85 during 2004 and 2006.

I compute a series named ROPR which is the ratio of the monthly standardized price of petrol to the monthly standardized price of E85. The relative cost-performance of petrol to E85 equals the inverse of ROPR. Apparently, the increase of series ROPR means that petrol becomes less economic than before compared with E85.

Availability is another factor which might influence bilpool's members' choice towards fuels. The number of E85 pumps reflects availability of E85. Unfortunately I did not find the monthly number of E85 pumps from 2004 to 2006, and what I obtain is yearly data in Sweden from 2001 to 2006 as indicated by Figure 2. Since I can not know monthly number of ethanol pumps, I consider the yearly number as the number of E85 pumps at the end of each year and estimate monthly data by assuming that the number of pumps is increased steadily during each year. Thereby I obtain a variable named PUMP to denote the estimated monthly number of E85 pumps.

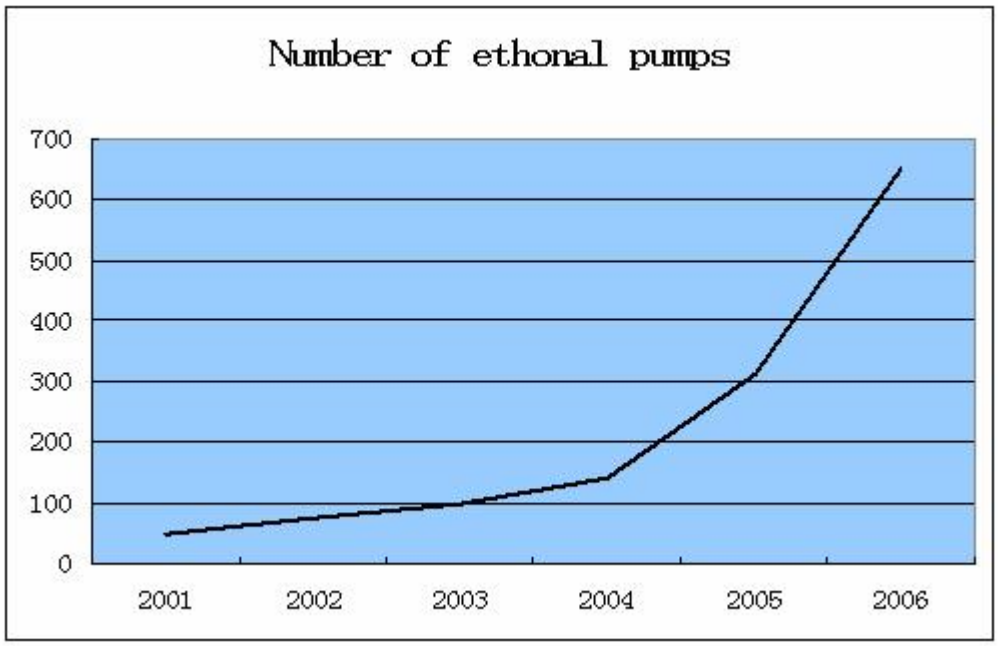

Figure 2 Number of ethanol pumps in Sweden from 2001 to 2006

From 2004 to 2006, at the bilpool, there have been 8 cars used; in which 7 cars appear in my dataset. They belong to 3 brands: Renault Megane, Ford Focus and Citroen C3. In which 4 Ford Focus Fuel-flexible vehicles can burn on ethanol, petrol or any proportion of mixture of two kind of fuel. While 2 Renault Megane cars can only burn on petrol and 1 Citroen C3 car can only burn on diesel. For this reason I delete all the records related to Renault Megane and Citroen C3.

In the data the refueling stations are recorded, in which the stations OKQ8 Falun, Volvo Falun and Statoil Falun Vasagatan appeared very frequently. From the website of SEKAB, the list of refueling stations with E85 can be found. At a station without E85 pump, consumers have no choice. For bilpool members, convenience is the top concern. There are 508 refueling records left in my final dataset, including the cases for members to choose between petrol and ethanol during 2004 and 2006, in which 271 records are of E85-refueling and 237 are of petrol-refueling.

From Table 3, we can obtain a general view of the distributions of the variables.

Analyzing the refueling data of Falu Bilpool provides a very good opportunity to find what factors affect people's choices between ethanol and petrol under the same condition. It helps to find an effective way to promote E85's consumption towards petrol, so as to reduce negative effects to the environment.

Table 3 Summary statistics of explanatory variables

Covariate	Min.	1st Quantile	Median	Mean	3rd Quantile	Max.
Temperature (°C)	-5.6	0.3	8.0	7.6	13.5	18.9
Ratio of prices	0.91	1.00	1.05	1.04	1.09	1.15
Quantity of refueling (L)	3.28	29.84	36.28	35.29	41.71	55.57
No. of E85 pumps	103.42	130.75	226.00	281.99	424.67	652.00

3. Model analysis and results

Different members have different tendency to refuel E85 rather than petrol. By calculating the proportion of volume of E85 in the sum of volume of fuel in my dataset for each member who had refueling records, I can find the tendency of refuel E85. This variable can be denoted as TEND. And this tendency is shown in Figure 3. In Figure 3, I find that the proportions of refueling E85 for different members are remarkably different and the proportions do not correspond to a normal distribution. Therefore, members' preferences may be an important causal factor to their choices between fuels, however, members' preferences are not the main interest of this study. Thus I use Generalized Linear Mix Models to analyze this problem and let the variable MEMBER be the random part.

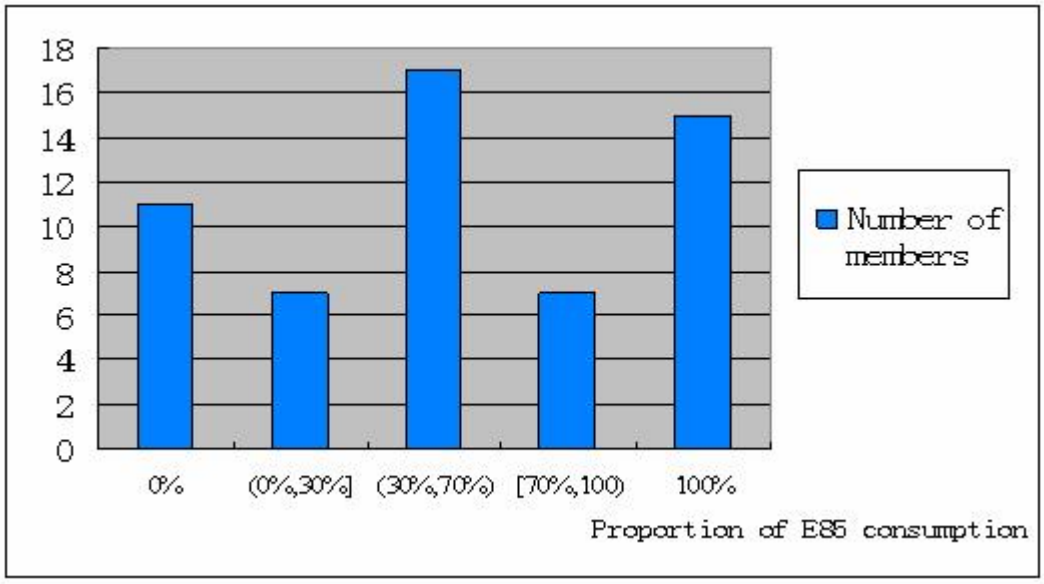

Figure 3 Proportion of groups grouped by tendency to refuel E85

I choose the series FUEL as response variable. At the same time, I choose 4 stimulus variables: ratio of prices ROPR, quantity of each refueling QUAN, temperature TEMP, and the number of E85 pumps PUMP. The variable MEMBER which denotes customer number is chosen as

random part. GLMM method can be used in this analysis. Criteria of 5% significant level is commonly used in linear models, while for GLM, the criteria should not be so strict, and 20% are commonly used, so I choose 20% significant level as the criteria.

The basic model of GLM models is that $\eta=\mathbf{X}\square$. And GLMM models add random effect **u** to GLM models. The basic model of GLMM models is that $\mathbf{E(Y|u)=X}\square\mathbf{+Zu}$.
And the conditional distribution of **y** given **u** is as follows:

$$y_i \mid u \sim indep. f_{Y_i|u}(y_i \mid u)$$

$$f_{Y_i|u}(y_i \mid u) = \exp\{[y_i\gamma_i - b(\gamma_i)]/\square^2 - c(y_i,\square)\}$$

The conditional mean of y_i given **u** is μ_i, which can be express as follows:

$$E[y_i \mid u] = \mu_i$$

$$g(\mu_i) = x_i'\square + z_i'u$$

I set up a GLMM models to analyze this problem. The equation of my models is that:

$$\eta = \mu + MEMBER_i + \square_j \bullet RP + \square_k \bullet TE + \square_l \bullet QUAN + \square_m \bullet PUMP + \varepsilon_{ijklm}$$

In the above model, $MEMBER_i$ is the random intercept.
Since FUEL is binary, I make use of GLMM model with binomial family to analyze it. For GLM model with binomial family, there are 3 type of common link function can be used: logit, probit, and cloglog. So I set up GLMM models with three link functions above respectively, and compare them by their AIC value, P-value and other criteria.

Then I find that the GLMM model with logit link function has the lowest AIC value which is 391.1, while the AIC values of models with other two link both exceed 12000. At the same time, this model is among the models with most significant variables and the estimate of parameter is reasonable. The estimates fulfill the main aim of the analysis as well. Therefore I make use of the model with logit link.

Table 4 Estimate result of the GLMM model

	estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.06703	2.003376	-1.0318	0.3022
TEMP	0.011957	0.014998	0.7973	0.4253
ROPR	1.272659	1.850746	0.6876	0.4917
QUAN	0.018118	0.014852	1.2199	0.2225
PUMP	0.001171	0.000657	1.7838	0.0745

By estimating this GLMM model in software R of version 2.4.1 with package “lme4”, I can

obtain estimates of the model. The result of estimation is put in Table 3. The model is acceptable because two casual variables are proved to be significant and no significant variable has unreasonable signal. Since the response variable of the model is binary, so it is very hard to make residuals distributed normally and the Q-Q Plot is not informative.

From the estimate result as shown in Table 4, I can find that the causal variable and PUMP are significant according to the criteria of 20% significant level. Other three causal variables QUAN, TEMP and ROPR are not significant under the criteria.

Table 5 Result of ANOVA

Variable	Sum of variance	Proportion
TEMP	4.3016	27.98%
ROPR	0.0381	0.25%
QUAN	4.7134	30.66%
PUMP	6.3188	41.11%

The result of ANOVA is shown in Table 5. I can find that the variable QUAN has the largest proportion of the sum of variances of the four stimulus variables which is 41.11%. It means the quantity of refueling is the most significant cause of the choice between fuels. The second largest proportion is of the variable QUAN which is 30.66%, so quantity of refueling might have a vary weak effect in the choice.

The estimate of PUMP is positive and significant, whose P-value of Z-test equals to 0.0745. It indicates that as the number of pumps grows up, the bilpool's members also tend to refuel more E85 in proportion of fuel because E85 becomes more convenient to acquire. This means that bring more E85 pumps to fuel station indeed promote the use of Ethanol. We expect to see the trend of increasing and spreading of E85 pumps continue steadily in the follows years.

The variable QUAN is also positive and very significant with Z-value equal to 1.2199. It means that with larger volume refueled, drivers prefer to purchase more proportion of E85 very much. However the result of ANOVA shows than QUAN has the second largest influence the total variance. So it may indicates that people do not care much about the choice between petrol and E85 in refueling a small volume of fuel, however, feels more responsible for protecting environment when refueling volume goes up.

The variable TEMP is not significant in this model, which means that the temperature does not remarkably influence the choice between petrol and ethanol.

It is very interesting to find that the relative cost-performance denoted by ratio of prices have no significant effect on choices between ethanol and petrol when the ratio of prices is between 0.91 and 1.15. It may means that the people in Sweden does no care about small change of fuel's cost-performance. For other models I tried containing ROPR, it has no significant influence either. It might due to the drawbacks of my dataset, the variance of ROPR is very small and it limits the analysis of regression.

4. Conclusions and suggestions

The choice of fuel between ethanol and petrol is mainly affected by the volume of refueling and the number of E85 pumps. The small proportion of change in relative cost-performance within 10% has no significant effect on the choice.

For the governments, fuel companies, they are responsible to promote the introduction of new ethanol pumps to make drivers more convenient to refuel ethanol. The increase of ethanol pumps is more effective than price concessions within 10 percent.

On the other hand, since the preferences of people are remarkably different, in order to promote the use of E85 rather than petrol, it would be better if governments, bilpools and environmental protection organizations pay more attention to promote people's environmental perspective.

Reference:

[1] Ulf Olsson (2002), *Generalized linear models: an applied approach*
[2] Charles E. McCulloch, Shayle R. Searle (2001), *Generalized, linear, and mix models*
[3] Peter McCullagh, J. A. Nelder (1989), *Generalized linear models*
[4] Peter Dalgaard (2002), *Introductory statistics with R*