Family name:	First name:	Matrikel number:

Examination for the Master's Course Methods of Econometrics Winter semester 2021/22

Problem 1 (25 points)

Given is a general discrete-choice situation with known deterministic utilities V_i and additive random utilities ϵ_i for all the options $i = 1, ..., I$. Indicate if the following statements are true **and justify your choice with a short sentence**

1. Without changing the choice probabilities, you can add or subtract to all V_i a common constant which also allows setting one $V_i = 0$.
2. Without changing the choice probabilities, you can multiply all V_i with a common nonzero factor.
3. If the random terms ϵ_i are independent between choices, you can set $\epsilon_1 = 0$ by subtracting ϵ_1 from all the other random utilities .
4. If there are I alternatives, only $I - 1$ deterministic and random utilities can be defined.
5. You can multiply *both* deterministic and random utilities for all alternatives with a common positive factor without changing the choice probabilities.
6. The expectation value of the random utilities must be =0.
7. For any correlated or uncorrelated ϵ_i, the choice probabilities can be expressed in terms of a distribution function if $I = 2$ (binomial case) while no analytical solution is possible for the general multinomial ($I > 2$) case except for the Logit model.

Family name:	First name:	Matrikel number:

Problem 2 (45 points)

Consider a revealed-choice survey of the mode choice for commuting trips home to work in larger cities. The people to be questioned are randomly drawn from the person registers of the cities and can answer the questions by telephone or via internet using a personal one-off link (Einmallink). Only employed people are included. The design includes as exogenous variables (i) the age, (ii) the gender, (iii) the selected mode, (iv) the total travelling time for this mode, (v) the ad-hoc travel costs, (vi) the weather, (vii) the income range (low, middle, high), (viii) whether this person has usable cars and/or rideable bicycles and/or public transport season tickets.

(a) What is the reason for selecting the interviewed persons from a register giving them one-off links instead of just offering a general link on a homepage or social network which surely will attract more persons?

(b) The study design intends to use four alternatives: pedestrian ($i = 1$), bicycle($i = 2$), public transport($i = 3$), and car($i = 4$). What is the problem with this choice set? (*hint:* discuss exclusivity and completeness.)

(c) Indicate shortly, *in this context*, the advantages and drawbacks of a revealed-choice vs a stated-choice study. For the disadvantage, also consider whether the characteristics of all alternatives are known, or not.

(d) Classify each of the variables (i) to (viii) as one of the following: (a) alternative-specific constant, (b) characteristic, (c) socioeconomic variable, (d) external variable, (e) endogenous variable.

(e) Discuss whether it is possible to formulate the gender as a factor

$$V_i^{\mathrm{g}} = \beta_0 g, \quad i = 1, ..., 4, \quad g = 0: \text{ male}, \quad g = 1: \text{ female}.$$

(f) The weather influence is modelled by a factor V_i^{w} with

$$V_1^{\mathrm{w}} = \beta_1 w, \quad V_2^{\mathrm{w}} = \beta_2 w, \quad V_3^{\mathrm{w}} = \beta_3 w, \quad w = \begin{cases} 0 & \text{nice} \\ 1 & \text{bad} \end{cases}$$

What sign do you expect for β_1 and β_2 (justify your answer)? Also give the utility increase of driving in a car relative to riding a bicycle if the weather changes from nice to bad.

(g) Some argue that the weather preferences increase with the total travel time. How would you model this?

Family name:	First name:	Matrikel number:

Problem 3 (50 points)

In order to determine what influences the road capacity y, defined as the *maximum number of vehicles per hour and lane that does not lead to traffic breakdowns*, following table has been generated from counting detector data (the road categories are s: city streets, r: roads outside of cities, and f: freeway/Autobahn

speed limit x_1 [km/h]	50	100	100	70	80	30	30	none	30	60
road category x_2	s	r	f	r	f	s	s	f	c	r
truck percentage x_3[%]	20	30	10	10	20	0	5	20	10	40
Capacity y [Veh./h/lane]	800	1800	2400	1500	2200	500	400	1800	550	600

(a) What could you do with the "none" entry for the speed limit in the 8$^{\text{th}}$ data column?

(b) Argue that it is not a good idea to model the speed limit linearly, $y = \beta_1 x_1$. Consider the fact that, at a limit of zero, the capacity is zero while, at very high speed limits, the difference between the speeds of the vehicles leads to perturbed traffic with reduced capacity.

(c) The analyzed model is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 \begin{cases} 1 & \text{road=s} \\ 0 & \text{otherwise} \end{cases} + \beta_4 \begin{cases} 1 & \text{road=r} \\ 0 & \text{otherwise} \end{cases} + \beta_5 x_3.$$

Give the meaning of the parameters β_0 to β_5.

(d) Why is it not a good idea to add an additional factor with a dummy =1 if the road is a freeway and zero, otherwise?

(e) Instead of $\beta_5 x_3$, one could also model the truck influence as an additional factor $(1 + \beta_5 x_3)$ multiplied to the factors with the parameters β_1 to β_4. What is the difference between these two approaches?

(f) The parameters are now given for the model in (c) as

$$\hat{\beta}_0 = 1000, \quad \hat{\beta}_1 = 20, \quad \hat{\beta}_2 = -0.1, \quad \hat{\beta}_3 = -900, \quad \hat{\beta}_4 = -500, \quad \hat{\beta}_5 = -5$$

Give the expected capacity for a road with following properties

(i) city street, speed limit 40 km/h, truck percentage 20 %

(ii) Freeway on Sunday (no trucks), speed limit 120 km/h.

Family name:	First name:	Matrikel number:

(g) Some of the estimated variances and co-variances are

$$\hat{V}_{00} = 1\,100, \quad \hat{V}_{11} = 20, \quad \hat{V}_{22} = 0.002, \quad \hat{V}_{01} = 15, \quad \hat{V}_{02} = -0.3, \quad \hat{V}_{12} = 0.02$$

Test the null hypothesis that the linear speed dependence is negative. Why is it not a good idea to test the hypothesis that it is positive?

(h) Give the confidence interval for the estimated capacity for a freeway with no trucks and a speed limit of 100 km/h ($\hat{y} = \hat{\beta}_0 + 100\hat{\beta}_1 + 10\,000\hat{\beta}_2$).

Tables

Quantiles $t_{n,p}$ of the Student t distribution with n degrees of freedom

n	$p = 0.60$	0.70	0.80	0.90	0.95	0.975	0.990	0.995	0.999	0.9995
1	0.325	0.727	1.376	3.078	6.315	12.706	31.821	63.657	318.31	636.62
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.153	3.707	5.208	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.154	4.587
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090	3.291