Airline Itinerary Choice (Boeing)

MNL with Generic Attributes

Files to use with BIOGEME:
Model file: *MNL_generic_boeing.mod*
Data file: *boeing.dat*

The choice set consists of the following three alternatives:

1. a non-stop flight,
2. a flight with one stop on the same airline,
3. a flight with one stop and a change of airline.

We define the deterministic part of the utility for the household by including the alternative specific constants (ASCs) and five attributes, namely Fare, $\mathsf{Legroom}$, $\mathsf{Total\_TT}$ and $\mathsf{SchedDE}$ and $\mathsf{SchedDL}$, with their respective generic coefficients β_{Fare}, β_{Legroom}, $\beta_{\mathsf{Total\_TT}}$, β_{SchedDE} and β_{SchedDL}:

$$
\begin{aligned}
V_1 &= ASC_1 + \beta_{\mathsf{Fare}} \cdot \mathsf{Fare}_1 + \beta_{\mathsf{Legroom}} \cdot \mathsf{Legroom}_1 + \beta_{\mathsf{Total\_TT}} \cdot \mathsf{Total\_TT}_1 \\
&\quad + \beta_{\mathsf{SchedDE}} \cdot \mathsf{SchedDE}_1 + \beta_{\mathsf{SchedDL}} \cdot \mathsf{SchedDL}_1 \\
V_2 &= ASC_2 + \beta_{\mathsf{Fare}} \cdot \mathsf{Fare}_2 + \beta_{\mathsf{Legroom}} \cdot \mathsf{Legroom}_2 + \beta_{\mathsf{Total\_TT}} \cdot \mathsf{Total\_TT}_2 \\
&\quad + \beta_{\mathsf{SchedDE}} \cdot \mathsf{SchedDE}_2 + \beta_{\mathsf{SchedDL}} \cdot \mathsf{SchedDL}_2 \\
V_3 &= ASC_3 + \beta_{\mathsf{Fare}} \cdot \mathsf{Fare}_3 + \beta_{\mathsf{Legroom}} \cdot \mathsf{Legroom}_3 + \beta_{\mathsf{Total\_TT}} \cdot \mathsf{Total\_TT}_3 \\
&\quad + \beta_{\mathsf{SchedDE}} \cdot \mathsf{SchedDE}_3 + \beta_{\mathsf{SchedDL}} \cdot \mathsf{SchedDL}_3
\end{aligned}
$$

One of the alternative specific constants (arbitrarily ASC_1) is normalized to zero for identification. The corresponding alternative is the reference alternative for the ASCs. This is important for the interpretation we will perform in the next paragraphs.

Given our specification, and everything being equal, an ASC with negative sign indicates a lower utility level for the corresponding alternative compared to the normalized one (i.e., the first one). As it can be observed in Table 1, this is the case for both other alternatives (ASC_2 and ASC_3 are negative and statistically significant). It means that alternative 1 is preferred to alternatives 2 and 3, i.e., alternative without stop is preferred to alternatives with stops.

The parameter related to leg room has a positive sign and it is significantly different from zero. It implies that more room for legs increases the utility of the alternative.

Generic MNL estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_2	-1.26	0.126	-9.95
2	ASC_3	-1.49	0.127	-11.72
3	β_{Fare}	-0.0194	0.000795	-24.37
4	$\beta_{Legroom}$	0.222	0.0266	8.35
5	$\beta_{SchedDE}$	-0.130	0.0161	-8.08
6	$\beta_{SchedDL}$	-0.0883	0.0145	-6.10
7	β_{Total_TT}	-0.326	0.0671	-4.85
		.	.	.
Summary statistics				
Number of observations $= 3609$				
$\mathcal{L}(0) = -3964.892$				
$\mathcal{L}(\hat{\beta}) = -2333.701$				
$\bar{\rho}^2 = 0.410$				

Table 1: Estimation results for the MNL model with generic attributes

For other parameters, like fare, delays and travel time, the sign is negative. It means that all these factors have a negative impact on utility: they make the alternative less likely to be chosen.

MNL with Alternative-Specific Coefficients

Files to use with BIOGEME:
Model file: *MNL_specific_boeing.mod*
Data file: *boeing.dat*

Next we present a model (unrestricted) with alternative-specific travel time coefficients and we compare it with the (restricted) model with generic coefficients presented in the previous section. We carry out a statistical test (likelihood ratio test) to assess if one specification is significantly better than the other. We will perform the analysis on the coefficient of the travel time. The deterministic utilities for this model with alternative-specific travel times are:

$$
\begin{aligned}
V_1 &= ASC_1 + \beta_{Fare} \cdot Fare_1 + \beta_{Legroom} \cdot Legroom_1 + \beta_{Total\_TT\_1} \cdot Total\_TT_1 \\
&\quad + \beta_{SchedDE} \cdot SchedDE_1 + \beta_{SchedDL} \cdot SchedDL_1 \\
V_2 &= ASC_2 + \beta_{Fare} \cdot Fare_2 + \beta_{Legroom} \cdot Legroom_2 + \beta_{Total\_TT\_2} \cdot Total\_TT_2 \\
&\quad + \beta_{SchedDE} \cdot SchedDE_2 + \beta_{SchedDL} \cdot SchedDL_2 \\
V_3 &= ASC_3 + \beta_{Fare} \cdot Fare_3 + \beta_{Legroom} \cdot Legroom_3 + \beta_{Total\_TT\_3} \cdot Total\_TT_3 \\
&\quad + \beta_{SchedDE} \cdot SchedDE_3 + \beta_{SchedDL} \cdot SchedDL_3
\end{aligned}
$$

Note that instead of only β_{Total_TT}, we have know $\beta_{Total\_TT\_1}$, $\beta_{Total\_TT\_2}$ and $\beta_{Total\_TT\_3}$.

Alternative-specific MNL estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_2	-1.43	0.183	-7.81
2	ASC_3	-1.64	0.192	-8.53
3	β_{Fare}	-0.0193	0.000802	-24.05
4	$\beta_{Legroom}$	0.226	0.0267	8.45
5	$\beta_{SchedDE}$	-0.139	0.0163	-8.53
6	$\beta_{SchedDL}$	-0.104	0.0137	-7.59
7	β_{Total_TT1}	-0.332	0.0735	-4.52
8	β_{Total_TT2}	-0.299	0.0696	-4.29
9	β_{Total_TT3}	-0.302	0.0699	-4.31
		.	.	.
Summary statistics				
Number of observations = 3609				
$\mathcal{L}(0) = -3964.892$				
$\mathcal{L}(\hat{\beta}) = -2320.447$				
$\bar{\rho}^2 = 0.412$				

Table 2: Estimation results for the MNL model with specific attributes

The results for the unrestricted model are reported in Table 2.

Generic vs Specific Test Under the null hypothesis:

$$H_0 : \beta_{Total\_TT\_1} = \beta_{Total\_TT\_2} = \beta_{Total\_TT\_3}$$

Reject null hypothesis (generic travel time coefficient) if :

$$-2(L_R - L_U) > \chi_{((1-\alpha),df}$$

Next we describe the standard steps to perform the test:

1. L_R and L_U represent the log-likelihood for both the restricted and the unrestricted models:

$$\begin{aligned} L_R &= -2333.701 \\ L_U &= -2320.447 \end{aligned}$$

2. The degree of freedom is given by the difference in the number of estimated parameters between the models:

$$df = K_U - K_R = 9 - 7 = 2$$

3. $-2(L_R - L_U) = -2(-2333.701 + 2320.447) = 26.508$
4. The critical value for $\chi_{(0.95,2)}$ is 0.103.
5. We conclude that we can reject the null hypothesis H_0 of generic coefficient in favor of alternative-specific coefficients.

Inclusion of Socio-Economic Characteristics

Files to use with BIOGEME:
Model file: *MNL_socioecon_boeing.mod*
Data file: *boeing.dat*

It is reasonable to assume that people make choices not only in relation to the attributes that characterize the alternatives but also depending on some personal characteristics or socioeconomic indicators. The availability of individual-specific information gives us the opportunity to model partly the heterogeneity present in the population. We modify the previous model by adding income of respondents into the utilities.

$$\begin{aligned} V_1 &= ASC_1 + \beta_{Fare} \cdot Fare_1 + \beta_{Legroom} \cdot Legroom_1 + \beta_{Total\_TT\_1} \cdot Total\_TT_1 \\ &\quad + \beta_{SchedDE} \cdot SchedDE_1 + \beta_{SchedDL} \cdot SchedDL_1 + \beta_{Inc_1} \cdot Income \\ V_2 &= ASC_2 + \beta_{Fare} \cdot Fare_2 + \beta_{Legroom} \cdot Legroom_2 + \beta_{Total\_TT\_2} \cdot Total\_TT_2 \\ &\quad + \beta_{SchedDE} \cdot SchedDE_2 + \beta_{SchedDL} \cdot SchedDL_2 + \beta_{Inc_2} \cdot Income \\ &\quad + \beta_{MI} \cdot MissingIncome \\ V_3 &= ASC_3 + \beta_{Fare} \cdot Fare_3 + \beta_{Legroom} \cdot Legroom_3 + \beta_{Total\_TT\_3} \cdot Total\_TT_3 \\ &\quad + \beta_{MI} \cdot MissingIncome \\ &\quad + \beta_{SchedDE} \cdot SchedDE_3 + \beta_{SchedDL} \cdot SchedDL_3 + \beta_{Inc_3} \cdot Income \end{aligned}$$

Since the variable of the income does not vary between the alternatives and only differences in utilities matter, we need to normalize one alternative to zero. We interpret the estimated coefficients for the remaining alternatives with respect to the reference alternative, which arbitrarily is alternative 1. It is similar to what we did when specifying alternative specific constants.

We assumed that the income of the respondent affects differently each alternative.

The estimation results of this model are reported in Table 3.

Socio-economic MNL estimation				
Parameter number	Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
1	ASC_2	-1.07	0.215	-4.96
2	ASC_3	-1.05	0.228	-4.61
3	β_{Fare}	-0.0195	0.000807	-24.18
4	β_{Inc_2}	-0.0419	0.0148	-2.83
5	β_{Inc_3}	-0.0755	0.0154	-4.90
6	β_{Legroom}	0.227	0.0268	8.49
7	β_{MI}	-0.578	0.159	-3.64
8	β_{SchedDE}	-0.139	0.0163	-8.50
9	β_{SchedDL}	-0.104	0.0139	-7.49
10	$\beta_{\text{Total\_TT1}}$	-0.335	0.0735	-4.56
11	$\beta_{\text{Total\_TT2}}$	-0.301	0.0696	-4.32
12	$\beta_{\text{Total\_TT3}}$	-0.304	0.0698	-4.36
Summary statistics				
Number of observations $= 3609$				
$\mathcal{L}(0) = -3964.892$				
$\mathcal{L}(\hat{\beta}) = -2307.488$				
$\bar{\rho}^2 = 0.415$				

Table 3: Estimation results for the MNL model with socioeconomic variables

Therefore we have specified 2 different β parameters associated with the attribute "income". β_{Inc} for alternative 1 has been normalized to zero. The two parameter estimates have negative signs, implying that the higher the income of the respondent, the lower the likelihood for choosing these two alternatives (with stops) compared to the first one (without stops).

In this model, we used the strategy proposed in Presentation 3 to deal with missing data. We defined "Income" as being the income variable without -1 and 99. We also defined another variable, called “MissingIncome”. “MissingIncome” is equal to 1 if the income variable is -1 or 99. We added both variables in the utilities.