Discrete-Choice Model Example: Travel Mode

You need to choose bus vs MRT. What variables affect your decision?

The utility of the traveler depends on the expected travel time, travel time uncertainty, and the fare as follows

$$U_{o,d}(z(t), t, i) = \alpha_{t,z} + \alpha_{i,z} + \beta_T(t)E\left[T_{o,d}(z(t), t)\right] + \beta\sigma(t)\sigma\left[T_{o,d}(z(t), t)\right] + \beta_C(t)C_{o,d}(z(t), t) + e_{o,d}(z(t), t), \quad (1)$$

where binary variable $z(t) \in \{B, R\}$ and it corresponds to mode choice, i.e., traveling by bus (B) or by rail (R) at time t, $\alpha_{t,z}$ and $\alpha_{i,z}$ are mode specific time and traveler fixed effects, $\sigma[X]$ is the standard deviation of X , $T_{o,d}(z(t),t)$ is the travel time from o to d by bus or rail when the travel starts at time t, $\beta_T(t)$ and $\beta_\sigma(t)$ are presumably negative coefficients for the mean travel time and the standard deviation, $\beta_C(t)$ is negative parameter for the cost, and $e_{o,d}(z(t),t)$ is an i.i.d. error term with extreme value distribution (see e.g. Embrechts et al. (1997)) driven by factors not included in our model.

The traveler i maximizes the utility by selecting z. That is,

$$V_{o,d}(t, i) = \max_{z(t)\in\{B,R\}} U_{o,d}(z(t), t, i). \quad (2)$$

Thus, the traveler uses bus between o and d at time t if and only if $U_{o,d}(B,t,i) \geq U_{o,d}(R,t,i)$.

The forecasted utility (or the expected utility):

$$\widehat{U}_{o,d}(z(t), t, i) = \alpha_{t,z} + \alpha_{i,z} + \beta_T(t)E\left[T_{o,d}(z(t), t)\right] + \beta\sigma(t)\sigma\left[T_{o,d}(z(t), t)\right] + \beta_C(t)C_{o,d}(z(t), t).$$

Since $e_{o,d}(z(t),t)$ has extreme value distribution, (2) corresponds to the chance of choosing bus:

$$P_{o,d}(B,t,i) = \frac{\exp\left(\widehat{U}_{o,d}(B,t,i)\right)}{\exp\left(\widehat{U}_{o,d}(B,t,i)\right)+\exp\left(\widehat{U}_{o,d}(R,t,i)\right)} =$$

$$\left[1+\exp\left(\widehat{U}_{o,d}(R,t,i)-\widehat{U}_{o,d}(B,t,i)\right)\right]^{-1}, \quad (3)$$

where $P_{o,d}(z(t),t,i)$ is the probability that traveler i choses $z(t) \in \{B, R\}$ at time t.

The logit corresponding to (3) is given by

$$\log\left(\frac{P_{o,d}(B,t,i)}{1-P_{o,d}(B,t,i)}\right) = \widehat{U}_{o,d}(B,t,i) - \widehat{U}_{o,d}(R,t,i), \quad (4)$$

where

$$\begin{aligned}\widehat{U}_{o,d}(B,t,i) - \widehat{U}_{o,d}(R,t,i) &= \alpha_{t,B} - \alpha_{t,R} \\ &+ \alpha_{i,B} - \alpha_{i,R} + \beta_T(t)E\left[T_{o,d}(B,t) - T_{o,d}(R,t)\right] \\ &+ \beta\sigma(t)\left(\sigma\left[T_{o,d}(B,t)\right] - \sigma\left[T_{o,d}(R,t)\right]\right) \\ &+ \beta_C(t)\left(C_{o,d}(B,t) - C_{o,d}(R,t)\right).\end{aligned}$$

Thus, the time fixed effect difference $\alpha_t = \alpha_{t,B} - \alpha_{t,R}$, the individual fixed effect difference $\alpha_i = \alpha_{i,B} - \alpha_{i,R}$, regression slope parameters $\beta_T(t)$, $\beta\sigma(t)$, and $\beta_C(t)$ are estimated by logistic regression or the corresponding ordinary least square (OLS) regression using (4).

By our model, we can also analyze, for instance, the effect of exogenous rainfall shocks on travelers' choices as follows. Heavy rainfall significantly affects bus travel time, because then road traffic slows down. However, the rain does not affect the rail travel time.

Therefore, by (3) and (4), the rainfall shock decreases the probability of taking bus:

$$\tilde{P}_{o,d}(B,t,i) - P_{o,d}(B,t,i) \approx$$
$$\left[1 + \exp\left(\widehat{U}_{o,d}(R,t,i) - \widehat{U}_{o,d}(B,t,i)\right)\right]^{-2} \beta_T(t) \, \Delta T_{o,d}(B,t) =$$
$$\left(P_{o,d}(B,t)\right)^2 \beta_T(t) \, \Delta T_{i,j}(B,t), \quad (5)$$

where $\Delta T_{i,j}(B,t)$ is the increase in bus travel time due to a heavy rain and $\tilde{P}_{o,d}(B,t,i)$ is traveler *i*'s probability of travelling by bus at time *t* between *o* and *d* under the heavy rain. Since the coefficient of the mean travel time $\beta_T(t) < 0$, equation (5) says that the rainfall shock indeed decreases the probability of taking bus.

Example: Use logistic regression (and/or tree models) to find which companies made offers to a sample of students.

Call:

glm(formula = choice ~ Industry + Date + Agent +…, family = "binomial", data = mydata)