What are the factors that drive passengers to pay more ?

Marius Agasse-Duval - Edouard Fourmaux - Louis Rigondet

ENAC, IENAC15 OPS,SITA - PREV

April 7, 2017

Table of contents

Story of our project :

- Collect data to learn about people's expectations regarding comfort.

Story of our project :

- Collect data to know people's requirements concerning comfort.
- Build an econometric model with this set of data.

Story of our project :

- Collect data to know people's requirements concerning comfort.
- Create an Econometrics model of these data.
- Interpret the results computed by Eviews and modelize them to try and find an answer to our problem

Table of contents

Substantial figures

- n = 176 answers
- Majority of students (60%)
- k = 34 variables

Table of contents

Characteristics of respondents

Observations

- Mean age : 32
- 100 french people and 60 spanish
- 120 respondents unmarried
- 135 people have studies between 4 and 6 years after high school

$\Rightarrow$ **Model migth be biaised**.

Variable PAY_MORE_2

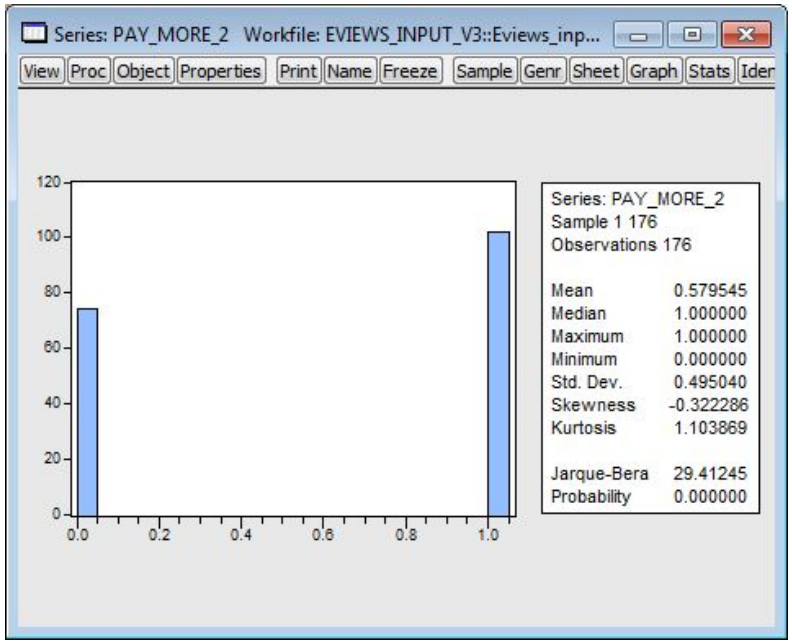


FIGURE – PAY_MORE_2 histogram

Questions

- Are people ready to pay more for the short haul flight ? (First answer with the mean)
- Is it correlated to other variables ?

Variable PRICE

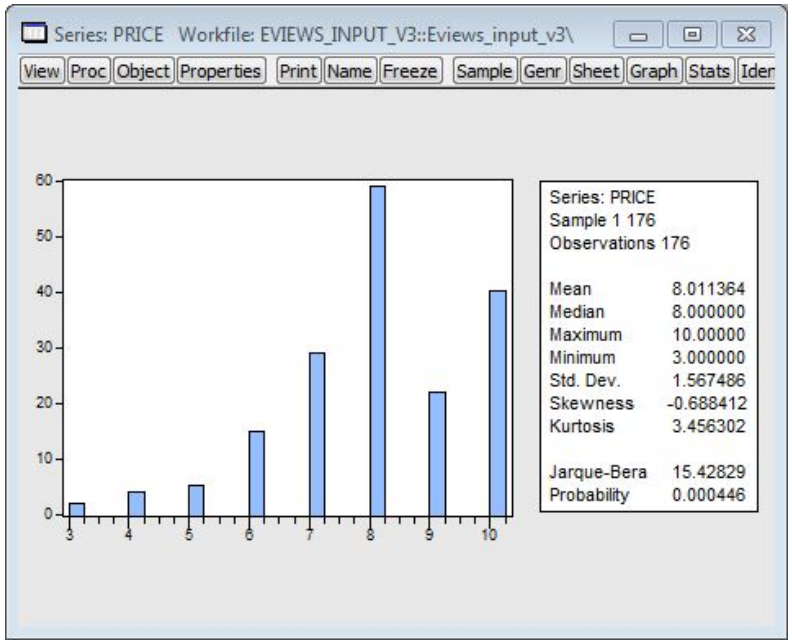


FIGURE – PRICE histogram

Observations

- Price is really significant : no responses below 3
- Most of values are beyond 7

Table of contents

Analysis of the correlation matrix

Variable	Correlated with	Coefficient
Age	Children	0.66
Age	Married	0.62
Married	Children	0.73
Crew's Kindness	Meal	0.51
Gender	Height	0.73
Luggage	Snack	0.57
Meal	Seat	0.52
Plugs	Wifi	0.60
Ramp	Snacks	0.58
Seat	Seat_2	0.61

TABLE – Main significant coefficients

How do we choose variables ?

- Step 1 : Look for variables which are not relevant during descriptive method (e.g. drinks).
- Step 2 : Analysis of the correlation matrix.
- Step 3 : Suppression of some variables to increase $\overline{R^2}$.
- Step 4 : Suppression of some variables to avoid multicollinearity.

What will be our Y ?

In regards to our topic, we observed that the most relevant variables giving us the best explanatory models are :

- pay_more for long haul
- pay_more_2 for short haul

Table of contents

Equation: UNTITLED Workfile: EVIEWS_INPUT_V3::Eviews_i...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PAY_MORE_1
Method: Least Squares
Date: 04/06/17 Time: 16:31
Sample: 1 176
Included observations: 168

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.416408	0.744633	1.902157	0.0592
CHOICE_1	-0.632090	0.070403	-8.978110	0.0000
DEPARTURE_HOUR	0.027808	0.017291	1.608226	0.1100
GENDER	0.112774	0.086500	1.303742	0.1944
RAMP	0.004403	0.012786	0.344335	0.7311
SEAT	0.001954	0.018703	0.104481	0.9169
PRICE	-0.015926	0.020974	-0.759329	0.4489
AGE	0.001004	0.003553	0.282515	0.7780
AIRPORT_PROXIMITY	-0.025012	0.030395	-0.822896	0.4119
BEST_PRICE	0.005032	0.064056	0.078563	0.9375
CHILDREN	-0.022737	0.049916	-0.455499	0.6494
DRINKS	0.001748	0.013693	0.127639	0.8986
EDUCATION	-0.003610	0.061956	-0.058269	0.9536
FLY_WHY	-0.014148	0.042443	-0.333327	0.7394
HEIGHT	-0.000848	0.004241	-0.199902	0.8418
INCOMES	0.025815	0.029405	0.877899	0.3815
LUGGAGE	-0.000926	0.015000	-0.061723	0.9509
MARIED	0.108122	0.102599	1.053830	0.2938
MORE_MONEY_1	0.000863	0.000121	7.131185	0.0000
NATIONALITY	-0.005958	0.020244	-0.294318	0.7689
OCCUPATION_CAT	0.012352	0.012117	1.019350	0.3098
PLEASURE	0.017269	0.015980	1.080636	0.2817
PLUGS	0.004521	0.014059	0.321592	0.7482
SNACKS	-0.025495	0.016115	-1.582052	0.1159
SCREEN	-0.007506	0.015114	-0.496664	0.6202
WIFI	-0.001684	0.015267	-0.110291	0.9123

R-squared	0.528022	Mean dependent var	0.654762
Adjusted R-squared	0.444928	S.D. dependent var	0.476867
S.E. of regression	0.355281	Akaike info criterion	0.909573
Sum squared resid	17.92391	Schwarz criterion	1.393043
Log likelihood	-50.40410	Hannan-Quinn criter.	1.105789
F-statistic	6.354468	Durbin-Watson stat	1.795819

Observations

- Prob(F-stat) = 0.000
- $R^2 = 0.53$
- Many insignificant variables (in red).

Analysis of the model

Observations

- We can formulate linear regression as : $Y = X\beta + u$
- Our model is statistically meaningful as a whole

How to improve the model ?

- Reduce the number of(insignificant) variables.
- Improve $\overline{R^2}$, to have a better explanation of why people agree with paying more.

Table of contents

The first linear regression

Equation: UNTITLED Workfile: EVIEWS_INPUT_V3::Eviews_i...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PAY_MORE_2
Method: Least Squares
Date: 04/06/17 Time: 16:21
Sample: 1 176
Included observations: 168

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.616933	0.498258	3.245173	0.0015
CHOICE_2	-0.816271	0.045516	-17.93376	0.0000
DEPARTURE_HOUR	0.020780	0.011089	1.873939	0.0630
GENDER	0.088814	0.055200	1.608940	0.1098
RAMP	0.015253	0.008439	1.807510	0.0728
SEAT_2	-0.018872	0.009779	-1.929827	0.0556
PRICE_2	-0.018465	0.013317	-1.386565	0.1677
AGE	-0.000785	0.002454	-0.319841	0.7496
AIRPORT_PROXIMITY	0.001051	0.019984	0.052584	0.9581
BEST_PRICE	-0.038031	0.042233	-0.900503	0.3694
CHILDREN	-0.007040	0.033214	-0.211950	0.8324
DRINKS	-0.007968	0.008652	-0.920928	0.3586
EDUCATION	0.020689	0.040023	0.516927	0.6060
FLY_WHY	-0.039354	0.027090	-1.452697	0.1485
HEIGHT	0.000816	0.002739	0.298116	0.7660
INCOMES	0.021456	0.019412	1.105283	0.2709
LUGGAGE	0.005283	0.009814	0.538322	0.5912
MARIED	0.010276	0.068435	0.150163	0.8808
MORE_MONEY_2	0.015032	0.002126	7.069300	0.0000
NATIONALITY	-0.002668	0.013004	-0.205189	0.8377
OCCUPATION_CAT	0.007786	0.007676	1.014303	0.3121
PLEASURE	-0.009623	0.010315	-0.932967	0.3524
PLUGS	-0.005003	0.007580	-0.659962	0.5103
SNACKS	0.011658	0.010710	1.088521	0.2782

R-squared	0.810575	Mean dependent var	0.577381
Adjusted R-squared	0.780320	S.D. dependent var	0.495453
S.E. of regression	0.232219	Akaike info criterion	0.049291
Sum squared resid	7.765286	Schwarz criterion	0.495571
Log likelihood	19.85958	Hannan-Quinn criter.	0.230413
F-statistic	26.79115	Durbin-Watson stat	2.147289
Prob(F-statistic)	0.000000		

Observations as previous

- Prob(F-stat)= 0.000
- R^2 = 0.81
- Many insignificant variables.

Analysis of the model

Observations

- Like for long haul study, we formulate : $Y = X\beta + u$
- Our model is statistically meaningful as a whole

How to improve the model ?

Our conclusion is the same as before : it is a different model, though with the same statistics results

Table of contents

The linear regression

Equation: UNTITLED Workfile: EVIEWS_INPUT_V3::Eviews_i...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PAY_MORE_1
Method: Least Squares
Date: 04/06/17 Time: 16:36
Sample: 1 176
Included observations: 176

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.934670	0.235265	3.972835	0.0001
CHOICE_1	-0.510769	0.063204	-8.081300	0.0000
DEPARTURE_HOUR	0.044985	0.015742	2.857564	0.0048
GENDER	0.109103	0.059489	1.833991	0.0684
SNACKS	-0.027425	0.011369	-2.412320	0.0169
PLEASURE	0.021169	0.014870	1.423563	0.1564

R-squared	0.346321	Mean dependent var	0.647727
Adjusted R-squared	0.327095	S.D. dependent var	0.479041
S.E. of regression	0.392961	Akaike info criterion	1.003285
Sum squared resid	26.25115	Schwarz criterion	1.111369
Log likelihood	-82.28906	Hannan-Quinn criter.	1.047123
F-statistic	18.01330	Durbin-Watson stat	1.853556
Prob(F-statistic)	0.000000		

FIGURE – Main equation

Observations

- $R^2 = 0.35$
- Like previously Prob(F-stat) = 0.000
- There is only one insignificant variable left

The final equation

Equation with coefficient of the model

$$PAY\_MORE\_1 = \beta_0 + \beta_1 \times CHOICE\_1 + \beta_2 \times DEPARTURE\_HOUR + \beta_3 \times GENDER + \beta_4 \times SNACKS + \beta_5 \times PLEASURE$$

Marginal effects : expectations vs observations(1)

Variable	Expected effect	Observed
Age	+	Irrelevant
Airport_proximity	+	Irrelevant
Best_price	-	Irrelevant
Children	?	Irrelevant
Choice_1	-	-
Departure_hour	+	+
Education	?	Irrelevant
Fly_why	+	Irrelevant
Gender	?	+
Heigth	?	Irrelevant
Incomes	+	Irrelevant

Marginal effects : expectations vs observations(2)

Variable	Expected effect	Observed
Married	?	Irrelevant
Meal	+	Irrelevant
Nationality	?	Irrelevant
Nb_fly	?	Irrelevant
Occupation_cat	+	Irrelevant
Pleasure	+	Irrelevant
Plugs	+	Irrelevant
Price	-	Irrelevant
Screen	+	Irrelevant
Seat	+	Irrelevant
Snacks	+	-
Wifi	+	Irrelevant

Normality of errors

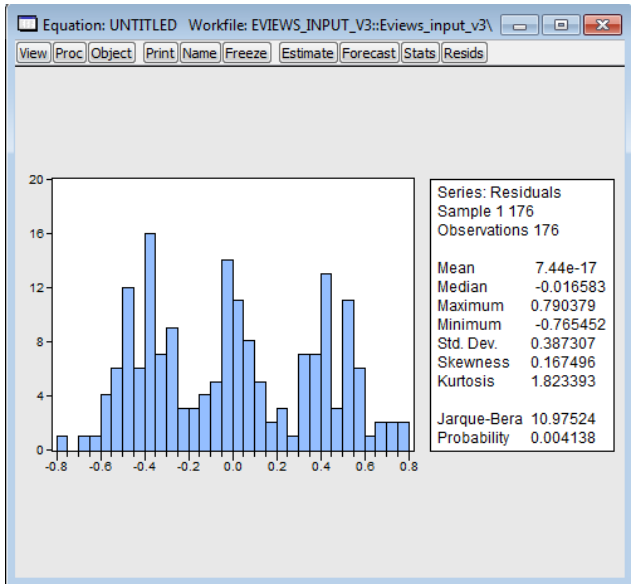


FIGURE – Normality of errors

Observations

- Jarque-Bera value is equal to 11.
- Kurtosis is far from 3
  $\Rightarrow$ Residuals don't follow a normal law.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	2.026531	(1, 170)	0.1564
Chi-square	2.026531	1	0.1546

FIGURE – Wald test

Analysis

To test the significance of PLEASURE, we compute a Wald test :

$$\begin{cases} H_0 : & \beta_5 = 0 \\ H_1 : & \backslash H_0. \end{cases}$$

$qF = 2.025 < \chi^2(q = 1) = 3.841.$
$\Rightarrow$ We do not reject H_0.

Final Equation

Equation: UNTITLED Workfile: EVIEWS_INPUT_V3::Eviews...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PAY_MORE_1
Method: Least Squares
Date: 04/06/17 Time: 16:54
Sample: 1 176
Included observations: 176

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.127351	0.193008	5.840951	0.0000
CHOICE_1	-0.521677	0.062926	-8.290386	0.0000
DEPARTURE_HOUR	0.043411	0.015751	2.756130	0.0065
GENDER	0.112086	0.059630	1.879687	0.0619
SNACKS	-0.027777	0.011400	-2.436534	0.0159

R-squared	0.338529	Mean dependent var	0.647727
Adjusted R-squared	0.323056	S.D. dependent var	0.479041
S.E. of regression	0.394139	Akaike info criterion	1.003771
Sum squared resid	26.56409	Schwarz criterion	1.093842
Log likelihood	-83.33189	Hannan-Quinn criter.	1.040304
F-statistic	21.87865	Durbin-Watson stat	1.844023
Prob(F-statistic)	0.000000		

FIGURE – Final Equation

What about heteroskedasticity ?

Heteroskedasticity Test: White

F-statistic	19.02485	Prob. F(12,163)	0.0000
Obs*R-squared	102.6851	Prob. Chi-Square(12)	0.0000
Scaled explained SS	37.32538	Prob. Chi-Square(12)	0.0002

FIGURE – White's test

Observations

$$\begin{cases} H_0 : & R^2 = 0 \\ H_1 : & R^2 = 1. \end{cases}$$

- nR^2 follows a χ^2 law with q = 18.
- nR^2 = 102.68 $> \chi^2(12)$ = 21.026

$\Rightarrow$ At the 95% level, there is heteroskedasticity.

Pros

- We obtained a meaningful model.

Cons

- Studied variables usually were brackets : not very accurate
- R^2 is acceptable but not close enough to 1. Some useful variables might have been forgotten in the survey.
- Classical assumptions fail.

Table of contents

The linear regression

Equation: EQ01 Workfile: EVIEWS_INPUT_V3::Eviews_inpu...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PAY_MORE_2
Method: Least Squares
Date: 04/06/17 Time: 15:02
Sample: 1 176
Included observations: 176

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.660899	0.169789	9.782158	0.0000
CHOICE_2	-0.771025	0.046653	-16.52671	0.0000
DEPARTURE_HOUR	0.022062	0.011487	1.920651	0.0565
GENDER	0.105898	0.040808	2.595026	0.0103
RAMP	0.018485	0.008174	2.261481	0.0250
SEAT_2	-0.026539	0.009737	-2.725510	0.0071
PRICE_2	-0.019994	0.012507	-1.598614	0.1118

R-squared	0.715567	Mean dependent var	0.579545
Adjusted R-squared	0.705469	S.D. dependent var	0.495040
S.E. of regression	0.268662	Akaike info criterion	0.248236
Sum squared resid	12.19831	Schwarz criterion	0.374334
Log likelihood	-14.84473	Hannan-Quinn criter.	0.299381
F-statistic	70.86068	Durbin-Watson stat	2.144317
Prob(F-statistic)	0.000000		

FIGURE – Main equation

Observations

- $R^2 = 0.72$
- Prob(F-stat) = 0.000
- Only one insignificant variable

The final equation

Equation with coefficient of the model

$$PAY\_MORE\_2 = \beta_0 + \beta_1 \times CHOICE\_2 + \beta_2 \times DEPARTURE\_HOUR + \beta_3 \times GENDER + \beta_4 \times RAMP + \beta_5 \times SEAT\_2 + \beta_6 \times PRICE\_2$$

Marginal effects : expectations vs observations(1)

Variable	Expected effect	Observed
Age	+	Irrelevant
Airport_proximity	+	Irrelevant
Best_price	-	Irrelevant
Children	?	Irrelevant
Choice_2	-	-
Crew_kindness	+	Irrelevant
Departure_hour	+	+
Drinks	+	Irrelevant
Education	?	Irrelevant
Fly_why	+	Irrelevant
Gender	?	+
Heigth	?	Irrelevant

Marginal effects : expectations vs observations(2)

Variable	Expected effect	Observed
Incomes	+	Irrelevant
Luggage	+	Irrelevant
Married	?	Irrelevant
Nationality	?	Irrelevant
Nb_fly	?	Irrelevant
Occupation_cat	+	Irrelevant
Pleasure	+	Irrelevant
Price	-	Irrelevant
Ramp	+	+
Seat_2	+	Irrelevant
Snacks	+	-

Normality of the errors

Observations

- We notice that Jarque Bera test gives 269, which is far from 0.
- Residuals don't follow a normal law.

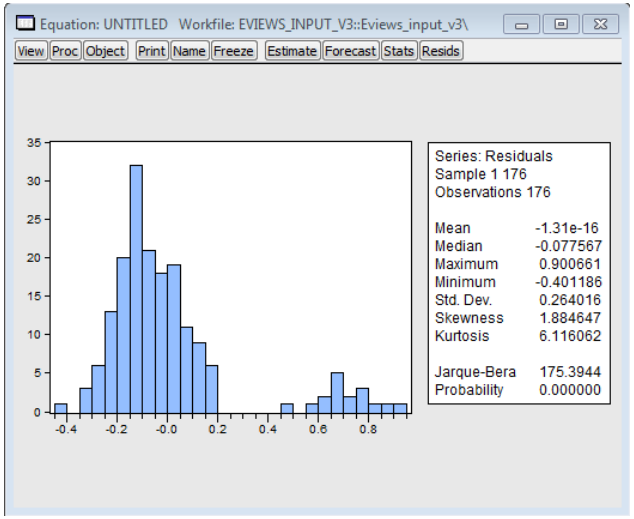


FIGURE – Residual normality for short haul

Analysis

To test the significance of PRICE, we compute a Wald test :

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	2.555566	(1, 169)	0.1118
Chi-square	2.555566	1	0.1099

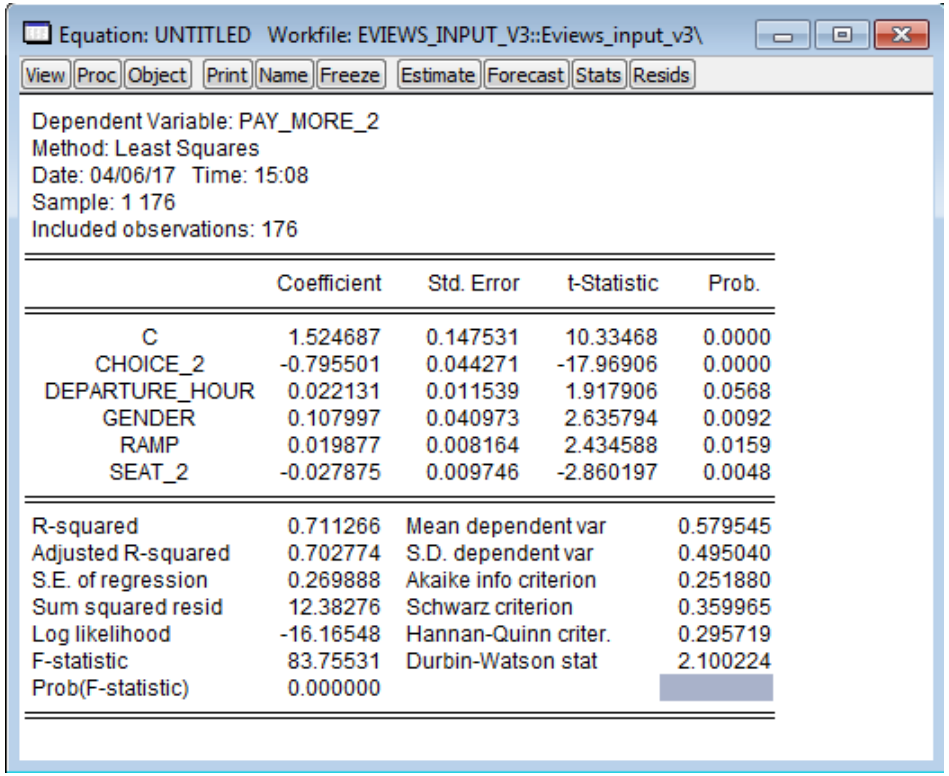
FIGURE – Wald test

$$\begin{cases} H_0 : & \beta_6 = 0 \\ H_1 : & \backslash H_0. \end{cases}$$

$qF = 2.556 < \chi^2(q = 1) = 3.841.$

$\Rightarrow$ We do not reject H_0.

Final Equation



Equation: UNTITLED Workfile: EVIEWS_INPUT_V3::Eviews_input_v3\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PAY_MORE_2
Method: Least Squares
Date: 04/06/17 Time: 15:08
Sample: 1 176
Included observations: 176

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.524687	0.147531	10.33468	0.0000
CHOICE_2	-0.795501	0.044271	-17.96906	0.0000
DEPARTURE_HOUR	0.022131	0.011539	1.917906	0.0568
GENDER	0.107997	0.040973	2.635794	0.0092
RAMP	0.019877	0.008164	2.434588	0.0159
SEAT_2	-0.027875	0.009746	-2.860197	0.0048

R-squared	0.711266	Mean dependent var	0.579545
Adjusted R-squared	0.702774	S.D. dependent var	0.495040
S.E. of regression	0.269888	Akaike info criterion	0.251880
Sum squared resid	12.38276	Schwarz criterion	0.359965
Log likelihood	-16.16548	Hannan-Quinn criter.	0.295719
F-statistic	83.75531	Durbin-Watson stat	2.100224
Prob(F-statistic)	0.000000		

FIGURE – Final Equation

Heteroskedasticity

F-statistic	3.321717	Prob. F(18,157)	0.0000
Obs*R-squared	48.54078	Prob. Chi-Square(18)	0.0001
Scaled explained SS	121.5627	Prob. Chi-Square(18)	0.0000

FIGURE – White's test

Observations

$$\begin{cases} H_0 : & R^2 = 0 \\ H_1 : & R^2 = 1. \end{cases}$$

- nR^2 follows a χ^2 law with q = 18
- $nR^2 = 48.54 > \chi^2(12) = 28.87$

$\Rightarrow$ At the 95% level : there is heteroskedasticity.

Pros

- We obtained a meaningful model.
- Better model than Long haul model.

Cons

- Studied variables usually were brackets : not very accurate.
- Classical assumptions fail.
- R^2 is acceptable but not close enough close to 1. Some useful variables might have been forgotten in the survey.

Conclusion

- A **satisfying** model but far from being perfect.
- Some surprising survey's results : people **don't answer** or don't **understand** questions as we expected them to.
- Most of the variables are **irrelevant**.
- Short haul model is **better** than the long haul one.
- We highlighted **heteroskedasticity** in both cases. Dispersion of errors increases when the number of variables increases.