EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire

Mathematical Modeling of Behavior
Fall 2020

Netherlands mode choice case

1 Model specification with generic attributes

Files to use with Biogeme:
Model notebook: *binary_generic_netherlands.ipynb*
Data file: *netherlands.dat*

In this first model, we assume that the total travel time (in-vehicle and out-of-vehicle) and travel cost of the modes are the only factors influencing the mode choice. We also assume that the coefficients of the explanatory variables are generic, i.e., they do not vary between alternatives. The expression of utility for this simple model can be written as:

$$
\begin{aligned}
V_{\text{car}} &= \text{ASC}_{\text{car}} + \beta_{\text{time}}\text{time}_{\text{car}} + \beta_{\text{cost}}\text{cost}_{\text{car}} \\
V_{\text{rail}} &= \beta_{\text{time}}\text{time}_{\text{rail}} + \beta_{\text{cost}}\text{cost}_{\text{rail}}
\end{aligned}
$$

The estimation results are shown in Table 1. All the estimated coefficients are statistically significantly different from zero. Looking at the alternative specific constant, the negative sign indicates that all else being equal, car is less preferred than rail. However, this may be due to the fact that the model is too simple and there are important variables left out of the model. The negative signs for the generic coefficients for travel cost and travel time indicate, as expected, that the utility perceived by the decision maker for any of the two alternatives decreases with an increase in the travel cost and travel time.

2 Model specification with alternative specific attributes

File to develop using the same dataset as before:
Model notebook: *binary_specific_netherlands.ipynb*

In the second specification, we relax the hypothesis of generic travel time coefficients. The alternative specific coefficients are more relevant if people perceive a minute spent in one mode

Estimation results			
Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
$\mathrm{ASC}_{\mathrm{car}}$	-0.798	0.275	-2.90
β_{cost}	-0.113	0.0241	-4.67
β_{time}	-1.33	0.354	-3.75
Summary statistics			
Number of observations = 228			
$\mathcal{L}(0) = -158.038$			
$\mathcal{L}(\hat{\beta}) = -123.133$			
$\bar{\rho}^2 = 0.202$			

Table 1: Estimation results with generic attributes (binary logit)

to be different than a minute spent in the other mode. To illustrate this idea, two different travel time coefficients are introduced for car and rail. The corresponding utility function is given below:

$$\begin{aligned} V_{car} = &\ \mathrm{ASC}_{\mathrm{car}} + \beta_{\mathrm{time\_car}}\mathrm{time}_{\mathrm{car}} + \beta_{\mathrm{cost}}\mathrm{cost}_{\mathrm{car}} \\ V_{\mathrm{rail}} = &\ \beta_{\mathrm{time\_rail}}\mathrm{time}_{\mathrm{rail}} + \beta_{\mathrm{cost}}\mathrm{cost}_{\mathrm{rail}} \end{aligned}$$

Estimation results			
Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
$\mathrm{ASC}_{\mathrm{car}}$	2.43	0.973	2.50
β_{cost}	-0.123	0.0256	-4.79
$\beta_{\mathrm{time\_car}}$	-2.26	0.485	-4.66
$\beta_{\mathrm{time\_rail}}$	-0.543	0.396	-1.37
Summary statistics			
Number of observations = 228			
$\mathcal{L}(0) = -158.038$			
$\mathcal{L}(\hat{\beta}) = -118.023$			
$\bar{\rho}^2 = 0.228$			

Table 2: Estimation results with alternative-specific coefficients (binary logit)

The estimation results are shown in Table 2. This model has a better adjusted likelihood ratio index than the model with generic travel time coefficients. However, the coefficient for the travel time of the rail alternative is not statistically significantly different from zero. The coefficient for the travel time of the car alternative is negative and significant as expected, and is also greater in absolute value than the generic one presented in the previous table (-2.26 vs. -1.33). As in the previous example, the negative sign indicates that the utility perceived by the decision maker for the car alternative decreases with the increase of travel time.

From the obtained results, it appears that travel time does not affect the car and rail alternatives in the same way. The results indicate that people have less negative utility for travel time in rail compared to car. This may be due to the fact that people can make better use of their time when traveling by rail. The alternative specific constant for the car alternative has now the reversed sign denoting increased preference for car (all else being equal), which is more intuitive. A likelihood ratio test can be performed to test whether or not there is a significant improvement in the goodness-of-fit in the modified specification with alternative specific coefficients for travel time.

Generic vs. specific test

The likelihood ratio test can be used to test the generic vs. the alternative-specific specification. The likelihood ratio test statistic for the null hypothesis of generic attributes is

$$-2(\mathcal{L}(\hat{\beta}_G) - \mathcal{L}(\hat{\beta}_{AS})),$$

where G and AS denote the generic and alternative-specific models, respectively. It is χ^2 distributed with the number of degrees of freedom equal to the number of restrictions ($K_{AS} - K_G = 1$). In this case, $-2(-123.133 + 118.023) = 10.22$. Since $\chi^2_{0.95,1} = 3.841$ at a 95% level of confidence, we can conclude that the model with the alternative-specific coefficients has a significant improvement in fit.

3 Model specification with socio-economic characteristics

File to develop using the same dataset as before:
Model notebook: *binary_socioec_netherlands.ipynb*

The previous two models only included variables that were attributes of the alternatives. We now introduce a socio-economic variable *gender* which indicates the respondent's gender. The variable is categorical and equals one if the gender is female and zero if male. Since this variable does not vary by alternative (recall that only difference in utility matters), we have normalized the term associated with the car alternative to zero. As is shown in the utility function below, the gender variable only enters the utility of the rail alternative. However, this is an arbitrary normalization, as we could have also normalized the rail alternative.

$$\begin{aligned} V_{car} &= \mathrm{ASC}_{\mathrm{car}} + \beta_{\mathrm{time\_car}}\mathrm{time}_{\mathrm{car}} + \beta_{\mathrm{cost}}\mathrm{cost}_{\mathrm{car}} \\ V_{rail} &= \beta_{\mathrm{time\_rail}}\mathrm{time}_{\mathrm{rail}} + \beta_{\mathrm{cost}}\mathrm{cost}_{\mathrm{rail}} + \beta_{\mathrm{gender}}\mathrm{gender} \end{aligned}$$

The estimation results are shown in Table 3. The results show that there is a slight improvement in the adjusted likelihood ratio index. The coefficient of the gender variable is positive and statistically significant, which indicates that women have higher probability than men of choosing the rail alternative with respect to the car alternative. The reader can verify that if we had included the gender variable in the utility of the car alternative instead of the rail alternative, the conclusion would remain unchanged. In fact, the results would be exactly the same. The only difference is that the coefficient would show the opposite sign. In our case, it would become negative. The interpretation would be that women would have lower probability than men of using the car alternative with respect to the train alternative, which is exactly the same result we had before. Regarding the coefficients of the other explanatory variables, they are almost unchanged with respect to the previous model.

Estimation results			
Parameter name	Parameter estimate	Robust standard error	Robust *t statistic*
$\mathrm{ASC}_{\mathrm{car}}$	2.85	1.02	2.80
β_{gender}	0.675	0.329	2.05
β_{cost}	-0.130	0.0265	-4.89
$\beta_{\mathrm{time\_car}}$	-2.34	0.495	-4.73
$\beta_{\mathrm{time\_rail}}$	-0.529	0.414	-1.28
Summary statistics			
Number of observations = 228			
$\mathcal{L}(0) = -158.038$			
$\mathcal{L}(\hat{\beta}) = -115.880$			
$\bar{\rho}^2 = 0.235$			

Table 3: Estimation results with socio-economic characteristics (binary logit)

mpp / th / rk / no / jp / mw