EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire

Mathematical Modeling of Behavior
Fall 2020

EXERCISE SESSION 13 (Solution)

Exercise 1 An analyst is interested in modeling a mode choice problem between car and public transportation (PT) in a large city taking into account the effect of attitudes such as discomfort of using public transportation (*AttAntiPT*) and environmental concern (*AttProEnv*). In this context, an integrated choice and latent variable model is appropriate in order to account for the effects of the attitudes on mode choice.

The analyst is designing the survey for the data collection. She will collect information about the characteristics of the individuals and the observed attributes of the alternative modes, which are exogenous explanatory variables, as well as *attitudinal indicators* for the latent variables model. To obtain these indicators, she proposes statements and asks the respondent to position himself about them, using a scale ranging from -2 (totally disagree) to 2 (totally agree). Two indicators related to each attitude are collected. The two indicators related to the attitude *AttAntiPT* are noted I_{1n} and I_{2n} and the two indicators related to *AttProEnv* are noted I_{3n} and I_{4n}.

1. In the specification of the choice model the analyst has considered the effect of: i) cost, ii) time, iii) number of children, iv) frequency of public transportation (Figure **??**). Complete the diagram in Figure **??** in order to integrate the effect of the two latent variables in the choice model. We assume that the latent variables are not explained by the variables i)-iv). Make sure to use appropriate shapes and arrows and explain the drawing convention, i.e., what do the different shapes and arrows represent.

   Solution:

   - The shapes indicate the nature of the variables: square shape for observable variables and oval shape for unobserved (latent) variables.
   - The arrows indicate the type of the relation between the model components. A solid arrow indicates "cause-effect" (structural) relations. A dashed arrow indicates a measurement (relationships between the underlying latent variables and their observable manifestations). A doted arrow corresponds to the errors.

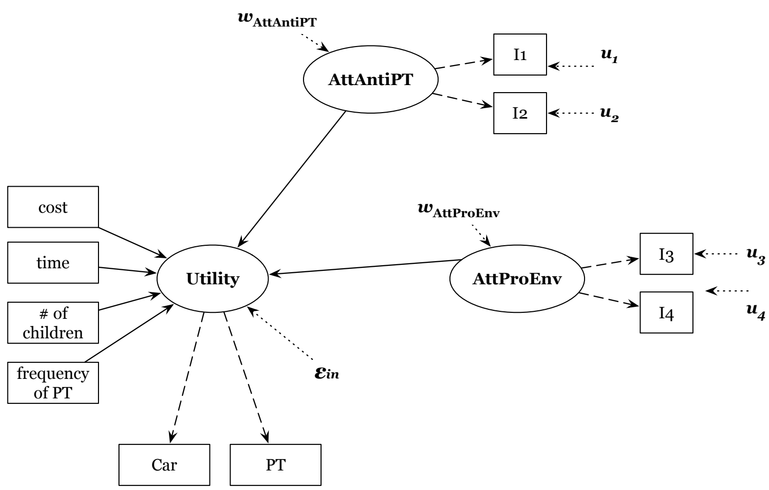


Figure 1: Specification diagram of the mode choice model with two latent variables.

2. Update your diagram once more (use Figure **??**), assuming that people with high education level and people who own a bike have a pro-environmental attitude, while people with low education level and owning more than one car have an attitude against PT.

Solution:

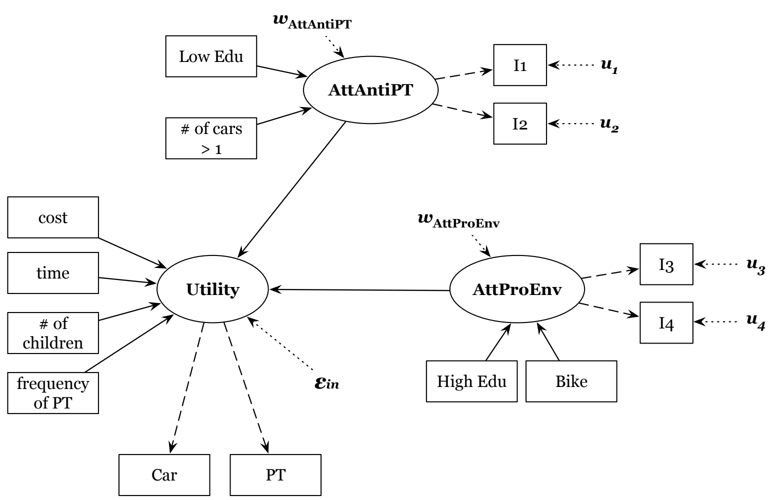


Figure 2: Specification diagram of the mode choice model with two latent variables (updated).

3. Write down the structural equations corresponding to the a) *choice* and b) *latent variable* models, accordingly. Consider the specification of the updated diagram obtained in sub-question 2.

Solution:

a) choice model

$$\begin{aligned} U_{\mathrm{PT},n} &= \mathrm{ASC}_{\mathrm{PT}} + \beta_{cost} \cdot \mathrm{cost}_{\mathrm{PT},n} + \beta_{time} \cdot \mathrm{time}_{\mathrm{PT},n} + \beta_{child} \cdot \mathrm{child}_n + \\ &\quad + \beta_{freqPT} \cdot \mathrm{freqPT}_n + \gamma_1 \cdot AttProEnv_{\mathrm{PT},n} + \gamma_2 \cdot AttAntiPT_{\mathrm{PT},n} + \varepsilon_{\mathrm{PT},n} \\ U_{\mathrm{CAR},n} &= \beta_{cost} \cdot \mathrm{cost}_{\mathrm{CAR},n} + \beta_{time} \cdot \mathrm{time}_{\mathrm{CAR},n} + \varepsilon_{\mathrm{CAR},n} \end{aligned}$$

b) latent variable model

$$\begin{aligned} AttProEnv_n &= \theta_0 \cdot 1 + \theta_{HighEdu} \cdot \mathrm{HighEdu}_n + \theta_{Bike} \cdot \mathrm{Bike}_n + w_{\mathrm{AttProEnv},n} \\ AttAntiPT_n &= \eta_0 \cdot 1 + \eta_{LowEdu} \cdot \mathrm{LowEdu}_n + \eta_{Car} \cdot \mathrm{Car}_n + w_{\mathrm{AttAntiPT},n} \end{aligned}$$

4. Write down the measurement equations corresponding to the a) *choice* and b) *latent variable* models, accordingly. Take into account that the collected indicators are discrete.

Solution:

a) choice model

$$y_{\mathrm{CAR},n} = \begin{cases} 1 & if\ U_{\mathrm{CAR},n} \geq U_{\mathrm{PT},n} \\ 0 & otherwise \end{cases} \tag{1}$$

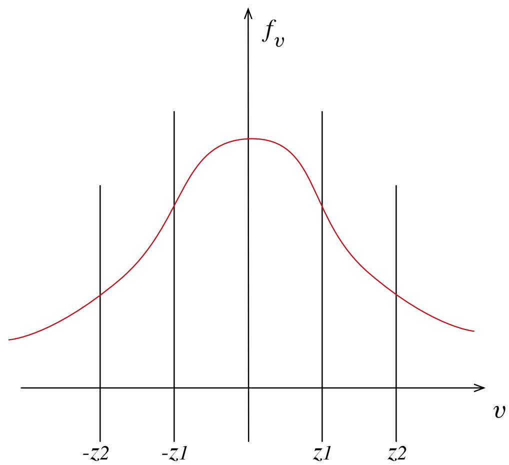
b) latent variable model

The indicators are discrete (they take value -2, -1, etc.). When discrete indicators are involved, the approach used for continuous indicators cannot be applied directly, and needs to be adapted. We consider the following procedure:

- We define a continuous (auxiliary) indicator $I^*_{k,n} = m(LV_n, \theta) + \nu$, where k denotes any of the indicators.
- We define the discrete indicator (the one observed and present in the data) from the continuous one $I^*_{k,n}$ as follows:

$$I_{k,n} = \begin{cases} -2 & -\infty < I^*_{k,n} \leq -z_2 \\ -1 & -z_2 < I^*_{k,n} \leq -z_1 \\ 0 & -z_1 < I^*_{k,n} \leq z_1 \\ 1 & z_1 < I^*_{k,n} \leq z_2 \\ 2 & z_2 < I^*_{k,n} \leq +\infty \end{cases} \tag{2}$$

where the values z_1 and z_2 need to be estimated. Note that the symmetry is used both to represent symmetry in the answers (in reality) and to simplify the number of parameters that needs to be estimated (see figure).



- Then, the probability of the indicator $I_{k,n}$ to take a certain value can be illustrated with the following example ($I_{k,n} = 0$):

$$P(I_{k,n} = 0) = P(-z_1 < I^*_{k,n} \leq z_1) =$$
$$P(I^*_{k,n} \leq z_1) - P(I^*_{k,n} \leq -z_1) =$$
$$P(m(\cdot) + v \leq z_1) - P(m(\cdot) + v \leq -z_1) =$$
$$P(v \leq z_1 - m(\cdot)) - P(v \leq -z1 - m(\cdot)) =$$
$$f_v(z_1 - m(\cdot)) - f_v(-z_1 - m(\cdot))$$

---

rk / th / no / mpp / jp / mw