EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire

Mathematical Modeling of Behavior
Fall 2020

EXERCISE SESSION 2

Exercise 1 We want to build a model that predicts the market penetration of electric vehicles (EV) as a function of the income level. We have a sample of 1000 individuals. The data is summarized in Table 1.

Table 1: Contingency table of EV ownership conditional on income level

	Income			
EV	low	medium	high	
yes	15	50	135	200
no	200	450	150	800
	215	500	285	1000

1. Estimate the parameters π_1, π_2 and π_3 using maximum likelihood estimation, where:

$$\begin{aligned} \text{Prob}(\text{EV} = \text{yes} \quad | \quad \text{income} = \text{low}) &= \pi_1 \\ \text{Prob}(\text{EV} = \text{yes} \quad | \quad \text{income} = \text{medium}) &= \pi_2 \\ \text{Prob}(\text{EV} = \text{yes} \quad | \quad \text{income} = \text{high}) &= \pi_3 \end{aligned} \tag{1}$$

   Hint: write the likelihood function and find its maximum.

2. Do the values of the parameters make sense?

3. Calculate the final log likelihood of the model.

4. Call the model described above M1. Consider a model with only two income categories: (i) low and medium income and (ii) high income. Call this M2. Now, consider another model with only one income category. Call this M0. Among the three models, which one would you choose as the best model? Why?
   Hint: Calculate the final log likelihood associated with each model and perform likelihood ratio tests.

5. Suppose now that after some economical growth, the income distribution is as follows: 75 individuals with low income, 400 individuals with medium income and 525 with high income. Use the best model obtained in question 4 to predict the market penetration of EV for this scenario.

6. Could we have used linear regression instead of discrete choice models in the context above? If so, what would have been the dependent variable?

Exercise 2 In a mode choice experiment the following utility functions are defined for private motorized modes (pmm) and public transportation (pt):

$$\begin{aligned} U_{pmm,n} &= -\beta_c \cdot \text{cost}_{pmm,n} - \beta_t \cdot \text{time}_{pmm,n} + \varepsilon_{pmm,n} \\ U_{pt,n} &= -\beta_c \cdot \text{cost}_{pt,n} - \beta_t \cdot \text{time}_{pt,n} + \varepsilon_{pt,n} \end{aligned} \tag{2}$$

where $\text{cost}_{pmm,n}$ and $\text{cost}_{pt,n}$ are the cost of the trip by private motorized modes and public transportation respectively for individual n in CHF, and $\text{time}_{pmm,n}$ and $\text{time}_{pt,n}$ are their travel times in minutes.

Our sample contains the following 10 observations:

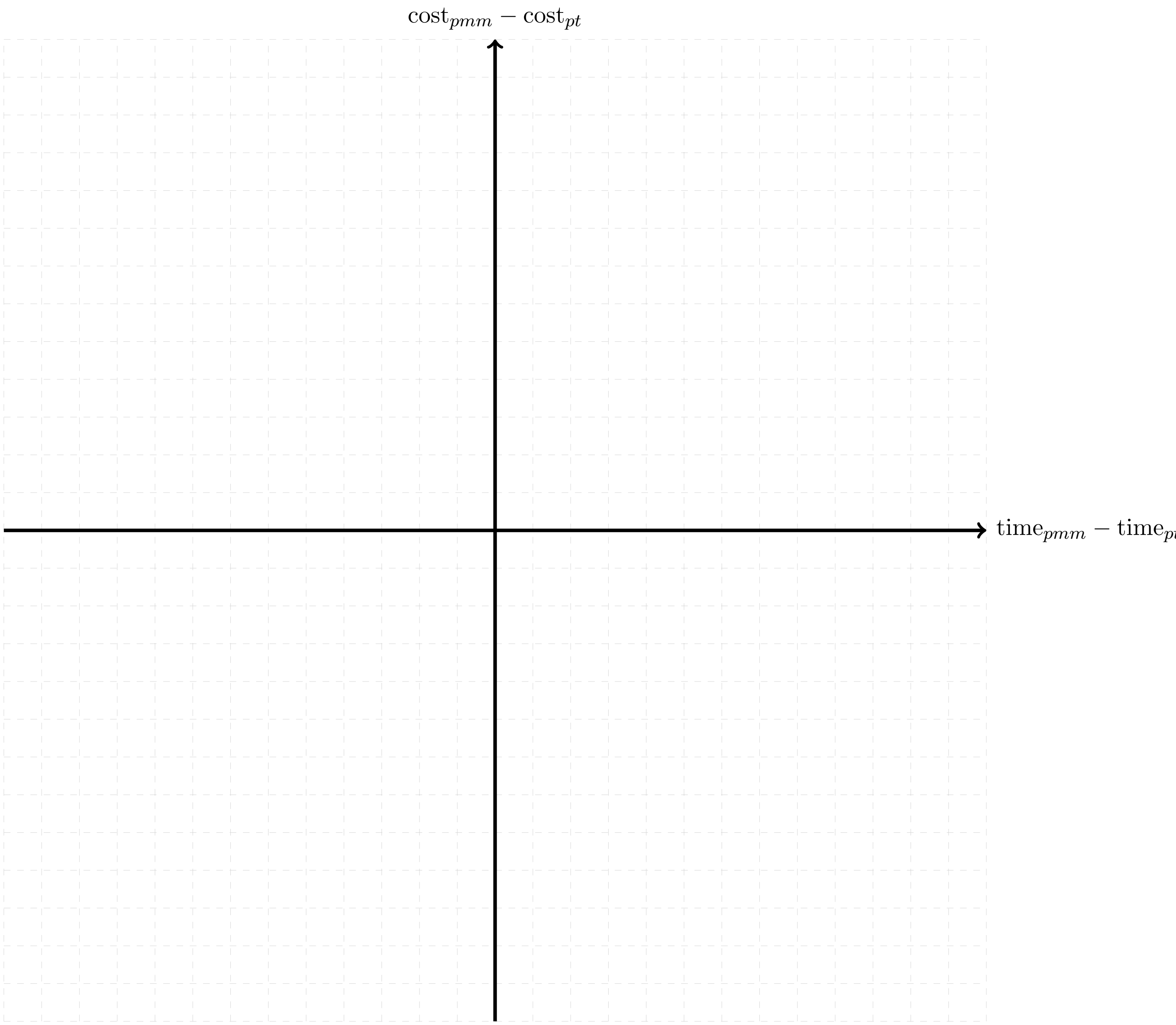
Individual	Choice	time_{pmm}	time_{pt}	cost_{pmm}	cost_{pt}
1	pmm	10	20	2.3	1
2	pt	5	10	2.3	0.5
3	pmm	35	30	9	12
4	pmm	20	22	1.5	2
5	pt	6	7.5	2	1.25
6	pt	10	15	5	3.5
7	pt	8	5	3	2
8	pt	19	18	4	5
9	pt	22	19	7	8.5
10	pmm	8	8.5	3	9

The parameter estimates are $\beta_c = 1.38$ and $\beta_t = 0.363$.

1. What is the value of time according to the model in [CHF/h]?

2. Plot these observations in the plot provided in the following page, where the x-axis is $\text{time}_{pmm} - \text{time}_{pt}$ and the y-axis is $\text{cost}_{pmm} - \text{cost}_{pt}$.

3. Add to the previous plot the line $-\beta_c \cdot \text{cost}_{pmm} - \beta_t \cdot \text{time}_{pmm} = -\beta_c \cdot \text{cost}_{pt} - \beta_t \cdot \text{time}_{pt}$. What does its slope represent?



mpp/th/rk/no/jp/mw